Received: 2 August 2019 | Revised: 12 November 2019 | Accepted: 20 January 2020

DOI: 10.1111/jcal.12426

REGISTERED REPORT STAGE 2: COMPLETED STUDY

Journal of Computer Assisted Learning WILEY

Homogeneous and heterogeneous multiple representations in equation-solving problems: An eye-tracking study

Sarah Malone[1] | Kristin Altmeyer[1] | Markus Vogel[2] | Roland Brünken[1]

[1]Department of Education, Saarland University, Saarbrücken, Germany

[2]Institute for Mathematics and Computer Science, Heidelberg University of Education, Heidelberg, Germany

Correspondence
Kristin Altmeyer, Department of Education, Saarland University, Campus A4 2, 66123 Saarbrücken, Germany.
Email: kristin.altmeyer@uni-saarland.de

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12426.

Abstract

Multiple external representations (MERs) play an important role in the learning field of mathematics. Whereas the cognitive theory of multimedia learning and the integrative text and picture comprehension model assume that the heterogeneous combination of symbolic and analogous representations fosters learning; the design, functions, and tasks framework holds that learning benefits depend on the specific functions of MERs. The current paper describes a conceptual replication study of one of the few studies comparing single representations, heterogeneous, and homogeneous MERs in the context of mathematics learning. In a balanced incomplete block design, the participants were provided single representations (a graphic, text, or formula) or a heterogeneous (e.g., text + graphic) or homogeneous (text + formula) combination of these to solve linear system of equations problems. In accordance with previous research, performance was superior in conditions providing MERs compared to single-representation conditions. Moreover, heterogeneous MERs led to time savings over homogeneous MERs which triggered an increase in cognitive load. Contrary to previous research, text was the least fixated representation whereas the graphical representation proved to be most beneficial. With regard to practical implications, experts should be fostered through more challenging homogeneous MERs whereas novices should be supported through the accessible graphic contained in heterogeneous MERs.

KEYWORDS

eye tracking, linear systems of equations, mathematics, multimedia effect, multiple external representations

1 | INTRODUCTION

There is consensus on the importance of multiple representations for learning in the field of mathematics. Accordingly, representation competency represents one of the seven fundamental mathematical capabilities of the PISA mathematics framework (Organization for Economic Cooperation and Development [OECD], 2013) and is characterized as the ability to employ, interpret, and translate between tables, diagrams, pictures, and specific symbolic representations like equations or formulas (Niss, 2015). Although the significance of mathematical representations in education is indisputable, teachers may

A study founded by the Department of Education of the Saarland University in cooperation with the Heidelberg University of Education.

not automatically recognize the crucial role of using multiple representations of a particular topic or task in mathematics (Dreher & Kuntze, 2015). This might be due to the ambiguous and complex impact of multiple representations: On the one hand, they can facilitate understanding, whereas on the other hand, they can be too complex, therefore hampering learning.

Several theoretical approaches have been used to explain the effect of *multiple external representations* (MERs). According to the *cognitive theory of multimedia learning* (CTML; Mayer, 2005) and the *integrative text and picture comprehension* (ITPC) model (Schnotz, 2005), one can assume that combining a symbolic and an analogous representation (mainly text and picture) works well for learning and problem solving (Hu, Chen, Li, & Huang, 2019) because the learners can profit from dual processing in working memory. The *Design*, *Functions*, *Tasks* (DeFT) framework (Ainsworth, 2006), however, purposes that learning benefits heavily depend on the specific functions of MERs. In contrast to the CTML and the ITPC model, MERs combining representations of similar kinds of code (e.g., both symbolic, such as formula and text) may still foster learning and problem solving.

Ott, Brünken, Vogel, and Malone (2018) conducted one of the few studies that has attempted to shed light on this controversy by systematically investigating the effects of and the mechanisms by which heterogeneous (different codes) and homogeneous (similar code) multiple representations are processed. The authors found evidence for the fostering effect of multiple symbolic representations (formula plus text) compared with single representations (formula or text). This effect on performance was comparable with the also-investigated common multimedia effect (text plus graphic). A subsequent eye-tracking study revealed differences regarding the use of the included representations in the problem-solving process, supporting the assumption that different functions are met by different combinations of representations.

The major purpose of the present study was to conduct a conceptual replication study to review Ott et al.'s (2018) results and thus evaluate the explanatory power of their assumptions.

1.1 | Theoretical and empirical background

Schnotz and Bannert (2003) defined all types of representations that are composed of symbols as *descriptive representations*. Examples of descriptions are texts and mathematical formulas, which are both symbolic despite using different symbol systems. Depictive representations, such as pictures, are analogous to the real-world phenomenon as they include iconic signs and often express contextual relations without the use of symbols.

To explain mathematical issues comprehensibly, combining representations instead of providing single representations to the learners has proved to be successful, as mentally transforming one representation into another is “at the heart of mathematical activity” (Duval, 2006, p. 107). There is empirical evidence that learners of all fields can benefit from the availability of MERs (Debellis & Goldin, 2006; Kaput, 1987; Van Someren, Reimann, & Boshuizen, 1998). However, research also indicates that in some cases, the usage of MERs can hinder learning (e.g., Ainsworth, 2006; English & Halford, 1995; Yerushalmy, 1991). There are two main factors that might be accountable for positive learning effects from MERs: learner characteristics, such as their prior knowledge, and the characteristics of the provided representations. Concerning the latter, several theoretical frameworks refer to the impact of combining different kinds of representations for learning and problem solving.

The CTML (Mayer, 2005), which explicitly refers to text–picture combinations, is primarily built on the multi-store memory model (Atkinson & Shiffrin, 1971), Paivio's (1986) dual-coding theory, Wittrock's (1974) view of learning as a generative process, the selection–organization–integration model (Mayer, 1996), and Baddeley's (1992) assumptions about the dual-channel working memory model. Mayer (2001, 2005, 2009) adopts Paivio's view of a separation between a verbal and a nonverbal cognitive system, both storing and processing information. According to the CTML, the separate channels are limited in processing capacity. During the multimedia learning process, the learner first selects relevant words and visual content; information is then organized into a coherent verbal and a coherent nonverbal model located in the working memory. Finally, textual and pictorial information are integrated and linked to prior knowledge. The main premise of the CTML is expressed through the multimedia effect, which describes the finding that learners benefit more from a combination of words and pictures than from words alone. This effect has been found to be most pronounced for novice learners (Butcher, 2014; Mayer, 2001). MERs that are composed of text and picture are processed in both channels of working memory, preventing overloading of one of the channels via dual coding and leading to more available and sophisticated integrated mental models than single representations.

The integrated model of text and picture comprehension (ITPC; Schnotz, 2005, 2014; Schnotz & Bannert, 2003) rests on the same theoretical basis as the CTML that verbal and nonverbal information are processed in two different channels with limited capacity. First, all incoming information is assumed to be processed on a perceptual level, and thereafter, on a cognitive level in the verbal and/or pictorial working memory channel. In contrast to the CTML, the authors emphasize that information deriving from the two channels is aligned into one coherent model from the very beginning of the multimedia learning process. Furthermore, the ITPC model refers to combinations of all kinds of descriptive and depictive representations (Horz & Schnotz, 2008).

A recent meta-analytic review on the multimedia effect in problem solving (Hu et al., 2019) investigated how the addition of illustrations to text-based problem tasks affects performance. The authors assumed that—analogous to the traditional multimedia effect in learning—pictures added to a problem-solving task would foster performance. As the authors expected mental model construction in problem solving to involve similar processes as in learning tasks, the theoretical basis for the multimedia effect in problem solving was deduced from multimedia learning theories. To give consideration to the specific requirements of problem-solving tasks, the different phases of problem solving (problem design, problem comprehension, and problem solution) were incorporated into the theoretical models on multimedia learning to explain the multimedia effect in problem solving. The addition of pictures was assumed to be particularly helpful in the solution phase, as they provide information in an easily accessible way, which facilitates recall and offloads working memory. Moreover, Lindner, Eitel, Strobel, and Köller (2017) suggested that illustrations in test items might support the students in making sense of ambiguous information in text-based problem descriptions.

Across various academic disciplines, pictures in problem-solving items were found to enhance performance, confirming a multimedia effect for problem solving. The authors of the meta-analysis detected a small to medium overall multimedia effect on accuracy (Hedges' $g = 0.25$) and a medium-sized effect on answer certainty (Hedges' $g = 0.48$), but no significant general effect was found on response times.

A limitation of the CTML and the ITPC model in explaining the effect of MERs in general is that they mainly focus on the combination of verbal and pictorial representations, described as the classical multimedia view. This view does not make any assumptions regarding combinations of representations that are based on the same coding system and are likely to be processed in the same working memory channel. Ott et al. (2018) address this by drawing a distinction between *homogeneous* and *heterogeneous MERs*. Homogeneous MERs are composed of either exclusively symbolic or exclusively analogous representations, whereas heterogeneous MERs combine symbolic and analogous representations. Homogeneous MERs are frequently used in mathematics education (e.g., a formula with corresponding verbalization). Furthermore, the included representations can be either informationally equivalent or contain different information. According to Larkin and Simon (1987), equivalent information is given if all of the information contained in one representation is also inferable from the other, and vice versa. In contrast, functionally equivalent representations can indeed be used to solve the same task but do not necessarily have to convey equivalent information. Regarding such equivalent information, the redundancy principle (Chandler & Sweller, 1996; Kalyuga & Sweller, 2014) states that learning material containing the same information may increase working memory load and thus obstruct learning (Kalyuga, Chandler, & Sweller, 2004). However, representations without any additional content do not always distract learners. Empirical evidence indicates that at least novice learners profited from text with informationally redundant illustrations (e.g., Mayer, 2009).

Compared with the abovementioned theories, Ainsworth's DeFT framework (Ainsworth, 2006) holds a broader view of learning with MERs. Ainsworth not only describes MERs through their design, such as the used representational codes, but also emphasizes their specific functions during learning and problem solving. Three basic functions of MERs are distinguished, which can be fulfilled at the same time (Ainsworth, Wood, & O'Malley, 1998). Which function or functions are actually fulfilled depends on the specific learning context and the learning goals. According to the first function, external representations can complement each other either by providing complementary information or by allowing for complementary approaches to processing information. The second function is that simultaneously presented representations can constrain each other's interpretation in two ways: the more familiar representations can constrain the interpretation of the less familiar one, or inherent properties of one representation can trigger the usage of the other representation. With the third function, Ainsworth (2006) assumes that MERs can lead to a deeper understanding by allowing integration of corresponding information of the provided representations, which is consistent with the CTML and the ITPC model.

In summary, the CTML, the ITPC model, and Ainsworth's functional approach lead to consistent as well as (at least partly) contradictory predictions. They are consistent in the assumption that heterogeneous MERs can have a positive effect on learning if they fulfill certain functions. Many studies confirm this common multimedia effect in general (e.g., Levin, Anglin, & Carney, 1987), whereas Ainsworth's approach, albeit theoretically compelling, has not yet been empirically well researched. Regarding homogeneous MERs, the CTML and the ITPC model would lead to the assumption that learners process information in one single channel, which would not foster learning compared with a single representation. Furthermore, informationally equivalent, and thereby redundant, homogeneous MERs could even hinder learning processes (Chandler & Sweller, 1996; Leahy, Chandler, & Sweller, 2003). In contrast, the DeFT framework holds the view that the positive effects of leaning with MERs are not dependent only on representational design but rather on their specific functions for learning and problem solving. Consequently, a multimedia effect could also emerge when homogeneous MERs are used. There is also some evidence for this point of view (Ainsworth, Bibby, & Wood, 2002; Rau, Aleven, & Rummel, 2013).

Building on these findings, Ott et al. (2018) aimed to investigate whether homogeneous and heterogeneous MERs have a positive impact on problem solving in propositional logic tasks and whether the multimedia effect is accentuated for one of the two kinds of MERs. The authors conducted a first experiment with six conditions. Two groups of participants worked with single representations (either text or formula). Three other groups were provided with dual representations, which were either homogeneous MERs (formula and text) or heterogeneous MERs (formula and graphic or text and graphic). The last group could make use of all three types of representations (text, formula, and graphic). Results indicated that all types of MERs were more helpful than single representations, as long as the text representation was included. Consequently, the condition using a combination of formula and graphic did not exceed the single text or formula conditions. Text plus formula (homogeneous MER) was as useful for problem solving as text plus graphic (heterogeneous MER). As it was assumed that, despite being equally helpful, the two kinds of MERs might foster different processing strategies, a subsequent within-subjects eye-tracking study was conducted to compare gaze behavior when working with a successful homogeneous MER (text and formula) and a successful heterogeneous MER (text and graphic). Gaze data revealed that text was attended most in both kinds of representations, and it was therefore considered the reference representation fulfilling a constraining function. More gaze switches were performed across text and graphic than across text and formula, which was assumed to reflect more attempts to integrate information across heterogeneous MERs than across homogeneous MERs.

The authors pointed out several methodical limitations of their study, including the lack of a prior knowledge test and the lack of time on task measurement in the first experiment. In addition, gaze data were only investigated for two out of six possible conditions. Further limitations concerned the applied material on propositional logic tasks, in particular the graphical representation. In contrast to text and formulas, the graphics were not functionally equivalent to the other representations and could therefore not be used in a single representation condition. Generally speaking, the graphics used were not as informationally rich as those in usual multimedia learning studies, which might have limited the effect of the heterogeneous MERs used in the experiments of Ott et al. (2018).

1.2 | Purpose of the study

The main objective of the present research was to conduct a conceptual replication study (Earp & Trafimow, 2015) of Ott et al. (2018) in order to confirm the newly established homogeneous multimedia effect of multiple symbolic representations, and beyond that, to extend Ott et al.'s investigation about heterogeneous multimedia effects by using functionally equivalent graphics.

In addition to eliminating the original investigation's concrete methodical limitations, this study evokes an urgent need for a replication and validation of the results for several reasons that are based on the current state of the research. First, the empirical evidence concerning some of Ott et al.'s results is inconsistent. For instance, there are mixed findings in respect to the effect of heterogeneous MERs. Furthermore, there is little research in general regarding the single or combined presentation of text and formula, even though these combinations are frequently used in mathematics.

The present study made use of altered graphical material containing richer spatial information. Every representation was functionally equivalent to the others and did not contain any mixtures of representational codes (e.g., text appearing in a graphic). To eliminate further limitations, gaze behavior was recorded for all conditions.

To realize these adjustments, the particular topic within the field of mathematics was shifted from propositional logic to formal algebra, or to be more specific, to linear systems of equations, which were represented as formula, text, and graphic (see Figure 1).

Equations containing terms and variables are composed of elements of what Vollrath (2003) calls a formula language. Like the processing of language expressions, the understanding of algebra is highly based on symbolic abstraction (Arcavi, 2005). An essential part of solving equations is to transform one mathematical representation into another, usually less complex mathematical representation, which is equivalent by following syntactical and semantical algebraic rules. This process is connected with element interactivity, which determines the extent to which information imposes cognitive load due either to its intrinsic characteristics or to the instructional design used (Sweller & Chandler, 1994; Sweller, 2010). When solving equations, people are assumed to be confronted with different levels of element interactivity: Novices have to consider each symbol of the equation (variables, parameters, and relating symbols such as the multiplication or equal sign) as well as the relation of each symbol to at least one other symbol, and further, they have to balance these elements of symbols and relations simultaneously in working memory. By contrast, experts are expected to be confronted with fewer elements of interactivity because their knowledge about the symbol–relation–organization holding in long-term memory comes into play (Chen, Kalyuga, & Sweller, 2016).

Despite its high requirements regarding symbolic abstraction and working memory capacity, teaching algebra is of enormous importance in mathematics education, as solving equations becomes important in every subfield of the mathematical domain ranging from analysis, algebra, number theory, graph theory, and stochastics to logic and set theory. The importance of equation solving is underlined by the fact that the national standards in many countries point out the significance of dealing with it using symbolic and graphical representations (e.g., for Germany: Kultusministerkonferenz, 2003).

However, whereas the importance of using multiple representations in solving equations seems undoubted, the research about transformation processes within and across different types of algebraic representations concerning the psychology of information processing is still sparse. The application of this mathematical domain offers the possibility to discover whether Ott et al.'s (2018) results can be generalized to the field of equation solving.

1.3 | Hypotheses

The present study combined Ott et al.'s (2018) hypotheses and results from both of their experiments. The first research question referred to the replication of the fostering effect of multiple representations compared with single representations. Although the formula–graphic version of the former material has not been found to be more helpful than the single representations, this combination was expected to lead to a multimedia effect in the present study, as the actual graphics conveyed more essential information.

Hypothesis 1 *Performance on the items on linear systems of equations improves if multiple, instead of single, representations are presented to the participants.*

The second research question aimed at comparing the effects of heterogeneous and homogeneous MERs on the use of the new material.

Hypothesis 2 *Performance in heterogeneous MERs is expected to be higher than in homogeneous MERs.*

Hypotheses 3 concerns variables that are expected to mediate and thus explain the effect of the kind of MER on performance. Subsequently, this hypothesis is differentiated in Hypotheses 3a through 3c.

The first mediator was cognitive load. Due to the possibility of processing information in two channels instead of only one, heterogeneous MERs were expected to spare working memory resources.

Hypothesis 3a *Cognitive load is assumed to be reduced in heterogeneous compared with homogeneous MERs, which is expected to result in higher performance in heterogeneous MERs.*

A second mediator is the number of transitions (direct gaze switches) across two representations. Though the combination of text and formula led to fewer transitions than the combination of text and graphic, Ott et al. (2018) found both homogeneous and heterogeneous MERs to be equally beneficial for problem solving. Still, due to the informationally enhanced graphics, it was assumed for the current study that the learners will profit more from stronger integration of heterogeneous than of homogeneous MERs.

Hypothesis 3b *Attempts to integrate information across representations are assumed to be particularly triggered in heterogeneous MERs, resulting in higher performance.*

The subsequent hypothesis concerns the relationship between the two mentioned mediators, assuming a serial mediation.

Hypothesis 3c *As heterogeneous representations are expected to spare cognitive resources (cognitive load, Mediator 1), the additionally available cognitive capacity may be used for integration of information across representations (transitions, Mediator 2). This is again assumed to result in better performance for heterogeneous than homogeneous MERs.*

According to the results of Ott et al. (2018), the text representation was the reference representation, as it was attended the most. The present study was expected to replicate the findings and, as eye-tracking was applied for all conditions, additionally identify whether there is a reference representation in formula–graphic MERs.

Hypothesis 4 regards the assumed role of text as the reference representation. Subsequently, this hypothesis is differentiated in Hypotheses 4a and 4b.

Hypothesis 4a *Text is expected to be attended to more often than formulas or graphics, reflected in higher mean fixation counts and visit counts for text.*

Hypothesis 4b *Text is expected to be attended to longer than formulas or graphics, reflected in longer mean total fixation durations and mean visit durations.*

2 | METHOD

The present study was preregistered before the start of the data collection at the Open Science Framework. Registered report protocol, raw data, materials, and laboratory log are available at https://osf.io/4x5qg/.

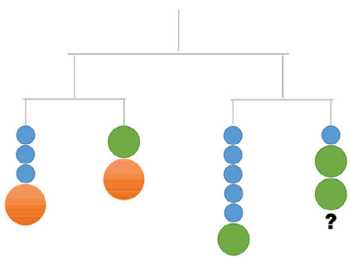
2.1 | Pilot study on material

A pilot study was conducted to pretest the developed material regarding its adequacy for the current research questions. The material included 13 computer-based items dealing with linear systems of equations. Every item was composed of three equivalent representations of the same linear system of equations: a formula, a graphic (a picture of a mobile), and a text expression (see Figure 1). Each system of equations contained three variables (*k*, *m*, and *g*), which were defined as being equivalent to a small (*k*), a medium (*m*), and a large (*g*) mobile ornament in the graphic representation, and a verbal term for each (e.g., "small disc") in the text representation. Each line of a

(1) $3k + g = m + g$

(2) $5k + m = k + 2m + ?$

(3) $3k + g + m + g = 5k + m + k + 2m + ?$



(1) A branch with three small and one large disc is as heavy as a branch with one medium and one large disc.
(2) A branch with five small and one medium disc is as heavy as a branch with one small, two medium and one **? disc.**
(3) A branch with three small and one large disc as well as one medium and one large disc is as heavy as a branch with five small and one medium disc as well as one small, two medium and one **? disc.**

☐ k or one small disc or ● ☐ m or one medium disc or ● ☐ g or one large disc or ●

FIGURE 1 Example of the math problems used in the pilot study [Color figure can be viewed at wileyonlinelibrary.com] system of equations was equivalent to a comparison of two balanced branches of the mobile and with a specific line of the text representation. One element of every system of equations was unknown and marked by a question mark in each of the three representations. Participants were instructed to find the missing element (one small, medium, or large disc) and choose their answer from the options at the bottom of each item slide. Additionally, they were provided an example of a solved item and a note that all discs are of the same thickness and that the branches of the graphic are defined as weightless.

Having solved the last item, participants were asked to rate to what extent they had used each type of representation, how much they had tried to integrate different representations, and the overall difficulty of the items on a 5-point Likert scale. Thereafter, they answered paper-and-pencil-based demographic questions and completed a prior knowledge test (three classical linear system of equations tasks, taken from an eighth-grade German mathematic school book). The prior knowledge test was not provided before the newly developed multi-representational material because by providing only formulas, it might trigger the use of formula-based approaches, which are typically used in school classes, to solve the items.

The sample size was $N = 63$. Most participants were German pupils and students (mean age: 18.17 years, $SD = 4.83$; 58.7% female). The participants were randomly assigned to one of two groups, differing only in the set of tasks they were provided. Both groups dealt with a set of 7 out of 13 items, and one item was identical across groups. For the first task set, reliability analysis showed moderate internal consistency ($\alpha = .64$), and discrimination indices were between .28 and .45. For the second task set, internal consistency was $\alpha = .44$, and discrimination indices reached from −.13 to .49. After excluding the three items whose discrimination indices were smaller than .1 from the task sets, internal consistencies were $\alpha = .57$ and $\alpha = .69$. As a result, 10 items were deemed appropriate for future research.

2.2 | Experimental design and moderator variables of the main study

To investigate the influence of different representations and their combinations on problem solving and the underlying cognitive processes, a one-factor within-subjects design was applied. The within-subjects factor was task presentation. Each individual was confronted with different conditions regarding the single representations or sets of representations they could use to solve the linear equation systems. Seven conditions were realized: three single representations (graphic, text, and formula), three possible combinations of two different representations, and one combining all three representations (see Table 1). In contrast to Ott et al.'s (2018) material, the graphical representation (mobile) was self-contained, intelligible, and functionally equivalent to the text and formula, which was given here by equation, and could therefore be used as a single representation condition.

Each individual worked on 10 items, which were presented in five of the seven conditions with two items per condition. The participants were not provided with all seven conditions because only 10 items had proved appropriate in the pilot study. Moreover, the pilot study had revealed that even doing only seven items had wearied some of the participants, implying that solving more than 10 equation systems would certainly be too many for most of the individuals.

It was ensured that two conditions were missing completely, at random, for each participant by applying a balanced incomplete block design. The design was incomplete because only five out of seven conditions are included in each block. According to all possibilities to draw five out of seven, the number of possible blocks for this design was 21. The design was balanced as each of the seven conditions appeared in exactly 15 blocks, and each pair of conditions occurred together in 10 blocks. For each participant, one block was drawn at random, without replacement. Further, the 10 items were assigned randomly to the conditions for every block.

Aside from the performance indicators mentioned in the hypotheses, three possible moderator variables were also measured, as they were considered to affect the mastering of the multi-representational tasks in this experiment. The first was prior knowledge, because it was assumed that performance in the newly developed items and the use of the different kinds of representations might be related to prior skills in solving linear equations systems the conventional way (formula-based approaches). In addition, spatial abilities were measured as they might affect the benefit of the provided visualizations and the respective textual representations (e.g., Kühl, Stebner, Navratil, Fehringer, & Münzer, 2018). As a third possible moderator variable, verbal memory was assessed as this skill might affect how the participants could make use of the provided text representations.

2.3 | Sample size and sample characteristics

The sample size of the current replication study was determined by power analysis, based on the smallest relevant effect size of the original study (Ott et al., 2018). This effect size ($d_{Cohen} = .70$) resulted from a Duncan post hoc test comparing the mean performance of single text and equation–text group. Alpha error probability of the current study was defined as $\alpha = .05$ and the power ($1 - \beta$) was aimed to be at 90%. The statistical one-tailed t test used in the power analysis considers the difference between two dependent means. On this basis, a minimum sample size of 19 participants was computed using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). Because every other effect to be replicated is based on higher original effect sizes (Ott et al., 2018), corresponding analyses are expected to be properly powered as well.

Taking into account that each pair of conditions occurred together in 10 out of 21 blocks, it was necessary to double the number of blocks in order to ensure the minimum sample size of $N = 19$ for comparisons between conditions. Consequently, 42 participants were randomly assigned to the resulting 42 blocks.

Participants mainly were German university students (95.24%), balanced by gender (50% female; age: $M = 23.78$ years; $SD = 3.49$). To avoid excess expertise, students whose courses of study were associated with the content of systems of equations (e.g., mathematics)

TABLE 1 Overview of the seven conditions of the within-subjects factor task presentation (based on Ott et al., 2018)

Condition	Representations	Quantity	Coding	Type of MER
1	Graphic	Single	Analog	
2	Equation	Single	Symbolic	
3	Text	Single	Symbolic	
4	Equation + text	Multiple	Symbolic	Homogeneous
5	Equation + graphic	Multiple	Symbolic + analog	Heterogeneous
6	Text + graphic	Multiple	Symbolic + analog	Heterogeneous
7	Equation + text + graphic	Multiple	Symbolic + analog	Heterogeneous

Abbreviation: MER, multiple external representation.

were not invited to participate. Nevertheless, basic prior knowledge in linear systems of equations (school lessons) and fluent German were prerequisites. A participant was considered an outlier and therefore was excluded from the data analysis if he or she performed more than two *SD*s above or below average in the prior knowledge test, task performance, or time on task. Missing data due to early termination of the experiment (e.g., because of technical errors or sickness) also led to the exclusion of subject data. A further reason for data exclusion was eye-tracking quality insufficient for valid analysis. Consequently, participants with more than 25% missing values in their gaze data due to factors such as eye makeup, thick lash lines, special glasses, further special features of a participant's eye area, or too many head turns, were excluded. Twenty one whole data sets had to be excluded because of the abovementioned criteria and were replaced by additional participants to reach the aimed sample size of 42 participants.

2.4 | Procedure and materials

Participants were randomly assigned to one of the 42 blocks. The experiment was conducted under the same laboratory conditions as the original study (Ott et al., 2018). First, the participants read general information about the study and data protection and provided written informed consent. The examiner gave a verbal task introduction and performed a nine-point grid calibration of the eye tracker. After that, participants read an on-screen instruction and then were shown a solved example item. Finally, the 10 items (selected from the pilot study) were presented in a random order and represented according to the study design in terms of type and number of the provided representations. The placing of single, dual, and triple representations varied randomly between three possible positions (see Figure 2). Analogous to the pilot study, subjects chose their answer from three options at the bottom of each item slide with a mouse click. Every item was followed by two questions measuring item difficulty and mental effort using the subjective rating scale by Paas (1992). After having solved the last item, the participants responded to the questions about usage and integration of representations as well as overall difficulty, the paper–pencil questionnaire on demographic data (sex, age, education, and school grades), and, analogous to the pilot study, the prior knowledge test. Furthermore, the paper-folding and card-rotation test to measure spatial skills (Ekstrom, French, Harmann, & Dermen, 1976) and the subtest "Meaningful Text" for verbal memory of the Berlin Intelligence Structure Test (BIS; Jäger, Süß, & Beauducel, 1997) were completed by the participants. Thereafter, the subjects received a monetary reward.

In line with Ott et al. (2018), eye-tracking was conducted on a 60-Hz-remote eye-tracking system (Tobii X2-60; averaging 0.4° accuracy) attached to a laptop (Dell Precision M6800 Mobile Workstation) with a 17.3″ screen. The laptop was placed about 65 cm (vertical gaze angle <31°) from the edge of the table. To make certain that the distance between the participant and the eye tracker did not change drastically, the examiner instructed the participants to keep their heads behind the edge of the table. Tobii Studio was used to run the experiment and analyze the eye-tracking metrics.

As in Ott et al. (2018), gaze behavior was analyzed via several measures using binocular data (averaged position of both eyes). These eye-tracking metrics were related to fixations (>100 ms) on one of the displayed representations or to transitions across representations. For this purpose, a rectangular fixed-size (800 × 283 pixels for a screen solution of 1920 × 1080) area of interest (AOI) was created for every provided representation using Tobii Studio eye-tracking software (see Figure 2). Fixation detection was performed based on the velocity-threshold identification algorithm (for example, Salvucci & Goldberg, 2000) operating with changes in the velocity of saccadic eye movements. Velocities below the velocity-threshold value of 30° per second indicate fixations, velocities above indicate saccades.

In the interest of replicating Ott et al.'s methods as exactly as possible, the same seven parameters (see Table 2) were calculated to analyze the participants' gaze behavior, related to the defined AOIs. As analyzing too many metrics in eye-tracking is regarded as a possible threat to the validity of the results (Orquin & Holmqvist, 2018), each metric is chosen carefully regarding its unique contribution to not only test the hypotheses but also help to validate the findings.

The durations of fixations were measured for each AOI. In general, the length of fixations can either reflect how relevant the respective stimulus is for the task to be performed (Orquin & Mueller Loose, 2013) or reflect how difficult it is to process (for math problems: Hegarty & Just, 1993; for text reading: Rayner, 1998). However, due to the design of the present study, the two possible interpretations can be separated. If single representations are given, longer fixation durations might speak for higher difficulty, and thus processing times, because the participants will have no choice but to work with the only

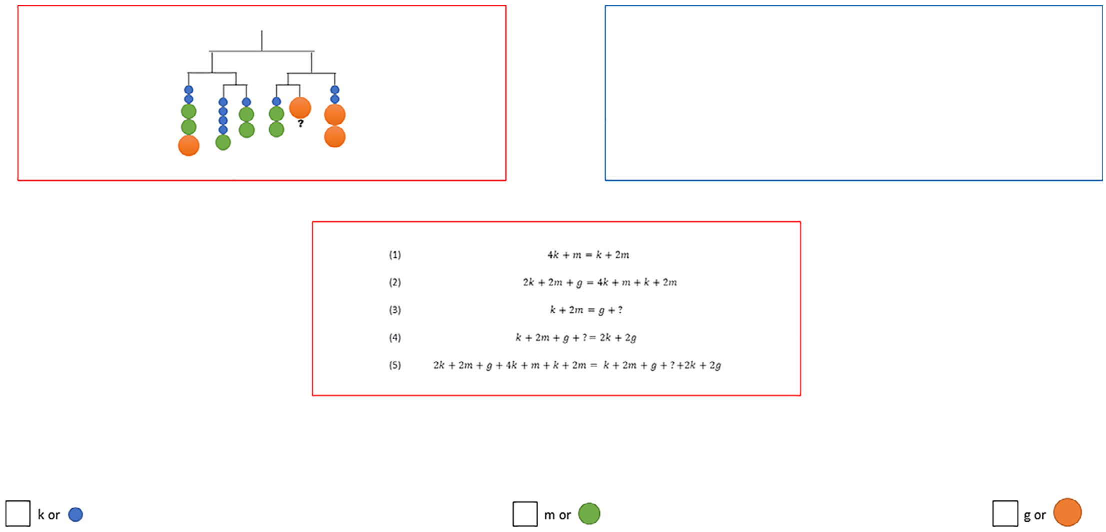


FIGURE 2 Example of an item of the heterogeneous equation–graphic condition, including visible areas of interest (AOIs). Red frames (left and lower one) mark actual AOIs for the two representations. The blue frame (on the upper right) indicates the third possible position for representations, randomly remaining empty for this item. During the experiment, AOIs were invisible for subjects [Color figure can be viewed at wileyonlinelibrary.com]

provided representation. In contrast, with multiple (informationally equivalent) representations, longer fixations on one representation compared with the others rather speak to its relevance to task performance: It can be assumed that the representation, which is fixated upon the longest, is the one primarily used to complete the task and can therefore be regarded as reference representation.

Unless being partly redundant, the durations of visits on the AOIs were also recorded. This seems important for the present kind of instructional material because the AOIs embraced whole representations, each entailing multiple stimuli, such as letters, operators, or mobile discs, that are considered high in element interactivity. Whereas the analysis of single fixations relates to the participants' dealing with these single stimuli, visit durations might reflect the participants' efforts to integrate several stimuli within one representation.

The number of fixations is considered another useful metric; more fixations are associated with more attention dedicated to a particular AOI, which classifies this area as either particularly visually salient, difficult, or relevant to the task (Orquin & Mueller Loose, 2013). Similar to their length, the number of fixations can be precisely interpreted due to the study design. In comparing the single representation conditions, more fixations might speak for higher difficulty, whereas in multiple representations, more fixations on one representation compared with the other(s) rather speak to its relevance to task performance or to its salience. No fixation on an AOI means that a stimulus is completely ignored for the task solution.

As a distinct measure of stimulus salience, the time to first fixation was recorded as it allows one to identify the one representation that attracts the students' attention first. As the location of the different types of representations varied randomly across the trials, early fixations were expected to reflect the salience of a stimulus. The representation, which holds fixation first can be assumed to be the most salient.

TABLE 2 Eye-tracking measures

Variable name	Variable description
Fixation duration	Lengths of fixations for a specific AOI
Visit duration[a]	Time from each first fixation on a specific AOI until the next fixation outside the AOI
Fixation count	Total number of fixations on a certain AOI
Visit count	Number of visits to a certain AOI
Time to first fixation	Timespan from the appearance of an item page to the participant's first fixation on a certain AOI
Transition count	Number of direct gaze switches between two AOIs
Last fixation	Last AOI that is fixated upon per item slide before choosing an answer

Abbreviation: AOI, area of interest.
[a]A visit lasts from the first fixation on an AOI until the next fixation outside that respective AOI, which means that a visit can involve multiple fixations on the same AOI.

The number of visits to a specific AOI was also measured, as more visits are related to frequent revisits of a stimulus (having inspected others in the meantime), which speaks to the relevance of the respective stimulus (Orquin & Mueller Loose, 2013) in multiple-representations' conditions.

Gaze switches from one AOI to another—often referred to as *transitions*—were counted for each pair of representations provided to the participants in the multiple-representations conditions. More transitions are assumed to reflect more intense actual (Andrá et al., 2015; Duval, 2006) or attempted (Mason, Tornatora, & Pluchino, 2013) integration processes taking place between the two stimuli.

The variable *last fixation* describes the last AOI a subject fixated upon per item slide. Immediately after their last fixation, subjects chose their answer via mouse click.

Previous research has indicated that measures for fixation time can also predict cognitive load and cognitive processing (Holmqvist et al., 2011; Korbach, Brünken, & Park, 2016). Therefore, they can also be used to validate the subjective cognitive load scale (Paas, 1992).

3 | RESULTS

First, some particulars concerning the following analyses will be pointed out.

The analysis plan of the current preregistered study can be found on the Open Science Framework (https://osf.io/tgf29/). Some analyses were slightly adjusted to fit the data structure but without any modification of their purpose.

Due to the balanced incomplete block design applied, data were not structured by participants as usual but by block and condition (seven alternative representations of equation-solving problems; see Table 1). The variable *block* was included as an additional factor in the multivariate and univariate analyses of variances. This variable merely takes into account that subjects were assigned to a specific combination of conditions and is of no predictive value itself. Interactions between block and other dependent variables were excluded from the models under the assumption that they did not contain information relevant to the present study.

Analyzing heterogeneous MERs, Conditions 5, 6, and 7 were included because they all contain symbolic as well as analogous representations (see Table 1). Homogeneous MERs are represented by Condition 4.

When focusing on transitions between particular representations of MERs, Condition 7 was excluded from the analyses as transitions between three representations are assumed to be influenced by different strategies of processing and holding information in working memory due to cognitive load issues. Thus, transitions between threefold representations cannot easily be compared with the two-representation conditions 4, 5, and 6. They will be discussed separately in the exploratory results section instead.

3.1 | Confirmatory analyses

Means and *SD*s for the dependent variables and mediator variables of Hypotheses 1 to 3 per condition are displayed in Appendix A.

The two-way independent MANOVA (Pillai's Trace) for Hypotheses 1 and 2 revealed a main effect for the factor block, $F(40, 366) = 2.04$, $p < .001$, $\eta_p^2 = .18$, as well as for condition, $F(12, 366) = 2.76$, $p = .001$, $\eta_p^2 = .08$, concerning the dependent variables' number of correctly solved items and time on task per item. These results indicate an overall effect of the way the systems of equations were represented on students' performance. Univariate analyses (one tailed) found an effect for condition regarding the dependent variable time on task per item, $F(6, 183) = 4.48$, $p < .001$, $\eta_p^2 = .13$, but not for number of correctly solved items, $F(6, 183) = 1.23$, $p = .146$.

3.1.1 | Hypothesis 1

The planned contrasts (one tailed) comparing single versus multiple task representations revealed that subjects solved more items correctly when presented with multiple than single representations, $F(1, 183) = 3.85$, $p = .026$, $\eta_p^2 = .02$, but there was no difference for time on task per item, $F(1, 183) = .41$, $p = .261$.

3.1.2 | Hypothesis 2

Contrast analyses (one tailed) that compared homogeneous and heterogeneous MERs showed that subjects performed significantly faster when they were provided heterogeneous MERs, $F(1, 183) = 6.35$ $p = .007$, $\eta_p^2 = .03$. There was no difference concerning the number of correctly solved items, $F(1, 183) = 1.58$, $p = .105$.

3.1.3 | Hypothesis 3

To answer Hypothesis 3, mediation analyses were conducted using PROCESS v3.3 (Hayes, 2017) by calculating 5,000 bootstrap resamples and testing for one-tailed confidence intervals (CIs) of 90%. A heteroscedasticity consistent standard error and covariance matrix estimator was used (HC3, Davidson-McKinnon). Regarding the subsequent path analyses that are displayed in Tables 3 and 4, homogeneous MERs were coded with 1 and heterogeneous MERs were coded with 0.

Hypothesis 3a concerns the mediator variables' mental effort and task difficulty, both indicating cognitive load. The simple mediation analysis investigating the relationship between the type of MER and performance (number of correctly solved items) mediated by mental effort revealed a significant indirect effect ($\beta = -.27$, BCa 90% CI [−.35, −.04]). Moreover, there was a corresponding significant indirect effect for the mediator task difficulty ($\beta = -.18$, BCa 90% CI [−.51, −.07]; see Table 3 for complete path characteristics). The indirect paths describe subjects reporting higher mental effort or task difficulty for homogeneous items, whereas performing worse when mental effort or task difficulty was high.

With regard to Hypothesis 3b, it was assumed that the effect of the MER type on performance would be mediated by the number of transitions across representations. However, the analysis revealed no significant indirect effect of MER type on performance (number of correctly solved items) through mean transition count ($\beta = .01$, BCa 90% CI [−.07, .11]; see Table 4 for complete path characteristics). Still, subjects showed more transitions for heterogeneous than for homogeneous MERs (see Path X on M in Table 4).

Hypothesis 3c stated that the type of MER would affect cognitive load, which in turn would have an effect on the number of transitions. No significant indirect effect of MER type on performance through the serial mediators cognitive load (indicated by mental effort) and mean transition count was found ($\beta = .00$, BCa 90% CI [−.01, .01]; see Table 4).

3.1.4 | Hypothesis 4

Hypothesis 4, containing the two sub-hypotheses Hypothesis 4a and 4b, considers Ott et al.'s (2018) finding that text serves as a reference representation in MERs (Ott et al., 2018). To investigate and compare the role of particular representations within MERs, gaze behavior was analyzed for the multiple represented conditions. The α-level for the main effects of the multiple ANOVAs concerning Hypotheses 4a and 4b was adjusted by Bonferroni correction (α = .025). As the subsequent corresponding contrasts are pairwise dependent, the α-level was adjusted by Bonferroni correction as well (α = .017). See Appendix B for descriptive data on eye-tracking measures.

A four-way independent MANOVA (Pillai's Trace) including the dependent eye-tracking variables' fixation count, visit count, total fixation duration, and visit duration revealed a significant effect for type of representation, $F(8, 502)$ 11.36, $p < .001$, $\eta_p^2 = .15$, and no differences for block, $F(56, 1{,}012) = 1.28$, $p = .082$.

With respect to Hypothesis 4a, a two-way independent ANOVA investigating the dependent variable fixation count revealed significant differences for the type of representation (text vs. graphic vs. equation), $F(2, 253) = 25.55$, $p < .001$, $\eta_p^2 = .17$, and no effect for the factor block, $F(14, 253) = .84$, $p = .631$. The planned contrasts comparing text versus graphic versus equation showed that fixation count for text was significantly lower than for graphic, $F(1, 253) = 46.03$, $p < .001$, $\eta_p^2 = .15$, and equation, $F(1, 253) = 28.52$, $p < .001$, $\eta_p^2 = .10$. Graphic and equation did not differ significantly in fixation count, $F(1, 253) = 2.09$, $p = .150$.

A second two-way ANOVA tested for differences in visit counts. There was a main effect for the type of representation, $F(2, 253) = 12.87$, $p < .001$, $\eta_p^2 = .09$, but no effect for block, $F(14, 253) = .50$, $p = .935$. The subsequent contrast analyses found a significantly lower visit count for text than for graphic, $F(1, 253) = 25.53$, $p < .001$, $\eta_p^2 = .09$, and, due to the Bonferroni adjustment, no difference for text and equation, $F(1, 253) = 4.51$, $p < .035$, $\eta_p^2 = .02$. In addition, visit count for graphic was higher than for equation, $F(1, 253) = 8.58$, $p = .004$, $\eta_p^2 = .03$.

With regard to Hypothesis 4b, a two-way ANOVA with the independent variable total fixation duration showed a main effect on the type of representation, $F(2, 253) = 31.42$, $p < .001$, $\eta_p^2 = .20$, and none for block, $F(14, 253) = 1.81$, $p = .037$. Planned contrasts found a significantly shorter total fixation duration for text compared with graphic, $F(1, 253) = 55.30$, $p < .001$, $\eta_p^2 = .18$, and to equation, $F(1, 253) = 37.17$, $p < .001$, $\eta_p^2 = .13$. No significant differences were found between graphic and equation, $F(1, 253) = 1.80$, $p = .181$.

A second two-way ANOVA concerned the dependent variable mean visit duration. There was a main effect for the type of representation, $F(2, 253) = 27.34$, $p < .001$, $\eta_p^2 = .18$, and none for block, $F(14, 253) = .85$, $p = .613$. Analysis of planned contrasts showed a significantly shorter mean visit duration for text than for graphic, $F(1, 253) = 49.12$, $p < .001$, $\eta_p^2 = .16$, as well as for equation, $F(1, 253) = 30.75$, $p < .001$, $\eta_p^2 = .11$. There were no differences in mean visit duration between equation and graphic, $F(1, 253) = 2.14$, $p = .145$.

TABLE 3 Path characteristics of the mediation models stated in Hypothesis 3a

	Mediator mental effort			Mediator task difficulty		
Path	*β*	*t*	*p*	*β*	*t*	*p*
X on Y	.46	2.19	.031	.55	2.88	.005
X on M	.44	2.10	.038	.53	2.36	.020
M on Y	−.40	−4.80	<.001	−.52	−6.10	<.001

TABLE 4 Path characteristics of the mediation models of Hypotheses 3b and 3c

	Hypothesis 3b			Hypothesis 3c		
Path	*β*	*t*	*p*	*β*	*t*	*p*
X on Y	.26	1.10	.274	.42	1.83	.070
X on M(1)	−.47	−2.55	.013	.40	1.79	.078
M(2) on Y	−.03	−.23	.818	.00	.05	.964
M1 on M2				.08	.69	.491

3.1.5 | Moderator variables

As regression analyses showed that the relationship between the type of MER (homogeneous vs. heterogeneous) and the number of correctly solved items was moderated by verbal memory (see Appendix C for the moderation model characteristics). Only when their verbal memory was high did subjects perform better in homogeneous MERs. In contrast, the variables' prior knowledge and spatial ability had no moderating function (see Appendix C). Nevertheless, prior knowledge positively correlates with the number of correctly solved items ($r = .29$, $p < .001$) and negatively with perceived item difficulty ($r = -.16$, $p = .020$).

3.2 | Exploratory analyses

3.2.1 | Further findings on performance

Bonferroni-adjusted pairwise post hoc tests (analogous to Ott et al., 2018) comparing each pair of the seven representational conditions (for respective MANOVA, see Hypotheses 1 and 2) revealed that neither one of the heterogeneous nor one of the homogeneous condition differed from a specific single-representation condition concerning the number of correctly solved items. Regarding time on task per item, the subjects performed faster when provided the single graphical condition than the single text condition ($p = .006$) and also faster than the homogeneous condition consisting of text and equation ($p = .002$). Furthermore, time on task per item for homogeneous MERs was higher than for the heterogeneous combination of equation and graphic ($p = .042$).

3.2.2 | Further findings on gaze behavior

Descriptive gaze data for the particular representations within the MERs are displayed in Appendix B.

Taking a closer look at the eye-tracking parameters (analogous to Ott et al., 2018; see Table 2), an exploratory MANOVA comparing text and equation within the homogeneous MERs (*Condition 4*) revealed significant differences, $F(6, 39) = 6.54$, $p < .001$, $\eta_p^2 = .50$. There was no effect for block, $F(84, 264) = 1.06$, $p = .362$. Subsequent multiple ANOVAs (Bonferroni-adjusted $\alpha = .008$; see Table 5 for numerical results) showed a significantly higher fixation count and marginally significantly higher visit count as well as a longer total fixation duration and visit duration for the equation representation. There was no difference for time to first fixation between text and equation. There was also no difference in the sum of the last fixations between the representations.

The explorative MANOVA comparing eye-tracking measures between the particular representations within the equation-and-graphic MERs (*Condition 5*) found no differences, $F(6, 39) = 1.54$, $p = .192$. Furthermore, there was no effect for block, $F(84, 264) = 1.00$, $p = .494$.Concerning the MERs consisting of the combination of text and graphic (*Condition 6*), an exploratory MANOVA showed differences in gaze behavior for the particular representations, $F(6, 39) = 13.05$, $p < .001$, $\eta_p^2 = .67$, and a main effect for block, $F(84, 264) = 1.42$, $p = .019$, $\eta_p^2 = .31$. Univariate comparisons (Bonferroni-adjusted $\alpha = .008$; see Table 5 for numerical results) revealed significantly higher fixation counts and visit counts as well as a longer visit duration and total fixation duration for the graphical representation. Time to first fixation did not differ between the representations. The number of last fixations was higher for graphic than for text.

When comparing the representations within the threefold MER combining text, graphic and equation (*Condition 7*), the exploratory MANOVA revealed differences for eye-tracking measures, $F(12, 138) = 4.74$, $p < .001$, $\eta_p^2 = .29$, and no main effect for block, $F(84, 438) = 1.19$, $p = .134$. Post hoc tests (Bonferroni) showed that the graphic was fixated upon more often than the equation ($p = .015$) and text ($p < .001$). Furthermore, the equation was fixated upon more often than text ($p = .015$). Moreover, visit count was higher for graphic than for text ($p < .001$). Concerning total fixation duration and visit duration, text was attended to a shorter extent than graphic ($p < .001$; $p = .001$) and equation ($p = .010$; $p = .006$). There were no differences between the pairs of representations in terms of time to first fixation. The number of last fixations was higher for graphic than for equation ($p < .001$) and text ($p < .001$).

The threefold MERs provided in Condition 7 allowed for three different kinds of transitions between the particular representations within the MERs: gaze switches between text and equation, between graphic and equation, and between text and graphic. To investigate whether these kinds of transitions differ in their frequency, an ANOVA was conducted. It revealed no effect for block, $F(14, 73) = 1.06$, $p = .410$, and a main effect for kind of transition, $F(2, 73) = 4.89$, $p = .010$, $\eta_p^2 = .12$. Post hoc tests showed a difference in number of transitions between graphic and text neither when compared with equation–text transitions ($p = .222$) nor when compared with graphic–equation transitions ($p = .593$). However, subjects made significantly more gaze switches between graphic and equation than between equation and text ($p = .008$).

In order evaluate the cognitive load rating scale, the relationship between cognitive load and total fixation duration was investigated. No significant correlations were found between task difficulty ($r = .125$, $p = .071$) or mental effort ($r = .112$, $p = .106$) and total fixation duration.

3.2.3 | Participants' rating

Results of the gaze behavior analysis can be supported by the questions about individual usage of the representations that every subject answered after solving all items. Focusing on the dependent variable helpfulness of the representations, rated by participants by means of a 5-point Likert scale, an ANOVA for the factors' block and type of representation was conducted. No differences were found for the factor block, $F(20, 103) = .79$, $p = .723$. There was a main effect for type of representation, $F(2, 103) = 66.85$, $p < .001$, $\eta_p^2 = .57$. As post hoc analyses showed, participants considered the graphic to be the most helpful representation ($M = 4.26$; $SD = .89$; $p < .001$ for every pairwise comparison), followed by the equations ($M = 3.19$; $SD = 1.40$; $p < .001$ for every pairwise comparison). Text was clearly rated as the least helpful representation ($M = 1.57$; $SD = .77$, $p < .001$ for every pairwise comparison).

TABLE 5 Univariate comparisons on eye-tracking measures between the particular representations within the MERs of Conditions 4 (text + equation) and 6 (text + graphic); Bonferroni-adjusted $\alpha = .008$

	Condition 4			Condition 6		
	$F(1, 44)$	p	η_p^2	$F(1, 44)$	p	η_p^2
Fixation count	23.00	<.001	.34	40.53	<.001	.48
Visit count	7.55	.009	.15	8.21	.006	.16
Fixation duration	32.78	<.001	.43	42.11	<.001	.49
Visit duration	20.06	<.001	.31	25.74	< 001	.37
Time to first fixation	0.06	.804		<.01	.998	
Last fixation	6.61	.014		15.88	< 001	.27

Abbreviation: MER, multiple external representation.

4 | DISCUSSION

As a conceptual replication of Ott et al. (2018), this study not only confirmed the general fostering effect of multiple representations compared with single representations but also emphasized, in line with Ott et al. (2018), that heterogeneous MERs did not by all counts outperform homogeneous ones. The results of the present study indicate that perceived cognitive load, depicted by item difficulty and mental effort, was lower for problem solving with heterogeneous than with homogeneous MERs; the latter required a higher investment of time but did not generally lead to losses in performance. Moreover, subjects scoring high in verbal memory could deal especially well with homogeneous MERs. Regarding the particular representations within MERs, the current study revealed that the text-based representation of linear systems of equations was inferior to the other representations for problem solving, whereas the graphical representation proved to be most beneficial. These findings do not correspond to Ott et al.'s (2018) assumption of a reference function for the representation text in MERs.

Regarding the distinct hypotheses, the current study provides confirming as well as contradicting evidence.

Hypothesis 1 of the present study concerned the potential benefit of presenting MERs instead of single representations of linear equations for problem solving. This hypothesis could be supported by the data, as the participants solved more MER items correctly than single-representation items. This finding regarding a small-sized fostering effect of MERs replicates the results of Ott et al. (2018) and matches the findings of a meta-analytic review on the general multimedia effect in problem solving (Hu et al., 2019). Another result in line with the findings of the meta-analysis is that the presentation of MERs instead of single representations did not affect time on task. According to that, problem-solving efficiency was neither reduced nor enhanced by the presentation of additional representations. In the case of adding pictures to a text-based problem description, this result could be due to a compensating effect: On the one hand, adding pictures to text increases the time to decode information, as more than one representation needs to be considered. On the other hand, pictures might hasten the decision process as they provide an easily and holistically accessible representation of the problem (Lindner et al., 2017; Saß, Schütte, & Lindner, 2017).

Relating to *Hypothesis* 2 it was assumed that heterogeneous MERs would foster problem solving better than homogeneous MERs. This hypothesis was partly confirmed as the outcomes of the present study indicated that problem solving with homogeneous instead of heterogeneous MERs required a higher investment of time to achieve equivalent performance. Consequently, this means that the problem space seems to be more easily accessible via heterogeneous representations.

The time savings for heterogeneous MERs could be due to the special benefits of the graphical representations, which were exclusively provided in heterogeneous MERs and whose direct accessibility is assumed to be especially useful during the solution phase of problem solving. Lindner et al. (2017) revealed that a picture added to a textual problem description in test items significantly reduced the time required to read the text. This result was assumed to indicate that the inclusion of pictures in test items fosters mental model construction.

Moreover and consistent with the findings of Ott et al. (2018), heterogeneous MERs triggered more gaze switches across the representations than homogeneous MERs, which can be interpreted as more attempts to integrate information in heterogeneous MERs. These outcomes support the dual-channel assumption of the CTML and the ITPC model as heterogeneous MERs allow for simultaneous processing of analogous and symbolic representations, which results in stronger integration processes and time saving compared with the processing of homogeneous MERs. Homogeneous representations would have to be processed sequentially in one single channel.

Hypothesis 3 of the current study aimed at specifying the effect of different types of MERs by the means of possible mediating variables: cognitive load and the number of transitions across the representations. The assumption that cognitive load mediates the effect of the type of MER on performance was confirmed by the present study: The homogenous representations were related to enhanced mental effort and perceived difficulty of the material, leading to losses in performance. This result confirms the hypothesis that dual-coded heterogeneous MERs spare cognitive resources. Following the CTML and the ITPC model, homogeneous MERs are expected to overload the verbal channel of working memory, and consequently, performance is affected negatively. Because the heterogeneous MERs always contain a graphical representation, the results further correspond to the findings of Yung and Paas (2015) on primary school children: The authors found higher learning performance on mathematical tasks and reduced cognitive load when the tasks were enriched with visual representations.

Replicating the results of Ott et al. (2018), heterogeneous MERs triggered more transitions across the representations than homogeneous MERs. However, the number of transitions was not confirmed to mediate the effect of MER type on performance. This could be due to the fact that gaze switches are assumed to be indicators for integration attempts, but the mere number of transitions cannot serve as an indicator for the quality of these attempts to integrate information. In addition, because all representations in this study were functionally equivalent, transitions were, in practice, not necessary and might offer no advantage for problem solving. Subjects rather chose their preferred representation (e.g., the easily accessible graphic) and then disregarded the other representations that were part of the respective MER.

Hypothesis 4 concerned the previous finding that text representations have a significant impact on performance in MERs. Ainsworth (2006) assumes that MERs are useful for learning and problem solving if they fulfill specific functions. One of these functions is the constraining function, which implies that one representation, which can be called the reference representation (Ott et al., 2018), is either easier to understand or more familiar to the learner and therefore is used to make sense of the other representation. The current study aimed to investigate whether text could be confirmed as the reference representation in another mathematical domain. The results contradicted Ott et al.'s (2018) findings: For the new material of the current study, text-based representations of linear systems of equations were clearly inferior to graphic and equation, whereas the graphical representation proved to be the most beneficial.

The following *exploratory results* provide a deeper insight into the usage of different representations within MERs.

Taking a closer look at the graphic, results showed that subjects not only performed faster for the single graphical representation than for the single text condition but also performed even faster than for the multiple homogeneous condition while mastering the same number of items. The advantage in efficiency of the single graphical representation over the multiple homogeneous MER, as the only MER without a graphical representation, highlights the striking role of the graphical-based representation for linear system of equation problems. The superior role of the graphic was also displayed in participants' ratings: It was clearly assessed as most helpful representation.

Concerning the representation equation within different MERs, results revealed that time investment for homogeneous MERs (equation + text) was higher than for the heterogeneous combination of equation and graphic. When combined with text, the equation was clearly favored by the participants to solve systems of equations. However, it could not compensate for the time saving effect of a graphical representation, which seemed to be especially helpful in an equation-containing MER. Findings merely about performance could lead to the conclusion that adding an equation to the graphical representation did not helpfully complement the graphic. However, gaze behavior revealed that subjects being provided with a combination of equation and graphic considered both representations to the same extent. Moreover, gaze switches were especially high for heterogeneous MERs (e.g., graphic + equation; see Hypothesis 3). These results indicate that subjects in fact actively worked with both representations and that translation processes took place. Following Ainsworth (2006), these findings lead to the assumption that the MER composed of equation and graphic fulfilled the function of a deeper understanding. Consistent with the CTML and ITPC model, translating processes between graphic and formula should lead to integration of corresponding information of the provided representations. Through integrating the concrete and accessible graphic with the abstract equation, a new mental construct could be established. Prior research also revealed that equation representations can have their specific advantages (Müller & Heise, 2006), particularly for complex problem solving (Nathan, Stephens, Masarik, Alibali, & Koedinger, 2002; complexity–representation interaction, Koedinger, Alibali, & Nathan, 2008).

With respect to the representation text, gaze data indicated that subjects avoided the text when working with homogeneous MERs, as they attended to it less often and for shorter times than to the equation. Concerning the combination of text and graphic, subjects attended to graphic more often and longer than to text, which matches the findings of Lindner et al. (2017). In combination with the holistically depicted, and thus easily accessible graphic, the complex and abstract text even seemed to be almost fully neglected. In addition, graphic was fixated last more often than text, which might simply be explained by the subject's prevalent attendance to the graphic. Text seemed to be avoided by the participants whenever possible, and if it became unavoidable as single representation, the necessary time investment increased. The neglect of text was supported by the subject's ratings as it was considered as the least helpful representation.

The MER combining all three representations at once (text, graphic, and equation) stands out from the other MERs. On the one hand, it can be logically classified as a heterogeneous MER because it contains analogous as well as symbolic representations. On the other hand, it can be seen as a hybrid condition providing homogeneous as well as heterogeneous MERs. With respect to this special MER, graphics and equations were attended to more often and longer than text. Hence, in line with the previous results, text was neglected. Graphic was checked last more often than text or equation and thus proved to be especially important to finally solve the task. Furthermore, there were more gaze switches between graphic and equation than between the symbolic representations, text and equation. These findings indicate that subjects preferred to work with the heterogeneous combination of equation and graphic rather than with text-including MERs, when they were given the possibility to choose between MERs within the threefold representation. Again, this underlines the assumed function of a deeper understanding (Ainsworth, 2006) for the MER being composed of graphic and equation.

All in all, these findings beyond the hypotheses lead to the assumption that, for linear system of equation tasks as particular kind of mathematical problem solving, graphic might be the most used and helpful representation. The superior role of the graphic over text is especially remarkable given that Ott et al. (2018) found evidence for text as a reference representation in their study using propositional logic tasks. On the one hand, these contradictory results emphasize the dependence of a representation's role on the specific material and problem-solving context (Ainsworth, 2006). On the other hand, the discrepancy could at least be partly due to the fact that Ott et al.'s graphical representations were not informationally rich and thus their effect might have been limited. As the authors stated, their material on propositional logic tasks differed from typical multimedia studies in that the graphic representations did not contain any relevant spatial information. Therefore, Ott et al. assumed that using material with informationally richer illustrations would have a more substantive effect when added to a symbolic representation. This assumption was confirmed in the current study. Because Ott et al.'s tasks could not be solved using only the informationally inferior graphic, subjects had to at least in part use the equation or text to find the solution. The symbolic language of the equations presented in Ott et al.'s experiments was completely new to the participants, as only novices lacking domain-specific knowledge (apart from the given explanations) took part in their study. Therefore, when choosing between text and equation, subjects most likely referred to the textual representations, which appeared most familiar. In contrast to that, the symbolic language of equations depicted in the present study was familiar to all of the participants, although math experts were excluded from the study as well. This was due to the fact that the respective kind of mathematical equation is frequently used in math class at school. Even participants having little experience in working with the specific kind of equation-solving problems were at least familiar with the appearance and structure of the used symbolic language. Consequently, when choosing a representation to solve the task, subjects of the present study might either have favored the equation as a more familiar representation, which they were often confronted with in school, or they preferred the easily accessible, colorful, and attention-catching graphic that also contained all the necessary information to solve the task. Contrary to Ott et al.'s study, the text representation inhered no perceived benefit over the other representations for the participants, which could explain the fact that it was mostly neglected.

Further exploratory analyses showed that not only the design of the representations influenced how subjects dealt with the equation-solving problems but also learner characteristics such as verbal memory and prior knowledge, which will be discussed below.

Subjects showing good performance in a verbal memory test mastered homogeneous MERs (text + equation) even better than the heterogeneous MERs. Following the ITPC model and CTML, the symbolic representations text and equation are both processed in the verbal channel of the working memory, thus leading to an overload of this channel. Subjects scoring high in verbal memory seem to cope better with a high load in the verbal working memory channel. This assumption complements the findings of Hypothesis 3, namely, that a high cognitive load might be the reason for losses in performance concerning homogeneous MERs. A high verbal memory capacity might compensate for the high extrinsic cognitive load caused by homogeneous MERs and thus prevent losses in performance. Furthermore, the results could possibly be explained through the assumption that the supportive function (Ainsworth, 2006) of homogeneous MERs might only come into effect when verbal memory was high. This could be due to the fact that high verbal memory capacity makes the textual representations appear more accessible and easier to deal with. Consequently, translating processes between the symbolic representations could be facilitated, and a function of deeper understanding could be met by the MER, resulting in benefits in performance. However, the particular function of the MER and underlying processes remained unclear. As such, the relationship between verbal memory and performance in homogeneous MERs requires further investigation.

Moreover, prior knowledge about systems of equations proved to be related to task difficulty and performance: The higher participants scored in the prior knowledge test, the lower the assessed item difficulty was and the better was the performance concerning the items. The negative relationship between rated task difficulty and prior knowledge matches the findings of Kalyuga, Ayres, Chandler, and Sweller (2003). Increased prior knowledge is associated with a reduction in intrinsic cognitive load, which could compensate for the higher extraneous cognitive load stemming from the homogeneous MER condition's more demanding task representation. This assumption relates to the finding that expertise moderated the relationship between type of representation and performance in other domains (e.g., Ainsworth, 2006; Cromley et al., 2017; Kozma, 2003; Rau, 2017).

Previous research (e.g., Holmqvist et al., 2011) led to the assumption of a positive relationship between total fixation duration and indicators of cognitive load. The fact that this relationship was not confirmed in the current study might not necessarily be due to a weakness of the cognitive load scale but rather in participants guessing or skipping very challenging items of high cognitive load. This would result in shorter fixation times for demanding items.

4.1 | Limitations and future research

The first limitation concerns differences in salience of the representations that depicted equation-solving problems in the present study. Whereas text and equation were illustrated black and white and depicted as simple as possible, the graphical representations contained discs in three different colors. Due to the attention-catching graphic, subjects might have focused on the analogous representation first. Further, because no other representation was needed to solve the task, participants might have then stuck to the graphic for the continued problem-solving process. Further research should make use of graphics designed in black and white to ensure similarly salient representations. It remains to be investigated whether the preference for the black-and-white graphical representations would still be as strong as depicted in the present study.

The second limitation concerns the sample. Because only novices of slightly different levels took part in our study, future research should compare how real experts and novices master homogeneous and heterogeneous MERs of linear systems of equations and thus further investigate the role of prior knowledge. Pronounced prior knowledge could also affect integration processes in MERs and reveal an interrelationship that remained hidden in the present study.

Third, the present study could only infer the composition of intrinsic and extraneous cognitive load during the equation-solving tasks because it was not assessed differentially. Taking into account the substantial impact of cognitive load on task performance regarding heterogeneous and homogeneous MERs, a future study should shed light on its composition during tasks providing different representations. Because prior knowledge should lead to a reduction in intrinsic cognitive load, the question arises if these free resources of an expert could be used to cope with higher extraneous load such as that caused by homogeneous representations.

4.2 | Practical implications and conclusion

In math education, the use of MERs is quite common. However, particularly with the topic of equation solving, graphical representations are used infrequently, and if they are, it is typically only in the introduction phase of equations. The current study could show that MERs are indeed more helpful than single representations to students who have basic knowledge in this domain but not much practice in solving linear systems of equations. However, heterogeneous MERs that include a graphical representation could even be more helpful for absolute novices. The mobile as a concrete graphical interpretation of linear systems of equations resulted in great appeal on the students' part, and it could be even more appealing for school students starting to work with systems of equations to provide them with an easily accessible and analogous representation of the topic. This assumption is supported by Yung and Paas' (2015) finding on the beneficial effect of visual representations added to mathematics tasks for primary school children.

Moreover, the current research gives some indication for applying the principle of desirable difficulty, that is, creating a challenging learning environment that fosters long-term retention and transfer (Bjork, 1994). When training experts (e.g., pupils in intensified math school courses or students taking math courses in college) to reach their maximum performance in linear system of equations tasks, the required investment of mental effort should be set high by the instructor. As the present study showed, this can be achieved by providing homogeneous MERs. Hence, experts would profit most from the tasks.

In conclusion, the results of the present study match previous research showing that MERs were superior to single representations. Homogeneous MERs triggered particularly high cognitive load resulting in performance losses compared with heterogeneous MERs. The easily accessible and holistic graphic proved to be the most helpful representation. The MER composed of graphic and equation probably fulfilled the function of a deeper understanding. Different types of MERs could be used to adapt the level of difficulty to the learner's skill level. Novices who start solving linear system of equation tasks should be supported with the graphical representation to foster translation processes and thus gain a deeper understanding of the abstract symbolic language.

DATA AVAILABILITY STATEMENT

The present study was preregistered before the start of the data collection at the Open Science framework. Preregistration, raw data, materials and laboratory log are available at: https://osf.io/4x5qg/.

ORCID

Sarah Malone https://orcid.org/0000-0001-8610-2611
Kristin Altmeyer https://orcid.org/0000-0003-1835-4432
Markus Vogel https://orcid.org/0000-0001-6402-3314

REFERENCES

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences*, *11*(1), 25–61. https://doi.org/10.1207/S15327809JLS1101_2

Ainsworth, S., Wood, D., & O'Malley, C. (1998). There is more than one way to solve a problem: Evaluating a learning environment that supports the development of children's multiplication skills. *Learning and Instruction*, *8*(2), 141–157. https://doi.org/10.1016/S0959-4752(97)00013-3

Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., & Sabena, C. (2015). Reading mathematics representations: An eye-tracking study. *International Journal of Science and Mathematics Education*, *13*(2), 237–259. https://doi.org/10.1007/s10763-013-9484-y

Arcavi, A. (2005). Developing and using symbol sense in mathematics. *For the Learning of Mathematics*, *25*(2), 42–47.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, *225*, 82–90. https://doi.org/10.1038/scientificamerican0871-82

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174–205). Cambridge: Cambridge University Press.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, *10*(2), 151–170. https://doi.org/10.1002/(SICI)1099-0720(199604)10:2<151::AID-ACP380>3.0.CO;2-U

Chen, O., Kalyuga, S., & Sweller, J. (2016). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction*, *45*, 20–30. https://doi.org/10.1016/j.learninstruc.2016.06.007

Cromley, J. G., Booth, J. L., Wills, T. W., Chang, B. L., Tran, N., & Madeja, M. (2017). … & Zahner, W. (2017). Relation of spatial skills to calculus proficiency: a brief report. *Mathematical Thinking and Learning*, *19*(1), 55–68. https://doi.org/10.1080/10986065.2017.1258614

Debellis, V. A., & Goldin, G. A. (2006). Affect and meta-affect in mathematical problem solving: A representational perspective. *Educational Studies in Mathematics*, *63*, 131–147. https://doi.org/10.1007/s10649-006-9026-4

Dreher, A., & Kuntze, S. (2015). Teachers' professional knowledge and noticing: The case of multiple representations in the mathematics classroom. *Educational Studies in Mathematics*, *88*(1), 89–114. https://doi.org/10.1007/s10649-014-9577-8

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics*, *61*(1), 103–131. https://doi.org/10.1007/s10649-006-0400-z

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*, 621. https://doi.org/10.3389/fpsyg.2015.00621

Ekstrom, R. B., French, J. W., Harmann, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

English, L., & Halford, G. (1995). *Mathematics education: Models and processes*. Hillsdale, NJ: Lawrence Erlbaum.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). New York, NY: Guilford Press.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*, 717–742. https://doi.org/10.1006/jmla.1993.1036

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. New York: Oxford University Press.

Horz, H., & Schnotz, W. (2008). Multimedia: How to combine language and visuals. *Language at Work - Bridging Theory and Practice*, *3*(4). https://doi.org/10.7146/law.v3i4.6200

Hu, L., Chen, G., Li, P., & Huang, J. (2019). Multimedia effect in problem solving: A meta-analysis. *Educational Psychology Review*. https://doi.org/10.1007/s10648-019-09490-4

Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test*. Hogrefe.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23–31. https://doi.org/10.1207/S15326985EP3801_4

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors*, *46*(3), 567–581. https://doi.org/10.1518/hfes.46.3.567.50405

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 247–262). Cambridge: Cambridge University Press.

Kaput, J. J. (1987). Towards a theory of symbol use in mathematics. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 159–196). Hillsdale, NJ: Lawrence Erlbaum.

Koedinger, K., Alibali, M., & Nathan, M. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science: A Multidisciplinary Journal*, *32*(2), 366–397. https://doi.org/10.1080/03640210701863933

Korbach, A., Brünken, R., & Park, B. (2016). Learner characteristics and information processing in multimedia learning: A moderated mediation of the seductive details effect. *Learning and Individual Differences*, *51*, 59–68. https://doi.org/10.1016/j.lindif.2016.08.030

Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction*, *13*(2), 205–226. https://doi.org/10.1016/S0959-4752(02)00021-X

Kühl, T., Stebner, F., Navratil, S. C., Fehringer, B. C. O. F., & Münzer, S. (2018). Text information and spatial abilities in learning with different visualizations formats. *Journal of Educational Psychology*, *110*(4), 561–577. https://doi.org/10.1037/edu0000226

Kultusministerkonferenz. (2003). Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (No. October 18, 2013): Kultusministerkonferenz. Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-Bildungsstandards-Mathe-Mittleren-SA.pdf

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–100. https://doi.org/10.1111/j.1551-6708.1987.tb00863.x

Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *17*(4), 401–418. https://doi.org/10.1002/acp.877

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration* (Vol. 1, pp. 51–86). New York: Springer.

Lindner, M. A., Eitel, A., Strobel, B., & Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An eye-movement analysis. *Learning and Instruction*, *47*, 91–102. https://doi.org/10.1016/j.learninstruc.2016.10.007

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*(1), 95–109. https://doi.org/10.1016/j.compedu.2012.07.011

Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, *8*(4), 357–371. https://doi.org/10.1007/BF01463939

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge, MA: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Müller, R., & Heise, E. (2006). Formeln in physikalischen Texten: Einstellung und Textverständnis von Schülerinnen und Schülern. *Physik Und Didaktik in Schule Und Hochschule*, *2*(5), 62–70.

Nathan, M. J., Stephens, A. C., Masarik, D. K., Alibali, M. W., & Koedinger, K. R. (2002). Representational fluency in middle school: A classroom based study. In D. Mewborn, P. Szatain, D. White, H. Wiegel, R. Bryant, & K. Nooney (Eds.), *Proceedings of the Twenty-fourth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Niss, M. (2015). Mathematical competencies and PISA. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy* (pp. 35–55). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-10121-7_2

OECD (Organization for Economic Cooperation and Development) (2013). Draft PISA Mathematics Framework. Retrieved from https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Mathematics%20Framework%20.pdf [July 19, 2018].

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645–1656. https://doi.org/10.3758/s13428-017-0998-z

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Ott, N., Brünken, R., Vogel, M., & Malone, S. (2018). Multiple symbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic. *Learning and Instruction*, *58*, 88–105. https://doi.org/10.1016/j.learninstruc.2018.04.010

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29*(4), 717–761. https://doi.org/10.1007/s10648-016-9365-3

Rau, M. A., Aleven, V., & Rummel, N. (2013). How to use multiple graphical representations to support conceptual learning? Research-based principles in the fractions tutor. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education: 16th international conference, AIED 2013. Lecture notes in computer science* (pp. 762–765). Berlin: Springer. https://doi.org/10.1007/978-3-642-39112-5_107

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037//0033-2909.124.3.372

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 symposium on eye-tracking research & applications ETRA'00*, 71–78.

Saß, S., Schütte, K., & Lindner, M. A. (2017). Test-takers' eye movements: Effects of integration aids and types of graphical representations. *Computers & Education*, *109*, 85–97. https://doi.org/10.1016/j.compedu.2017.02.007

Schnotz, W. (2005). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–69). New York: Cambridge University Press.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 72–103). Cambridge: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*(2), 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*(3), 185–233.

Van Someren, M. W., Reimann, P., & Boshuizen, H. (1998). *Learning with multiple representations*. In *Advances in learning and instruction series*. New York, NY: Elsevier science Inc.

Vollrath, H.-J. (2003). Algebra in der Sekundarstufe (2nd edition). In *Mathematik Primar- und Sekundarstufe*. Heidelberg, Berlin: Spektrum, Akad. Verl.

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, *11*(2), 87–95. https://doi.org/10.1080/00461527409529129

Yerushalmy, M. (1991). Student perceptions of aspects of algebraic function using multiple representation software. *Journal of Computer Assisted Learning*, *7*(1), 42–57. https://doi.org/10.1111/j.1365-2729.1991.tb00223.x

Yung, H. I., & Paas, F. (2015). Effects of computer-based visual representation on mathematics learning and cognitive load. *Journal of Educational Technology & Society*, *18*(4), 70–77.

How to cite this article: Malone S, Altmeyer K, Vogel M, Brünken R. Homogeneous and heterogeneous multiple representations in equation-solving problems: An eye-tracking study. *J Comput Assist Learn*. 2020;36:781–798. https://doi.org/10.1111/jcal.12426

APPENDIX A: | Means (M) and standard deviations (SDs) per representational condition for the dependent and mediator variables regarding Hypotheses 1 to 3

	M (SD)						
Condition	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Time on task per item(s)	104.73 (57.33)	146.08 (67.67)	164.58 (72.21)	169.58 (86.27)	118.83 (44.32)	127.93 (51.07)	150.78 (73.20)
Number of correctly solved items[a]	1.40 (0.81)	1.13 (0.82)	1.23 (0.77)	1.60 (0.62)	1.37 (0.67)	1.50 (0.57)	1.40 (0.67)
Mental effort[b]	5.65 (1.30)	6.47 (1.46)	7.20 (1.06)	6.67 (1.33)	6.10 (1.25)	6.17 (1.36)	5.95 (1.37)
Task difficulty[b]	5.30 (1.49)	6.50 (1.70)	7.15 (1.08)	6.60 (1.59)	5.72 (1.42)	5.88 (1.42)	5.87 (1.33)
Transition count				3.45 (3.05)	5.09 (4.92)	6.40 (5.82)	

[a]Out of two per condition.
[b]Rated on a 9-point Likert scale.

APPENDIX B: | Means (M) and standard deviations (SDs) of eye-tracking measures for the particular representations within the MERs of conditions 4 to 7

M (SD)									
Condition	**4**		**5**		**6**		**7**		
	Formula	Text	Formula	Graphic	Text	Graphic	Formula	Text	Graphic
Total fixation duration	75.42 (39.17)	19.04 (23.64)	25.63 (31.11)	41.59 (30.01)	11.95 (11.75)	61.13 (41.11)	32.49 (48.68)	5.53 (7.96)	52.16 (35.39)
Visit duration	14.71 (9.94)	5.01 (6.06)	7.06 (8.46)	10.67 (8.77)	2.20 (2.37)	13.00 (11.02)	6.70 (8.73)	1.12 (.93)	10.12 (7.66)
Fixation count	282.10 (186.25)	91.85 (120.90)	90.07 (91.20)	154.45 (8.77)	59.02 (55.13)	220.10 (126.66)	105.77 (127.08)	26.55 (33.93)	184.70 (122.59)
Visit count	10.05 (6.82)	6.15 (3.93)	7.08 (8.46)	10.15 (6.46)	7.83 (5.41)	11.53 (6.27)	6.95 (5.33)	4.78 (2.98)	9.73 (5.57)
Time to first fixation	910.89 (393.42)	888.08 (383.99)	970.02 (450.29)	958.51 (454.79)	902.74 (348.35)	902.66 (346.70)	878.72 (374.06)	879.97 (378.11)	883.71 (376.45)
Last fixation	1.17 (.65)	0.70 (.60)	0.80 (0.55)	0.99 (0.67)	0.53 (0.63)	1.30 (0.70)	0.50 (0.51)	0.30 (0.47)	1.20 (.66)

Abbreviations: MER, multiple external representation.

APPENDIX C: | Moderation analyses

	Verbal memory			Prior knowledge			Paper-folding test[a]			Card-rotation test[b]		
	b	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*
Constant	1.46	25.15	<.001	1.47	25.37	<.001	1.47	25.07	<.001	1.47	24.79	< .001
MER type	0.17	1.27	0.208	0.16	1.16	0.249	0.17	1.27	0.208	0.17	1.26	0.211
Moderator	−0.12	−0.81	0.421	0.19	2.86	0.005	0.05	1.83	0.069	0.00	0.62	0.534
MER type × moderator	0.11	2	0.048	−0.00	−0.02	0.982	0.02	0.29	0.774	0.01	0.96	0.338

Abbreviations: MER, multiple external representation.

[a]Score was calculated by the number of correctly solved items minus one fifth number of incorrect items.

[b]Score was calculated by the number of correctly solved items minus number of incorrect items.