

Lindner, Marlit Annalena; Eitel, Alexander; Strobel, Benjamin; Köller, Olaf  
**Identifying processes underlying the multimedia effect in testing: An eye-movement analysis**

*Learning and instruction* (2017) 47, S. 91-102



Quellenangabe/ Reference:

Lindner, Marlit Annalena; Eitel, Alexander; Strobel, Benjamin; Köller, Olaf: Identifying processes underlying the multimedia effect in testing: An eye-movement analysis - In: *Learning and instruction* (2017) 47, S. 91-102 - URN: urn:nbn:de:0111-pedocs-127949 - DOI: 10.25656/01:12794

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-127949>

<https://doi.org/10.25656/01:12794>

**Nutzungsbedingungen**

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**Terms of use**

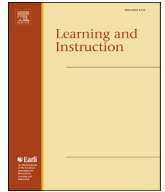
This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor.

By using this particular document, you accept the above-stated conditions of use.



**Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)



# Identifying processes underlying the multimedia effect in testing: An eye-movement analysis



Marlit Annalena Lindner<sup>a,\*</sup>, Alexander Eitel<sup>b,c</sup>, Benjamin Strobel<sup>a</sup>, Olaf Köller<sup>a</sup>

<sup>a</sup> Leibniz Institute for Science and Mathematics Education (IPN), Kiel, Germany

<sup>b</sup> University of Freiburg, Freiburg, Germany

<sup>c</sup> Leibniz-Institut für Wissensmedien (IWM), Tübingen, Germany

## ARTICLE INFO

### Article history:

Received 1 March 2016

Received in revised form

14 October 2016

Accepted 17 October 2016

Available online 16 November 2016

### Keywords:

Multimedia effect in testing

Multiple external representations

Eye-tracking

Cognition in assessment

## ABSTRACT

Test items become easier when a representational picture visualizes the text item stem; this is referred to as the multimedia effect in testing. To uncover the processes underlying this effect and to understand how pictures affect students' item-solving behavior, we recorded the eye movements of sixty-two schoolchildren solving multiple-choice (MC) science items either with or without a representational picture. Results show that the time students spent fixating the picture was compensated for by less time spent reading the corresponding text. In text-picture items, students also spent less time fixating incorrect answer options; a behavior that was associated with better test scores in general. Detailed gaze likelihood analyses revealed that the picture received particular attention right after item onset and in the later phase of item solving. Hence, comparable to learning, pictures in tests seemingly boost students' performance because they may serve as mental scaffolds, supporting comprehension and decision making.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Visualizations are frequently integrated into text-based test items in large-scale assessments (LSA; e.g., *Programme for International Student Assessment* [PISA]; OECD, 2013), yet little is known about how multimedia elements affect cognitive processing in item solving. Recent studies have shown that adding a *representational picture* (i.e., a picture that displays task-relevant information stated in the text; in the following referred to as *picture*) to a text-based item stem reduces item difficulty (Hartmann, 2012; Lindner, Ihme, Saß, & Köller, in press; Saß, Wittwer, Senkbeil, & Köller, 2012). Similar to the findings of many studies in the learning context (see, e.g., Butcher, 2006 for a review), items become easier when representational pictures are added to the text. Thus, it seems that the so-called multimedia effect (e.g., Mayer, 2009) can be transferred from learning to testing situations (i.e., multimedia effect in testing; cf. Lindner et al., in press). However, not much is known about the cognitive processes that underlie a decrease in item difficulty due to the addition of a picture to text-based

multiple-choice (MC) items. In the following, we refer to multimedia theories of *learning* to derive informed hypotheses about how pictures may affect item processing in the *testing* context.

### 1.1. Learning with text and picture

A large number of studies provide evidence that students' performance improves when they learn with text and pictures rather than learning with text alone (see e.g. Ainsworth, 2006; Butcher, 2014; Mayer, 2009 for reviews). To explain why a combination of text and picture is effective for learning, the *Cognitive Theory of Multimedia Learning* (CTML; Mayer, 2009, 2014) postulates that information from text and picture is processed in two separate channels in working memory, while each one is limited in its capacity. When using both channels simultaneously, for example, by learning with (spoken) text and pictures, the limited working-memory capacity is better exploited. Furthermore, the CTML refers to a more active processing of information when pictures are added to a text, because of the necessity to integrate the information from the two resulting mental models (a verbal and a pictorial model) with the help of prior knowledge that is retrieved from long-term memory. Moreover, the necessity to integrate multiple representations can foster the abstraction of general rules and can,

\* Corresponding author. Leibniz Institute for Science and Mathematics Education, Olshausenstraße 62, 24118 Kiel, Germany.

E-mail address: [mlindner@ipn.uni-kiel.de](mailto:mlindner@ipn.uni-kiel.de) (M. A. Lindner).

therefore, facilitate conceptual understanding at a higher level of organization (Ainsworth, 2006).

According to the *Integrated Model of Text and Picture Comprehension* (ITPC; Schnotz & Bannert, 2003; Schnotz et al., 2014), when processing a text, the learner first constructs a mental representation of the text's surface structure, from which a propositional representation of the semantic content is generated (cf. Van Dijk & Kintsch, 1983). A mental model can be constructed from a text only by generating a propositional representation as an intermediate step. This requires students to infer relations that are sometimes only implicitly expressed in the text (cf. Glenberg & Langston, 1992), which is subject to erroneous interpretations, so that the mental model may inadequately reflect the contents or situations described. In contrast, when processing pictures, the perceptual representation of the picture's visuospatial relations can be mapped onto semantic relations to provide the structure of the mental model (analogical structure mapping; Schnotz & Bannert, 2003). This has unique benefits for comprehension, for instance, because the picture as a second representation can disambiguate potentially ambiguous text (i.e., constraining interpretation function; Ainsworth, 2006). Hence, presenting a picture in addition to a text means that the mental model construction (or interpretation) from the text can be facilitated or even replaced by a mental model construction based on the picture. Especially early attention on the picture—which has been found in studies with simultaneous text-picture presentation (e.g., Lenzner, Schnotz, & Müller, 2013)—can provide learners with the structure of a mental model so that part of the mental model construction is already completed based on the picture (cf. scaffolding assumption; Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013). As a consequence, subsequent processes of mental model construction from the text are facilitated. This assumption is supported by studies that have provided evidence for a link between early picture processing, facilitated text processing, and better learning outcomes (see e.g., Eitel & Scheiter, 2015, for a review).

However, while learning with text and picture has great potential, it can also be too demanding to yield the desired outcomes. In particular, successful learning with text and picture requires students to integrate information both within and between representations. Therefore, because students often tend to treat representations in isolation, beneficial outcomes do not always occur (see Ainsworth, 2006, for a review). Fortunately, integration behavior can be facilitated and, thus, learning supported when key principles of effective multimedia design are applied (see Mayer, 2014, for an overview) — for instance, when split attention is avoided because text and pictures are presented in close spatial and temporal proximity (e.g., Ayres & Sweller, 2005; Brünken, Plass, & Leutner, 2004).

Even though these assumptions were originally proposed for *learning* with text and picture, they may extend to the case of *testing* with text and picture. We will discuss this issue in the next section.

## 1.2. Testing with text and picture

Reading to understand is an important process in which pictures can be helpful – not only in learning but also in testing situations. In testing situations, students usually have to understand the problem that is presented in the test item stem to be able to solve the item correctly. Especially when test items require conceptual problems to be solved by transferring declarative domain knowledge to a more applied situation, which is often the case in LSA (cf. e.g., IEA, 2013; Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009; OECD, 2013), it is crucial to develop a coherent mental model of the problem in order to draw the right inferences and find the correct solution. Thus, it is assumed that theoretical assumptions

from learning research about how pictures facilitate text understanding (e.g., Schnotz & Bannert, 2003) can be generalized to item solving (see also Jarodzka, Janssen, Kirschner, & Erkens, 2015; Lindner et al., in press; Saß et al., 2012). If multimedia principles are accounted for (cf. Mayer, 2014), presenting a picture in addition to a text probably facilitates comprehension of the problem situation because it may help students to disambiguate or abstract information from a text and facilitate mental model construction, also in testing situations.

Empirical studies that investigated younger test-takers from the fourth to the ninth grade so far tentatively support these assumptions. On the one hand, there is accumulating evidence that pictures reduce item difficulty in the context of multiple-choice assessment (e.g., Hartmann, 2012; Saß et al., 2012), and this effect has moreover been shown to be highly stable across students and items by Lindner et al. (in press). On the other hand, response time measures in a computer-based study showed that students did not spend more time on solving items that contained a picture in the item stem compared to corresponding text-only items (Saß et al., 2012). In this initial result on the processing of text and pictures, the comparable response time measures for text-only and text-picture items indicate that the time used to process a picture is compensated for by an acceleration of the item-solving process. However, the question concerning the stage at which this takes place remains open. Potential answers to this question are presented in the following section.

## 1.3. Processing text and picture in MC testing

In MC testing, eye-tracking is a suitable method for revealing cognitive processes because the components of MC items are typically displayed in distinct locations on the screen – with the item stem and question on top, and the answer options below (see Fig. 1). Accordingly, Lindner et al. (2014) found that students mainly processed the question at the beginning and the answer options, particularly the subjectively preferred and chosen answer options, towards the end of the decision-making process. According to their findings, the item-solving process can be roughly divided into two phases: (1) an *information-acquisition phase*, in which students construct a mental representation of the problem or situation described in the item stem, and (2) a *decision-making phase*, in which students evaluate the answer options with respect to their likelihood of being the correct answer, followed by the actual choice of an answer.

Referring to multimedia learning theories (e.g., Mayer, 2009; Schnotz & Bannert, 2003), the picture should especially help in the information-acquisition phase; that is, when trying to construct a representation of the problem or situation stated in the item stem. Since the picture represents the problem as the text does, both media can be used to construct a mental model of the situation or problem, but a mental model can be constructed more directly based on the picture (analogical structure mapping; Schnotz & Bannert, 2003). Therefore, the processing of the text item stem should be facilitated by the picture, which can be expected to receive initial attention (cf. Lenzner et al., 2013). In consequence, less time should be required to read the text. A reduction in text-processing time should, however, go hand in hand with early picture processing, allowing students to initially extract the gist of the item and to thereby reduce the time required to encode and understand the corresponding verbal information (i.e., scaffolding; Eitel et al., 2013). Nevertheless, the processing time required for the picture might still outweigh any time saved by a shorter processing of the text, resulting in a longer overall engagement time with the item stem information in text-picture compared to text-only items. This could be desirable, as spending more time and using different

media to evaluate the presented problem may further promote students' understanding and the consolidation of central concepts and may prevent them from overlooking important task aspects.

Due to a more elaborated mental representation of the problem, as well as the externalization of important information (cf. Schnotz & Kürschner, 2008), pictures may also aid logical reasoning and decision making, thereby speeding up the answering process. This assumption is supported by the findings of Lindner et al. (2014), which revealed that, although the attention devoted to the MC item stem decreased across time, the item stem still received considerable attention in the later phase of decision making. Such reviewing behavior seems to support students in choosing an answer option, because re-checking whether the internal problem representation and the preferred answer option fit to the described problem is certainly beneficial in making a decision. In accordance with this, a picture may be particularly helpful in a later reviewing process, because its contents can be extracted more easily from its visual appearance than re-reading the text of the item stem (Ainsworth, 2006). Such a process is described by Schnotz et al. (2014) as mental model updating. Hence, we assume that if a picture is presented in addition to the text in the item stem, students review the picture rather than the text in the course of the decision-making phase prior to choosing an answer option.

Furthermore, the so-called *gaze bias effect* from decision-making research (i.e., longer fixation on subjectively preferred decision options; Shimojo, Simion, Shimojo, & Scheier, 2003) was recently successfully transferred to the context of MC testing, which reveals that there is a valuable link between students' proficiency levels and their time spent on fixating objectively correct and incorrect answer options in MC items (Lindner et al., 2014). Thus, if pictures enhance students' level of understanding in testing (i.e., *subjectively* preferring *objectively* correct answers), students' should be more capable of dismissing the incorrect answer options. According to this, distractors should receive less attention in text-picture items compared to text-only items.

#### 1.4. Present research and hypotheses

The main goal of this study was to investigate students' processing of text and pictures in assessment situations in order to better understand processes that may underlie a multimedia effect in testing. Therefore, we analyzed students' response correctness, response times, and their eye movements during the MC item-solving process as indicators for attentional and potentially cognitive operations that may provide insights into the way in which students interact with the multimedia material (cf. e.g., Eitel, 2016; Mason, Tornatora, & Pluchino, 2013, 2015). Our hypotheses are formulated on both an outcome and a process level:

- (1) *Multimedia effect*: Adding a picture to a text-based item reduces the item's difficulty.
- (2) *Facilitated text processing*: Adding a picture to a text-based item reduces the processing time of the item stem's text, but increases the time students spend engaged with the overall text and picture information in the item stem (problem representation).
- (3) *Faster decision making*: Decision making is facilitated by the picture, and this is reflected in a shorter processing of the answer options (especially the incorrect answer options).
- (4) *Picture processing across solution time*: Students are expected to direct their attention to the picture shortly after item onset, in the information-acquisition phase, but to review the picture in the decision-making phase in order to identify the correct answer option.

## 2. Method

### 2.1. Participants and study design

The sample comprised  $N = 62$  students from the fifth and sixth grades ( $M_{\text{age}} = 10.8$ ; 36% female) from both higher track and lower track schools in the northern part of Germany who participated voluntarily in their free time. Students were randomly assigned to one of two groups that answered items with an inverse alternation of text-only and text-picture conditions in order to implement a within-subject design. Pictures were integrated into even items in the first group ( $n = 31$ ) and into uneven items in the second group ( $n = 31$ ). The sequence of the test items remained the same, so that both groups solved the same tasks. This design was used to reduce error variance, which could stem from general differences in students' viewing and item-solving habits, but also from potential differences in their background variables, such as domain knowledge, gender, and motivation or reading abilities. The design thus allowed for an optimal comparison regarding effects resulting from the experimental manipulation (*text-only* vs. *text-picture*) because irrelevant background variance at a person level was reduced to a minimum. Furthermore, the alternation of experimental conditions across the tasks ensured that a student did not answer an item twice under both conditions, while still providing data for text-only and text-picture items in the study.

### 2.2. Materials and measures

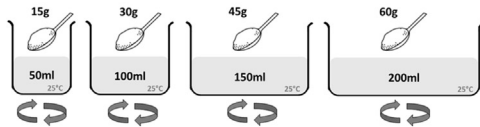
#### 2.2.1. Multiple-choice test

Students answered 24 MC items in the domain of science. The items dealt with basic concepts from biology (e.g., food chains), physics (e.g., magnetism) and chemistry (e.g., solubility of substances) and in some cases also required knowledge about more general science principles (e.g., basic experimental principles). In several items, students needed to integrate their knowledge in order to understand how different concepts are related to each other (e.g., mass and volume). All items required students to *apply* their declarative domain knowledge to everyday situations or to simple experimental settings that were outlined in the item stem. Thus, students needed to understand the situation described in the item stem in order to be able to correctly answer the corresponding question. To ensure that the materials had sufficient content validity, we constructed test items according to the framework of the *Trends in International Mathematics and Science Study* (TIMSS; Mullis et al., 2009) and adapted several science items that were applied in the TIMSS 2011 study (cf. IEA, 2013). The item adaption was necessary in order for all items to have the same formal structure, i.e., all items were presented in an MC format that comprised a short item stem, a one-sentence question, and four answer options with one correct option (single-choice format). An example item is displayed in Fig. 1.

As the experimental factor, we either added a representational picture to the item stem, or we did not. Most pictures stemmed from TIMSS, while some additional pictures were self-designed in accordance with the following criteria: (a) All the pictures were schematic images in grey shades. (b) Every item was constructed so that the picture *illustrated* the information given in the text, but never provided solution-relevant information beyond the text information (i.e., *multiple representations*; cf. Ainsworth, 2006). Thus, both text-only and text-picture items had equivalent information and were, apart from the added pictorial element, perfectly parallel. (c) To avoid split attention in the text-picture condition, the picture was always presented in close spatial proximity to the item stem text (placed between the text and the question; cf. Fig. 1) and, therefore, was also in close spatial proximity to the answer options.

## TEXT-PICTURE ITEM

Maria designed an experiment using salt and water. She experienced that 15g salt dissolved in 50ml water, 30g salt in 100ml water, 45g salt in 150ml water and 60g salt in 200ml water. The water's temperature was always 25°C. Maria stirred every mixture several times.



What was Maria studying in her experiment?

- Ⓐ How much salt will dissolve in different volumes of water.
- Ⓑ How much salt will dissolve at different temperatures.
- Ⓒ If stirring increases how fast salt will dissolve.
- Ⓓ If stirring decreases how fast salt will dissolve.

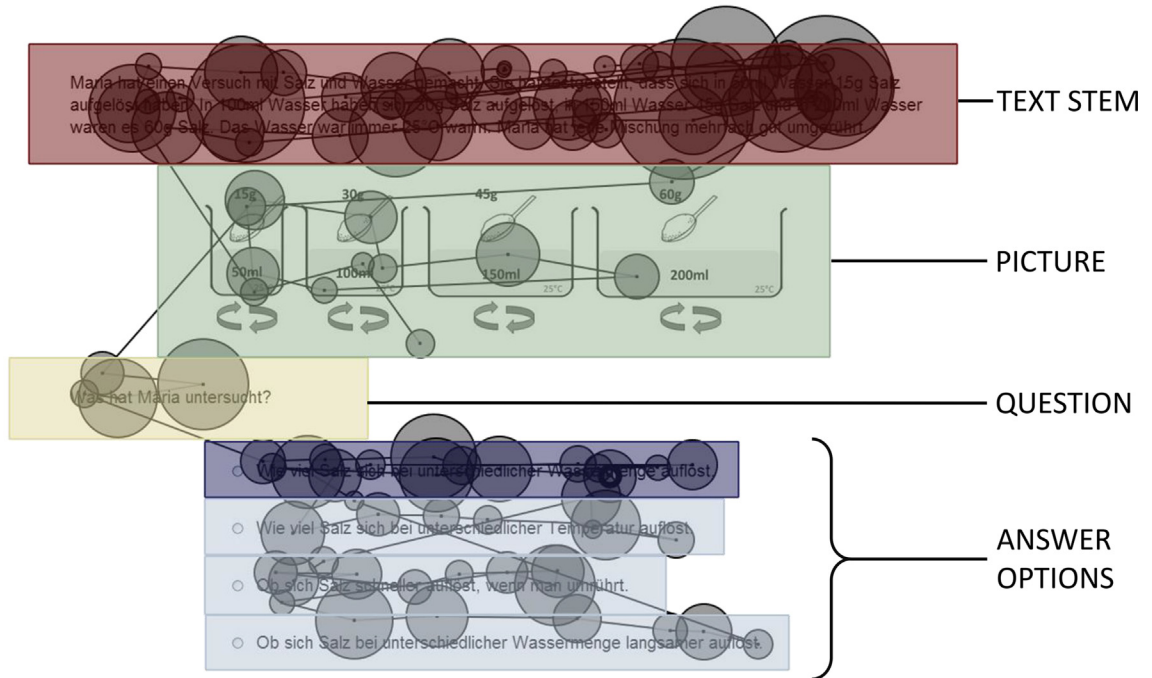
## TEXT-ONLY ITEM

Maria designed an experiment using salt and water. She experienced that 15g salt dissolved in 50ml water, 30g salt in 100ml water, 45g salt in 150ml water and 60g salt in 200ml water. The water's temperature was always 25°C. Maria stirred every mixture several times.

What was Maria studying in her experiment?

- Ⓐ How much salt will dissolve in different volumes of water.
- Ⓑ How much salt will dissolve at different temperatures.
- Ⓒ If stirring increases how fast salt will dissolve.
- Ⓓ If stirring decreases how fast salt will dissolve.

## SCANPATH EXAMPLE



**Fig. 1.** Left: Example of an item (material adapted from IEA, 2013) with the experimental manipulation (text-only vs. text-picture); Right: Areas of interest and an example of eye movements (scanpath) from one participant solving the item as a text-picture version.



To assure that the pictures did not contain any solution-relevant information that was not also stated in the text, three experts with a professional background in education rated the items independently from each other regarding this criterion before the items were included in the study. Focusing on the information equivalence between the text and the picture, the raters concordantly confirmed that all items met the study's criteria.

### 2.2.2. Background variables

We controlled for students' age, science grade, general cognitive abilities, and reading abilities (cf. Table 1). Background variables were collected in a demographic questionnaire. General cognitive abilities were assessed by the N2 subtest of the KFT 4–12 + R (Heller & Perleth, 2000) and students' reading ability was assessed by using the German reading comprehension test ELFE 1–6 (Lenhard & Schneider, 2006) in a computer-administered version.

### 2.3. Apparatus

Test items were presented on a 22-inch screen with a 1680 × 1050 pixel resolution, using the software Experiment Center 3.1 from SensoMotoric Instruments (SMI; Teltow, Germany). Every item was presented on a single page on the screen. While working on the test, participants sat in front of the screen at a distance of approximately 70 cm from the screen. Participants' eye movements were recorded using a video-based eye-tracking system (SMI iView X<sup>TM</sup> RED; 120 Hz sampling rate) and the corresponding SMI Software iView X<sup>TM</sup>. The system was calibrated using an animated 8-point calibration image and subsequent validation. Mean calibration accuracy was  $M_x = 0.46$  degrees of visual angle for  $x$  and  $M_y = 0.36$  degrees of visual angle for  $y$  coordinates ( $min = 0.11$ ;  $max = 0.84$ ).

### 2.4. Procedure

Students were tested in single sessions. Before the test, they answered a paper-pencil demographic questionnaire. After completing the questionnaire, they were familiarized with the procedure and the eye-tracking system and completed a short example test item to see how to log in their answers. Students were informed that they would not be able to return to an earlier question and had no time constraints for the task. Furthermore, they were instructed to choose exactly *one* answer to each item and to make a well-considered guess if they did not know the correct answer. The eye-tracking system was calibrated before the MC test started. To avoid placing excessive demands on the students, they were invited to a separate session to complete the cognitive ability test and the reading abilities test. All students participated in both sessions and provided complete data sets.

## 2.5. Analyses

### 2.5.1. Eye-movement data pre-processing

Eye movements were analyzed with a dispersion-based algorithm. A fixation was detected when eye movements lasted for at least 80 ms in a position with a maximum dispersion of 100 pixels (equals about 2° visual angle, considering the specifics of our hardware and the software settings). These thresholds were chosen to achieve a valid separation of eye-movement events (cf. e.g., Holmqvist et al., 2011). For each item, seven separate rectangular *areas of interest* (Aols) were placed, encompassing the item stem, the picture, the question, each of the four answer options, and the remaining white space (see Fig. 1). A margin was added to the text to account for data inaccuracy and the failure of participants to look directly at the text. We used total fixation time as gaze data. Fixation time is defined as the sum of all consecutive fixations on an area of interest (Aol), indicating the time for which attention was directed at this area, while total fixation time cumulates the duration of all Aol fixations from task onset until task end (cf., Holmqvist et al., 2011). All reported fixation times are specified in units of milliseconds, if not indicated otherwise.

### 2.5.2. Mixed-effects models

We applied linear mixed-effects models (LMMs; for introductions, see, e.g., Barr, 2008; Snijders & Bosker, 2012; Quené & van den Bergh, 2008) when testing the significance of fixation time differences between Aols in text-only and text-picture items. We used this analysis approach to account for the cross-classified multilevel structure of our data: The multilevel structure was present because the implemented within-subject design forced the students from the two experimental groups to answer the same items (under varying conditions, i.e., *text-picture* vs. *text-only*). Therefore, responses from a person to different items and also responses from different persons to the same item were not completely independent. More precisely, the measures were *nested* within students and also within items (cf. e.g., Baayen, Davidson, & Bates, 2008; Snijders & Bosker, 2012). Thus, the data structure was not strictly hierarchical but cross-classified, which means that measures can be classified by both factors (items and students) at the same time. In such situations, LMMs deliver more valid standard error estimates than common analysis of variance or *t*-tests because they explicitly take the dependence in the data into account and, as a consequence, deliver more valid statistical tests. This is because LMMs allow *fixed effects* and *random effects* to be modeled separately. While fixed effects aim to identify *typical* rates of change in the outcome variable (e.g., following an experimental manipulation), random effects aim to identify *unsystematic* variance related to sample characteristics: here, for example, at the level of items and students.

We modeled students' fixation time (1) as a function of dummy-

**Table 1**  
Means and standard deviations for several background variables in the two experimental groups.

	Group 1 <i>N</i> = 31; <i>M</i> ( <i>SD</i> )	Group 2 <i>N</i> = 31; <i>M</i> ( <i>SD</i> )	<i>p</i> -value ( <i>t</i> -value <sub>df</sub> ) <sup>a</sup>	Cohens' <i>d</i>
Age	10.81 (0.83)	10.84 (0.66)	.87 (0.17 <sub>60</sub> )	0.04
Science Grade <sup>a</sup> (max. 6)	2.40 (0.67)	2.30 (0.87)	.62 (0.49 <sub>58</sub> )	0.12
General Cognitive Abilities (max. 25)	18.03 (5.41)	17.29 (6.19)	.61 (0.50 <sub>60</sub> )	0.13
General Reading Abilities <sup>b</sup> (max. 100)	56.96 (24.06)	57.91 (21.20)	.87 (0.16 <sub>59</sub> )	0.04
Reading Speed (max. 100)	72.97 (34.75)	80.21 (33.48)	.42 (0.82 <sub>60</sub> )	0.21
Overall MC Test Score (max. 24)	13.90 (3.92)	15.19 (4.61)	.24 (1.18 <sub>60</sub> )	0.30
Solution Time per Item (sec)	39.94 (8.14)	41.75 (13.28)	.53 (0.63 <sub>60</sub> )	0.16
Overall Testing Time (min)	15.92 (3.24)	16.71 (5.33)	.49 (0.70 <sub>60</sub> )	0.18

<sup>a</sup> Two students did not provide science grade information.

<sup>b</sup> One measure of reading abilities was lost due to a technical problem.

coded variables representing the Aols, and (2) by a variable that coded the picture manipulation (both as fixed effects). This model allowed for a comparison of students' fixation times on Aols in the experimental conditions. Every model contained a fixed intercept  $\gamma_{(00)}$  as well as a random intercept  $u_{(j0)}$  for students and a random intercept  $v_{(0k)}$  for items, in order to address the data structure. Accordingly, the expected fixation time (i) for a certain student (j) in a specific item (k) without predictors can be represented as (cf. Quené & van den Bergh, 2008):

$$(1) \text{FixTime}_{i(jk)} = \gamma_{(00)} + (u_{(j0)} + v_{(0k)} + e_{i(jk)}).$$

To model the effects of interest, we added the experimental condition (picture) and dummy-coded variables for the different Aols as fixed effects to the models. To investigate both basic Aols (text, question, correct and incorrect options) and aggregated Aols (overall item stem and answer options), we tested two separate models to ensure independence of the explanatory factors. Model M1 was specified as:

$$(2) \text{FixTime}_{i(jk)} = \gamma_{(00)} + \gamma_{\text{Picture}(00)}\text{Picture} + \gamma_{\text{Text}(00)}\text{Text} + \gamma_{\text{Picture*Text}(00)}\text{Picture*Text} + \gamma_{\text{Question}(00)}\text{Question} + \gamma_{\text{Picture*Question}(00)}\text{Picture*Question} + \gamma_{\text{Correct}(00)}\text{Correct} + \gamma_{\text{Picture*Correct}(00)}\text{Picture*Correct} + \gamma_{\text{Incorrect}(00)}\text{Incorrect} + \gamma_{\text{Picture*Incorrect}(00)}\text{Picture*Incorrect} + (u_{(j0)} + v_{(0k)} + e_{i(jk)})$$

Correspondingly, Model M2 was specified as:

$$(3) \text{FixTime}_{i(jk)} = \gamma_{(00)} + \gamma_{\text{Picture}(00)}\text{Picture} + \gamma_{\text{ItemStem}(00)}\text{ItemStem} + \gamma_{\text{Picture*ItemStem}(00)}\text{Picture*ItemStem} + \gamma_{\text{Answers}(00)}\text{Answers} + \gamma_{\text{Picture*Answers}(00)}\text{Picture*Answers} + (u_{(j0)} + v_{(0k)} + e_{i(jk)})$$

It should be noted that the principle for coefficient (fixed effect) interpretation is comparable to common regression analyses (interpretation of one predictor when all other predictors equal zero). We chose the Aol *white space* as a reference category in the dummy coding for both models, because this area was not expected to be influenced by the experimental manipulation and should usually receive the least amount of attention. Thus, for example, the main effect of the picture represents the difference in fixation time on the white space between items with and without a picture (i.e., when all other dummy variables equal zero). Similarly, the interaction effects denote how the picture's effect differs across Aols. However, the effects we were most interested in—the difference in fixation time on an Aol between items with and without a picture—are not represented in the models by a single coefficient, but rather as a combination of main effects and interaction effects (i.e., *picture + picture × Aol*). Therefore, another analysis step was needed, in which we applied the delta method (Fox & Weisberg, 2011) to test the contrasts of interest by defining linear constraints on the estimated parameters (i.e., *picture + picture × Aol*). By doing this, we were able to compare the fixation time on the Aols in text-only and text-picture conditions in the models we defined.

### 2.5.3. Gaze likelihood analysis

To determine how much attention students devoted to different item areas across the solution time, we programmed an algorithm to divide the item-solving process into ten equal time intervals for every item and every student according to the seven Aols (see Fig. 1). This made it possible to conduct an adapted gaze likelihood analysis (cf. Lindner et al., 2014) to compare the mean time course of item processing across all students and items. For each time

interval in each item, the mean total fixation time was determined. This procedure was first carried out at a student level; the student means were then aggregated separately for text-only items and text-picture items.

### 2.5.4. Software

We used the software BeGaze™ 3.0 (SMI) to determine eye-movement parameters and SPSS® 19 for further data preparation and descriptive analyses. The 'lme4' package (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2015) was applied to estimate LMM models and to test linear constraints (*delta method*; package 'car'; Fox et al., 2015; Fox & Weisberg, 2011), while R was also used to implement the gaze likelihood analysis.

## 3. Results

### 3.1. Control variables

As a randomization check, we compared performance-related background variables, overall solution rates, item solution times, and overall testing time in the two randomized groups of students, using independent *t*-tests (see Table 1). All comparisons were nonsignificant ( $t(60) < 1.18$ ), while descriptive means varied in an acceptable range, which was also reflected by overall small effect sizes ( $M_d = 0.15$ ;  $\text{Range}_d = 0.04\text{--}0.30$ ).

### 3.2. Test outcome

Item difficulty was calculated for every item, in accordance with classical test theory, as the proportion (*p*) of participants who got the item correct. The reliability of the test (Cronbach's  $\alpha$ ) was sufficiently high in both of the groups ( $\alpha_{G1} = 0.69$ ;  $\alpha_{G2} = 0.79$ ). The overall item difficulty ranged between  $p_{\min} = 0.10$  and  $p_{\max} = 0.87$ , with a mean difficulty of  $M = 0.61$  ( $SD = 0.16$ ). Comparing text-only (TO) and text-picture (TP) items, we found the expected decrease in item difficulty (i.e., *multimedia effect in testing*) for items that contained a picture in the item stem compared to parallel text-only items ( $M_{TO} = 0.56$ ,  $SD_{TO} = 0.17$ ;  $M_{TP} = 0.65$ ,  $SD_{TP} = 0.18$ ;  $t(23) = 2.94$ ,  $p = .007$ ), with a medium effect size of Cohen's  $d = 0.66$ . There was no significant correlation between students' testing time and their test outcome at a test level ( $r = .15$ ;  $p = .25$ ), while there was a negligibly small though significant correlation between students' item-solution time and their response correctness at an item level ( $r = -.07$ ;  $p < .05$ ).

### 3.3. Overall attention distribution

Students' overall solution times were not significantly different in the text-only and text-picture condition ( $M_{TO} = 40.43$  seconds,  $SD_{TO} = 9.28$ ;  $M_{TP} = 41.22$  seconds,  $SD_{TP} = 11.07$ ;  $t(23) = 0.68$ ;  $p = .51$ ). In text-picture items, students spent an average of  $M_{\text{Picture}} = 5937.0$  milliseconds ( $SD_{\text{Picture}} = 4822.5$ ;  $SE_{\text{Picture}} = 176.8$ ) on fixating the picture.

To investigate the influence of adding a picture on the attention devoted to different MC item information units, we fitted one LMM (M1) for basic Aols (text stem, question, correct answer, incorrect answers) and one LMM (M2) for aggregated Aols (overall item stem, answer options) to test fixation time differences between the experimental conditions for significance. Model parameters are summarized in Table 2; descriptive means are visualized in Fig. 2.

Because the LMM specification with dummy-coded variables did not allow for a direct interpretation of the interaction terms from Table 2 (although it was close to the tested parameters because fixations on the reference category 'white space' were near zero), we tested the actual interaction effects of interest, as

**Table 2**

Model parameters from two estimated LMM models that predicted fixation times on Aols using a dummy coding of the Aols ('white space' is the reference category in both models).

Model 1 (M1)	Fixed Effects			Random Effects			
	Estimate ( <i>SE</i> ) <sup>b</sup>	<i>t</i> (14)/ <i>p</i> -level <sup>c</sup>		Intercept Variance Students ( <i>SD</i> )	Intercept Variance Items ( <i>SD</i> )	Residual Variance ( <i>SD</i> )	REML Criterion
Intercept	1014.7 (474.3)	2.1 *		3,965,606 (1991)	2,740,603 (1655)	34,742,179 (5894)	150,299
Picture <sup>a</sup>	27.7 (305.7)	0.1					
Text Stem	19373.4 (305.8)	63.4 ***					
Picture:Text Stem	−3966.9 (432.3)	−9.2 ***					
Question	2540.3 (305.8)	8.3 ***					
Picture:Question	−171.1 (432.3)	−0.4					
Correct	2746.9 (305.8)	9.0 ***					
Picture:Correct	−269.0 (432.3)	−0.6					
Incorrect	7582.7 (305.8)	24.8 ***					
Picture:Incorrect	−1034.1 (432.3)	−2.4 **					
Model 2 (M2)	Estimate ( <i>SE</i> ) <sup>b</sup>	<i>t</i> (18)/ <i>p</i> -level <sup>c</sup>		Intercept Variance Students ( <i>SD</i> )	Intercept Variance Items ( <i>SD</i> )	Residual Variance ( <i>SD</i> )	REML Criterion
Intercept	1018.3 (762.3)	1.3		9,259,074 (3043)	8,538,614 (2922)	59,800,135 (7733)	92685
Picture <sup>a</sup>	24.1 (401.1)	0.1					
Item Stem	19373.4 (401.2)	48.3 ***					
Picture:Item Stem	1970.2 (567.2)	3.5 ***					
Answers	11340.8 (401.2)	28.3 ***					
Picture:Answers	−1271.8 (567.2)	2.2 *					

<sup>a</sup> Dummy coding of the experimental conditions (0 = *text-only*; 1 = *text-picture*).

<sup>b</sup> Standard error similarity within the models resulted from the highly balanced study design and is not incorrect; REML = restricted maximum likelihood; degrees of freedom (*df*) were estimated conservatively using a rule of thumb from Snijders & Bosker (2012, p. 95);  $df = N - q - 1$ , with *N* = number of investigated units (*N* = 24 items) and *q* = explanatory model parameters (*q* = 9 in M1; *q* = 5 in M2) results in 14 *df* (M1) and 18 *df* (M2).

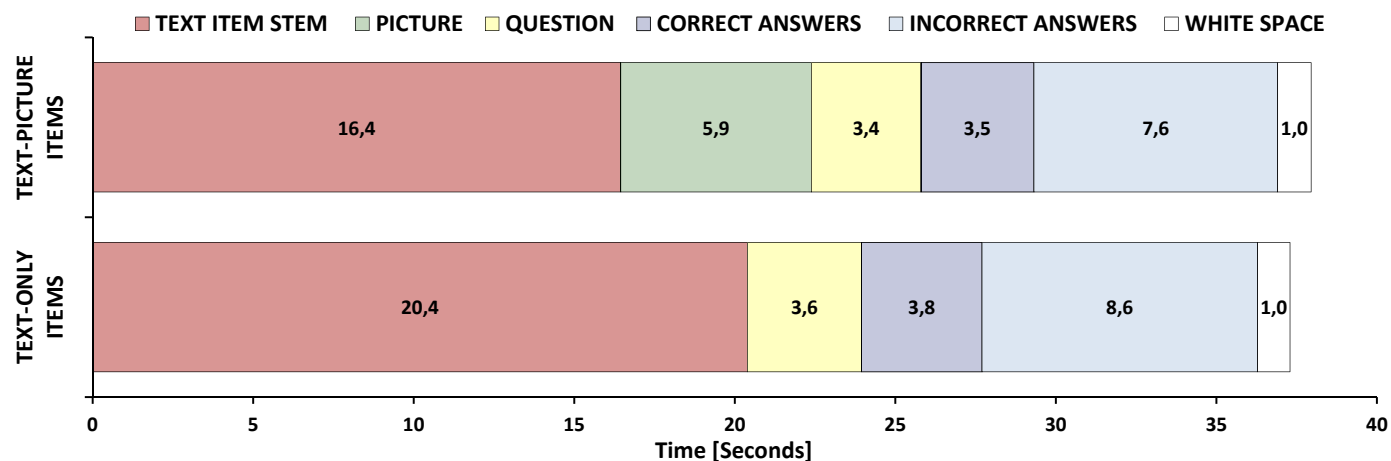
<sup>c</sup>  $p \leq .05$  = \*;  $p \leq .01$  = \*\*;  $p \leq .001$  = \*\*\*.

mentioned, with the delta method by defining linear constraints (contrasts). Accordingly, focusing on the contrast of the *picture* + *picture* × *text stem* interaction, mean fixation times on the text stem were significantly lower in text-picture than in text-only items ( $M_{diff} = -3935.6$ ;  $SE_{diff} = 305.71$ ;  $p < .001$ ,  $d = 0.44$ ). Nevertheless, overall, students spent significantly more time processing the item stem (text + picture) in text-picture items than in text-only items (text stem), reflected by a significant *picture* + *picture* × *overall item stem* interaction ( $M_{diff} = 2001.4$ ;  $SE_{diff} = 401.08$ ;  $p < .001$ ,  $d = 0.16$ ). While there was no significant difference between the time students spent fixating the question statement (*picture* + *picture* × *question*) in both item types ( $M_{diff} = -139.9$ ;  $SE_{diff} = 305.71$ ;  $p = .32$ ,  $d = 0.02$ ), students spent less time on the answer options in text-picture items than in text-only items, reflected by a significant *picture* + *picture* × *answers* interaction ( $M_{diff} = -1247.7$ ;  $SE_{diff} = 401.08$ ;  $p < .001$ ,  $d = 0.10$ ). More specifically, students spent significantly less time on incorrect answer options in text-picture items than in text-only items (*picture* + *picture* × *incorrect answers*;  $M_{diff} = -1002.8$ ;

$SE_{diff} = 305.71$ ;  $p = .002$ ,  $d = 0.11$ ), while the reduction in the time spent on the correct answer options was not significant (*picture* + *picture* × *correct answers*;  $M_{diff} = -237.75$ ;  $SE_{diff} = 305.71$ ;  $p = .22$ ,  $d = 0.03$ ). The time students spent looking at white space (i.e., they fixated a place on the blank screen), was near zero and did not significantly differ between the conditions ( $M_{diff} = 31.27$ ;  $SE_{diff} = 305.71$ ;  $p = .46$ ;  $d < 0.01$ ). Serving as the reference category in the LMM specifications, the white space effect difference was reflected by the *picture* main effect in both models M1 and M2.

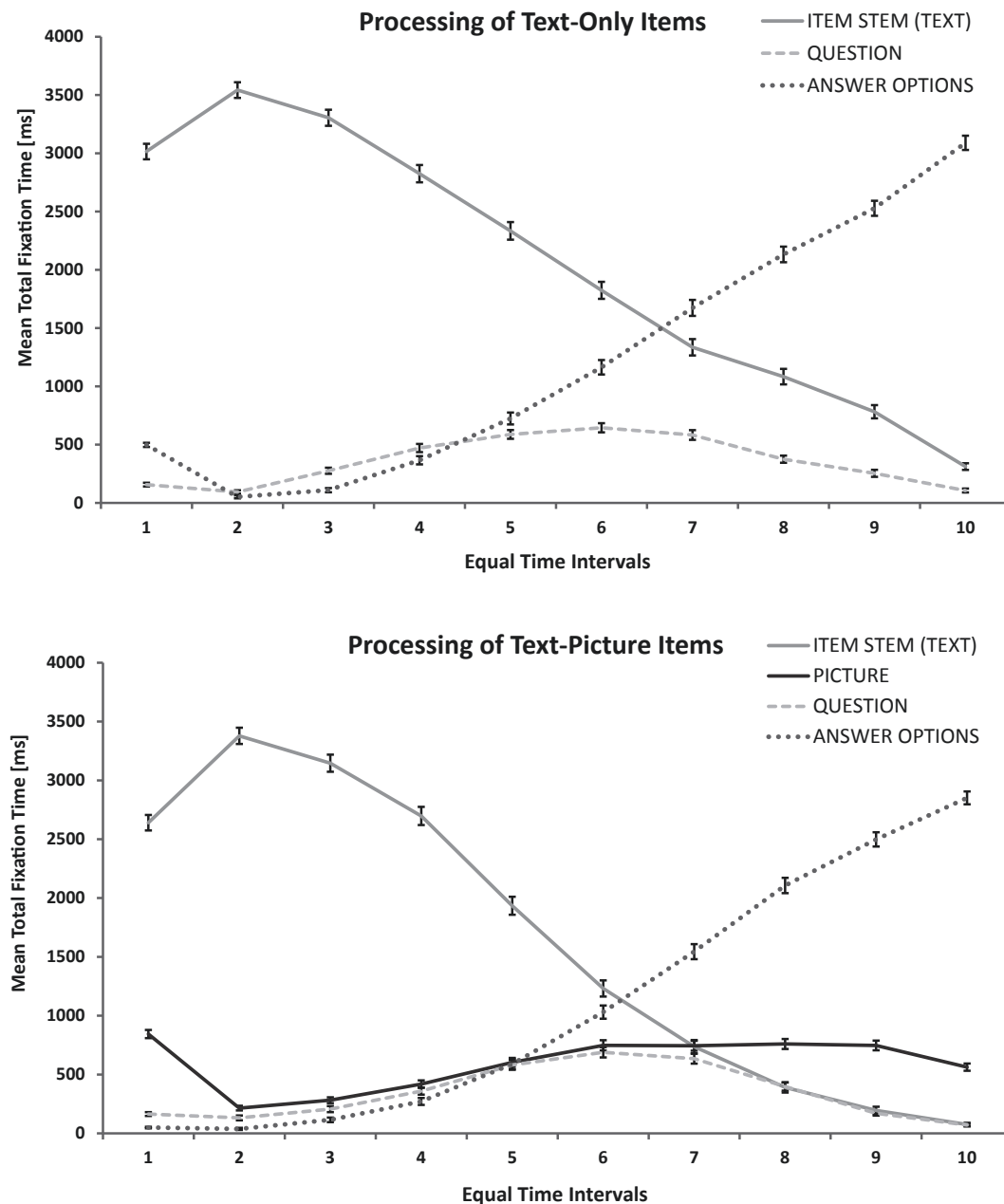
### 3.4. Information processing across solution time

Gaze likelihood analyses provided insights into the average time students spent on different Aols in each of ten equal time intervals that cover the whole item-solution process (Fig. 3). Because inference statistics for this data would require too many single comparisons (inflation of alpha error), only the general patterns of information processing across time are described and depicted in



**Fig. 2.** Comparison of mean total fixation time (in seconds) students spent on different MC information units (*text stem*, *picture*, *question*, *correct* and *incorrect answers*, *white space*) for text-only and text-picture items.





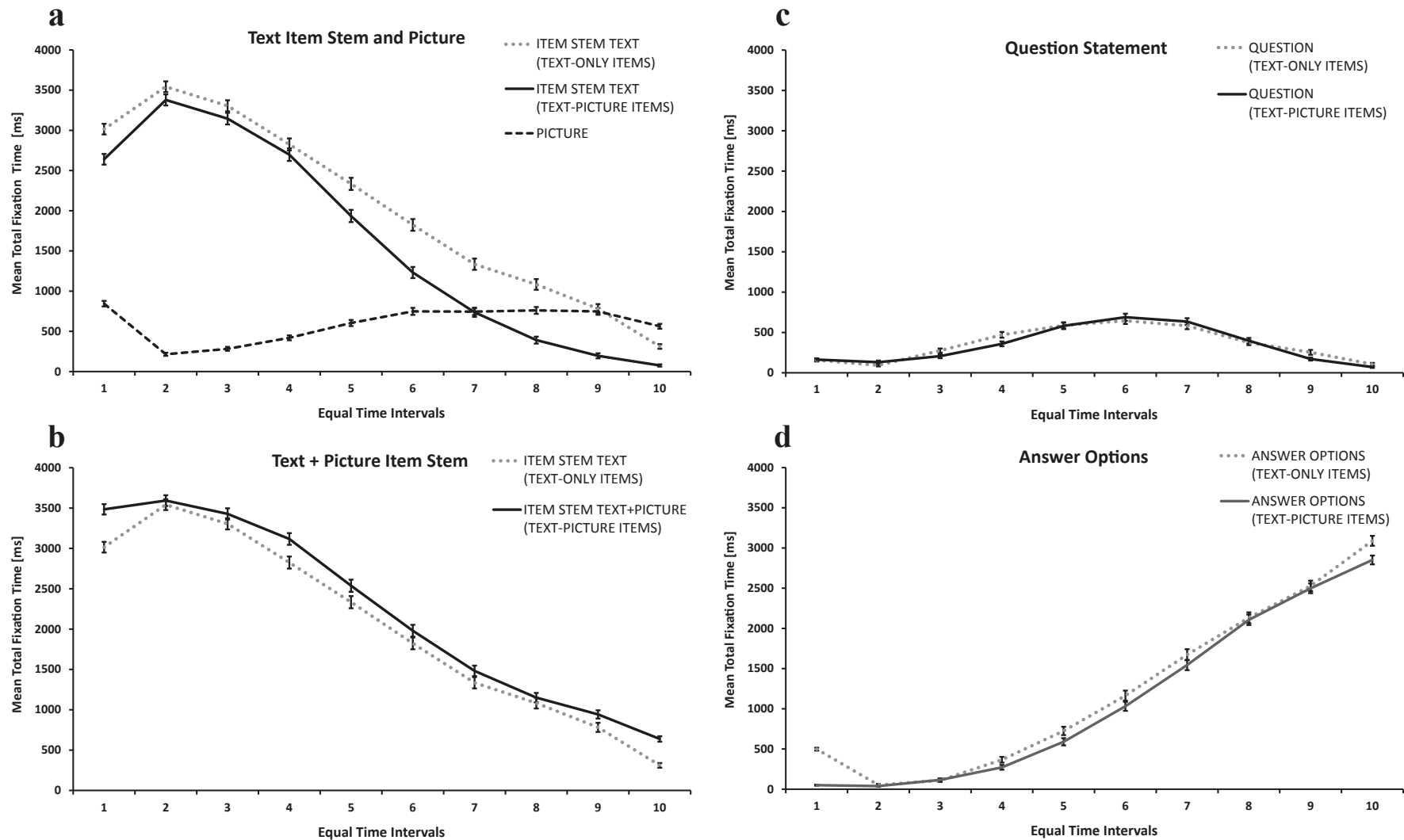
**Fig. 3.** Average fixation duration on MC information units (*text stem, picture, question, answer options*) in ten equal time intervals that encompass the whole item solution process, displayed separately for text-only and text-picture items.

Figs. 3 and 4. A first visual inspection of the curves (with standard errors) in Figs. 3 and 4 shows that the processing of item information across solution time (Interval 1–10) was generally highly comparable for text-only and text-picture items.

In the first five or six time intervals, roughly encompassing the *information-acquisition phase*, students spent most of their time reading the text stem in both item types, while in text-picture items, students initially inspected the picture in the first time interval to a notable extent ( $M_{M1;Picture} = 884.2$ ;  $SE_{M1;Picture} = 34.5$ ). After this early picture processing, the text received much more attention at the expense of the picture. However, the attention that the picture received increased again when students started to focus on the question and the answer options, entering the *decision-making phase*. In this later phase, students still allocated a considerable amount of time to the picture in text-picture items; in

contrast, for text-only items, they directed their attention to the text. Thus, students reviewed the information in the item stem for both item types; this indicates the relevance of the item stem for decision making.

The separate diagrams for different Aols in Fig. 4 provide a direct comparison of time-course differences between text-only and text-picture items. As can be seen, the amount of time spent on the question statement as well as on the answer options in the course of the solution process was almost identical when processing text-only and text-picture items. However, the likelihood of directing attention to the answer options was slightly higher (on a descriptive level) in several time intervals in text-only items than in text-picture items, while the difference was especially evident in the first and the last time interval. This is most probably the cause for the significant difference obtained from the total fixation time



**Fig. 4.** Average fixation duration (with standard errors) in ten equal time intervals encompassing the item solution process, that shows the time course of students' attention distribution towards (a) the *text stem* and the *picture*, (b) the *overall item stem* (text + picture), (c) the *question*, and (d) the *answer options*, displayed separately for text-only (dotted grey lines) and text-picture items (solid black lines).

comparison.

While the time students devoted to the text stem was relatively comparable in the first five time intervals of both conditions, it appeared that students spent less time on the text in the later solution process in text-picture items (see Fig. 4a). However, taking the overall time into account that students spent directing their attention to the picture and the text stem, the course of the curves was highly comparable for both item types (see Fig. 4b). Nevertheless, considering the standard errors, students spent more time processing the overall item-stem information in about half of the time intervals in text-picture items compared to text-only items, while the differences were largest in the first and last interval, when students devoted a lot of attention to the picture in text-picture items, while they inspected the answer options in text-only items instead.

#### 4. Discussion

The present experiment aimed to identify processes that may underlie recent findings of a multimedia effect in testing. Specifically, we examined how students make use of a representational picture in the item stem and adapt their solution behavior compared to text-only items. Relying on theories of multimedia learning (Mayer, 2009; Schnotz & Bannert, 2003), we expected that adding a picture to an item would not only reduce its difficulty (*multimedia effect hypothesis*), but also reduce the processing time of the visualized text, while prolonging the time spent on processing the overall text + picture item stem (*facilitated processing hypothesis*). We further assumed that pictures shorten the decision-making process, reflected in the reduced time spent on incorrect answer options in particular (*faster decision-making hypothesis*). This relates to our expectation that pictures (1) serve as an immediate scaffold for mental model construction in the information-acquisition phase, but (2) also help students to review important item stem information when choosing an answer option (*picture-processing hypothesis*).

First, in line with our *multimedia effect hypothesis*, we found a significant decrease in item difficulty with a medium effect size in text-picture items compared to corresponding text-only items ( $d = 0.66$ ). This replicates earlier findings (Hartmann, 2012; Lindner et al., in press; Saß et al., 2012) and means that our data fulfill an important prerequisite for an interpretation of the eye-movement parameters as potential indicators that may reflect cognitive processes that underlie the multimedia effect in testing.

Second, in agreement with our *facilitated text-processing hypothesis*, we found that, students needed overall significantly less time to read the text stem when a picture was present. This suggests that—comparable to learning situations (cf. Eitel et al., 2013; Schnotz & Bannert, 2003)—the picture facilitated the construction of a mental model about the problem or situation described in the text item stem. This might have prevented students from making erroneous interpretations of the text or from missing important task aspects, in turn increasing their test performance. These findings suggest that prominent effects found in multimedia learning may also transfer to the testing situation. Moreover, this supports the idea that theories from learning (CTML, Mayer, 2009; ITPC, Schnotz & Bannert, 2003) are applicable to testing. They may also partially explain why similar solution times were found for text-only and text-picture items in previous research (Saß et al., 2012); that is, additional encoding time spent on the picture was compensated for by shorter (facilitated) processing of the text. However, the average time saving on processing the item stem text was smaller than the average picture processing time, which resulted in a longer engagement with the overall item stem (text + picture) in text-picture items compared to text-only items,

as expected. Hence, increasing the time spent on studying the problem might have eventually even mediated students' better test performance by promoting a better understanding. However, this is only a speculative assumption and should be addressed in future studies.

Third, supporting our *faster decision-making hypothesis*, we found that pictures not only facilitated item-stem processing, but also supported the answering process, reflected in the reduced processing time of incorrect (but not of correct) answer options. Thus, with pictures, participants were not only better but seemingly also faster in dismissing distractors in the answering process. On the one hand, this certainly results from a better understanding of the problem in the item stem, which was achieved by processing the picture; but, on the other hand, it may also result from a better on demand recall of the given problem, because the picture (as opposed to the text) can be used as a more easily accessible external representation for updating the problem-related mental model (Schnotz et al., 2014). These interpretations are carefully supported by a significant correlation ( $r = -.59$ ) that was found between the time students directed their attention to incorrect answer options and their test performance.

Fourth, our time-course analyses yielded support for both of these explanations in favor of the *picture processing hypothesis*. The analyses revealed that students directed their attention to the picture in the initial information-acquisition phase and also in the early decision-making phase. Against the background of findings from learning studies (Eitel & Scheiter, 2015; Eitel et al., 2013; Schnotz et al., 2014), for the information-acquisition phase, we argue that inspecting the picture for a short time at the beginning of item processing probably enabled students to extract the gist of the picture (cf. Eitel, Scheiter, & Schüler, 2012) and, thus, the gist of the problem representation from the item stem. Accordingly, the picture may have served as a scaffold to support understanding of the corresponding item-stem text, and in turn, to construct a more coherent mental model of the problem at hand.

Whereas there was a rather short peak in picture inspection at the beginning, there was a plateau in picture inspection across the last phases prior to decision announcement (i.e., the decision-making phase). It seems likely that this is due to the fact that students tried to use the picture when evaluating the answer options with respect to their potential of being the correct answer. Interestingly, the descriptive pattern of results revealed that students directed more attention to the picture than to the item-stem text in the final phases of the decision-making process, suggesting that students might have preferred pictures when referring back to the item stem while trying to reach a decision. Again, this is in line with the assumption from Schnotz et al. (2014) that pictures, rather than a text, are used for mental model updating when solving a specific task. Together with the finding of better test performance due to the addition of a picture, our results indicate that the picture as an external representation might have supported students in making inferences and in the reasoning processes used to reach a better decision compared to the case of an internal (mental) representation that was constructed from a text only (cf. Schnotz & Kürschner, 2008).

In general, however, there were no major changes in the time courses of processing the other MC elements (text item stem, question, answer options) due to the addition of the picture. Regardless of whether there was a picture present or not, the item stem received most of the attention at the beginning, which may reflect reading to understand the problem. The attention directed at the item stem decreased across time, probably reflecting that students were then finished with gaining a first impression of the information. In both item types, the attention devoted to the question had a peak midway through the process and the attention

devoted to the answer options increased in the decision-making phase until the choice announcement. However, an important finding is that, across all phases of the item-solving process, it seemed that the attention directed to the picture in text-picture items replaced attention that was otherwise devoted to the text stem in text-only items (see Figs. 3 and 4).

#### 4.1. Limitations and future directions

In the following, we discuss limitations that should be taken into account when interpreting the results of the present study. Moreover, we consider several aspects that need to be addressed in future research.

First, we investigated young students from the fifth and sixth grades. This might limit the generalizability of our findings, because their proficiency to read, understand, and extract the important information from text and pictures may not be fully developed yet. Pictures might thus potentially have other effects on older students' processing and understanding of MC test items; thus, our findings should be reconsidered in other samples in the future. However, in this study, we deliberately chose to use an ecological valid sample within the age range of the students that participated in the TIMSS study (4th grade and 8th grade), thereby ensuring that the abilities of the students in our sample approximately matched the difficulty of the science items that we adapted for the study from this LSA program (cf. IEA, 2013).

Second, the testing material comprised relatively short texts and only representational pictures. Longer texts might influence the findings and should therefore be used in future studies. But also the use of different types of pictures in testing, such as decorative pictures or graphs and diagrams, would be worth investigating at a processing level, as different findings could be expected due to the different functions of the pictures. Related to this aspect, also the generalizability of the effects of pictures on multiple-choice assessment to the case of open-response items could be a subject to study, as it cannot be taken for granted that the reported multimedia effect in testing easily transfers from one test format to another. This is mostly due to the fact that both item types demand different cognitive processes when answering a question (i.e., *choosing* vs. *constructing* a response), even though the initial phase of the problem representation (i.e., information-acquisition phase), and in this regard also the role of the picture as a mental scaffold for understanding the text, is probably still highly comparable between both formats.

Third, although general motivational effects were balanced by our study design, we cannot exclude the possibility that starting with a text-picture item might have been a design-implemented motivational advantage for the second group. However, as the groups did not significantly differ in their performance or response time, strong influences that could have biased our findings are unlikely. Nevertheless, motivational aspects of the multimedia effect in testing would be another interesting area for studies using eye-tracking analyses in the future, as there is evolving evidence for interplay between the cognitive and motivational functions of pictures in testing (Lindner et al., in press).

#### 4.2. Conclusions and educational implications

The present study successfully replicates previous findings of a reduced item difficulty when pictures are added to a verbal item stem (e.g., Hartmann, 2012; Lindner et al., in press; Saß et al., 2012), and thus supports the growing body of evidence in favor of a *multimedia effect in testing*. Furthermore, this study is—to the best of our knowledge—the first to empirically investigate the impact that pictures have on item-solving behavior by experimentally

varying the presence and absence of pictures in test items and by analyzing effects via eye-tracking at a process level. By this, we were able to show that the processes identified in multimedia *learning* are also found in multimedia *testing*; e.g., processing of the picture facilitated processing of the corresponding text. This effect, together with a reduced viewing time spent on incorrect answers in text-picture items, could also explain why adding pictures to text-based MC items fostered performance without increasing students' time on task. Moreover, it suggests that theories on multimedia learning can, in fact, contribute to a better theoretical understanding of the processes underlying the multimedia effect in testing. As the overall eye-movement pattern revealed predictable and welcome changes in students' item and information processing, namely, it became more efficient, we would carefully encourage educational practitioners such as item constructors in LSA or teachers to consider integrating representational pictures into testing material more often in the future.

#### Acknowledgements

We thank Simon Grund for his support with the mixed-effects analyses and Gráinne Newcombe for language editing.

#### References

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16, 183–198. <http://dx.doi.org/10.1016/j.learninstruc.2006.03.001>.
- Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135–146). New York, NY: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <http://dx.doi.org/10.1016/j.jml.2007.12.005>.
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474. <http://dx.doi.org/10.1016/j.jml.2007.09.002>.
- Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using Eigen and S4 (R Package Version 0.999999-0)* [Computer software]. Retrieved from <https://CRAN.R-project.org/package=lme4>.
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, 32, 115–132. <http://dx.doi.org/10.1023/B:TRUC.0000021812.96911.c5>.
- Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, 98, 182–197. <http://dx.doi.org/10.1037/0022-0663.98.1.182>.
- Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 174–205). New York, NY: Cambridge University Press.
- Eitel, A. (2016). How repeated studying and testing affects multimedia learning: Evidence for adaptation to task demands. *Learning and Instruction*, 41, 70–84. <http://dx.doi.org/10.1016/j.learninstruc.2015.10.003>.
- Eitel, A., & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. *Educational Psychology Review*, 27, 153–180. <http://dx.doi.org/10.1007/s10648-014-9264-4>.
- Eitel, A., Scheiter, K., & Schüler, A. (2012). The time course of information extraction from instructional diagrams. *Perceptual and Motor Skills*, 115, 677–701. <http://dx.doi.org/10.2466/22.23.PMS.115.6.677-701>.
- Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*, 28, 48–63. <http://dx.doi.org/10.1016/j.learninstruc.2013.05.002>.
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks CA: Sage <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.
- Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., et al. (2015). *car: Companion to applied regression (R Package Version 2.1-1)* [Computer software]. Retrieved from <https://cran.r-project.org/web/packages/car/index.html>.
- Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language*, 31, 129–151. [http://dx.doi.org/10.1016/0749-596X\(92\)90008-L](http://dx.doi.org/10.1016/0749-596X(92)90008-L).
- Hartmann, S. (2012). *Die Rolle von Leseverständnis und Lesegeschwindigkeit beim Zustandekommen der Leistungen in schriftlichen Tests zur Erfassung naturwissenschaftlicher Kompetenz [The role of reading comprehension and reading speed in text-based assessments of scientific inquiry skills]*. Doctoral dissertation. University of Duisburg-Essen. Retrieved from [http://duepublico.uni-dueburg-essen.de/servlets/DerivateServlet/Derivate-33260/hartmann\\_diss.pdf](http://duepublico.uni-dueburg-essen.de/servlets/DerivateServlet/Derivate-33260/hartmann_diss.pdf).



- Heller, K. A., & Perleth, C. (2000). *KFT 4–12+ R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. [Cognitive Abilities Test for students from grade 4 to 12+ (CogAT; Thorndike, L. & Hagen, E., 1954–1986) German adapted version/author]. Göttingen: Beltz.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & De Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.
- International Association for the Evaluation of Educational Achievement [IEA]. (2013). *TIMSS 2011 assessment*. Chestnut Hill, MA/Amsterdam, NE: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College/International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat. Retrieved from [http://nces.ed.gov/timss/pdf/TIMSS2011\\_G4\\_Science.pdf](http://nces.ed.gov/timss/pdf/TIMSS2011_G4_Science.pdf).
- 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, 803–817. <http://dx.doi.org/10.1111/bjet.12174>.
- Lenhard, W., & Schneider, W. (2006). *ELFE 1–6: Ein Leseverständnistest für Erst- bis Sechstklässler [ELFE 1–6: A reading comprehension test for students from grade 1 to 6]*. Göttingen, Germany: Hogrefe.
- Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science*, 41, 811–831. <http://dx.doi.org/10.1007/s11251-012-9256-z>.
- Lindner, M. A., Eitel, A., Thoma, G.-B., Dalehefte, I. M., Ihme, J. M., & Köller, O. (2014). Tracking the decision making process in multiple-choice assessment: Evidence from eye movements. *Applied Cognitive Psychology*, 28, 738–752. <http://dx.doi.org/10.1002/acp.3060>.
- Lindner, M. A., Ihme, J. M., Saß, S., & Köller, O. (in press). How representational pictures enhance students' performance and test-taking pleasure in low-stakes assessment. *European Journal for Psychological Assessment*. Advance online publication. <http://dx.doi.org/10.1027/1015-5759/a000351>
- 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, 95–109. <http://dx.doi.org/10.1016/j.compedu.2012.07.011>.
- Mason, L., Tornatora, M. C., & Pluchino, P. (2015). Integrative processing of verbal and graphical information during re-reading predicts learning from illustrated text: An eye-movement study. *Reading and Writing*, 28, 851–872. <http://dx.doi.org/10.1007/s11145-015-9552-5>.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511811678>.
- Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139547369>.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Amsterdam, NE: International Association for the Evaluation of Educational Achievement (IEA). Retrieved from <http://files.eric.ed.gov/fulltext/ED512411.pdf>.
- Organisation for Economic Co-operation and Development [OECD]. (2013). *PISA 2012 Assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris, FR: OECD Publishing. <http://dx.doi.org/10.1787/9789264190511-en>.
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413–425. <http://dx.doi.org/10.1016/j.jml.2008.02.002>.
- R Core Team. (2015). *R: A language and environment for statistical computing and graphics [Computer software]*. Retrieved from <http://www.R-project.org>.
- Saß, S., Wittwer, J., Senkbeil, M., & Köller, O. (2012). Pictures in test items: Effects on response time and response correctness. *Applied Cognitive Psychology*, 26, 70–81. <http://dx.doi.org/10.1002/acp.1798>.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13, 141–156. [http://dx.doi.org/10.1016/S0959-4752\(02\)00017-8](http://dx.doi.org/10.1016/S0959-4752(02)00017-8).
- Schnotz, W., & Kirschner, C. (2008). External and internal representations in the acquisition and use of knowledge: Visualization effects on mental model construction. *Instructional Science*, 36, 175–190. <http://dx.doi.org/10.1007/s11251-007-9029-2>.
- Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *Journal of Educational Psychology*, 106, 974–989. <http://dx.doi.org/10.1037/a0037054>.
- Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6, 1317–1322. <http://dx.doi.org/10.1038/nn1150>.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: SAGE.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.