Received: 10 November 2023 | Revised: 17 July 2024 | Accepted: 28 July 2024

DOI: 10.1111/jcal.13051

ARTICLE

Journal of Computer Assisted Learning WILEY

# Measuring (meta)cognitive processes in multimedia learning: Matching eye tracking metrics and think-aloud protocols in case of seductive details

Lisa Stark[1] | Andreas Korbach[1] | Roland Brünken[1] | Babette Park[2]

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

[2]Institute for Psychology, University of Education Freiburg, Freiburg, Germany

**Correspondence**
Lisa Stark, Department of Education, Saarland University, Campus A5 4, 66123 Saarbrücken, Germany.
Email: lisa.stark@uni-saarland.de

## Abstract

**Background:** In recent years, eye tracking has become a prominent method for learning research as it is assumed to indicate (meta)cognitive processes. However, there is little empirical evidence for hypothesized relations between eye tracking indicators and specific (meta)cognitive processes so that construct validity of used metrics can be questioned.

**Objectives:** The main goal was to provide validity hints in order to create an empirical basis for interpreting specific eye tracking indicators in terms of respective (meta) cognitive processes of multimedia learning.

**Methods:** $N = 60$ students learned with multimedia instructional material. Referring to a process model of multimedia learning, correlations between theoretically deduced eye tracking indicators with verbalized (meta)cognitive processes of multimedia learning captured by think-aloud protocols were examined. In addition, the sensitivity of both process measures of (meta)cognitive processes was regarded considering the well-investigated seductive details effect of an established multimedia instruction in a two-group design. Finally, serial mediations were calculated in order to investigate whether both process measures complement one another in a joint explanation of the seductive details effect.

**Results and Conclusions:** Eye tracking indicators and verbalized (meta)cognitive processes did only partly correspond as it was shown by correlation and serial mediation analyses. However, both measures were sensitive to indicate the seductive details effect. Thus, even though the study provided insights in how validation could be possible, further systematic research will be needed for validating eye tracking indicators of specific (meta)cognitive processes in multimedia learning.

KEYWORDS

eye tracking, learning processes, multimedia learning, process measures, seductive details, think-aloud protocols

# 1 | INTRODUCTION

Theoretical frameworks on multimedia learning (e.g., cognitive-theory of multimedia learning; CTML; Mayer, 2021) postulate diverse cognitive and metacognitive processes, which are necessary for successful learning with multimedia instruction. Thus, the question arises how to capture these learning processes during multimedia learning.

In the last decade, eye tracking has become a popular means for measuring learning processes in multimedia learning. Several eye tracking studies pointed out the advantage of using this method as eye tracking indicators unveil specific aspects of learning processes with multimedia instruction (e.g., Jarodzka et al., 2017). Recent developments on how eye tracking can be used in empirical studies diversified the research field, also for multimedia learning research. On the one hand, there is an expanding body of studies applying eye tracking data as pedagogical interventions in the form of feedback (Kok et al., 2022) or as eye movement modelling examples (EMME, e.g., Salméron et al., 2020; Van Marlen et al., 2022; Xie et al., 2021). On the other hand, there exist multi-method approaches combining eye tracking with other process measures like, for example, EEG or electrodermal activity (e.g., Baceviciute et al., 2021; Liberman & Dubovi, 2022). Such mostly bimodal approaches culminate in research on multimodal data analytics (Di Mitri et al., 2018). In those studies, eye tracking and other process measures are applied for prognosing dependent measures like learning outcomes in a data-driven approach. Mostly, eye tracking data have been interpreted as general indicators of visual attention (e.g., Alemdag & Cagiltay, 2018; Kok et al., 2022). Yet, eye tracking studies are typically based on the established assumption that eye movements and cognitive or metacognitive processes are closely intertwined (see below; eye mind hypotheses). Thus, the question on whether specific eye tracking indicators can be interpreted in terms of particular cognitive or metacognitive processes still stands to reason.

Eye tracking indicators have scarcely been interlinked *empirically* with theoretically postulated cognitive or metacognitive processes during learning, but the hitherto interpretations of eye tracking indicators have mainly been based on *theoretical* assumptions (Alemdag & Cagiltay, 2018). As a consequence, there is still a lack of empirical support for these postulated relations. Altogether, validated eye tracking metrics are needed not only for explaining multimedia learning effects (which is one core focus of multimedia learning research) but also for supporting the interpretation of findings in related research fields allowing to use eye tracking indicators in an evidence-based way.

The think-aloud method (Ericsson & Simon, 1993) has usually been applied to catch learners' (meta)cognitive processes during learning in a lot of research fields besides multimedia learning as it can be coded regarding learners' (meta)cognitive processing during task performance. As think-aloud protocols are assumed to be very closely interlinked with actual (meta)cognitive processing, think-aloud measures might provide a means to shed light on the relation of eye tracking indicators and the postulated multimedia learning processes from an empirical perspective.

Referring to a process model of multimedia learning, which introduces cognitive and metacognitive processes of learning with multimedia instruction, the present study aimed at examining the relation between theoretically deduced indicators for multimedia learning processes captured by eye tracking and (meta)cognitive processes identified in think-aloud protocols. Thereby, interpretations of eye tracking indicators concerning (meta)cognitive processes were examined from an empirical point of view. Thus, the present study elaborated on validity aspects of eye tracking and think-aloud data. Therefore, the present study exploited the well-investigated seductive details effect (Park, Flowerday, & Brünken, 2015; Park, Korbach, & Brünken, 2015; Sundarajan & Adesope, 2020) in order to obtain more information about the sensitivity of both measures regarding instructional design effects. In addition, it was examined how these process measures complement one another in an integrated serial mediation analysis on mechanisms of action for explaining instructional design effects.

# 2 | THEORETICAL BACKGROUND

## 2.1 | Multimedia learning

Multimedia learning can be defined as knowledge acquisition from an instructional message, which comprises words and pictures (Mayer, 2021). The cognitive theory of multimedia learning (CTML; Mayer, 2021) combines several assumptions on cognitive architecture (dual channel processing, Baddeley, 1986; dual coding; Paivio, 1986; limited cognitive capacity; Baddeley, 1986) with the claim for active information processing in a theoretical framework. This framework is used for explaining instructional design effects when learning with multimedia instructional messages by mainly focusing on cognitive architecture. However, a process perspective on multimedia learning can provide complement and advance explanations.

### 2.1.1 | A process model of multimedia learning

In CTML different cognitive processes of multimedia learning have been introduced, that describe learning-relevant cognitive activities of schema construction, which were based on a basic model of information processing during learning (selecting-organizing-integration [SOI] model; Mayer, 1996). Related models (e.g., cognitive-affective theory of learning with media, CATLM, Moreno, 2006; Moreno & Mayer, 2007; integrated model of text and picture comprehension, Schnotz & Bannert, 2003; Processes of coherence formation, Seufert, 2003) have mentioned slightly different categorizations of multimedia learning processes. In particular, these models added extraneous and metacognitive processes (see especially, Moreno & Mayer, 2007). These theoretical considerations have recently been integrated into a process model of multimedia learning, which draws a rather complete picture on information processing during multimedia learning (Stark et al., 2018; see Stark et al., 2020 for a more detailed description of the process model). The process model of multimedia learning distinguishes five major categories, whereas the first three build on each other in line with the SOI model (Mayer, 1996) and CTML (Mayer, 2021).

(1) *Selecting* processes refer to picking relevant verbal and pictorial information respectively, which will be processed further (cf. Mayer, 1996, 2021). (2) After having selected the relevant information, this information has to be understood and contextualized by *organizing* processes (cf. Mayer, 1996, 2021), which include three types of coherence formation (cf. Seufert, 2003): (2a) *Local coherence formation* represents cognitive processes of developing the meaning of verbal or pictorial information parts of the learning instruction respectively (Seufert, 2003). (2b) *Horizontal global coherence formation* refers to matching processes of corresponding verbal and pictorial information, e.g., on one slide of the learning instruction. (2c) *Vertical global coherence formation* represents cognitive processes of bringing several information units of the whole multimedia instruction together (e.g., from different slides of the learning instruction). The terms *horizontal* and *vertical*, which differentiate processes of global coherence formation, were chosen on the basis of the alignment of presented information on a perpendicular timeline. (3) In order to encode organized information, *elaboration* processes are needed, which can regard the integration of organized information with prior knowledge (i.e., *integrating* in CTML; e.g., Mayer, 2021) and, or the use of further (genuine) elaboration strategies (e.g., building memory hooks). (4) With reference to theories of self-regulated learning (e.g., Schmitz & Wiese, 2006; Zimmerman, 2008) and CATLM (Moreno, 2006; Moreno & Mayer, 2007), *metacognitive processes* of planning, monitoring, and regulating are comprised in the process model due to their overarching function to manage the other kinds of processes. (5) Cognitions that are irrelevant for achieving the learning objective are referenced as *extraneous processes* (cf. Moreno & Mayer, 2007). Extraneous processes can occur due to private thoughts or due to suboptimal design elements (e.g., seductive details) of the learning environment, for example (cf. Cognitive Load Theory, e.g., Sweller et al., 2019). From a theoretical perspective (e.g., SOI model; Mayer, 1996), it can be assumed that processes of selecting, organizing, and elaborating build on each other and that extraneous processes are independent from other cognitions. However, metacognitive processes can be assumed to cut across the other processes or even overlapping them by guiding and managing the whole learning process.

This process model has been confronted with empirical data in form of cued retrospective think-aloud protocols and proofed to be able to capture differences in verbalized (meta)cognitive processes that were related to differences in learning outcomes (Stark et al., 2020): The study showed in a comparison of extreme groups that high performing learners verbalized a smaller percentage of metacognitive and selecting processes but a larger percentage of elaborating processes than low performing learners. There were no differences in organizing or extraneous processes. However, organizing processes were not differentiated into the subcategories of local, horizontal global, and vertical global coherence formation, so that potential effects might have been confounded. Applying the process model also enabled detecting differences in (meta)cognitive processing, when instructional design features of multimedia instruction were varied (Stark et al., 2018), so it proofed to be sensitive regarding such variations in multimedia instruction.

### 2.1.2 | The seductive details effect

The seductive details effect describes the detrimental effect on learning outcomes, which is caused by additional, non-redundant and interesting information in a learning instruction that is related to the learning content but irrelevant with regard to the learning goal (i.e., seductive details; Garner et al., 1989; Harp & Maslich, 2005; Harp & Mayer, 1998; Park, Flowerday, & Brünken, 2015; Sundarajan & Adesope, 2020). Even though there were inconsistent findings on the seductive details effect (e.g., Rey, 2012), there exists consent that adding seductive details to a learning instruction affects visual information processing with significant effects on learners' direction of visual attention and engagement in mental model construction (Korbach et al., 2017).

Here, the seductive details effect was exploited as a means to vary (meta)cognitive processing while learning with multimedia instruction in the present study. Previous studies that investigated learners' eye movement in relation to the seductive details effect showed that seductive details altered the visual focus during learning and affected diverse eye tracking indicators (e.g., Korbach et al., 2017; Rey, 2014; Strobel et al., 2019). Thus, the seductive details effect was assumed to increase the demands for cognitive as well as for metacognitive processes (Eitel et al., 2020) and, thereby, to provide the possibility to test the sensitivity of process measures to create hints for their validity.

## 2.2 | Eye tracking indicators of (meta)cognitive processes

Eye tracking data provide information about the allocation of visual attention. For the analysis of eye movements, it is assumed that the information in the focus of visual attention is cognitively processed (Just & Carpenter, 1976; Rayner, 1998). Hence, the analysis of the amount and the distribution of visual attention on a learning instruction, especially when related to learning performance, can help to understand learners' perceptual information processing (Jarodzka et al., 2017; Mayer, 2010). There is a large variety of eye movement measures that are assumed to be related to cognitive processing (e.g., Alemdag & Cagiltay, 2018; Coskun & Cagiltay, 2022; Johnson & Mayer, 2012).

Oftentimes eye tracking analyses refer to the number or duration of fixations. Further measures are often created with reference to the number and duration of fixations or in relation to time on task, as for example mean fixation duration or fixation frequency (Canham & Hegarty, 2010; Scheiter et al., 2019). For these measures, in line with the so-called eye-mind hypothesis (Rayner, 1998), a longer fixation duration or a higher number of fixations on specific Areas of Interest (AOIs) is generally interpreted as a higher amount of visual attention and higher cognitive engagement in the processing of the respective information (e.g., Mayer, 2010).

Furthermore, transitions, which are shifts in visual attention between corresponding AOIs, have already been used as a promising indicator specifically in multimedia learning research. Especially, a larger number of transitions from text to corresponding picture AOIs is assumed to represent higher cognitive engagement in cognitive processes for the integration of textual and pictorial information (Scheiter et al., 2019). Mason and colleagues (Mason et al., 2015, 2017) showed that specifically the time spent on an AOI subsequent to a transition predicted learning performance and can be assumed to indicate deeper information processing. Consequently, specifically transitions with a subsequent long visit duration can be assumed to indicate productive integrating processes whereas transitions with a subsequent short visit duration can be assumed to indicate processes of redirecting visual attention (i.e., effort regulation), specifically when the learning instruction includes additional distractive information (Korbach et al., 2017). Until now, no common approach has been identified about how to differentiate transitions by means of subsequent visit duration. The present study applied an approach using 1 s of visit duration (Potter, 2012) to distinguish between transitions that indicate deep processing (subsequent long visit duration) and transitions that indicate redirecting processes (subsequent short visit duration).

As the present study aimed at finding empirical support for relations between specific eye tracking indicators and (meta)cognitive processes in multimedia learning, these processes have to be captured by a second process measure in order to infer hints at the validity of the eye tracking indicators. In contrast to recent research applying multimodal learning analytics the data from both process measures are not fused in the sense of a data-driven approach (Chango et al., 2022; Drachsler & Schneider, 2018). The focus of the present study was to underpin interpretations of theoretically deduced eye tracking metrics by comparing data from different modalities (i.e., eye tracking and think-aloud protocols) with respect to (meta)cognitive processes instead of combining them for a holistic analysis of the learning process (Di Mitri et al., 2018).

## 2.3 | Think-aloud protocols indicating (meta) cognitive processes

An established method for capturing information on learning processes utilizes think-aloud protocols (Ericsson & Simon, 1993; Van Gog et al., 2009). For obtaining these protocols, participants are asked to verbalize everything that comes to their mind while working with a learning instruction (i.e., concurrent think-aloud procedure). Even though this method has oftentimes be implemented in other research fields and researchers of other disciplines have already succeeded in successfully combining eye-tracking with the concurrent think-aloud procedure to identify cognitive processes and strategies during task execution (Andaloussi et al., 2021; Baß et al., 2024; Elling et al., 2012; Van Gog et al., 2005), this method has only scarcely been implemented in studies on multimedia learning (exceptions: Jarodzka et al., 2010; Park et al., 2020; Stark et al., 2018) and has even more seldomly been interlinked with eye tracking data (exceptions: Gegenfurtner & Seppänen, 2013, Stark et al., 2018).

There is a large body of research pointing at potential problems when using the concurrent think-aloud method (e.g., reactivity with task performance, lack of completeness of verbalizations; containing manipulative statements instead of verbalized cognitions; cf. Bannert & Mengelkamp, 2008; Hansen, 1991; Hyrskykari et al., 2008). Nevertheless, until now no other method has been introduced to get to know what learners are thinking while learning and how they engage in schema construction from a qualitative point of view with such a close assumed linkage to actual information processing.

In most cases, think-aloud protocols are analysed referring to a category system which is derived from prior cognitive task analysis (Chi, 1997; Holmqvist et al., 2011; Renkl, 1997) supporting a normative perspective on required learning processes to reach a learning goal. In the present study, the category system is based on theoretical considerations concerning those processes that occur during multimedia learning (see the process model described above) so that results are comparable between different multimedia learning studies (e.g., with Stark et al., 2018, 2020).

Stark et al. (2018) investigated effects of an emotional text design on learning outcomes, eye tracking data and cued retrospective think-aloud protocols. The think-aloud protocols were coded using the process model of multimedia learning mentioned above. Results showed that the applied process model was sensitive to group differences caused by the variations in the instructional design of the learning program. However, eye tracking and think-aloud data were not related to each other. An experimental study by Park et al. (2020) showed that the seductive details effect did not interfere with the think-aloud method. Even though effects of the seductive details on the content of the think-aloud protocols were not regarded in that study, results underline that the seductive details effect can be used as instructional design effect for the purpose of the present study.

## 2.4 | Deducing eye tracking indicators of specific (meta)cognitive processes

Even though recent eye tracking studies used the analysis of scanpaths or gaze patterns to investigate (meta)cognitive processes (e.g., Bühler et al., 2024; Stark et al., 2024; Tjon et al., 2023), the use of simple eye tracking indicators is widespread in research on multimedia learning as they are easy to calculate. In this research field, metrics based on fixations, visits and transitions are thereby assumed to indicate specific (meta)cognitive processes (Alemdag & Cagiltay, 2018; Coskun & Cagiltay, 2022). However, their validity has not yet been investigated, which might be one reason why eye tracking researchers face the critique to choose arbitrary and sometimes inappropriate measures (Orquin & Holmqvist, 2018). The present study displays one attempt to theoretically deduce eye tracking indicators of specific (meta)cognitive processes and to generate hints at the validity of these variables. On the basis of the process model of multimedia learning (see above), the specific multimedia learning processes were regarded in the present research. The choice for specific eye tracking indicators for (meta)cognitive processes of multimedia learning was based on theoretical considerations and deductions. In addition, results of previous studies that investigated effects of instructional elements in multimedia learning on eye tracking indicators provided hints at potential eye tracking indicators for specific (meta)cognitive processes.

As introduced above, *selecting* processes refer to picking relevant elements, which are recognized and potentially further processed into working memory. Even though previous studies often referred to measures like time to first fixation for particular AOIs (cf. Alemdag & Cagiltay, 2018), it can be argued that selecting processes do also occur after a first fixation, so that from our point of view all potential selecting processes can only be indicated by regarding the *number of all fixations* (cf. Alemdag & Cagiltay, 2018) on relevant information as a fixation indicates (by definition) that attention is allocated at some point.

Alemdag and Cagiltay (2018) showed that hitherto used measures for organizing processes in general mostly referred to total fixation durations on AOIs. These measures were chosen in line with the eye-mind hypothesis that longer fixation times refer to more deep processing of the fixated information. For the present study, which differentiated between different kinds of organizing processes, *fixation duration* on relevant AOIs was used to match *processes of local coherence formation* because it is a local measure referring to information processing of specific areas.

In line with previous studies (e.g., Scheiter et al., 2019), transitions were considered to indicate *horizontal coherence formation*. With respect to the work of Mason and colleagues (Mason et al., 2015; Mason et al., 2017), only *transitions with a subsequent long visit duration* (>1 s) were associated with higher learning gains. Consequently, only those transitions were used for measuring processes of horizontal coherence formation in the present study.

The process model of multimedia learning assumes in line with the SOI model (Mayer, 1996) a hierarchical model with selecting being a prerequisite for organizing processes. In the same line of argumentation, it can be argued that *vertical coherence formation* processes can only follow successful processes of local coherence formation and, especially, horizontal coherence formation. Thus, the *visit duration subsequent to a transition* (cf. horizontal coherence formation) was considered to indicate cognitive activity for contextualizing information in a larger nexus by bringing several information units together (Malone et al., 2020) that is, vertical coherence formation processes.

Elaborating processes indicate deep cognitive processing. In previous studies, mean fixation duration on relevant information has been used as an indicator for diverse aspects of information processing like for local coherence formation or even cognitive load measurement (Alemdag & Cagiltay, 2018). Even though, there are critical issues concerning the validity of *mean fixation duration* and the way it indicates cognitive processing, however, this measure seems appropriate as indicator for overall cognitive activity for task elaboration (Amadieu et al., 2009).

Controlling and redirecting information processing as well as self-monitoring are active *metacognitive processes* to activate and maintain cognitive effort for learning activities (De Bruin et al., 2020). Thus, metacognition can be related to the regulation and redirection of visual attention (Usher & Schunk, 2018). As outlined above, *transitions with a subsequent short visit duration* (<1 s) represent quick changes in the allocation of visual attention to potentially redirect visual attention to essential processing (Korbach et al., 2017). Thus, they can be assumed to indicate metacognitive activities during learning.

Finally, the fixation time outside the relevant information in relation to total gaze time, i.e., the *percentage of irrelevant fixation time*, is assumed to indicate *extraneous processing* in contradiction to relevant fixation time, which has already been used in other studies as indicator of generative (i.e., relevant; Moreno & Mayer, 2007) processing (e.g., Knörzer et al., 2016). With respect to the eye-mind hypothesis (Just & Carpenter, 1976) the fixation time outside the relevant AOIs shows the amount of time when the visual attention is not actively directed to cognitive processing of learning-relevant information.

## 2.5 | Research questions and hypotheses

The aim of the present study was to provide empirical support for the linkage of eye tracking indicators with (meta)cognitive processes of multimedia learning. Therefore, besides learning performance as dependent measure and eye tracking as process measure also think-aloud protocols were used for measuring (meta)cognitive processes during multimedia learning, which were deduced from the above-mentioned process model of multimedia learning (Stark et al., 2018). Further, exploiting the well-established seductive details effect, a two-group design with learners studying a multimedia instruction that either contained seductive details or did not, aimed at examining the sensitivity of both process measures. In addition, the linkage of both process measures as mediators of the seductive details effect as an instructional design effect was investigated in order to shed light on whether these process measures complement one another for explaining mechanisms of action of instructional design effects. This procedure also allowed to test the robustness of postulated relations between both measures. The following research questions and hypotheses were formulated:

> **RQ1.** Relationship between eye tracking and think-aloud data. To what extent are the eye tracking indicators related to specific verbalized (meta)cognitive processes in think-aloud protocols? Positive correlations between eye tracking indicators and corresponding think-aloud categories were postulated. In addition, it was hypothesized that these relations were robust against instructional design effects.

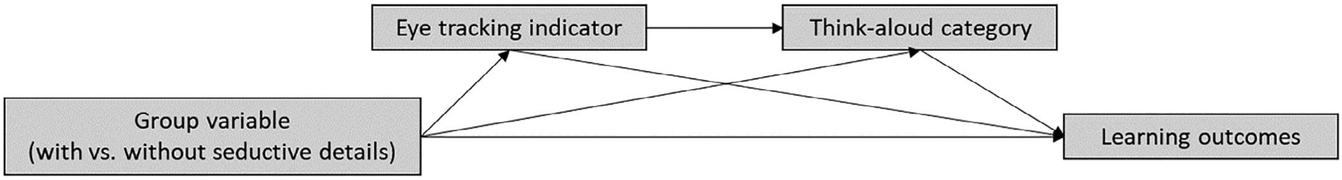


**FIGURE 1** Postulated serial mediation model.

> **RQ2.** Sensitivity regarding the seductive details effect. To what extent is the seductive details effect reflected not only in learning outcomes, but also in eye tracking and think-aloud data?

It was assumed that learners who received the multimedia instruction with seductive details showed lower learning success. Concerning visual attention, it was assumed that learners who received the multimedia instruction with seductive details showed less visual attention for the processing of relevant information and higher visual attention for extraneous processing reflected in lower respectively higher values of the related eye tracking indicators. Concerning the verbalized (meta)cognitive processes it was assumed that learners who received the multimedia instruction with seductive details showed a smaller percentage of verbalized selecting, elaborating, organizing and metacognitive processes and a larger percentage of verbalized extraneous processes.

> **RQ3.** Serial mediation of the seductive details effect by process data. To what extent do eye tracking and think-aloud data complement one another to mediate the seductive details effect? Therefore, corresponding indicators of both methods were regarded in serial mediation analyses, whereof the conceptual serial mediation model is displayed in Figure 1.

In general, it is assumed that the seductive details effect is mediated by eye tracking indicators as a first stage mediator and their corresponding verbalized (meta)cognitive processes obtained from the think-aloud protocols as a second stage mediator. This order of mediating variables was chosen by referring to Helle (2017), who pointed out that speech production is delayed to visual processing. Even though, the significance of mediating effects is hypothesized, the question whether one process measure or the combination of process measures will contribute more to the mediation is investigated in an explorative manner.

## 3 | METHOD

### 3.1 | Sample and design

The present sample originally consisted of $N = 60$ university students from the department of psychology (age: $M = 21.7$ years, $SD = 2.8$, 88% female) who took part voluntarily in the present study. Five participants had to be excluded from the analyses because their think-aloud protocols were not recorded properly or their sample quality of the eye tracking procedure was too low. Finally, $N = 55$ data sets (age: $M = 21.5$ years, SD $= 2.8$, 89% female) were considered for the analyses.

A one-factorial design with two groups was applied by varying the factor *seductive details* (with vs. without).

### 3.2 | Material

The multimedia instruction dealt with the structure and function of the ATP synthase molecule and has already been used in previous studies (e.g., Korbach et al., 2017; Park et al., 2020; Stark et al., 2018). It consisted of 11 slides presenting a combination of text and static picture on 10 of the 11 slides, whereas the first slide comprised only text information. The multimedia instruction was presented on a 17-inch laptop computer screen with a resolution of 1920 × 1080 pixel.

Learning time was partly standardized: A minimum time for each slide was predefined from data of previous studies using the same learning instruction. A forward button appeared after the minimum time in the lower right corner of the respective slide. There was no maximum time for avoiding interferences with the think-aloud procedure. It was not possible to go backwards during working with the multimedia instruction.

The seductive details group received additional text and picture information on 4 of the 11 slides (slides 2, 3, 4, and 9) about the usage and benefits of ATP for humans and animals (cf. Korbach et al., 2017; Park et al., 2020). This information was neither necessary nor helpful to understand the learning content or to reach the learning goal (see Figure 2). The seductive details were chosen by interestingness, irrelevance, concreteness, conciseness, emotionality and by reference to the relevant topic (Garner et al., 1989) and they have already been applied in previous studies using the same learning instruction (e.g., Park et al., 2020).

### 3.3 | Measures

*Learning performance*. Learning performance was measured by a learning performance test consisting of 27 items (Cronbach's $\alpha = 0.83$) with item difficulties in the range of $0.20 < p_i < 0.80$. The test included 15 closed-ended (e.g., "The matrix is …"—the inside of the

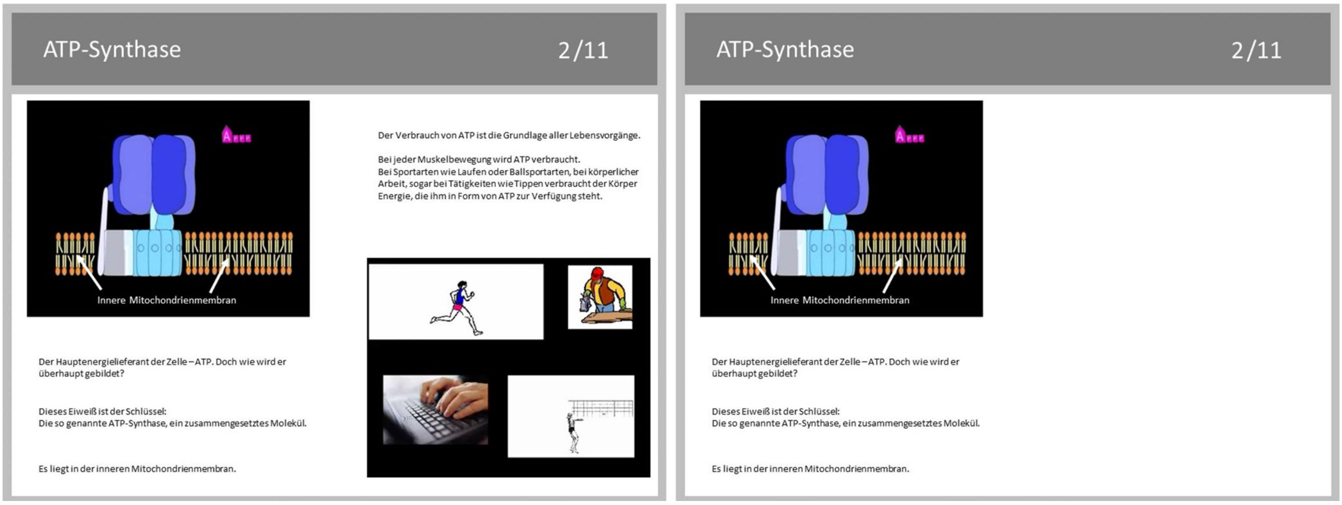


**FIGURE 2** Screenshot of the learning program, with and without seductive details.

mitochondrium; the intermembrane space; a united cell structure in tissues; the space outside the mitochondrium) and 12 open-ended questions (e.g., Describe the term "proton-motive force."). The closed-ended tasks were scored with one point for each correct answer and the open-ended response tasks with one to three points according to the number of aspects that were necessary for a correct response (max. 30.5 points).

*Eye tracking*. A Tobii ×2-60-compact eye tracker was used for recording the gaze behaviour during the learning phase. A remote eye tracking setting was used in order to avoid eye tracking data from being affected by movements of the participants' during working on the multimedia instruction. The system was calibrated immediately before the recording started, using a nine-point calibration. The calibration results were visually checked and only participants with proper hits on all nine calibration points and a deviation below 35 pixels were included. One participant had to be excluded because the system could not be calibrated with sufficient accuracy. The analysis of gaze behaviour was conducted using Tobii-Studio software and fixations were identified using the Tobii fixation filter settings with a duration threshold of 100 ms, a velocity threshold of 35 pixel and a distance threshold of 35 pixel. The total sample quality of the gaze recordings was analysed to ensure high data quality and served as control variable in the present study to ensure comparability of recordings between the groups. *Sample quality* was calculated as the ratio of gaze duration and learning time with a total sample quality over 78% ($M = 93.0$, $SD = 2.9$). For the analysis of gaze behaviour Areas of Interest (AOIs) were set for the learning relevant textual and pictorial information on each slide. Because time on task was not constant, indicators were calculated per second as frequency or relativized by learning time or total gaze. Table 1 provides an overview over the computed eye tracking indicators in line with the theoretical considerations.

**TABLE 1** Deduced eye tracking indicators for specific (meta) cognitive processes.

| (Meta) cognitive process | Eye tracking indicator | Computed eye tracking variable |
|---|---|---|
| Selecting | Number of fixations | Total number of fixations relativized by time-on-task that is, fixations per second |
| Organizing: Local coherence formation | Fixation duration | Sum of all fixation durations relativized by time-on-task |
| Organizing: Horizontal coherence formation | Number of transitions with long subsequent visit duration | Total number of transitions with a subsequent visit duration >1 s relativized by time-on-task that is, transitions (>1 s) per second |
| Organizing: Vertical coherence formation | Visit duration subsequent to transitions | Sum of all visit durations subsequent to transitions relativized by time-on-task |
| Elaborating | Mean fixation duration | Mean value of all fixation durations |
| Metacognitive processes | Number of transitions with short subsequent visit duration | Total number of transitions with a subsequent visit duration <1 s, relativized by time-on-task that is, transitions (<1 s) per second |
| Extraneous processes | Percentage of irrelevant fixation time | Sum of all fixation durations outside the learning relevant AOIs relativized by total gaze duration (i.e., total gaze time of a participant) |

*Think-aloud protocols*. All participants were instructed following Hak et al. (2004) to verbalize their thoughts during the learning process (cf. Park et al., 2020). As part of the instruction, participants received an example of an appropriate verbalization of thoughts and had to verbalize their own thoughts for an example as well. If necessary, participants had the opportunity to practice verbalization. The think-aloud protocols were recorded together with the gaze behaviour by Tobii Studio software and transcribed for qualitative analyses. In case that participants did not verbalize any thoughts for 30 s, the instructor reminded them to do so. The protocols were analysed applying deductive qualitative content analysis (Mayring, 2015). As coding units, sense-making segments of different lengths from few words to a full sentence were identified.

*Control measures*. The following learning prerequisites and process variables served as control measures: (1) *Prior knowledge* was measured by nine items about the topic of ATP (*Cronbach's* $\alpha = 0.67$). Five of them were in closed-ended format (e.g., "What is the main function of ATP?"—storage of phosphate; generation of an electrochemical gradient; storage of Adenosine; storage of energy) and four of them as open-ended questions (e.g., Please explain the meaning of the term "Synthesis"). (2) *Visuospatial working memory capacity* was measured by the Corsi Block tapping task (Schellig & Hättig, 1993). The length of the block sequence increased until participants produced three fails in row. The maximal sequence length with at least two correct trials indicated the individual visuospatial block span. (3) *Spatial ability* was measured by a standardized paper-folding and card-rotation test (Ekstrom et al., 1976). (4) *Learning motivation* was measured with the subscales intrinsic motivation (four items; Cronbach's $\alpha = 0.83$) and external regulation (four items; Cronbach's $\alpha = 0.85$) from the Situational Motivation Scale (SIMS; Guay et al., 2000). (5) In addition, as learning time was not restricted, *time on task* served as control measure.

## 3.4 | Category system

The category system (cf. Stark et al., 2020) consisted of the following categories, which were considered for deductive analysis of the think-aloud protocols. These categories were directly inferred from the process model of multimedia learning (see Section 2.1.1). Five of the think-aloud protocols were analysed by two independent coders reaching a reliability of $\kappa = 0.72$.

*Selecting*. The first step for information processing is selecting relevant information in words and pictures that will be further processed. At this stage, relevant information is perceived (e.g., by reading text passages) but not yet related to other chunks of information (e.g., "Now, on the picture, I'm seeing the stiff part and the rotating part."; [participant]16:[paragraph]8; "…and the rotation enforces changes in the head part [*citing the written text*]"; 28:17).

*Organizing*. Understanding and contextualizing selected information can be subsumed under the category organizing. Three different organizing processes were distinguished: (1) Processes of *local coherence formation* referred to cognitive processes dedicated at comprehending information of verbal or pictorial information respectively ("So, synthesis means the formation of ATP [rephrasing the text]"; 21:9; "There you see, how it is built"; 07:8). (2) Matching processes between verbal and pictorial information were considered as processes of *horizontal global coherence formation* ("and, thereby, the movable parts are opened, the axis and that part, there below" 15:41–42). (3) *Vertical global coherence formation* referred to contextualizing information of the learning program with reference to the whole topic like bringing information from different screens of the learning program together ("Okay, so, O- L-, and T-conformation [*summarizing information from three different slides*]"; 21:18). In contrast to previous research (Stark et al., 2018), the three organizing categories were analysed distinctively as they refer to qualitatively different processes.

*Elaborating*. Information that has been selected and organized is to be encoded and stored in long-term memory by elaborating the information. Here, processes of connecting the learnt information to prior knowledge ("Yes, that is what I remember from school: Mitochondria are the powerhouses of the cells…"; 21:16–17) as well as genuine elaboration processes or strategies like building memory hooks ("Looks like a flower"; 12:16) were considered for this category.

*Extraneous processes*. All processes irrelevant to achieve the learning objective can be regarded as extraneous processes ("still a bad graphic"; 25:90; "I just had to sneeze"; 27:43). In case of the present study also all processes regarding the seductive (i.e., irrelevant) information of the learning program were subsumed under this category.

*Metacognitive processes*. Processes of monitoring as well as planning behaviour were coded using this category ("First, I'll read the text to have and overview"; 7:3; "But I don't understand, what this change of conformation is"; 26:13). Whereas the other categories cannot overlap by definition, metacognitive processes often refer to other cognitive processes (e.g., [un]successful organizing processes) so that overlapping was allowed with this category (e.g., "[…] now I understand where the F0 component should be"; 7:18; "True, it always has three Ps. I just forgot. […]"; 35:25).

In order to keep the analyses independent from different learning times and the respective length of think-aloud protocols, proportions of verbalized processes were computed (number of codings in one category divided by total number of codings). The *total number of codings* was used as additional control measure.

## 3.5 | Procedure

The participants were tested in individual settings in the eye tracking lab. First, they were informed about the experimental procedure and informed consent was obtained before they started to participate in the study. The study started with a demographic questionnaire, followed by the tests of visuospatial working memory capacity and of spatial ability as well as by the questionnaire for learning motivation. After the prior knowledge test, the think-aloud method was instructed. Then, the eye tracking system was calibrated and the recording of gaze behaviour and verbalizations started immediately after the calibration and before the presentation of the learning instruction. Participants were able to proceed individually in the learning instruction when forward buttons appeared after a minimum reading time for each slide with learning time not being restricted. The recording of gaze behaviour and verbalizations was stopped after the last slide of the learning instruction and before the post test for learning performance was conducted.

### 3.6 | Methodical remarks

Especially for variables which are bounded by zero (e.g., think-aloud data), normality of the data could not be assumed so that robust test statistics or bootstrapping was used. Thus, for all analyses, conventionally, $\alpha = 0.05$ served as level of significance when applicable or bootstrapped bias corrected and accelerated (Bca) 95% confidence intervals (CI) were regarded. To provide the possibility of replicating the results, the same starting point for the Mersenne twister (5289, random choice) was implemented.

All statistical procedures were calculated using the IBM SPSS Statistics (Version 27). In order to test the postulated relations between eye tracking indicators and think-aloud data (RQ1), the bivariate Pearson correlation coefficients between indicators for specific (meta)cognitive processes of both methods were computed and interpreted referring to 1000 bootstrap samples and Bca 95% CI. Then, the sensitivity regarding the seductive details effect (RQ2) was examined. Independent t-tests with 1000 bootstrap samples were used for detecting potential differences in control measures and learning performance. For correlated data, Hotelling's trace statistics in multivariate analyses were calculated for eye tracking and think-aloud data respectively. These tests were followed-up by independent t-tests with 1000 bootstrap samples regarding each dependent measure for comparing the two experimental conditions with each other.

For investigating the potential of both process measures for explaining mechanisms of work of the seductive details effect (RQ3), the regression-based approach for conditional process modelling (Hayes, 2018) was applied in IBM SPSS Statistics (Version 27). Separate serial mediation models were computed with two mediators in a row (model 6). Here, the eye tracking indicators served as first stage mediator and the corresponding think-aloud indicator as second stage mediator according to methodological considerations that the verbalization follows the (meta)cognitive processes indicated by gaze behaviour (Helle, 2017). For these analyses, all outcome and mediating variables were z-standardized and the group variable was contrast coded with $-1$ and $+1$. Similar to previous analyses, 1000 bootstrap samples were used to test direct paths of the mediation models in IBM SPSS Statistics (Version 27). A total of 10,000 bootstrap samples was used to test indirect paths of mediation in line with recommendations by Hayes (2018) using his macro. In order to have all mediation models based on the same bootstrapped sample, seed was set to 5289 (random choice). The second regression model as part of the mediation analyses with the think-aloud category as criterium and the eye tracking indicator and the group variable as predictors also provided evidence regarding the second part of RQ1 referring to the robustness of correlations.

## 4 | RESULTS

### 4.1 | Relationships of eye tracking indicators and think-aloud categories

The assumptions concerning the relations of eye tracking indicators and specific verbalized (meta-)cognitive processes of think-aloud protocols were only partially confirmed by the results of the correlation analyses (see Table 2). In accordance with the assumptions, the eye tracking indicators for horizontal coherence formation and vertical coherence formation as well as for extraneous processes showed positive correlations with the respective think-aloud variables. In contrast to the assumptions, the number of fixations was positively but not significantly related to verbalized selecting processes, fixation duration was not related to verbalized local coherence formation processes, and eye tracking indicators for elaboration and metacognitive processes showed no correlations with their corresponding think-aloud categories.

Further, it has to be noted that the intercorrelations of eye tracking indicators were mostly significant with correlation coefficients with absolute values in the range $0.15 < |r| < 0.85$ (Table 2). As a consequence, whereas some indicators shared almost no variance others shared up to 72% of variance. The intercorrelations of the think-aloud variables were partly significant. Whereas local coherence formation had no significant relations to the other variables, selecting had substantial negative correlations to vertical coherence formation, elaborating and metacognitive processes sharing up to 34% of variance. In addition, horizontal coherence formation was negatively related to elaborating and metacognitive processes, which in turn were related positively.

The think-aloud categories vertical coherence formation, elaboration, and metacognitive processing had substantial positive correlations and extraneous processes a substantial negative correlation with learning performance. Regarding the eye tracking indicators, horizontal and vertical coherence formation (i.e., number of transitions with long subsequent visit duration and visit duration subsequent to transitions) showed significant positive correlations and for extraneous processes a negative correlation with learning performance.

### 4.2 | Sensitivity regarding the seductive details effect

There were no significant between-group differences in the control measures prior knowledge, spatial ability, and working memory capacity. In addition, the two groups did not differ significantly in time on

**TABLE 2** Bivariate correlations between all variables with significance tested using 1000 bootstrap samples and 95% bias corrected and accelerated confidence intervals.

| Variable | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Prior knowledge | −0.19 | 0.11 | −0.20 | −0.11 | 0.15 | −0.03 | −0.06 | 0.38* | 0.04 | −0.04 | 0.28* | 0.32* | −0.04 | 0.07 | 0.02 | −0.01 | −0.15 | 0.06 | 0.02 | 0 | 0.1 | −0.12 |
| 2. Time-on-task | 1 | −0.15 | −0.01 | 0.22 | −0.16 | 0.23 | 0.60* | 0.21 | −0.42* | −0.31* | −0.07 | 0.13 | 0.15 | −0.24 | 0.06 | −0.19 | −0.05 | −0.35* | 0.37* | 0.32* | 0.08 | 0.11 |
| 3. Spatial ability | | 1 | 0.25 | 0.1 | 0.08 | −0.27* | −0.12 | 0.12 | 0.02 | −0.03 | 0.01 | −0.02 | −0.08 | 0.02 | −0.10 | −0.13 | −0.07 | 0.18 | 0.06 | 0.17 | −0.05 | 0 |
| 4. Working memory capacity | | | 1 | 0.2 | −0.41* | −0.10 | −0.11 | −0.09 | −0.12 | −0.12 | −0.17 | −0.27 | −0.03 | 0.04 | −0.12 | 0.03 | 0.12 | 0.01 | 0.15 | −0.14 | −0.08 | −0.08 |
| 5. Learning motivation (intrinsic) | | | | 1 | −0.53* | −0.28 | 0.21 | −0.14 | −0.26 | −0.12 | 0.03 | −0.00 | 0.15 | −0.02 | 0 | −0.20 | 0.06 | 0.09 | −0.01 | 0.15 | 0.02 | 0.1 |
| 6. Learning motivation (extrinsic) | | | | | 1 | 0.06 | −0.19 | 0.1 | 0.32* | 0.26 | 0.04 | 0.12 | −0.08 | 0.11 | −0.08 | 0.19 | −0.09 | 0.07 | −0.04 | −0.13 | −0.11 | −0.04 |
| 7. Sample quality | | | | | | 1 | 0.08 | 0.01 | −0.00 | 0.02 | 0.08 | 0.05 | 0.05 | −0.04 | −0.03 | −0.18 | 0.21 | −0.10 | 0.32* | −0.02 | 0.04 | −0.08 |
| 8. Number of codings | | | | | | | 1 | 0.18 | −0.31* | −0.12 | 0.07 | 0.19 | 0.30* | −0.08 | 0.07 | −0.40* | 0.08 | −0.53* | 0.2 | 0.52* | 0.47* | −0.04 |
| 9. Learning performance | | | | | | | | 1 | 0.01 | 0.01 | 0.40* | 0.52* | 0.05 | 0.22 | −0.24* | −0.19 | −0.26 | −0.01 | 0.30* | 0.34* | 0.28* | −0.40* |
| 10. Number of fixations | | | | | | | | | 1 | 0.69* | 0.26 | 0.19 | −0.25 | 0.50* | −0.48* | 0.17 | 0.01 | 0.39* | −0.04 | −0.08 | −0.21 | −0.33* |
| 11. Fixation duration | | | | | | | | | | 1 | 0.47* | 0.40* | 0.48* | 0.19 | −0.59* | 0.04 | 0.03 | 0.29 | 0.02 | −0.05 | −0.01 | −0.40* |
| 12. Number of transitions with long subsequent visit duration | | | | | | | | | | | 1 | 0.85* | 0.37* | 0.35* | −0.32* | −0.41* | −0.24 | 0.36* | 0.31 | 0.07 | 0.34* | −0.40* |
| 13. Visit duration subsequent to transitions | | | | | | | | | | | | 1 | 0.40* | 0.23* | −0.31* | −0.34* | −0.31* | 0.13 | 0.30* | 0.22 | 0.38* | −0.35* |
| 14. Mean fixation duration | | | | | | | | | | | | | 1 | −0.35* | −0.15 | −0.13 | −0.02 | −0.14 | 0.1 | 0.03 | 0.24 | −0.09 |
| 15. Number of transitions with short subsequent visit duration | | | | | | | | | | | | | | 1 | −0.15 | −0.20 | −0.06 | 0.45* | 0.05 | 0 | 0.07 | −0.28* |
| 16. Percentage of irrelevant fixation time | | | | | | | | | | | | | | | 1 | 0.05 | −0.04 | −0.22 | −0.29* | −0.22 | −0.08 | 0.73* |
| 17. Selecting | | | | | | | | | | | | | | | | 1 | −0.17 | −0.17 | −0.33* | −0.35* | −0.59* | 0.04 |
| 18. Organizing: LCF | | | | | | | | | | | | | | | | | 1 | −0.06 | 0.12 | −0.06 | −0.21 | −0.12 |
| 19. Organizing: HCF | | | | | | | | | | | | | | | | | | 1 | −0.04 | −0.26* | −0.25* | −0.17 |
| 20. Organizing: VCF | | | | | | | | | | | | | | | | | | | 1 | 0.2 | −0.13 | −0.32* |
| 21. Elaborating | | | | | | | | | | | | | | | | | | | | 1 | 0.23* | −0.23* |
| 22. Metacognitive processes | | | | | | | | | | | | | | | | | | | | | 1 | −0.20 |
| 23. Extraneous processes | | | | | | | | | | | | | | | | | | | | | | 1 |

*Note*: * Significant due to zero not being included in bootstrapped 95% Bca CI with 1000 bootstrap samples.

TABLE 3 Means and standard deviations for both groups as well as results of group comparisons using *t* tests with 1000 bootstrap samples and 95% bias corrected and accelerated confidence intervals.

| Variable | noSD ($n = 27$) *M* | noSD ($n = 27$) SD | SD ($n = 28$) *M* | SD ($n = 28$) SD | $t(53)$ | *p* | Bootstrap Mean difference | Bootstrap 95% CI [LL, UL] | *d* |
|---|---|---|---|---|---|---|---|---|---|
| Control measures | | | | | | | | | |
| Prior knowledge | 3.2 | 2.2 | 2.4 | 1.9 | 1.57 | n.s. | 0.87 | −0.18, 2.03 | 0.42 |
| Time-on-task (s) | 778.3 | 265.7 | 932.9 | 345.3 | −1.86[a] | n.s. | −154.52 | −326.21, 33.03 | −0.50 |
| Spatial ability | 69.5 | 15.1 | 69.2 | 13.1 | 0.08 | n.s. | 0.30 | −8.30, 8.21 | 0.02 |
| Working memory capacity | 5.9 | 0.8 | 5.7 | 0.8 | 1.15 | n.s. | 0.25 | −0.18, 0.66 | 0.31 |
| Learning motivation (intrinsic) | 4.1 | 1.2 | 4.3 | 1.0 | −0.74 | n.s. | −0.23 | −0.88, 0.35 | −0.20 |
| Learning motivation (extrinsic) | 2.4 | 1.4 | 2.3 | 1.3 | 0.40 | n.s. | −0.15 | −0.61, 0.96 | 0.11 |
| Sample quality (%) | 93.1 | 2.6 | 92.8 | 3.3 | 0.42 | n.s. | 0.33 | 0.98, 1.73 | 0.11 |
| Number of codings | 44.7 | 25.4 | 54.5 | 25.5 | −1.43 | n.s. | −9.80 | −22.68, 3.53 | −0.38 |
| Learning performance | 16.8 | 5.8 | 13.1 | 5.1 | 2.55 | 0.014 | 3.76 | 0.75, 6.91 | 0.69 |
| Eye tracking indicators | | | | | | | | | |
| Number of fixations | 2.4 | 0.4 | 2.0 | 0.3 | 3.95 | <0.001 | 0.40 | 0.19, 0.58 | 1.07 |
| Fixation duration | 80.8 | 14.5 | 68.3 | 9.3 | 3.83 | <0.001 | 12.52 | 5.42, 18.61 | 1.03 |
| Number of transitions with long subsequent visit duration | 0.04 | 0.02 | 0.03 | 0.01 | 3.17[a] | 0.003 | 0.01 | 0.005, 0.02 | 0.86 |
| Visit duration subsequent to transitions | 14.1 | 5.8 | 10.5 | 4.1 | 2.66[a] | 0.011 | 3.60 | 0.80, 6.59 | 0.72 |
| Mean fixation duration | 0.3 | 0.1 | 0.3 | 0.1 | −0.11 | n.s. | −0.002 | −0.03, 0.03 | −0.03 |
| Number of transitions with short subsequent visit duration | 0.06 | 0.02 | 0.05 | 0.02 | 2.82 | 0.007 | 0.02 | 0.01, 0.03 | 0.76 |
| Percentage of irrelevant fixation time | 5.5 | 4.6 | 19.1 | 5.6 | −9.86 | <0.001 | −13.59 | −16.22, −10.96 | −2.66 |
| Think-aloud category | | | | | | | | | |
| Selecting (%) | 0.26 | 0.18 | 0.28 | 0.15 | −0.30 | n.s. | −0.01 | −0.09, 0.07 | −0.08 |
| Organizing: LCF (%) | 0.08 | 0.08 | 0.06 | 0.06 | 0.96 | n.s. | 0.02 | −0.02, 0.05 | 0.26 |
| Organizing: HCF (%) | 0.25 | 0.11 | 0.18 | 0.08 | 2.65 | 0.010 | 0.07 | 0.01, 0.13 | 0.72 |
| Organizing: VCF (%) | 0.13 | 0.10 | 0.09 | 0.07 | 2.00 | 0.050 | 0.04 | 0.001, 0.09 | 0.54 |
| Elaborating (%) | 0.06 | 0.07 | 0.04 | 0.05 | 1.12 | n.s. | 0.02 | −0.01, 0.05 | 0.30 |
| Metacognitive processes (%) | 0.20 | 0.17 | 0.18 | 0.14 | 0.33 | n.s. | 0.01 | −0.06, 0.10 | 0.09 |
| Extraneous processes (%) | 0.02 | 0.03 | 0.17 | 0.06 | −11.60[a] | <0.001 | −0.15 | −0.17, −0.13 | −3.09 |

Abbreviations: CI [LL, UL], confidence interval with lower and upper limit; HCF, horizontal coherence formation; LCF, local coherence formation; noSD, group without seductive details; SD, group with seductive details VCF, vertical coherence formation; VD, visit duration.
[a]Adjusted degree of freedom due to significant Levene test of homogeneity of variances.

task, sample quality, and the total number of coded segments of the think-aloud protocols (see Table 3).

As predicted, the group learning without seductive details outperformed the group learning with seductive details significantly(see Table 3).

Regarding eye tracking variables as dependent measures, Hotelling's trace statistics revealed a significant difference between the two groups, $T = 2.63$, $F(7, 47) = 17.63$, $p < 0.001$, $\eta^2 = 0.72$. Subsequent *t*-tests showed significant results for the eye-tracking indicators with an overall lower visual processing activity for the seductive details group except for mean fixation duration as an indicator for elaborating, for which the between-group difference was not significant (see Table 3).

Hotelling's trace statistics revealed a significant effect, $T = 3.24$, $F(7, 47) = 21.77$, $p < 0.001$, $\eta^2 = 0.76$ regarding the think-aloud indicators of (meta)cognitive processes as dependent variables and group as between-subject factor. Subsequent *t*-tests (see Table 3) showed significant between-group differences in the variables horizontal coherence formation, vertical coherence formation, and extraneous processes. There were no significant between-group differences in the dependent measures selecting, local coherence formation, elaborating, and metacognitive processes. Hence, seductive details seemed to not affect those processes. However, learning with seductive details led to less processes of horizontal and vertical coherence formation as well as to more extraneous processes in comparison to learning without seductive details.

**TABLE 4** Results of linear regression models regarding serial mediation analyses using 1000 bootstrap samples and bias corrected and accelerated 95% confidence intervals.

| | Dependent measures | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1st mediator | | | 2nd mediator | | | Learning outcomes | | |
| *Model 1: Selecting* | Number of fixations | | | Selecting | | | | | |
| Model summary | $R^2 = 0.23$, $F(1, 53) = 15.6$, $p < 0.001$ | | | $R^2 = 0.05$, $F(2, 52) = 1.4$, *n.s.* | | | $R^2 = 0.16$, $F(3, 51) = 3.2$, $p = 0.031$ | | |
| | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ |
| Seductive details | −0.95 | −1.39, −0.45 | 0.002 | 0.30 | −0.28, 0.85 | n.s. | −0.79 | −1.38, 0.18 | 0.008 |
| Fixation count | | | | 0.25 | −0.06, 0.55 | n.s. | −0.15 | −0.43, 0.23 | n.s. |
| Selecting | | | | | | | −0.15 | −0.48, 0.09 | n.s. |
| *Model 2: Local coherence formation* | Fixation duration | | | Organizing: LCF | | | | | |
| Model summary | $R^2 = 0.22$, $F(1, 53) = 14.7$, $p < 0.001$ | | | $R^2 = 0.02$, $F < 1$ | | | $R^2 = 0.23$, $F(3, 51) = 5.1$, $p = 0.003$ | | |
| | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ |
| Seductive details | −0.92 | −1.33, −0.46 | 0.003 | −0.29 | −0.98, 0.63 | n.s. | −0.92 | −1.43, −0.26 | 0.003 |
| Fixation duration | | | | −0.04 | −0.35, 0.46 | n.s. | −0.20 | −0.38, 0.15 | n.s. |
| Organizing: LCF | | | | | | | −0.31 | −0.56, 0.002 | 0.015 |
| *Model 3: Horizontal coherence formation* | Number of transitions with long subsequent visit duration | | | Organizing: HCF | | | | | |
| Model summary | $R^2 = 0.16$, $F(1, 53) = 10.2$, $p = 0.002$ | | | $R^2 = 0.18$, $F(2, 52) = 5.7$, $p = 0.006$ | | | $R^2 = 0.24$, $F(3, 51) = 5.3$, $p = 0.003$ | | |
| | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ |
| Seductive details | −0.75 | −1.23, 0.31 | 0.004 | −0.46 | −0.95, 0.02 | n.s. | −0.52 | −1.19, 0.06 | *n.s.* |
| Number of transitions with long subsequent visit duration | | | | 0.29 | 0.03, 0.53 | 0.031 | 0.40 | 0.10, 0.73 | 0.008 |
| Organizing: HCF | | | | | | | −0.24 | −0.50, 0.03 | n.s. |
| *Model 4: Vertical coherence formation* | Visit duration after transitions | | | Organizing: VCF | | | | | |
| Model summary | $R^2 = 0.12$, $F(1, 53) = 7.1$, $p = 0.010$ | | | $R^2 = 0.12$, $F(2, 52) = 3.6$, $p = 0.036$ | | | $R^2 = 0.31$, $F(3, 51) = 7.6$, $p < 0.001$ | | |
| | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ |
| Seductive details | −0.68 | −1.16, −0.25 | 0.012 | −0.36 | −0.85, 0.14 | n.s. | −0.30 | −0.89, 0.27 | n.s. |
| Visit duration after transitions | | | | 0.24 | −0.09, 0.56 | n.s. | 0.43 | 0.15, 0.69 | 0.003 |
| Organizing: VCF | | | | | | | 0.13 | −0.09, 0.36 | n.s. |
| *Model 5: Elaborating* | Mean fixation duration | | | Elaborating | | | | | |
| Model summary | $R^2 = 0.00$, $F < 1$ | | | $R^2 = 0.03$, $F < 1$ | | | $R^2 = 0.19$, $F(3, 51) = 4.1$, $p = 0.011$ | | |
| | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ | $\beta$ | CI [LL, UL] | $p$ |
| Seductive details | 0.03 | −0.48, 0.53 | n.s. | −0.32 | −0.89, 0.21 | n.s. | −0.56 | −1.05, −0.04 | 0.036 |
| Mean fixation duration | | | | 0.03 | −0.15, 0.29 | n.s. | 0.04 | −0.18, 0.31 | n.s. |
| Elaborating | | | | | | | 0.29 | 0.06, 0.54 | 0.010 |
| *Model 6: Metacognitive processes* | Number of transitions with short subsequent visit duration | | | Metacognitive processes | | | | | |
| Model summary | $R^2 = 0.13$, $F(1, 53) = 7.9$, $p = 0.007$ | | | $R^2 = 0.01$, $F < 1$ | | | $R^2 = 0.17$, $F(3, 51) = 3.5$, $p = 0.022$ | | |

**TABLE 4** (Continued)

| | Dependent measures | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1st mediator | | | 2nd mediator | | | Learning outcomes | | |
| | *β* | CI [LL, UL] | *p* | *β* | CI [LL, UL] | *p* | *β* | CI [LL, UL] | *p* |
| Seductive details | −0.62 | −1.09, −0.21 | 0.013 | −0.08 | −0.68, 0.48 | n.s. | −0.56 | −1.11, −0.06 | 0.053 |
| Number of transitions with short subsequent visit duration | | | | 0.06 | −0.24, 0.39 | n.s. | −0.12 | 0.7 | n.s. |
| Metacognitive processes | | | | | | | 0.23 | 1.7 | n.s. |
| *Model 7: Extraneous processes* | Percentage of irrelevant fixation time | | | Extraneous processes | | | | | |
| Model summary | $R^2 = 0.65$, $F(1, 53) = 97.2$, $p < 0.001$ | | | $R^2 = 0.72$, $F(2, 52) = 67.8$, $p < 0.001$ | | | $R^2 = 0.17$, $F(3, 51) = 3.5$, $p = 0.022$ | | |
| | *β* | CI [LL, UL] | *p* | *β* | CI [LL, UL] | *p* | *β* | CI [LL, UL] | *p* |
| Seductive details | 1.54 | 1.24, 1.85 | 0.001 | 1.45 | 0.80, 2.045.2 | 0.001 | −0.13 | −1.11, 1.19 | n.s. |
| Percentage of irrelevant fixation time | | | | 0.15 | −0.14, 0.47 | n.s. | 0.16 | −0.34, 0.54 | n.s. |
| Extraneous processes | | | | | | | −0.46 | −0.91, −0.07 | 0.071. |

*Note*. *β*s are standardized values, CI [LL, UL] = Confidence interval with lower and upper limit.

### 4.3 | Serial mediation with data from eye tracking and think-aloud protocols

Serial mediation models (cf. Figure 1) were computed to examine to what extent eye tracking indicators and verbalized (meta)cognitive processes complement one another when investigating mechanisms of work of the seductive details effect. Table 4 shows the results of the included regressions with 1000 bootstrap samples. Results regarding tests of the indirect regression paths are displayed in Table 5.

Replicating the results of the group comparisons, there were significant regression models with the eye tracking indicators as dependent measures except for mean fixation duration as the indicator of elaborating processes (Table 4). Results of the regressions with think-aloud variables as dependent measures were not as consistent as expected: Only direct paths between the eye tracking and think-aloud variables of horizontal coherence formation as well as the direct effect of the group variable on extraneous processes substantial. Regressions on learning outcomes were significant showing that the group variable in combination with both process measures accounted for substantial variance in learning outcomes whereas a look at the direct paths revealed inconsistent findings: The direct effect of the group variable on learning outcomes was significant in the models of selecting, local coherence formation, and elaborating. The path from the eye tracking variable to learning outcomes was only significant in the model for horizontal and vertical coherence formation respectively. The direct path from the second stage mediator that is, the think-aloud variable to learning outcomes was substantial for local coherence formation and elaborating.

Tests of indirect paths of the regression models revealed no serial mediation of the seductive details effect when regarding corresponding eye tracking and think-aloud variables as serial mediators for all confidence intervals included zero. Only the following indirect paths in simple mediations were substantial with zero not being included in the confidence interval: The eye tracking indicators for horizontal and vertical coherence formation, mediated the seductive details effect, respectively. The think-aloud variable for extraneous processes also led to a substantial simple mediation of the seductive details effect. Thus, the seductive details effect could be explained by the eye tracking indicators for horizontal and vertical coherence formation as well as verbalized extraneous processes: Seductive details caused lower learning outcomes due to less transitions with long subsequent fixation duration lower visit duration after transitions, and more verbalized extraneous processes. Total effect sizes of indirect paths were only significant for the model comprising the process measures of vertical coherence formation.

## 5 | DISCUSSION

The main goal of the present study was to examine the relation of eye tracking metrics and (meta-)cognitive processes in order to provide an empirical basis for interpreting specific eye tracking indicators in terms of particular (meta)cognitive processes of multimedia learning.

**TABLE 5** Tested indirect effects for the serial mediation models.

| Model | SD → 1st m → LO $\beta$ | SD → 1st m → LO 95% CI [LL, UL] | SD → 2nd m → LO $\beta$ | SD → 2nd m → LO 95% CI [LL, UL] | SD → 1st m → 2nd m → LO $\beta$ | SD → 1st m → 2nd m → LO 95% CI [LL, UL] | Total indirect path $\beta$ | Total indirect path 95% CI [LL, UL] |
|---|---|---|---|---|---|---|---|---|
| Model 1: Selecting | 0.14 | [−0.17, 0.47] | −0.05 | [−0.30, 0.04] | 0.04 | [−0.02, 0.22] | 0.13 | [−0.17, 0.46] |
| Model 2: Local coherence formation | −0.19 | [−0.05, 0.43] | 0.09 | [−0.07, 0.36] | −0.01 | [−0.12, 0.10] | 0.27 | [−0.01, 0.58] |
| Model 3: Horizontal coherence formation | **−0.30** | **[−0.69, −0.07]** | 0.11 | [−0.01, 0.37] | 0.05 | [−0.00, 0.27] | −0.14 | [−0.50, 0.18] |
| Model 4: Vertical coherence formation | **−0.29** | **[−0.67, −0.06]** | −0.05 | [−0.22, 0.02] | −0.02 | [−0.13, 0.01] | **−0.36** | **[−0.76, −0.10]** |
| Model 5: Elaborating | −0.01 | [−0.01, 0.03] | −0.09 | [−0.34, 0.04] | −0.00 | [−0.01, 0.03] | −0.09 | [−0.32, 0.06] |
| Model 6: Metacognitive processes | −0.07 | [−0.34, 0.12] | −0.02 | [−0.22, 0.10] | −0.01 | [−0.10, 0.03] | −0.10 | [−0.41, 0.12] |
| Model 7: Extraneous processes | 0.24 | [−0.62, 0.77] | **−0.66** | **[−1.45, −0.05]** | −0.11 | [−0.50, 0.06] | −0.53 | [−1.53, 0.34] |

*Note*: significant indirect paths are printed in bold letters.
Abbreviations: CI [LL, UL], confidence interval with lower and upper limit; LO, learning outcomes; m, mediator; SD, seductive details (with vs. without); $\beta$, standardized values.

Therefore, with reference to a process model of multimedia learning, eye tracking indicators were theoretically deduced. The study examined relations between eye tracking indicators and verbalized (meta) cognitive processes of multimedia learning captured by think-aloud protocols. As instructional design effects constitute a primary aspect in multimedia learning research, the sensitivity of both process measures of (meta)cognitive processes was regarded considering the well-investigated seductive details effect of an established multimedia instruction. In a last step, serial mediations were calculated in order to investigate whether both process measures complement one another in a joint explanation of the seductive details effect.

There were no significant between-group differences in the control measures pointing at a high internal validity of the study. In sum, it has to be admitted that the results regarding the postulated hypotheses were rather disillusioning. However, they can be turned into implications for future research further that pursues the goal to triangulate process measures of (meta)cognitive processes of multimedia learning.

## 5.1 | Relations between eye tracking indicators and verbalized (meta)cognitive processes

The results only partly confirmed the assumed relations between eye tracking indicators and verbalized (meta)cognitive processes. The eye tracking indicator and corresponding think-aloud variable for extraneous processes showed a substantial high positive correlation as assumed. There were substantial positive correlations between the eye tracking indicators of horizontal and vertical coherence formation and their respective corresponding think-aloud variables. However, it has to be noted for both measures, that even higher relations were detected between these eye tracking indicators and other than the assumed corresponding think-aloud measures. Thus, a potential non-correspondence of eye tracking indicators and think-aloud categories can be discussed. Also for the eye tracking indicators of all other processes except for mean fixation duration, which did not have any substantial correlations to think-aloud variables, substantial correlations were detected with other than the assumed corresponding think-aloud variables. For instance, the number of transitions with a short subsequent fixation duration did not correlate substantially with metacognitive processes as expected but showed a very high correlation with horizontal coherence formation processes, which contradicts the assumption that only transitions with longer subsequent fixation times would be beneficial to such processes (cf., Mason et al., 2015). One explanation for the mismatch of eye tracking indicators and think-aloud protocols in contrast to previous studies (cf., Baß et al., 2024) might be that the operationalization for verbal reports and observable gaze was on a different behavioural level. With respect to the complexity of the learning task, it is questionable whether participants could introspect and verbalize their cognitions on the same level and with the same granularity as gaze behaviour was recorded, so there might be a gap for unconscious cognitive (Jarodzka et al., 2015) as well as metacognitive processes (Wirth et al., 2020). However, most notably, from an empiricist's perspective post-hoc explanations for detected relations might be a way to re-think the postulated relations, however, this procedure would not be in line with good scientific practice. In addition, then, operationalizations of multimedia learning processes and eye tracking indicators might risk to lack a theoretical basis.

The results of the correlation analysis can also be interpreted in a more positive light. The analyses revealed high intercorrelations between eye tracking indicators, which is quite usual for eye tracking studies that refer to more than one eye tracking variable. However, the relations of the eye tracking indicators to learning performance and think-aloud variables were not as highly consistent as one would assume when referring to the high intercorrelations that might have suggested that all eye tracking indicators would measure similar processes due to the shared variance and, thus, showing similar correlations to other measures. Thus, the results suggest that the eye tracking indicators indeed did measure qualitatively different (meta) cognitive processes and the diverse relations can help to shed light on the essence of what the eye tracking variables might actually indicate in future studies. Such studies could exploratively search for relations (e.g., by means of data mining) or systematically investigate differences by further exploiting multimedia learning effects (e.g., split-attention effect) with clear assumptions on (meta)cognitive mechanisms at work in order to be able to infer causal relations. Especially the latter procedure, which straightforwardly goes in line with the paradigm of multimedia learning research methodology, provides a means to generate validity hints of eye tracking indicators besides a correlative convergent validity approach using another process measure of the same constructs, which was chosen in the present study.

Regarding the robustness of relations of eye tracking and think-aloud data, i.e., the second regression models of the mediation analyses with the think-aloud variables as criterium, only for horizontal coherence formation the eye tracking indicator predicted the think-aloud variable besides group affiliation. Thus, this relationship can be regarded to be robust against the seductive details effect as one representative of instructional design effects in multimedia learning. However, it has to be kept in mind that most correlations were not substantial even when they were not controlled for the impact of variations of the learning instruction. Thus, it is difficult to infer interpretations regarding the robustness of relations when relations remain similarly around zero when controlling for the seductive details effect. Future studies with larger sample sizes might then also refer to further hints at the robustness of relations (e.g., residual analyses).

## 5.2 | Sensitivity regarding the seductive details effect

The results confirmed the seductive details effect with a lower learning performance for the seductive details group, which is in line with most studies on seductive details (e.g., Kienitz et al., 2023; Park et al., 2020). In addition, the eye-tracking indicators were sensitive to the seductive details effect, which was also the case in hitherto research (e.g., Korbach et al., 2016). There was an overall lower visual processing activity for the seductive details group except for mean fixation duration as an indicator of elaboration processes, where no between-group difference was detected. It must be considered that with respect to the different explanations for the seductive details effect (Rey, 2012) not all (meta)cognitive processes must be affected. However, in line with the null correlation with learning outcomes, it can be argued, that the eye tracking indicator for elaboration processes (i.e., mean fixation duration) might not function as it had been assumed. Future studies can disentangle these speculations by investigating another instructional design effect with an assumed impact on elaboration processes.

Concerning the sensitivity of the think-aloud method regarding the seductive details effect, assumptions were only confirmed regarding effects for extraneous processing and partially confirmed for organizing processes. The results indicated no differences in verbalized selecting, local coherence formation, elaborating and metacognitive processes. The large difference in the amount of verbalized extraneous processes can be explained by having declared the seductive details as irrelevant information following their definition. As a consequence, this between-group difference can be regarded as manipulation check underlining the effectiveness of the seductive information. There were no effects of the seductive details on selecting processes and on local coherence formation maybe because those local processes can refer to either the textual or the pictorial information of the learning program. Even though a category split in line with the multimedia processes mentioned in CTML (Mayer, 2021) might have been possible from a coder's point of view, the category system would not remain balanced considering size of the categories, which should be the case in qualitative content analysis (e.g., Mayring, 2015). Finally, the seductive details did not impact the amount of verbalized elaboration processes. On the one hand and in line with the previous argumentation, a consolidation of subcategories (integrating processes with prior knowledge vs. genuine elaboration processes) might have hided a potential seductive details effect. On the other hand, this result corresponds to the non-effect on the eye tracking indicator for elaboration. Indicators for global organizing processes (horizontal and vertical) were affected by the seductive details. Thus, the seductive details impaired integration processes concerning text-picture integration on one slide of the learning instruction as well as with information integration from different slides of the learning instruction.

Mediation analyses were conducted in order to investigate to what extent eye tracking and think-aloud data complement one another regarding the explanation of the seductive details effect. As mediation analyses refer to a regression-based, and thus correlative approach, the results were comparable of those of the correlation analyses, where for most corresponding variables non-substantial correlations were detected. Thus, the results showed no serial mediation with both process measures in a row and only single mediations with the eye tracking indicators of horizontal and vertical coherence formation as well as with the think-aloud variable extraneous processes. As outlined above, a discussion of these results would remain speculative as long as the validity of the eye tracking indicators is still questionable.

## 5.3 | Methodical remarks on using eye tracking data for measuring (meta)cognitive processes

As outlined above, eye tracking researchers might face the critique of choosing inappropriate metrics (Orquin & Holmqvist, 2018). Of course, also the indicators chosen in the present study must be reflected against the background of this critique, especially because the results were not as consistent as expected. For some of the indicators further theoretical reflections regarding their validity and further systematic investigations of their validity are needed. In sum, the results showed that eye-tracking indicators can be related to verbalized cognitive processes but at the same time these relations seem to be less specific as assumed (Alemdag & Cagiltay, 2018). There were no unique correlations between eye-tracking indicators and verbalized cognitive processes and the results of the mediation models suggested that cognitive processes were rather partially represented by the chosen indicators of gaze behaviour. Moreover, both methods share a similar problem that is the differentiation of mandatory and higher order cognitive processes respectively successful or unsuccessful processing activity within one indicator or category. For example, fixation frequency and verbalized horizontal coherence formation processes both were reduced by seductive details, but both showed a correlation around zero to learning performance and the partial correlations rather suggested that a lower fixation frequency as well as a lower percentage of verbalized processes was related to higher learning performance. With respect to exact process modelling the operationalization of these measures seems to be insufficient. As the present study used comparatively global eye movement indicators in accordance with the global category system and the given drawbacks, a separate analysis for text and picture processing might provide better information about local processing activities. However, using global indicators and categories paid attention to possible delays between visual information processing and speech production (Helle, 2017). As a consequence, the cued retrospective think-aloud method (Jarodzka et al., 2010) could be an optimization of the present study because it does alter neither the learning task, nor the needed time for task accomplishment, nor the visual focus of attention with respect to the learning task (Stark et al., 2020). From a different perspective, learning instructions that are designed to enhance specific (meta)cognitive processes could also solve the problem. Specifically, adequately designed AOIs, which directly provide information on the kind of which specific (meta)cognitive process was needed for their understanding, could allow interpreting established metrics like fixation duration or transitions also on a local level (Fan et al., 2022). Aligning gaze behaviour and verbal reports for a more detailed or even event related analyses could be achieved by implementing less complex, smaller, or small sequenced learning instructions in combination with the experience sampling method (Bühler et al., 2024). On a more holistic level and in correspondence with the nature of global processes, the analysis of partial or full scanpath data based on visits, saccades and revisits (Tjon et al., 2023) could be a promising approach, specifically when (meta)cognitive processes can be operationalized by a unique sequence of visits. With respect to the results of the present study, it seems also plausible that (meta)cognitive processing should be captured as a combination of different metrics or the combination of different states in one metric that can be identified as process related gaze patterns (Stark et al., 2024).

### 5.4 | Limitations and future research

As already outlined above, the use of the seductive details effect for testing the sensitivity of the process measures did also bring along a major limitation to the present study. Following the definition of seductive details, they are assumed to induce extraneous processing. Thus, effects and relations of indicators for extraneous processes measured either by the think-aloud procedure or eye tracking must be regarded as manipulation check and further generalizing interpretations remain speculations.

Regarding the use of the category system, it has to be noted that the total number of codings was relatively low when linked with time-on-task. Thus, there may be an underestimation of (meta)cognitive processes. Some participants had problems either to verbalize their thoughts generally or to keep their verbalization level high throughout the whole learning time. In addition, a lot of cognitions during learning occur rather unconsciously and thinking-aloud always faces the critique of a potential loss of information due to filtering by participants. Especially local coherence formation might underlie these restrictions. As a consequence, these processes might even be more underestimated which might explain that the percentage of codings for local coherence formation is lower than for global (horizontal and vertical) coherence formation. In addition, it is an inherent limitation of the applied process model that metacognitive processes overlap with cognitive processes per definition. Thus, it can be questioned to what extent cognitive processes (e.g., selecting) can be analysed separate from metacognitive processes, which are dedicated at guiding learners' attention and regulating their effort invested.

With regard to verbalized processes it has to be noted that proportions of codings in particular categories were integrated in the analyses for good reason (non-standardized learning time). However, another pattern of results could have been detected if absolute values were considered. The same issue accounts for eye tracking measures where frequency values and other relativized measures were used. Regarding data of the think-aloud protocols, even deeper insights could have been enabled by qualitative analyses of individual cases as with quantification of coded segments a loss of information cannot be avoided.

As a limitation of the methodological approach of serial mediation analysis for integrating eye tracking and think-aloud data, it has to be noted that it does not reach the noble goal of really putting together eye movements and speech production. A time stamp analysis could be one of the next steps in a research program to further shed light on direct links between eye tracking data and specific cognitions in general. Such a kind of analysis would provide more fine-grained results and could be another validation approach in future studies. Regarding the dissatisfying results of the mediation analyses it also has to be stated that even if eye tracking indicators and think-aloud categories corresponded, results of the mediation analyses could potentially show a similar picture because the mediating variables would share a larger amount of variance and might be confounded. This is a major limitation to the use of serial mediation as drawn inferences often remain speculative.

In addition, the validation approach applied in the present study must be discussed. Here, the theoretical deduction of the eye tracking indicators was used in order to ensure construct validity specifically face validity and content validity. In the empirical part, correlations with learning outcomes and especially the aspect of convergent validity was regarded. Therefore, the think-aloud method was used as a second process measure of (meta)cognitive processes. It has to be argued that also the think-aloud method and the applied category system, even though already used in multimedia research, had not been validated from a strict empirical perspective in advance. Thus, the question arises how much inference can be drawn from the given results as both process measures should still undergo validation processes.

Further research is needed to identify differential indicators that help to link eye tracking and data of think-aloud protocols in validation studies. One point to start from would be to use a less complex learning instruction and a learning task that allows precise operationalization of necessary cognitive processes with regard to the achievement of the learning goal on a more basic level. Moreover, the design of the learning instruction should allow the analysis of eye movements with regard to all cognitive processes, including global processes like vertical coherence formation. However, not only experimental research is needed in order to systematically investigate whether established multimedia effect (e.g., the split-attention effect) have corresponding effects on eye tracking and verbal data, but also correlational validation studies can generate evidence for discriminant validity of corresponding eye tracking and verbal data.

### 5.5 | Implications

The study mainly provides implications for research as it can be regarded to be a baseline study with more focus on internal than external validity. Besides already addressed research desiderata, the study showed how innovative ways of validating can be implemented. This implication cannot only be applied to multimedia learning but also other fields of educational research, as well as for further developments regarding multimodal learning analytics. In addition, the study underlines the necessity of incorporating a process perspective as it helps to understand how learners deal with learning instruction. Multimodal learning analytics provide a promising approach to align different kind of process data for a holistic analysis of learning processes in relation to observable learning behaviour. However, it also has been considered how different process measures provide information on diverse levels of granularity, that there always is a loss of information when analysing multimodal data, and the extent to which analyses remain speculative as interpretations are needed because of rather indirect indicators for learning processes.

Even though implications for practice were not the major intent of the present study, it can be concluded that also for practitioners, it is crucial to analyse learning processes by means of using as much process information as possible. In practice, mainly observations but also verbal reports like think-aloud procedures can be applied. Analysing these data can help identifying learning gaps, ill designed aspects of learning materials and settings, and, therefore, optimizing and fostering learning. However, practitioners should also keep in mind the limitations of process data analysis as they can never draw a real complete picture of learning.

## AUTHOR CONTRIBUTIONS

**Lisa Stark:** Writing – original draft; writing – review and editing; formal analysis; visualization; investigation; methodology; data curation. **Andreas Korbach:** Investigation; writing – original draft; methodology; visualization; writing – review and editing; formal analysis; data curation. **Roland Brünken:** Conceptualization; supervision; writing – review and editing. **Babette Park:** Conceptualization; supervision; project administration; writing – review and editing.

## ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

*Lisa Stark* https://orcid.org/0009-0009-6450-9504
*Andreas Korbach* https://orcid.org/0000-0003-1044-3406

## REFERENCES

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction*, *19*, 376–386. https://doi.org/10.1016/j.learninstruc.2009.02.005

Andaloussi, A. A., Zerbato, F., Burattin, A., Slaats, T., Hildebrandt, T. T., & Weber, B. (2021). Exploring how users engage with hybrid process artifacts based on declarative process models: A behavioral analysis based on eye-tracking and think-aloud. *Software and Systems Modeling*, *20*, 1437–1464. https://doi.org/10.1007/s10270-020-00811-8

Baceviciute, S., Lucas, G., Terkildsen, T., & Makransky, G. (2021). Investigating the redundancy principle in immersive virtual reality environments: An eye-tracking and EEG study. *Journal of Computer Assisted Learning*, *38*(1), 120–136. https://doi.org/10.1111/jcal.12595

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted: Does the verbalisation method affect learning? *Metacognition and Learning*, *3*, 39–58. https://doi.org/10.1007/s11409-007-9009-6

Baß, J., Winter, M., Pryss, R., & Reichert, M. (2024). Exploring comprehension strategies of modular process models: A combined eye-tracking and concurrent think-aloud study. *Brain Sciences*, *14*(4), 303. https://doi.org/10.3390/brainsci14040303

Bühler, B., Bockig, E., Deininger, H., Gerätes, P., Trautwein, U., & Kasneci, E. (2024). On task and in sync: Examining the relationship between gaze synchrony and self-reported attention during video lecture learning. *Proceedings of the ACM on Human-Computer Interaction*, *8*, 1–18. https://doi.org/10.1145/3655604

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*, 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. *WIREs Data Mining and Knowledge Discovery*, *12*(4), e1458. https://doi.org/10.1002/widm.1458

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, *6*(3), 271–315. https://doi.org/10.1207/s15327809jls0603_1

Coskun, A., & Cagiltay, K. (2022). A systematic review of eye-tracking-based research on animated multimedia learning. *Journal of Computer Assisted Learning*, *38*(2), 581–598. https://doi.org/10.1111/jcal.12629

de Bruin, A. B. H., Roelle, J., Carpenter, S. K., Baars, M., & EFG-MRE. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educational Psychology Review*, *32*(4), 903–915. https://doi.org/10.1007/s10648-020-09576-4

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349. https://doi.org/10.1111/jcal.12288

Drachsler, H., & Schneider, J. (2018). JCAL special issue on multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 335–337. https://doi.org/10.1111/jcal.12291

Eitel, A., Endres, T., & Renkl, A. (2020). Self-management as a bridge between cognitive load and self-regulated learning: The illustrative case of seductive details. *Educational Psychology Review*, *32*, 1073–1087. https://doi.org/10.1007/s10648-020-09559-5

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

Elling, S., Lentz, L., & De Jong, M. (2012). Combining concurrent think-aloud protocols and eye-tracking observations: An analysis of verbalizations and silences. *Transactions on Professional Communication*, *55*(3), 206–220.

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

Fan, Y., Lim, L., an der Graaf, J., Kilgour, J., Rakovic, M., Moore, J., Molenaar, I., Banjert, M., & Gašević, D. (2022). Improving the measurement of self-regulated learning using multi-channel data. *Metacognition and Learning*, *17*, 1025–1055. https://doi.org/10.1007/s11409-022-09304-z

Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction*, *6*, 41–57. https://doi.org/10.1207/s1532690xci0601_2

Gegenfurtner, A., & Seppänen, M. (2013). Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education*, *63*, 393–403. https://doi.org/10.1016/j.compedu.2012.12.021

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The situational motivation scale (SIMS). *Motivation and Emotion*, *24*, 175–213. https://doi.org/10.1023/A:1005614228250

Hak, T., van der Veer, K., & Jansen, H. (2004). The three-step test-interview (TSTI): An observational instrument for pretesting self-completion questionnaires (No. ERS-2004-029-ORG). ERIM Report Series Research in Management. Retrieved from http://hdl.handle.net/1765/1265

Hansen, J. P. (1991). The use of eye mark recordings to support verbal retrospection in software testing. *Acta Psychologica*, *76*, 31–49. https://doi.org/10.1016/0001-6918(91)90052-2

Harp, S. F., & Maslich, A. A. (2005). The consequences of including seductive details during lecture. *Teaching of Psychology*, *32*(2), 100–103. https://doi.org/10.1207/s15328023top3202_4

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*, 414–434. https://doi.org/10.1037/0022-0663.90.3.414

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Helle, L. (2017). Prospects and pitfalls in combining eye-tracking data and verbal reports. *Frontline Learning Research*, *5*(3), 1–12. https://doi.org/10.14786/flr.v5i3.254

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

Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K., & Lehtinen, M. (2008). Gaze path stimulation in retrospective think-aloud. *Journal of Eye Movement Research*, *2*, 1–18. https://doi.org/10.16910/jemr.2.4.5

Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, *10*(1), 1–18. https://doi.org/10.16910/jemr.10.1.3

Jarodzka, H., Janssen, N., Kirschner, P. A., & Erkennt, G. (2015). Avoiding split attention in computer-based testing: Is neglecting additional information facilitative? *British Journal of Education*, *46*(4), 803–817. https://doi.org/10.1111/bjet.12174

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*, 146–154. https://doi.org/10.1016/j.learninstruc.2009.02.019

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *18*(2), 178–191. https://doi.org/10.1037/a0026923

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kienitz, A., Krebs, M. C., & Eitel, A. (2023). Seductive details hamper learning even when they do not disrupt. *Instructional Science*, *51*, 595–616. https://doi.org/10.1007/s11251-023-09632-w

Knörzer, L., Brünken, R., & Park, B. (2016). Facilitators or suppressors: Effects of experimentally induced emotions on multimedia learning. *Learning and Instruction*, *44*, 97–107. https://doi.org/10.1016/j.learninstruc.2016.04.002

Kok, E., Hormann, O., Rou, J., Van Saase, E., van der Schaaf, M., Kester, L., & Van Gog, T. (2022). Re-viewing performance: Showing eye-tracking data as feedback to improve performance monitoring in a complex visual task. *Journal of Computer Assisted Learning*, *38*(4), 1087–1101. https://doi.org/10.1111/jcal.12666

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

Korbach, A., Brünken, R., & Park, B. (2017). Measurement of cognitive load in multimedia learning: A comparison of different objective measures. *Instructional Science*, *45*, 515–536. https://doi.org/10.1007/s11251-017-9413-5

Liberman, L., & Dubovi, I. (2022). The effect of the modality principle to support learning with virtual reality: An eye-tracking and electrodermal activity study. *Journal of Computer Assisted Learning*, *39*(2), 547–557. https://doi.org/10.1111/jcal.12763

Malone, S., Altmeyer, K., Vogel, M., & Brünken, R. (2020). Homogeneous and heterogeneous multiple representations in equation-solving problems: An eye-tracking study. *Journal of Computer Assisted Learning*, *36*, 781–798. https://doi.org/10.1111/jcal.12426

Mason, L., Scheiter, K., & Tornatora, M. C. (2017). Using eye movements to model the sequence of text–picture processing for multimedia comprehension. *Journal of Computer Assisted Learning*, *33*(5), 443–460. https://doi.org/10.1111/jcal.12191

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. https://doi.org/10.1007/s11145-015-9552-5

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

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, *20*(2), 167–171. https://doi.org/10.1016/j.learninstruc.2009.02.012

Mayer, R. E. (2021). Cognitive theory of multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (pp. 57–72). Cambridge University Press.

Mayring, P. (2015). Qualitativec Inhaltsanalyse: Grundlagen und Techniken. In *Qualitative content analysis* (12th ed.). Beltz.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, *22*, 149–158. https://doi.org/10.1111/j.1365-2729.2006.00170.x

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

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

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

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior*, *44*, 267–278. https://doi.org/10.1016/j.chb.2014.10.061

Park, B., Korbach, A., & Brünken, R. (2015). Do learner characteristics moderate the seductive-details-effect? A cognitive-load-study using eye-tracking. *Journal of Educational Technology & Society*, *18*, 24–36.

Park, B., Korbach, A., & Brünken, R. (2020). Does thinking-aloud affect learning, visual information processing and cognitive load when learning with seductive details as expected from self-regulation perspective? *Computers in Human Behavior*, *111*, 106411. https://doi.org/10.1016/j.chb.2020.106411

Potter, M. C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in Psychology*, *3*, 32. https://doi.org/10.3389/fpsyg.2012.00032

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

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*(1), 1–29. https://doi.org/10.1016/S0364-0213(99)80017-2

Rey, G. D. (2012). A review of research and meta-analysis of the seductive details effect. *Educational Research Review*, *7*, 216–237. https://doi.org/10.1016/j.edurev.2012.05.003

Rey, G. D. (2014). Seductive details and attention distraction: An eye tracker experiment. *Computers in Human Behavior*, *32*, 133–144. https://doi.org/10.1016/j.chb.2013.11.017

Salméron, L., Delgado, P., & Mason, L. (2020). Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic. *Journal of Computer Assisted Learning*, *36*(6), 1038–1051. https://doi.org/10.1111/jcal.12458

Scheiter, K., Schubert, C., Schüler, A., Schmidt, H., Zimmermann, G., Wassermann, B., Krebs, M. C., & Eder, T. (2019). Adaptive multimedia: Using gaze-contingent instructional guidance to provide personalized processing support. *Computers & Education*, *139*, 31–47. https://doi.org/10.1016/j.compedu.2019.05.005

Schellig, D., & Hättig, H. A. (1993). Die Bestimmung der visuellen Merkspanne mit dem block-board (assessment of visual memory span with the block board). *Zeitschrift für Neuropsychologie*, *4*, 104–112.

Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology*, *31*(1), 64–96. https://doi.org/10.1016/j.cedpsych.2005.02.002

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

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*, 227–237. https://doi.org/10.1016/S0959-4752(02)00022-1

Stark, L., Brünken, R., & Park, B. (2018). Emotional text design in multimedia learning: A mixed-methods study using eye tracking. *Computers & Education*, *120*, 185–196. https://doi.org/10.1016/j.compedu.2018.02.003

Stark, L., Brünken, R., & Park, B. (2020). Cued retrospective recall method reveals processes of multimedia learning. In E. Klopp, J. F. Schneider, & R. Stark (Eds.), *Thinking aloud: The mind in action* (pp. 105–127). Bertuch.

Stark, P., Bozkir, E., Sójka, W., Huff, M., Kasneci, E., & Göllner, R. (2024). The impact of presentation modes on mental torsion processing: A comparative analysis of eye movements and performance. *Scientific Reports*, *14*, 12329. https://doi.org/10.1038/s41598-024-60370-6

Strobel, B., Grund, S., & Lindner, M. A. (2019). Do seductive details do their damage in the context of graph comprehension? Insights from eye movements. *Applied Cognitive Psychology*, *33*(1), 95–108. https://doi.org/10.1002/acp.3491

Sundarajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review*, *32*, 707–734. https://doi.org/10.1007/s10648-020-09522-4

Sweller, J., VanMerriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292.

Tjon, J. K., Jarodzka, H., Linskens, I. H., van der Knoop, B. J., & De Vries, J. I. P. (2023). Eye-tracking visual patterns of sonographers with and without fetal motor assessment expertise. *Early Human Development*, *177*, 105722. https://doi.org/10.1016/j.earlhumdev.2023.105722

Usher, E. L., & Schunk, D. H. (2018). Social cognitive theoretical perspective of self-regulation. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 19–35). Routledge/-Taylor & Francis Group. https://doi.org/10.4324/9781315697048-2

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, *25*(3), 785–791. https://doi.org/10.1016/j.chb.2009.02.007

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., Raijmakers, M., & Van Gog, T. (2022). Looking through sherlock's eyes: Effects of eye movement modelling examples with and without verbal explanations on deductive reasoning. *Journal of Computer Assisted Learning*, *38*(5), 1497–1506. https://doi.org/10.1111/jcal.12712

Van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, *19*, 205–221. https://doi.org/10.1002/acp.1112

Wirth, J., Stenger, F., Trypke, M., Schuster, C., & Leutner, D. (2020). An interactive layers model of self-regulated learning and cognitive load.

*Educational Psychology Review*, *32*, 1127–1149. https://doi.org/10.1007/s10648-020-09568-4

Xie, H., Zhao, T., Deng, S., Peng, J., Wang, F., & Zhou, Z. (2021). Using eye movement modelling examples to guide visual attention and foster cognitive performance: A meta-analysis. *Journal of Computer Assisted Learning*, *37*(4), 1194–1206. https://doi.org/10.1111/jcal.12568

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. https://doi.org/10.3102/0002831207312909

**How to cite this article:** Stark, L., Korbach, A., Brünken, R., & Park, B. (2024). Measuring (meta)cognitive processes in multimedia learning: Matching eye tracking metrics and think-aloud protocols in case of seductive details. *Journal of Computer Assisted Learning*, *40*(6), 2985–3004. https://doi.org/10.1111/jcal.13051