CrossMark

ORIGINAL RESEARCH

There's more to the multimedia effect than meets the eye: is seeing pictures believing?

Magnus Ögren[1,2,3] · **Marcus Nyström**[4] · **Halszka Jarodzka**[5]

Abstract Textbooks in applied mathematics often use graphs to explain the meaning of formulae, even though their benefit is still not fully explored. To test processes underlying this assumed multimedia effect we collected performance scores, eye movements, and think-aloud protocols from students solving problems in vector calculus with and without graphs. Results showed no overall multimedia effect, but instead an effect to confirm statements that were accompanied by graphs, irrespective of whether these statements were true or false. Eye movement and verbal data shed light on this surprising finding. Students looked proportionally less at the text and the problem statement when a graph was present. Moreover, they experienced more mental effort with the graph, as indicated by more silent pauses in thinking aloud. Hence, students actively processed the graphs. This, however, was not sufficient. Further analysis revealed that the more students looked at the statement, the better they performed. Thus, in the multimedia condition the graph drew students' attention and cognitive capacities away from focusing on the statement. A good alternative strategy in the multimedia condition was to frequently look between graph and problem statement, and thus to integrate their information. In conclusion, graphs influence where students look and what they process, and may even mislead them into believing

✉ Marcus Nyström
marcus.nystrom@humlab.lu.se

Magnus Ögren
magnus@ogren.com

Halszka Jarodzka
halszka.jarodzka@ou.nl

1   Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark

2   Department of Chemistry, Nano-Science Center, University of Copenhagen, Copenhagen, Denmark

3   School of Science and Technology, Örebro University, 701 82 Örebro, Sweden

4   Lund University Humanities Laboratory, Helgonabacken 12, 221 00 Lund, Sweden

5   Open University, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands

accompanying information. Thus, teachers and textbook designers should be very critical on when to use graphs and carefully consider how the graphs are integrated with other parts of the problem.

Introduction

Mathematical textbooks often include different forms of pictures (such as illustrations, graphs, diagrams, etc.). The reasons are twofold. On the one hand, teachers and textbook designers generally believe that pictures would be helpful for students to better understand the material. On the other hand, cognitive theories of information processing recommend to enrich scientific texts with pictures to support students in building a rich and coherent mental model of the subject matter (i.e., multimedia effect). However, there are two critical points to this view. First, theories underlying the multimedia effect make statements about perceptual processes that have not been verified directly. Second, recent empirical research questions this general beneficial effect of pictures and even suggests that pictures may bias people into being uncritical towards scientific texts (i.e., picture bias effect) (McCabe and Castel 2008). Thus, in this study, we (1) investigate the multimedia effect and its under-lying cognitive and perceptual processes directly with *think-aloud protocols* and *eye tracking* and (2) we test the picture bias effect. Both investigations are in the context of mathematical education at a university level (vector calculus).

Basic assumptions of learning with multimedia

Material that presents information in different formats, such as text, pictures, diagrams, and formulae, is referred to as *multimedia*. Two leading theories describe how the human cognitive system processes multimedia material, namely the *Cognitive Theory of Multi-media Learning* (CTML; Mayer 2005a) and the *Cognitive Load Theory* (CLT; Sweller et al. 1998). Both theories assign a central role to working memory (Baddeley et al. 1992). They make three assumptions on the functioning of working memory.

First, for information to be learned and successfully stored, it has to be *actively pro-cessed* in working memory. Mayer (2005a) describes active processing in three steps: Information has to be selected from a source by means of attention to enter working memory. Next, this information has to be organized into mental models. Last, these mental models have to be integrated with each other and prior knowledge from long-term memory. Only information that has been processed in such a way can be stored in the long-term memory. The 'select' and 'integrate' processes refer to *perceptual* processes, however, these were only theoretically deduced, but not directly tested.

Second, working memory is of *limited capacity*, which must not be exceeded. Sweller et al. (1998) proposed that working memory capacity can be filled with three types of loads, namely load caused by active processing of the information (germane load), load stemming from the difficulty of the task (intrinsic load), and by load stemming from other unnecessary cognitive processes that do not contribute to executing the task at hand (ex-traneous load). The amount of cognitive load posed upon working memory (i.e., mental

effort) can be measured with different methods, such as subjective rating scales (Paas 1992) or silent pauses in thinking aloud (Yin and Chen 2007; Jarodzka et al. 2015).

Third, both theories assume that *two separate channels* exist for processing verbal and pictorial information (Baddeley et al. 1992; Paivio 1986). Both channels are in earlier steps of processing information independent, hence they are loaded with information separately. In later processing steps this information is integrated and leads to a richer mental model than when based on one modality (either pictures or words) only.

Based on these three assumptions both theories (CTL, CTML) provide guidelines on how multimedia material should be designed to optimize cognitive processing of information, as will be described in the next section. It is important to note that both theories and also their resulting guidelines refer to *learning*. Nevertheless, we argue that these theories as well as their resulting guidelines can be adapted to task performance without a specific learning intention, because they are built upon general assumptions of the human cognitive system: First, the assumption that the human cognitive system is limited in capacity with respect to how much information it can process at a time (not in long-term memory storage, though) dates back to early research on the structure and functioning of human working memory (e.g., Baddeley et al. 1992; Miller 1956). Hence, this assumption holds true not only for learning, but also for general task performance. Second, the active processing assumption of information is based on Atkinson's and Shiffrin's information processing model (Atkinson and Shiffrin 1968), which again is not a specific learning model, but instead describes general information processing. Thus, in this study we applied these principles to task performance i.e., solving a problem in vector calculus.

Guidelines for designing multimedia material: the multimedia principle

One of the basic guidelines of the above mentioned multimedia theories (CTL, CTML) is the *multimedia principle*, which assumes that "people learn better from words and pictures than from words alone (Fletcher and Tobias 2005; Mayer 2001a). The main idea is that text and pictures evoke different cognitive processes resulting in different mental models which, when later integrated, result in a richer mental model compared to one of the models alone. Moreover, when information is presented both in a pictorial and a textual manner, students can use both processing channels in parallel and more efficiently use their working memory. This enables an active processing of information.

A long history of research provides evidence for the multimedia principle (for instance, see the research conducted by the research group of Richard Mayer). Mayer (2001b) reports nine of his own studies, all with beneficial learning effects when pictures accompany text. Confirming this positive effect of pictures, Carney and Levin (2002) present a review with 18 articles from the 90ies reporting beneficial learning-effects of pictures accompanying texts.

In one of his articles, Mayer (1989) showed that learning about car mechanics improved when pictures were accompanied with text compared to text only (or pictures only). He explains the multimedia effect by the fact that such illustrations helped students to "focus their attention" and "organize the information into useful mental models" . However, these conclusions were not directly tested.

Thus, both theory and empirical research state that pictures accompanying texts in mathematical problem solving reduce mental effort and help the students to focus their attention, although these assumptions were often deduced from an improved task performance, but not directly tested.

Limitations and restrictions of the multimedia effect

Several empirical studies provide challenges for the multimedia principle. Often, students do not make use of pictures as was intended. For instance, Berends and van Lieshout (2009) found that school children do not benefit from pictures in mathematical problem solving as much as intended. The authors concluded that integrating two information sources probably required more working memory capacity than available (for similar findings in school exams see Crisp and Sweiry 2006). In line with these findings, Holsanova et al. (2009) found in a naturalistic newspaper reading study that if pictures and text are given in a standard format, i.e., where they are presented separately in a 'split' format, readers often do not make the effort to integrate these information sources (as shown by little visual integration between both information sources indicated with eye tracking). Thus, providing additional information in graphs—irrespective of whether it is relevant to the task—requires additional cognitive resources. If these resources are not available or not allocated correctly graphs can even be harmful.

Bias towards believing

Other researchers sees the additional use of pictures even more critically. Lenzner et al. (2013), for instance, showed that pictures reduce the *perceived difficulty* of a learning material. This could be very dangerous as students might put too little mental effort into understanding the text so that they do not process all information actively (i.e., select all relevant information from all possible information sources, organize it into coherent mental models, and integrate it), which in turn would result in a poorer task performance.

Other lines of research unrelated to learning or instruction also critically investigate the effect of pictures. Isberner et al. (2013) found that graphs increased the *perceived plausibility* of conflicting information in science text. Again, this is problematic as it could result in students overlooking logical flaws in a text and thus not being able to build a coherent mental model of the task at hand, again, resulting in poorer task performance. McCabe and Castel (2008) showed that the mere presence of an illustration increased the *perceived credibility* of a scientific text. The readers were less critical against the arguments of a scientific text, when it was accompanied by a scientific illustration. As with the other examples, this uncritical attitude towards a text prevents students from building a coherent mental model of its content. Therefore, they are not able to draw the correct conclusions from this mental model, when it has to be applied to perform a particular task. Hence, pictures that are of a scientific nature may easily be perceived as a proof of the accompanying text and mislead students into believing it—irrespective of whether they do add to its arguments or not. Only a careful integration of both information sources could prevent someone from making this mistake.

Vector calculus as an exemplary mathematical domain

In the present study the multimedia- and picture bias effects were investigated in the domain of vector calculus. We chose this domain for two reasons. First, vector calculus is a crucial foundation for studies in mathematics and is used in many branches of physics and engineering (for details on the Swedish curriculum in vector analysis see Griffiths 1999; Ramgard 1996; Persson and Böiers 1988). Second, vector calculus is a very visual topic where an abstract mathematical formula often can be accompanied with a direct graphical

representation. One of the authors of this article has been teaching courses in vector calculus and has discussed the topic among several colleagues from different countries. It is a common belief among teachers we have talked to that a key to understanding vector calculus is to be able to switch between different representations of a problem, and successfully integrate the information from all representations into one coherent mental model. This is referring to a deeper form of understanding, necessary for instance to be able to apply relevant knowledge to new applications.

The present study

In this study we investigate whether we can find a general multimedia effect for mathematical problem solving in vector calculus by comparing problem solving tasks with and without accompanying graphs. An example problem is shown in Fig. 2. Furthermore, we test whether these graphs bias students into believing their accompanying texts by asking students to reject or confirm statements about the task. To better understand the processes underlying the multimedia effect, we use two process-tracing measures: eye tracking (Holmqvist et al. 2011) and verbal reporting (Ericsson and Simon 1993). Eye tracking tells us which areas students visually select information from and how they visually integrate these areas. Concurrent verbal reporting may provide insight into the amount of mental effort invested by students (Yin and Chen 2007; Jarodzka et al. 2015). Moreover, it can deliver qualitative information about the underlying processes by serving as a dual-task measure of mental effort (e.g., Brunken et al. 2003; Park et al. 2015). We hypothesize the following with respect to performance (H_1 and H_2) and processes ($H_{3a,b,c}$).

H_1 Performance (i.e., correctly confirming or rejecting a problem statement) is higher with than without graphs, that is, we expect a multimedia effect.

H_2 Confirming the problem statement is more likely with than without graphs, that is, we expect a picture bias effect. As a results of the picture bias, we also expect higher performance in the multimedia condition when statements are to be confirmed, compared to the control condition (without a graph).

H_3 Students process information differently depending on whether a graph is present or not. In particular, we expect:

H_{3a} If a graph is present, students *search and select* information from it. This shows in time spent looking at the graph. Furthermore, as we expect a multimedia effect, we consequently expect that search and selecting information from the graph is positively related to task performance. In addition, we explore to which extent this shift of attention towards the graph happens at the expense of the other information areas (text and formula input and problem statement) and to which extent attending to these is related to performance; we predict a higher performance the more the graphs are attended.

H_{3b} If a graph is present, students *integrate* information from it with information from other sources, such as the input (text and formula) and the problem statement. This shows in the amount of transitions between the graph and the other information sources. Furthermore, as we expect a multimedia effect, we consequently expect that integrating from the graph with other sources is positively related to task performance.

H_{3c} In problems with graphs students use more *mental effort* than in problems without graphs, because they need to process more information. This becomes evident in the

overall proportion of silence calculated directly from the recorded sound file. A higher proportion of silence is predicted when graphs are present, as a result of the increased mental effort.

Moreover, as an open research question (RQ_1), we investigate in two contrasting cases the extent to which participants follow the processes predicted by the CTMML (search information, build a mental model, activate prior knowledge, integrate information, and form a problem solution). In addition, we investigate their meta-cognitive and off-topic statements.

## Method

Participants and design

Thirty-six students (three females) with an average age of 21.5 years (SD=3.0) took part in the experiment. They studied engineering physics (F) at the Lund Institute of Technology, and were 2 weeks into a basic course in vector calculus. Hence they should be considered as a fairly uniform population with respect to their study background. All students had normal or corrected-to-normal (i.e., with glasses or lenses) vision. They were randomly assigned to one of two conditions in a between-subject design: one solving eight problems without graphs ($N = 16$), and one solving the same problems with graphs ($N = 20$).

Stimuli

The stimuli consisted of eight problems dealing with basic concepts in vector calculus. They concerned, for example, simple cases of integration along curves in a two-dimensional domain, the interpretation of the gradient for a function of two variables, and Gauss formula in three dimensions. Each problem was composed of a text and a formula that described a general context, a problem statement that was to be confirmed or rejected and, in the multimedia condition, a graph. In this article the word *graph* is used in a broader context than it has in mathematics texts on, e.g., graph theory or functions, which we include, but are not restricted to. In three of the problems, the correct answer was to confirm the problem statement. In the remaining five, a rejection of the statement would provide a correct answer.

The graphs were designed by a lecturer in vector calculus to support students by describing a particular problem-related concept visually. In fact all the graphs used in the study could naturally be part of a textbook in vector calculus. Importantly, the students had not seen any of the problems before. They were interpretational in nature, and should therefore have a substantially positive effect on problem solving. All problems could be solved without having access to the graph. For example, many of the problems can be solved algebraically without using a mental geometric representation. Example graphs can be found in Fig. 1a, which shows level curves of a function of two variables and Fig. 1b, which depicts a curve that is restricted to a sphere. Note that the vectors are not labeled, so it was left for the students to identify and potentially use them when testing a particular solution strategy.

Each problem was saved as a grayscale png-image with a resolution of 1680×1050 pixels. This resulted in a total of 16 stimuli images, eight for each group. The problems were presented in a random order.
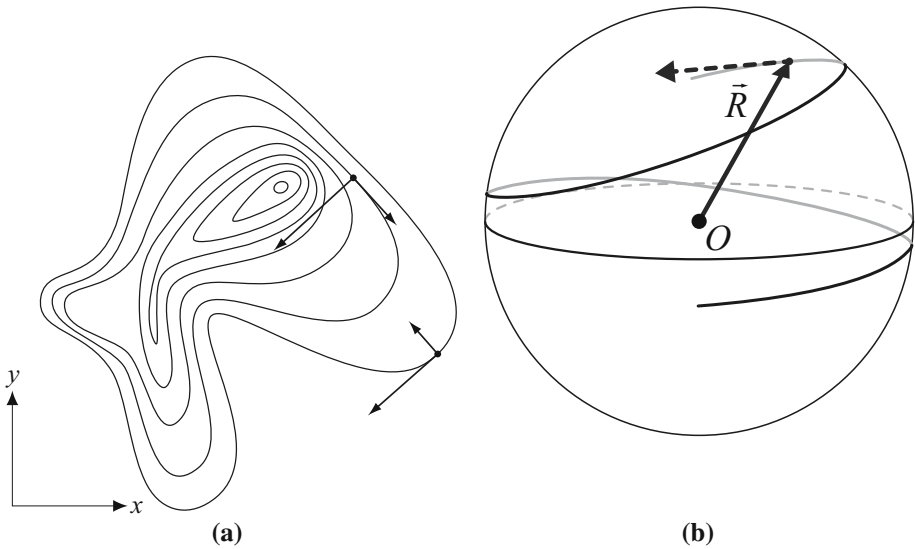
Fig. 1 Some examples of graphs used in the problems. **a** Level curves of a function of two variables. Several vectors are included to illustrate tangents of the level curves at different arbitrary chosen positions. Also corresponding perpendicular vectors, which are proportional to the gradient, are shown. **b** This graph, that was used in problem P3, had the strongest effect on performance of all the problems in our study. As is discussed in the text, it strongly supports one of the two major possible solution strategies used by the participants involving to depict the dashed vector (not shown for the students), and this solution strategy was very rare in the group not having access to this graph.

Apparatus

The experiment was performed with a Dell laptop (Intel Core i7 CPU 2.67GHz, 2.98 GB RAM) running Windows XP Professional (v. 2002, SP 3). Stimuli were presented with Experiment Center (v. 3.0.128) on a Dell 22 inch computer screen with a resolution of 1680 × 1050 pixels and a refresh rate of 60 Hz. Eye movements were recorded at 250 Hz with the RED250 eye tracker from SensoMotoric Instruments (Teltow, Germany) running iView X (v. 2.7.13). Data from the left and right eyes were averaged during recording, and therefore only one gaze coordinate represented the data for both eyes at each time instant.

Procedure

After an introduction to the experiment and after viewing an example problem not included in the actual test, participants were calibrated with a five point calibration followed by a four point validation of the calibration accuracy. Recalibrations were initiated when the operator—watching the eye image in iView X and the stimulus with overlayed gaze position in Experiment Center—judged that it was necessary. The average accuracy from all accepted calibrations reported by iView X was $0.5°$ ($SD = 0.18$) horizontally and $0.55°$ ($SD = 0.29$) vertically.

Each trial started with a centrally located fixation cross that was presented until the software detected a fixation within a $1°$ square centered on the cross. Then the problem appeared, and the participants were free to inspect the problem for a maximum of 120 s. If they felt that they were ready to provide an answer sooner, they could do so by pressing the

spacebar to answer two questions: first, participants were asked whether they thought the statement in the problem was *true* or *false* and, second, how certain they were in their answer on a scale from 1 (very unconfident) to 7 (very confident). Throughout the eye-tracking experiment, participants were asked to verbalize their thoughts as they solved the problems, according to the methodology described in [Ch. 3] Holmqvist et al. (2011) concerning training, instruction, and prompting.

Written consent was given by all participants, who got two movie theater tickets as a compensation for participating.

Data analysis

Fixations and saccades were calculated from raw data samples with BeGaze (v. 3.1 Build 152) using default settings.

Eye tracking data were analyzed by means of specific *areas of interest* (AOIs) which we defined for each problem. AOIs are coherent parts of the screen, for which eye tracking parameters were summarized. Figure 2 depicts a multimedia problem with input, problem statement, and graph, where AOIs are outlined by black rectangles and the name of the AOI is found in the upper left corner of the AOI. AOI names and rectangles were not shown to the participants. Specifically, we calculated total dwell time (sum of all time spent looking inside an AOI) from raw data samples and transitions between the AOIs from fixation and saccade data.

The proportion of speech was computed from the recorded speech signal, which was sampled at 44 kHz. A student was considered to speak when the amplitude (A) of the signal exceeded a threshold and when two consecutive speech samples above this threshold were located less than a given number of samples (n_s) apart. Limits for A and n_s were set to 0.015 (relative intensity) and 440 samples (i.e., 10 ms), respectively. Every part of the speech signal that was not detected as speech by the above definition was considered to be "silence".

The recorded speech was further analyzed by first transcribing it to text format, and then coding it into 'idea units' according to the scheme in Appendix (Table 10). The main categories in the coding scheme are based on Mayer's CTML (2005b) and thus refer to the cognitive processes assumed by this theory: searching and selecting of information from input and graph, activating prior knowledge, integrating information from different sources, and the final problem solution. In line with van Gog et al. (2005), who also
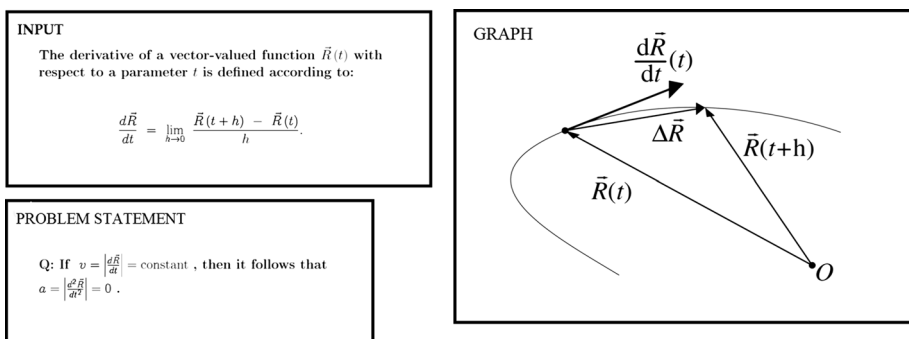
Fig. 2 An example of a stimulus (P7) used in the multimedia condition of the experiment, which has three overlayed areas of interest (AOIs): input, problem statement, and graph

investigated cognitive processes involved in problem-solving, we included meta-cognitive processes. The actual coding was conducted by two raters for 10 % of all data. Their interrater-reliability was above 70 %, calculated as the number of matching codes with respect to the total number of codes in this 10 % of the data. Since the inter-rater reliability was sufficiently high (i.e., higher than 0.70, van Someren et al. 1994), one of the raters coded the remaining data.

Data were analyzed with linear mixed effects models using R 2.15.2 (R Development Core Team 2008) and the packages lme4 (Bates et al. 2012) and languageR (Baayen 2011). Participants and problems were modelled as random factors in all analyses.

Results

The results are presented in the order of the hypotheses in Sect. 1.5.

Multimedia effect (H_1)

The participants used on average 90.8 s ($SD = 28.5$) to solve a problem. There was no significant difference in solution time between participants with ($M = 91.1$, $SD = 28.1$) and without ($M = 90.5$, $SD = 28.9$) graphs (two-sample Kolmogorov-Smirnov test, $p = 0.76$). The difference in solution times were small also when comparing problem where students answer correctly ($M = 91.0$, $SD = 27.0$) and incorrectly ($M = 90.6$, $SD = 30.2$), (two-sample Kolmogorov-Smirnov test, $p = 0.61$).

Participants solving problems with graphs answered correctly to 56 % of the problems compared to 52 % for participants without graphs. Table 1 shows the result of a multi-level logistic regression predicting a correct answer based on the presentation condition. As can be seen from the table, there is no statistically significant effect of presentation condition on students' abilities to answer the problems correctly.

Picture bias (H_2)

To test whether there was a confirmation bias when graphs were present, information about whether the correct answer is true or false was included in the regression. The output can be seen in Table 2.

The analysis reveals that participants were more likely to answer the problem statement correctly if it is true and, interestingly, there is a significant interaction between presentation condition and whether the answer is true or false. As illustrated in Fig. 3, it appears as if the students were more likely to answer correctly when the answer was true and a

Table 1 Result of the multi-level logistic regression predicting whether the answer was correct or incorrect

|	Estimate	SE	z value	Pr (>|z|)
(Intercept)	0.3019	0.4314	0.700	0.484
Withoutgraph	−0.1641	0.2622	−0.626	0.531

Here 'withoutgraph' refers to the problems without graphs. The sign of the 'Estimate' tells us that the condition with graphs led to a higher proportion of correct answers. However, the effect is not significant since the value of 'Pr(>|z|)' is above 0.05

Table 2 Result of the multi-level logistic regression predicting whether the answer was correct or incorrect

	Estimate	SE	z value	Pr (>|z|)
(Intercept)	−0.3132	0.5121	−0.612	0.540
Withoutgraph	0.3537	0.3285	1.077	0.281
AnswerTrue	1.7331	0.8578	2.020	0.043*
Withoutgraph:answerTrue	−1.4762	0.5625	−2.624	0.008**

Whether the answer is true or false has been included as a factor. Here '*' and '**' indicate significance on the 0.05 and 0.01 levels, respectively

Fig. 3 Illustration of the interaction between whether the problem statement is true or false in the multimedia (with graph) and control (without graph) conditions. *Error bars* represent standard errors
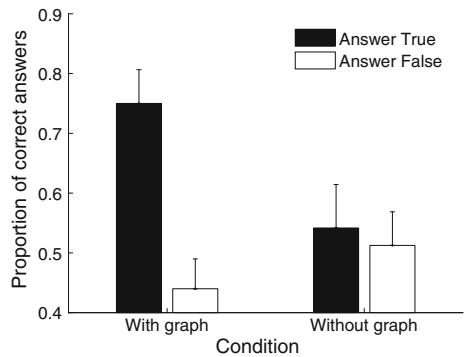
graph was present, compared to when the answer was false. On the contrary, whether the answer was true or false had no influence when the graph was not present. A post-hoc multiple comparison[1] revealed only one marginally significant difference, which occurred between the two conditions when the answer is true ($p = 0.056$). Additional support for a picture bias is provided in Table 3, which shows that presentation condition is a significant predictor for providing a confirmatory answer.

Since the nature of the answer (true or false) turned out to significantly predict the proportion of correct answers, this predictor was included in all further statistical models.

Information processing

Search and selection (H_{3a})

The overall small effect graphs had on comprehension raises the question of how the students utilize the additional graphical information. Given the similar performance results, it is tempting to believe that they did not spend much time on the graphs but, as in the non-illustrated condition, inspected only the text and the equations in the input and problem statement areas. At the same time, the interaction between whether the answer was true or false and the presentation condition (with or without graph) suggests that the graph influenced the students' problem solving processes.

Overall, the students spend a fairly large proportion of the total time viewing the graphs (19.0 ± 10.2 %). As can be seen from Fig. 4, it appears as if the graph is inspected at the

---

[1] With the R-package `multcomp` and the function `glht` using Tukey contrasts to compare the means.

Table 3 Result of the multi-level logistic regression predicting the answer (confirm/reject) based on presentation condition

	Estimate	SE	z value	Pr (>|z|)
(Intercept)	0.7224	0.4358	1.658	0.0974
Withoutgraph	−0.6366	0.2859	−2.226	0.0260*

The sign of the estimate for 'withoutgraph' tells us that more confirmatory answers were given in the multimedia condition

expense of the input and the problem statement, in such a way that equal amounts of time is taken from each of these regions. The proportion of total dwell time on both the input and problem statement was significantly shorter when the graph was present, according to a two sample t test ($p < 0.001$). Moreover, a similar test between the quotients of 'input' and 'problem statement' for problems with and without graphs did not come out significant ($p > 0.05$) for any of the problems.

Given that we know that a significant portion of time is spent visually inspecting the graph, does a longer inspection time also lead to better performance? On average, there were small differences in total dwell time on the graph when answering correctly ($M = 19.9, SD = 11.0$ %) compared to incorrect answers ($M = 18.6, SD = 10.8$ %) and, as seen in Fig. 5, there was no relationship between whether participants answered correctly and how much time they spent looking at the graph. This is confirmed statistically by the results reported in Table 4.

Previous research has shown that a good problem solving strategy is to read the problem formulation carefully, before moving on to other parts of the problem Andrà et al. (2009). However, Fig. 6 shows that performance is inversely proportional to the proportion of dwell time on the input area in a problem. Students who answered correctly looked at the input 42.6 % ($SD = 11.1$) of the time whereas those who answered incorrectly spent 47.0 % ($SD = 12.1$) of the time inspecting the input. A smaller proportion of dwell time on the input significantly predicts an increase in performance (*cf.* Table 5).

As shown in Fig. 7, the students dwelled proportionally longer at the problem statement when giving a correct answer ($M = 42.6, SD = 12.8$ %) compared to an incorrect answer ($M = 38.3, SD = 11.3$ %), and the total dwell time on the statement was a significant predictor for a correct answer (*cf.* Table 6).

Integration (H_{3b})

It could be that a long dwell time on the graph by itself does not help students' problem solving, but rather how they integrate the graph with other parts of the problem, i.e., the regions labeled as *input* and *problem statement* before. Figure 8 illustrates how performance is related to the number of transitions between different areas in the problem.

As shown in Table 7, there is a marginally significant effect ($p = 0.08$) that the number of transitions between the graph and the problem statement were higher for students that answered a problem correctly. No significant differences were found for the other transitions in Fig. 8.

Table 4 Result of the multi-level logistic regression predicting whether the answer was correct or incorrect

	Estimate	SE	z value	Pr (>|z|)
(Intercept)	−0.5382	0.6945	−0.775	0.43840
Logit (propDwellTime)	−0.1280	0.2754	−0.465	0.64195
AnswerTrue	3.6997	1.2825	2.885	0.00392**
Logit (propDwellTime):answerTrue	1.1224	0.4765	2.355	0.01850*

The proportion of dwell time on the graph is included as a factor. Note that the proportion of dwell time is logit-transformed before being used in the model

Fig. 5 The relationship between total dwell time on the graph and the proportion of correct answers. Each × corresponds to one of the 20 participant in the multimedia condition, and the line represents a linear fit of the data
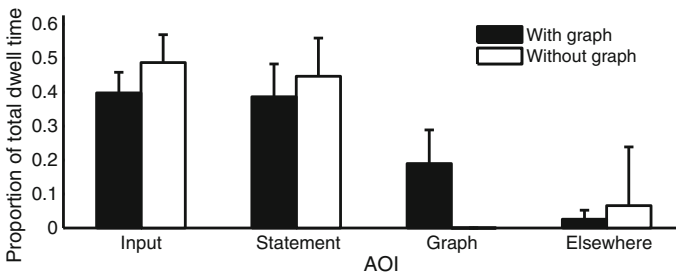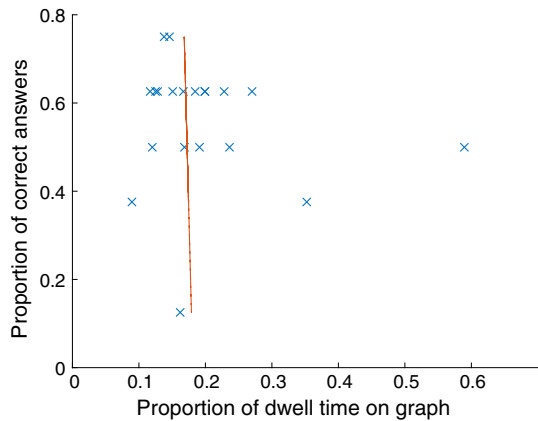

Fig. 4 Proportion of total dwell time with and without graphs. *Error bars* represent standard errors

Mental effort measured with verbal data (H_{3c})

To estimate whether the illustrated problems required more mental effort, the proportion of silence was calculated from the verbal data. Figure 9 shows that participants consistently speak less when the problem includes a graph; the proportion of silence increases from 62.3 % (SD = 7.0) to 66.3 % (SD = 9.1) for participants in the multimedia condition. As shown in Table 8, there is a marginally significant effect of presentation condition on the proportion students speak ($p = 0.07$).

Fig. 6 The relationship between total dwell time on the input and the proportion of correct answers. Each × corresponds to one of the 36 participants, and the *line* represents a linear fit of the data
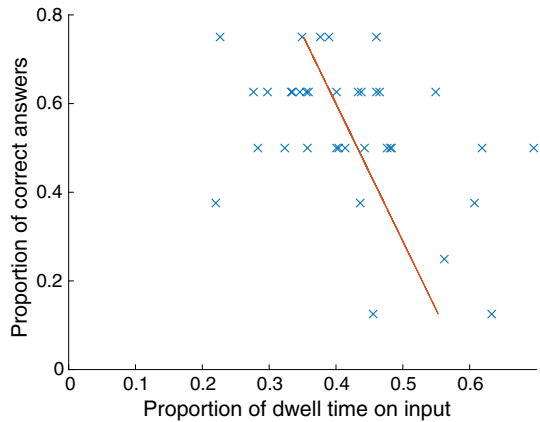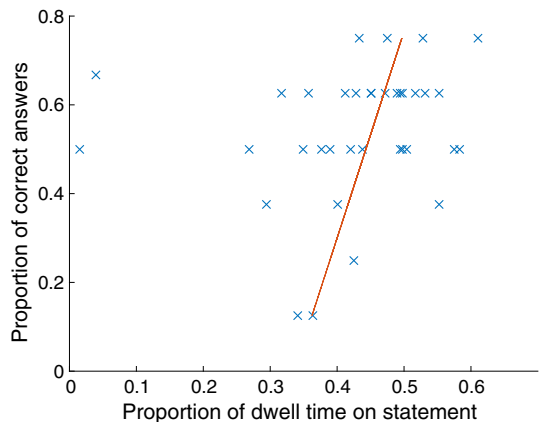
Table 5 Result of the multi-level logistic regression predicting whether the answer was correct or incorrect

	Estimate	SE	z value	Pr (>|z|)
(Intercept)	−0.4317	0.5543	−0.779	0.4361
Logit(propDwellTime)	−0.5694	0.2618	−2.175	0.0296*
AnswerTrue	1.2926	0.9021	1.433	0.1519
Logit(propDwellTime):answerTrue	−1.0773	0.5259	−2.049	0.0405*

* and ** indicate significance on the 0.05 and 0.01 levels, respectively

The proportion of dwell time on the input is included as a factor. Note that the proportion of dwell time is logit-transformed before being used in the model

Fig. 7 The relationship between total dwell time on the problem statement and the proportion of correct answers. Each × corresponds to one of the 36 participants from both conditions, and the *line* represents a linear fit of the data

Results verbal data - two contrasting cases (RQ$_1$)

In this section we compare the two most extreme cases from our experiment with respect to verbal data:problem P3 for which the presence of a graph improved the results the most, and problem P4 for which the results for the group having access to the graph was the
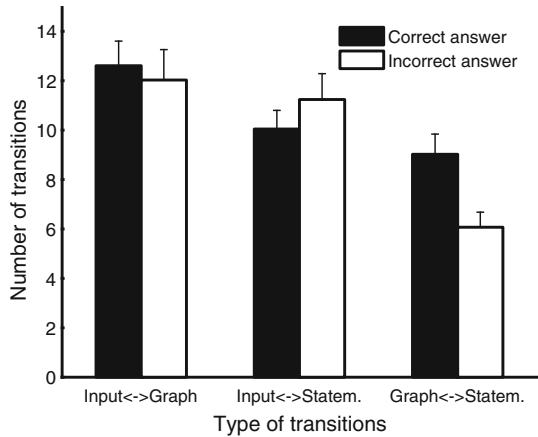
Table 6 Result of the multi-level logistic regression predicting whether the answer was correct or incorrect

	Estimate	SE	z value	Pr (>|z|)
(Intercept)	−0.01989	0.49089	−0.040	0.96768
Logit (propDwellTime)	0.55821	0.21766	2.565	0.01033*
AnswerTrue	0.56485	0.82891	0.681	0.49560
logit (propDwellTime):answerTrue	−1.09686	0.41480	−2.644	0.00818**

* and ** indicate significance on the 0.05 and 0.01 levels, respectively

The proportion of dwell time on the problem statement is included as a factor. Note that the proportion of dwell time is logit-transformed before being used in the model

Fig. 8 Number of transitions between the different AOIs: input, graph, and problem statement. Transitions in both direction are included. *Error bars* represent standard errors
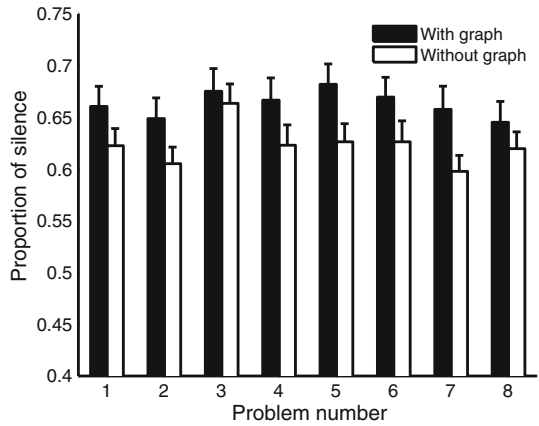
Table 7 Result of the multi-level logistic regression predicting whether the answer was correct or incorrect

	Estimate	SE	z value	Pr (>|z|)
(Intercept)	−0.84588	0.51767	−1.634	0.1023
nTransitions	0.06811	0.03928	1.734	0.0829
AnswerTrue	1.48012	0.85270	1.736	0.0826
Withoutgraph:answerTrue	0.01943	0.07887	0.246	0.8054

The number of transitions between graph and problem statement is included as a factor. Here '.' indicates that an effect is marginally significant

worst. These problems are shown in Fig. 10. In P3, the graph increased the proportion of correct answers from 0.25 ($SD = 0.44$) to 0.70 ($SD = 0.47$), whereas the graph in P4 decreased the proportion of correct answers from 0.38 ($SD = 0.50$) to 0.10 ($SD = 0.31$). Here we report analyses of verbal data based on the coding schema described in Appendix (Table 10). Table 9 shows the normalized frequency of each code in relation to the two

Fig. 9 Proportion of silence used as a measure of mental effort. *Error bars* represent standard errors

Table 8 Result of the multi-level logistic regression predicting the proportion of silence

	Estimate	SE	t value	Pr (>|*t*|)
(Intercept)	0.662	0.016	39.31	<2e−16
Withoutgraph	−0.044	0.023	−1.86	0.071
AnswerTrue	0.002	0.012	0.18	0.863
Withoutgraph:answerTrue	0.012	0.012	0.246	0.319

contrasting problems. In addition to the verbal analysis, we report how confident students were in their answers.
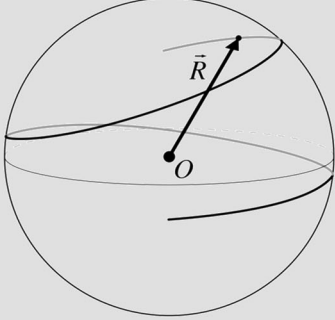
Effect of graph presence

To estimate the effect of the graph, we have calculated the difference between the codes in the multimedia and the control condition across both problems. Then, we picked the ten largest differences between these. Four of these differences were related to the graph. In that, we found that when a graph was present participants selected more information from the graph (1.25), they integrated more information from the graph with the statement (0.50) and with the input (0.90). Furthermore, they built more mental models based on the graph (0.65). Hence, the participants made more active use of the graphs. However, the presence of a graph did not only influence its use, but also the use of the other problem elements. In that, we found positive and negative influences of the graph. On the positive side for the performance in the multimedia condition, participants selected more information from the statement when a graph was present (0.66). Since we found that a proportionally longer dwell time on the statement was related with higher performance, more information selection from the statement can be seen as a positive effect of the graph. Moreover, the participants evaluated to a higher extent whether or not the statement (0.29) and the input were correct (0.58). At the same time, they used less prior knowledge (−1.09) and they integrated information with the input and the statement less frequently (−0.50), which in

A vector-valued function $\vec{R}(t)$ has the property:

$$\left|\vec{R}(t)\right|^2 = \vec{R}(t) \cdot \vec{R}(t) = \text{constant}.$$

Q: Then it follows that $\vec{R} \perp \frac{d\vec{R}}{dt}$.
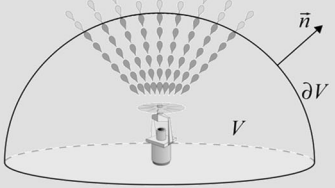
(a)

Let $\vec{v}$ be a smooth vector-field defined in $\mathbb{R}^3$. The compact volume V has a boundary ∂V, composed of one or several smooth surfaces, and oriented with an outbound normal $\vec{n}$. Then:

$$\iint_{\partial V} \vec{v} \cdot \vec{n}\, dS = \iiint_V \nabla \cdot \vec{v}\, dV .$$

Q: Then the so called Gauss formula have the interpretation that the production of a scalar quantity integrated over the volume V is equal to the surface integral of the corresponding gradient.

(b)

Fig. 10 Two contrasting cases. **a** P3, including the graph that improved the performance the most, and **b** P4, including the graph that helped the least. For improved readability, the text in the stimuli has been reproduced in the figure captions

turn probably is a negative effect of the graph. Furthermore, participants evaluated their own knowledge more positively (0.31), which may also be problematic. In summary, adding a graph seems to have both positive and negative effects on the processes underlying problem solving.

Effect of helpfulness of the graph

To further investigate the effects of a graph, we compared the use of graphs for two contrasting cases: when the graph was most helpful and when it was most harmful. Therefore we calculated the difference between the two problems in the multimedia condition. Again, we chose the ten biggest differences. Three of these differences were directly related to the use of graphs. We found that when the graph was helpful, it was

Table 9 Results of the verbal coding for problems 3 and 4

Code	P3			P4			Total (P3 + P4)			$P3_{Graph} - P4_{noGraph}$
	Graph	noGraph	Diff	Graph	noGraph	Diff	Graph	noGraph	Diff	
SIT	3.00	3.38	−0.38	5.80	5.56	0.24	8.80	8.94	−0.14	−2.80
SIG	0.90	0.00	0.90	0.35	0.00	0.35	1.25	0.00	1.25	0.55
SIS	1.50	1.31	0.19	5.10	4.62	0.47	6.60	5.94	0.66	−3.60
MMVT	0.45	0.69	−0.24	0.10	0.06	0.04	0.55	0.75	−0.20	0.35
MMVS	0.05	0.00	0.05	0.00	0.00	0.00	0.05	0.00	0.05	0.05
MMG	0.60	0.00	0.60	0.05	0.00	0.05	0.65	0.00	0.65	0.55
PK	3.30	4.06	−0.76	3.05	3.38	−0.33	6.35	7.44	−1.09	0.25
ITG	0.40	0.00	0.40	0.50	0.00	0.50	0.90	0.00	0.90	−0.10
ITS	0.20	0.56	−0.36	0.55	0.69	−0.14	0.75	1.25	−0.50	−0.35
IGS	0.50	0.00	0.50	0.00	0.00	0.00	0.50	0.00	0.50	0.50
EPS	0.95	0.69	0.26	0.90	0.88	0.03	1.85	1.56	0.29	0.05
SS	0.35	0.12	0.22	0.30	0.38	−0.08	0.65	0.50	0.15	0.05
TS	0.40	0.19	0.21	0.55	0.19	0.36	0.95	0.38	0.58	−0.15
GS	0.05	0.00	0.05	0.00	0.00	0.00	0.05	0.00	0.05	0.05
MCT+	0.20	0.25	−0.05	0.35	0.25	0.10	0.55	0.50	0.05	−0.15
MCT−	0.00	0.12	−0.12	0.00	0.06	−0.06	0.00	0.19	−0.19	0.00
MSO+	0.40	0.06	0.34	0.10	0.12	−0.02	0.50	0.19	0.31	0.30
MSO−	1.75	1.69	0.06	1.15	1.25	−0.10	2.90	2.94	−0.04	0.60
GO	0.25	0.19	0.06	0.30	0.31	−0.01	0.55	0.50	0.05	−0.05
OT	0.35	0.19	0.16	0.55	0.69	−0.14	0.90	0.88	0.03	−0.20

Each number represents the number of codes for all participants divided by the number of participants. 'Diff' represents the difference WithGraph–WithoutGraph for Problem 3, Problem 4, and Problem 3 + Problem 4. The numbers in the last column represent differences between the two problems in the multimedia condition (with a graph). For definitions and explanations of the codes, cf. Appendix. Color-coded codes represent the largest differences between the multimedia and control condition. Blue color indicates differences unrelated to graphs whereas red color indicates differences related to graphs. To make the table more readable and compact, WithGraph is denoted 'Graph' and WithoutGraph is denoted 'noGraph'

selected to a higher extent (0.55), more integrated with prior knowledge (0.5), and participants built more mental models from it (0.55) compared to when the graph was harmful. Hence, participants made a more active use of the graph, when it was helpful. Moreover, we found also impacts of the graph on processing the other information sources: when the graph was helpful, participants selected less information from the input (-2.8) and from the statement (-3.6), integrated less information from the statement and the input (-0.35), but built more mental models based on the input (0.35). Thus, when the graph was helpful, participants extracted less information from other sources, but still used these more actively. Moreover, when the graph was helpful, participants evaluated their own knowledge more—both in positive (0.3) and negative (0.6) terms.

Furthermore, we conducted a more qualitative analysis of these two contrasting cases.

Problem 3: When the graph was most helpful. By quantifying how often the keyword *sphere* (or *circle*) occurs in the verbal data, it seems as if the graph [see Fig. 1(b)] directly supported the most common way of solving this problem. That is to mentally picture the dashed arrow in Fig. 1(b) (or the opposite oriented counterpart), resulting from the vector $R(t)$ moving along the trajectory, and to finally recognize that the dashed vector is tangential to the sphere and hence orthogonal to its radius. In the multimedia condition 50 % of the participants used the keyword while reasoning about the problem, and 90 % of these gave the correct answer 'true'. In the control group only 6 % uttered the keyword.

The confidence scores for problem P3 support the view that a majority of the participants who answers true actually has solved the problem correctly; among the students having access to the graph, there was a higher confidence ($M = 5.3$) for students who answered true, which is the correct answer, than for those who answered false ($M = 4.3$). Similarly, for the group not having access to the graph, the confidence ($M = 4.6$) was also higher for those who answered the problem correctly than those who did not ($M = 4.0$). In summary, it seems that participants in the multimedia condition to a large extent actively used the graph to solve this problem, and were also confident about their solutions.

Problem 4: When the graph was most harmful. The graph in P4, related to Gauss formula, illustrates that material being created within a volume is equal to the flow of material through the boundary of that volume. Hence, this interpretation of the Gauss formula is expected to be clearer for the group having access to the graph. In the verbal analysis we found that 25 % of the participants in the multimedia condition commented on this interpretation, while in the control group this number decreased to 19 %. More interestingly, we found that 60 % in the multimedia condition said (something similar to) *this must be correct*, while in the control condition such statements were uttered only by 19 % of the participants.

Turning to the confidence scores we found that, for students in the multimedia condition, there was a higher confidence ($M = 4.4$) among students who answered true, which is the wrong answer, than for those who correctly answered false ($M = 3.0$). On the contrary, in the control condition, the confidence ($M = 4.6$) was higher for those who answered the problem correctly (i.e., false), than those who did not ($M = 4.0$).

Taken together, participants seems to be more likely—and confident—to confirm a statement, when a difficult problem is accompanied by a graph.

Discussion

In this study we investigated the multimedia effect in problem solving at the university level with examples taken from the field of vector calculus. We found no support for an overall multimedia effect (H_1). Instead, graphs had a beneficial effect on performance only when problem statements were to be confirmed (instead of rejected), which is referred to as the picture bias effect (H_2). With respect to H_3, analyses of eye movement data showed that the graphs attracted students' visual attention at the expense of fewer looks toward other parts of the problem. Moreover, spending a proportionally long time inspecting the problem statement as well as frequently moving the eyes between the graph and the problem statement correlated with a higher performance.

Finally, analyses of verbal data provided further insights into why graphs can be both helpful and harmful. It was hypothesized (H_3) that the students would actively use the graph in terms of utterances relating to the graph (H_{3a}) and integration between the graph and other information sources (H_{3b}). Results showed that when a graph was present participants indeed made active use of it, both in terms of selection and integration, and even more so when the graph was helpful. Interestingly, the presence of the graph also influenced the use of other information sources: participants made more use of the statement and evaluated the other data sources more. When the graph was particularly helpful, participants made a more focused (i.e., selecting less information from), but at the same time more efficient use (i.e., building mental models) of the other data sources. Moreover, with a graph, participants evaluated their own knowledge as being higher, confirming a picture biasing effect. When the graph was particularly helpful, though, they reflected more on their own knowledge. Finally, there was a systematic increase of silence in the multimedia condition (H_{3c}), suggesting that students use more mental effort when solving problems that contain graphs.

Beneficial or biasing picture effect?

The graphs we used in the current study were designed to fulfill an interpretational function, that is, to represent complex information presented in text or formulae pictorially, and thereby support students' problem solving processes Levin et al. (1987). We therefore expected to find an overall beneficial effect of adding graphs to problems, but no such effect was present in our data. From a theoretical point of view, the stimuli used in this study were designed in line with the *temporal contiguity principle*, that is, that pictorial and the textual material were presented at the same time. However, it is not fully in line with the *spatial contiguity principle* Mayer (2005b) (also known as the *split attention effect*, Chandler and Sweller 1992). The graph and the explanatory input were given on different parts of the screen and hence might have caused unnecessary visual search of related information, and therefore the absence of a multimedia effect. Split attention may have resulted in that the students invested more mental effort into integrating different parts of the problem, as suggested by the higher proportion of silence in the multimedia condition.

An alternative explanation for not having found a multimedia effect is that participants did not process the textual information, in particular the formula, in the phonological channel. In this way, they would have bypassed the benefits of the dual-processing assumption in working memory. However, post-hoc inspections of the eye-tracking recordings accompanied by the verbal reports of the participants, showed that a vast majority of the participants verbally described what the mathematical formulas contained;

many even read the formulas out loud. Consequently, it is likely that most participants indeed processed the textual information phonologically. Nevertheless, future research should explore when and under which circumstances textual information is actually processed phonologically.

Our results suggest that when seeing a graph, students are more likely to believe in the correctness of the accompanying statements. Students may recognize parts of the input and the graph, and parts (maybe only keywords) of the problem statement and they then say something like "yes this is [*for example*] the triangle inequality, so this must be true". These results are in line with McCabe and Castel (2008), who found that including brain images in an article increased the scientific credibility of the results. They argue that this may be because the brain images "provide a physical basis for abstract cognitive processes". In this study, the graphs rather provide concrete physical interpretations of abstract mathematical formulae. Still, the graphs seem to have a similar persuasive power to affect whether a statement is believed or not.

Processes underlying text-picture integration

An important aspect of processing multimedia material is to select and integrate information relevant for the task (Mayer 2005b). We used eye tracking to investigate how information was visually selected, i.e., where the students looked, for how long, and how information was integrated, that is, how often they transitioned between different problem areas. When the graph was present, students spent about 20 % of their time looking at it. As a result, they paid proportionally less attention to the input and the to-be-confirmed or rejected problem statement. The proportion of time looking at the graph was not related to performance. Interestingly, the more students looked at the problem statement, and the less they looked at the input, the better they performed. Furthermore, the more students switched their attention between the problem statement and the graph, the better they performed. Thus, the mere presence of a graph that is related to the input is not necessarily helpful. Instead, the graph needs to be integrated with the to-be-confirmed or rejected statement.

Analyses of verbal data revealed that in the multimedia condition participants were often more silent in comparison to the control condition without graphs. As silent pauses are indicators of increased mental effort (Yin and Chen 2007; Jarodzka et al. 2015), adding graphs to these problems could have increased the amount of mental effort for students. One explanation to this is the fact that the amount of elements in the task increased (i.e., the intrinsic load). A qualitative analysis of two contrasting problems revealed that in the problem where the graph was beneficial it provided students with a representation that was helpful to solve the problem. In the problem where it was most harmful, the graph itself was correct, but the problem statement was not. Still, the graph convinced the students to confirm the statement.

Implications for theory and educational practice

As a practical consequence, we can conclude that when including graphs in textbooks, it should be ensured that students first and foremost know exactly what their task is (here: confirm or reject problem statements) to know how to use these graphs. Next, they should always ensure to keep the task itself in mind by integrating the task formulation and the graph. Thus, when designing textbooks, it could be important to consider these integration processes. Future work should investigate different way to facilitate integration by e.g., referring to the graph in the statement and maybe even back from the graph to the problem statement.

Furthermore, implications can be also drawn for theory. Mayer (2005b) theory of processing information of multimedia clearly describes an optimal scenario, where students actively process all given information, by selecting the relevant information, organizing and integrating it. However, in line with other research (e.g., Holsanova et al. 2009), our study showed that students may simply not take the effort to actively process information and instead use a rather shallow processing strategy (e.g., assuming that when the graph is correct, the rest of the task must also be). In that, pictures could even support such a shallow and misleading processing. The CTML does acknowledge that this optimal way of processing can be hampered by different layout decisions and has thus formulated several design guidelines. Based on the findings in this paper, we suggest that the influence of a picture bias effect should be considered carefully alongside such guidelines.

Limitations and conclusions

It is evident from discussions we had with students after the test, that the experiment does not precisely reflect how they normally work with problems of this type at home, in the classroom, or at examinations. First, the time to solve a problem was limited and rather short. Such time pressure may lead to more shallow information processing, and therefore a greater picture bias. Second, they were not allowed to use pen and paper to scribble formulas and figures to organize their problem solving processes. Finally, these students are typically not exposed to problems where statements need to be falsified, in particular when the information is not presented in their native language. The implications of using this rather uncommon format for providing the answer need further investigation. This makes it challenging to construct suitable problems and graphs for these types of studies. Nevertheless, the format of the test is still common in other domains.

From the eye-tracking data and the verbal reports, examples of deep processing, such as building a rich mental model, of the information included in the problems were observed. However, the current data analysis does not allow for concrete evidence. Future research should investigate this issue in a qualitative manner.

In summary, graphs were not found to be beneficial per se in the experiment. Only when they were carefully framed and integrated with the problem statement they had a beneficial effect on performance. Otherwise, when the graphs were correct by themselves, they mislead the students to trust the problem statements. Either way, the graphs produced an increase in mental effort. Before including graphs in mathematical texts, teachers and textbook designers should very carefully consider their function and how they integrate with other parts of the information in the problem.

Appendix: Coding scheme of verbal data

See Table 10.

Table 10 Coding scheme for verbal data

Main category	Sub-category	Code	Description of code	Exemplary statement
Search of information from...	Input	S1T	Looking searching selecting something specific in the input or reading the input. Nothing more is uttered than information given in the input	"A vector function R ..." (Reading input)
	Statement	S1S	Looking / searching/selecting something specific in the statement. Nothing more is uttered than information given in the statement	"Then it follows..." (Reading the statement)
	Graph	S1G	Looking/searching/selecting something specific in the graph. Simply describing literally what is presented in the graph. Nothing more is uttered than information given in the graph	"The tangent to the sphere"
Building of mental model	Verbal (input)	MMVT	Integrating information within the input	"Because it is the same vector"
	Verbal (statement)	MMVS	Integrating information within the statement	"If R is perpendicular to the derivative"
	Graphical	MMG	Integrating information within the graph	"Or a curve on the surface"
Prior knowledge activation		PK	New things/issues/facts are mentioned that are not displayed on the computer screen	"The derivative can be decomposed into..."
Integration of information	Input and graph	ITG	Connecting/integrating information given in the input with information given in the graph	"Since the derivative is constant it has to be..."
	Input and statement	ITS	Connecting/integrating information given in the input with information given in the statement	"The square of the length is constant, such that..."
	Graph and statement	IGS	Connecting/integrating information given in the graph with information given in the statement	"The radius is perpendicular to..."
Problem solution	Input	EPS	Uttering, whether the entire statement "Q" is true or false	"So it is true"
	Statement	TS	Uttering, whether something about the input is true or false	"That is natural"
	Statement	SS	Uttering, whether something about the statement is true or false	"It seems to be a correct formula"
	Graph	GS	Uttering, whether something about the graph is true or false	"So it has to be..."

Table 10 continued

Main category	Sub-category	Code	Description of code	Exemplary statement
Metacognitive statements ('thinking about one's own thinking')	Evaluating difficulty of task (Positive)	MCT+	Expressing judgment about high complexity of the task	"This is difficult"
	Evaluating difficulty of task (Negative)	MCT−	Expressing judgment about low complexity of the task	"This was not strange at all"
	Evaluating own knowledge (Positive)	MSO+	Expressing positive judgment about own abilities and knowledge or about correctness of own task approach	"Now I get it"
	Evaluating own knowledge (Negative)	MSO−	Expressing judgment about lack of one's own knowledge and skills or about mistakes made	"I have no idea"
	Goal orientation	GO	Uttering the goal of the task.	"Then the question is…"
Off-topic / not understandable		OT	Anything that does not fit the coding schema or is not understandable	"Large graph!"

References

Andrà, C., Arzarello, F., Ferrara, F., Holmqvist, K., Lindström, P., Robutti, O. & Sabena, C. (Eds.). (2009). How students read mathematical representations: An eye tracking study: *Proceedings of the 33rd Conference of the International Group for the psychology of mathematics education* (Vol. 2, pp. 49–56).

Atkinson, R. C. & Shiffrin, R. M. (1968). The psychology of learning and motivation: Advances in research and theory. In K. W. Spence & J. T. Spence (Eds.), (pp. 89–191). New York: Academic Press.

Baayen, R. H. (2011). languager: Data sets and functions with "analyzing linguistic data: A practical introduction to statistics". [bibcomputer software manual]. http://CRAN.R-project.org/package=languageR R package version 1.4

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using s4 classes [Computer software manual]. http://CRAN.R-project.org/package=lme4 R package version 0.999999-0

Berends, I. E., & van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem solving: Effects of increased cognitive load. *Learning and Instruction*, *19*, 345–353.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*, 53–61.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Research Review*, *14*, 5–26.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*, 233–246.

Crisp, V., & Sweiry, E. (2006). Can a picture ruin a thousand words? the effects of visual resources in exam questions. *Educational Research*, *48*, 139–154.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge: MIT Press.

Fletcher, J. D. & Tobias, S. (2005). The cambridge handbook of multimedia learning. In R. E. Mayer (Ed.), (pp. 117–134). Cambridge: Cambridge University Press.

Griffiths, D. (1999). *Introduction to electrodynamics*. New Jersey: Prentice Hall.

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

Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology*, *23*, 1215–1226.

Isberner, M.-B., Richter, T., Maier, J., Knuth-Herzig, K., Horz, H., & Schnotz, W. (2013). Comprehending conflicting science-related texts: graphs as plausibility cues. *Instructional Science*, *41*, 849–872.

Jarodzka, H., Janssen, N., Kirschner, P. A., & Erkens, G. (2015). Avoiding split attention in computer-based testing: Is neglecting additional information facilitative? *British Journal of Educational Technology*, *46*, 803.

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science*, *41*, 811–831.

Levin, J. R. , Anglin, G. J. & Carney, R. N. (1987). The psychology of illustration. In D. M. Willows & H. A. Houghton (Eds.), (Vol. 1, pp. 51–85). New York: Cambridge University Press.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, *81*, 240–246.

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

Mayer, R. E. (2001). *Multimedia learning*. University Press. http://books.google.nl/books?id=ymJ9o-w_6WEC

Mayer, R. E. (2005). The cambridge handbook of multimedia learning. R. E. Mayer (Ed.), (pp. 31–48). Cambridge: Cambridge University Press.

Mayer, R. E. (2005). The cambridge handbook of multimedia learning. R. E. Mayer (Ed.), (pp. 183–200). Cambridge: Cambridge University Press.

McCabe, D., & Castel, A. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*, 343–352.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

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

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

Park, B., & Brünken, R. (2015). The rhythm method: A new method for measuring cognitive load–an experimental dual-task study. *Applied Cognitive Psychology*, *29*, 232–243.

Persson, A. & Böiers, L. (1988). *Analys i flera variabler*. Studentlitteratur.

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. http://www.R-project.org (ISBN 3-900051-07-0)

Ramgard, A. (1996).*Vektoranalys* (2nd ed.). Teknisk högskolelitteratur i Stockholm.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychological Review*, *10*, 251–296.

van Gog, T., Paas, F., van Merriënboer, J. J., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, *11*, 237.

van Someren, M., Barnard, Y., & Sandberg, J. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.

Yin, B. & Chen, F. (2007). Towards automatic cognitive load measurement from speech analysis. *Proceedings of the 12th International Conference on Human-computer Interaction: Interaction design and usability* (pp. 1011–1020). Berlin: Springer