Received: 20 April 2021 | Revised: 11 February 2022 | Accepted: 19 March 2022

DOI: 10.1111/jcal.12671

ORIGINAL ARTICLE

Journal of Computer Assisted Learning WILEY

Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning

Anna-Sophia Dersch[1,2] | Alexander Renkl[2] | Alexander Eitel[1]

[1]Department of Educational Psychology, University of Giessen, Giessen, Germany

[2]Department of Psychology, University of Freiburg, Freiburg, Germany

Correspondence
Anna-Sophia Dersch, Justus-Liebig-Universität Gießen, FB06 Lehren und Lernen mit Medien, Otto-Behaghel-Straße 10 D 35394 Gießen, Germany.
Email: anna-sophia.dersch@psychol.uni-giessen.de

[Correction made on 17 August 2022, after first online publication: The Article Category has been corrected from Review Article to Original Article in this version.]

Abstract

Background: Previous research has shown that teachers hold misconceptions about multimedia learning (e.g., multimedia instruction needs to be adapted to students' learning styles), which may be at odds with evidence-based teaching.

Objectives: Refutation texts are a classical method to reduce misconceptions and thus to stimulate conceptual change. We wanted to know whether making use of a computer algorithm to personalize refutation texts would best initiate teachers' conceptual change.

Methods: We designed an online experiment, in which $N = 129$ in-service teachers read either (1) expository texts (without direct refutation), (2) common refutation texts, or (3) personalized refutation texts. The teachers filled in a misconception questionnaire pre and post to assess their conceptual change.

Results and Conclusions: Statistical analyses revealed that personalized refutation texts initiated the strongest conceptual change, which was driven by increased feelings of guilt and shame. Common refutation texts did not foster teachers' conceptual change as compared to expository texts. These findings indicate that refutation texts should be personalized for experienced practitioners such as teachers.

Takeaways: Personalized refutation seems to be promising in the context of online teacher training programs. Further research should test to which extent the present findings also apply to other groups of experienced learners or practitioners.

KEYWORDS
refutation texts, multimedia learning, adaptive instruction, teacher training, teacher education, technology-enhanced instruction

1 | INTRODUCTION

"Students learn better when instructional materials are adapted to their learning style (e.g., visual or verbal)." Such beliefs about multimedia learning are very common in teachers (e.g., Dekker et al., 2012; Krammer et al., 2019). However, research does not support such beliefs (Rohrer & Pashler, 2012; Kirschner, 2017). The conception that effective teaching should be adapted to stable differences in students' learning styles (e.g., "visualizer or verbalizer") qualifies as a misconception. Misconceptions are subjective assumptions that are incompatible with scientific conceptions (Chi & Roscoe, 2002; Hughes et al., 2013; Vosniadou, 1994). Holding such misconceptions—in this case about multimedia learning—as a teacher can negatively affect teaching practice (e.g., McElvany et al., 2012). Therefore, such misconceptions should be reduced. In former research, misconceptions have been successfully reduced by applying refutation texts (Tippett, 2010). Refutation texts describe a commonly

held misconception, state that this assumption is incorrect, then introduce a scientific explanation as an alternative, "correct" conception (Sinatra & Broughton, 2011). In this study, we analyzed how to address and reduce such misconceptions by different types of refutation texts in a teacher sample, making use of computer algorithm allowing for the efficient personalization of refutation texts.

1.1 | Refutation text and conceptual change

Misconceptions are usually embedded in naïve framework theories about complex scientific principles – often acquired early in knowledge construction (Chi & Roscoe, 2002; Vosniadou, 1994). Misconceptions sometimes also originate from "folk knowledge," which is robust to revision within society (Hughes et al., 2013). In any case, misconceptions are incompatible with scientific findings (Chi & Roscoe, 2002; Hughes et al., 2013; Vosniadou, 1994). The idea that learning materials should be adapted to learning styles is such a misconception that is scientifically disproven (Kirschner, 2017). Further, the idea of adapting learning materials to learning styles may be acquired early in life, as students in school may be confronted with it through their teachers applying materials adapted to learning styles. Furthermore, the idea of learning styles seems to be "folk knowledge" that is common amongst teachers and persistent despite being disproven (Eitel et al., 2021; Krammer et al., 2019). But even though misconceptions are scientifically disproven, they may nevertheless be useful in everyday knowledge (Smith III et al., 1994). An example would be "stone is colder than wood." This is scientifically disproven as both materials have the same temperature in the same room. Yet, as wooden floor often feels warmer on the feet than stone floor, this scientific misconception can promote useful behaviour (e.g., wearing socks on stone floor). Misconceptions can be a starting point for conceptual change by refining, revising, bridging and building upon them for facilitating knowledge construction (Larkin, 2012). However, misconceptions can also hinder the acquisition of scientifically accurate conceptions because they are subjectively highly plausible and well integrated into knowledge structures (see "wood is warmer than stone"). Simply presenting scientifically accurate information is usually insufficient to foster conceptual change and to revise misconceptions (Chi & Roscoe, 2002). Specific forms of instruction such as refutation texts are needed to initiate conceptual change (e.g., Guzzetti et al., 1993; Posner et al., 1982; Tippett, 2010).

Refutation texts name a commonly held misconception, state that this assumption is incorrect, then introduce a scientific explanation as an alternative, "correct" conception (Sinatra & Broughton, 2011). Therefore, refutation texts stimulate co-activation of learners' prior beliefs (i.e., the misconception) and the related correct information in learners' associative memory. This direct comparison can foster conceptual change. This process includes the experience of a cognitive conflict, the evaluation of one's (prior) knowledge, the dissatisfaction with one's (prior) knowledge and the establishment of coherence in one's (prior) knowledge (e.g., Ariasi & Mason, 2011; Kendeou & O'Brien, 2014; McCrudden & Kendeou, 2012). In order to initiate conceptual change, however, the correct conception provided in the refutation text must be plausible enough to question persistent, yet incorrect ideas (Posner et al., 1982). In addition, one needs to be sufficiently motivated to change ones' prior concepts (Dole & Sinatra, 1998; Gregoire, 2003).

Empirical research revealed positive effects of refutation texts on conceptual change compared to presenting common expository texts. Research has mostly been conducted with novice learners in natural science domains (e.g., Tippett, 2010). A growing number of studies have found effects on conceptual change in the topics of psychology and empirical educational research (e.g., Lassonde et al., 2016; Menz et al., 2021; Prinz et al., 2021; Vosniadou et al., 2020).

1.2 | Refutation text and teachers' conceptual change

Refutation texts typically foster conceptual change amongst novices (e.g., Tippett, 2010). However, when the goal is fostering conceptual change for experienced practitioners, as in knowledge about multimedia learning for teachers (Eitel et al., 2019), two issues may arise that limit the effectiveness of refutation texts. First, teachers with experience in (multimedia) learning might not feel addressed when reading the introductory sentence of common refutation texts: "many people believe that [...]. However, this does not align with scientific evidence." Subjectively being experienced in this topic already, why should teachers count themselves to the people with erroneous assumptions and thus be further taught on this topic? In terms of the Cognitive-Affective Model of Conceptual Change (CAMCC; Gregoire, 2003), one would argue that teachers do not feel sufficiently self-implicated and thus not motivated to systematically process the text in order to achieve conceptual change. Refutation texts may be insufficient to induce the impasse experience needed for conceptual change (Sánchez et al., 2009). Secondly, even if teachers feel addressed, they may react with anger or repulse when they are lessoned on *their* topic, because they might consider it offensive. Teachers seem to be highly committed to their professional knowledge (Jacob et al., 2017). More specifically, teachers' pedagogical knowledge forms part of their self-concept (Beijaard et al., 2000; Friedman & Farber, 1992). If teachers' pedagogical knowledge is questioned, teachers' self-concept might be endangered (Molloy et al., 2012) and cognitive dissonance might occur (Festinger, 1962), leading to the typical reaction of anger and rejection of new information, which might hinder conceptual change (Chin & Brewer, 1993; Kluger & DeNisi, 1996; Molloy et al., 2012; Trevors et al., 2016).

Previous empirical studies support the aforementioned reasoning. Research showed that as soon as students became more knowledgeable on a topic, they used refutation texts in a more inefficient way (Poehnl & Bogner, 2013). In addition, there is very little research experimentally testing the effectiveness of refutation texts with in-service teachers. We know of just one recent study by Ferrero et al. (2020). This study showed that refutation texts indeed reduced misconceptions among in-service teachers. However, Ferrero et al. (2020) did not compare the applied refutation texts to an expository text control group. Rather, the control group received no text at all. Therefore, it remains

to be investigated whether and under which circumstances refutation text benefits teachers' conceptual change, compared to the typical presentation of new information by means of expository text.

1.3 | Personalized refutation text and teachers' conceptual change

The typical design of refutation texts might trigger teachers' conceptual change to a reduced degree (for teaching-related topics) or not at all, because teachers might either not feel addressed by the refutation texts and/or react with anger and rejection because of potential self-concept threat. A potential means to both avoid anger and increase motivation for conceptual change is to design *personalized* refutation texts. Personalization means that teachers receive a refutation text for a supposed misconception only if their answers in a previous test indicate that they endorse this misconception. Refutation texts thus appear as feedback for previous answers, which has been found beneficial across various learning settings (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Moreover, personalized refutation texts make teachers aware of their own answers in a previous test using conversational, personalized speech (e.g., "your answers indicate that…"). Such personalized messages can lead to higher personal involvement, and therefore to more systematic and focused information processing (Mayer, 2014; Renkl, 2015). The latter is especially relevant in the context of conceptual change as, according to influential models such as the CAMCC (Gregoire, 2003) or dual-process models (Dole & Sinatra, 1998; Maheswaran & Chaiken, 1991), conceptual change requires deliberate, systematic processing. Therefore, personalized messages can boost conceptual change via increased motivation for systematic processing (Alvermann & Hynd, 2015; Asterhan & Dotan, 2018; Sinatra, 2005). More specifically, such personalized messages seem to be especially effective when they draw the reader's attention towards the discrepancy between their own beliefs and the presented text (Asterhan & Dotan, 2018; Gill et al., 2020), potentially creating an impasse experience (Sánchez et al., 2009). In previous research, either a human instructor (Asterhan & Dotan, 2018; Gill et al., 2020) or a computer algorithm that scores data (e.g., Okoye, 2015; Sánchez et al., 2009) provided such feedback. The latter has been developed and tested with novice learners only. Here we developed and tested a computer algorithm with in-service teachers – a professional group with limited amounts of time to spare. Thus, a short and digitally available intervention promoting conceptual change during (online) teacher training programs would be a key asset. Here we test its effectiveness empirically. Moreover, we examine the processes underlying effective conceptual change considering the role of affect.

1.4 | Affect and teachers' conceptual change

The CAMCC (Gregoire, 2003) argues that teachers' conceptual change is driven by affective processes and that a certain degree of discomfort is required for teachers' conceptual change. If teachers are comfortable with what they know about teaching and learning, they are unlikely to engage in effortful conceptual change. However, if teachers are directly challenged in their ideas, for example, because corrective feedback to their answers (provided by personalized refutation texts) is discrepant to their pre-existing beliefs, they may sense enough discomfort to initiate conceptual change. Such discomfort may have two opposing effects. On the one hand, teachers may *blame themselves* for having held "incorrect" beliefs on the topic corrected by the feedback. Such internal blame is typically associated with feelings of shame and guilt (Nabi, 1999). Up to a moderate degree, guilt is known to drive attitude changes (Nabi, 1999), for example, because subsequent information is processed more systematically. Therefore, feelings related to internal blame can be expected to foster conceptual change. On the other hand, teachers may *blame others*, such as the learning system, for providing erroneous feedback to their subjectively correct beliefs. External blame is associated with anger and rejection of new information, which may hinder conceptual change. Therefore, personalization could backfire as found in previous research (e.g., Trevors et al., 2016). Sherman and Cohen (2002) argue that individuals may resist evidence when it threatens self-worth, motivating defense behaviours.

Personalized refutation texts, however, enable balanced feedback, meaning that both correct and incorrect answers to previous questions are appreciated as such. Thus, unlike in a typical refutation instruction, refutation texts appear only when previous answers indicate the presence of a misconception. This balanced feedback may reduce feelings of anger and blaming of the learning system, as learners more likely accept feedback providing the opportunity for improvement and learning (Molloy et al., 2012). Therefore, personalized refutation texts might lead to reduced anger compared to common refutation texts facilitating conceptual change.

2 | PRESENT STUDY AND HYPOTHESES

In-service teachers answered a misconception questionnaire on the topic of multimedia learning once before studying texts about multimedia learning (pretest) and once after studying texts about multimedia learning (posttest), to assess the degree of conceptual change in terms of improvement between pretest and posttest. Moreover, teachers took a transfer test assessing their selection of learning materials according to the underlying theories about multimedia learning. Depending on the experimental condition, teachers studied either expository texts, refutation texts, or personalized refutation texts (with the help of the developed computer algorithm, see 3.3). We postulate the three following hypotheses regarding whether and how conceptual change differs as a function of the experimental condition.

2.1 | Personalized refutation hypothesis

Only refutation texts that are personalized both directly address misconceptions and increase personal involvement leading to more systematic information processing. Therefore, personalized refutation texts should stimulate teachers' conceptual change more than both common

refutation texts and expository texts. Furthermore, we investigate whether common refutation texts would foster conceptual change compared to expository texts. Due to a lack of previous research using experienced practitioners, we conduct an exploratory analysis.

2.2 | Shame-drives-change hypothesis

We expect that stronger conceptual change due to personalized refutation texts is driven by a moderate increase in the specific negative affect of shame and guilt, resulting from blaming oneself for incorrect prior beliefs. This affects should, in turn, lead to more systematic processing as expressed by higher conceptual change scores compared to common refutation texts and expository texts. By contrast, very strong increases in shame and guilt should hinder conceptual change.

2.3 | Anger-hinders-change hypothesis

We expect that stronger conceptual change due to personalized compared to common refutation texts is accompanied by reduced anger, resulting from the more balanced feedback. We expect this effect especially in the case when teachers neither endorse all nor no misconception as identified by the pretest so that their feedback by the system is variable. Hence, especially for these teachers, personalized refutation texts should reduce anger, and therefore lead to stronger conceptual change.

3 | METHOD

3.1 | Participants and design

We conducted the study online. Participation was voluntary. Initially, 238 participants provided informed consent; 131 participants finished this online experiment. Participants who did not finish the study were excluded from the sample (45% dropout overall). Moreover, two participants, not being teachers, were excluded from the sample. Thus, we used a final sample of $N = 129$ in-service teachers (101 women, 28 men, 0 divers, $M_{age} = 43.74$). This sample size was large enough to meet the power requirements ($\alpha = 0.05$, $1 - \beta \geq 0.80$) for an expected medium-sized effect ($f = 0.25$) according to G*Power (Faul et al., 2007). Participants from all of the 16 German federal states, from German-speaking regions (Austria, Switzerland, South Tyrol) as well as from German schools in non-German-speaking countries (Portugal, Spain) took part in the study. Most participants spoke German as their native language ($n = 124$), five participants had another native language; the results of the latter participants were comparable with those of the other participants.

We recruited participants through acquaintances, research project partners working with

in-service teachers, and through the German online teachers' magazine News4Teachers (www.news4teachers.de). We randomly assigned participants to one of three experimental conditions in a pretest-posttest design: Expository texts ($n = 48$), refutation texts ($n = 41$), or personalized refutation texts ($n = 40$). Chi-square tests showed that dropouts were similarly distributed among the three conditions, $\chi^2(2) = 2.516$, $p = 0.284$.

3.2 | Materials and manipulations

3.2.1 | Expository texts

Teachers in this condition sequentially read four short expository texts (text 1: 167 words, text 2: 159 words, text 3: 200 words and text 4: 146 words). Expository texts were based on research findings on cognitive processes in learning with multimedia. More specifically, the expository texts were based on influential cognitive theories about learning with multiple representations such as the design, functions and task framework (DeFT; Ainsworth, 2006), the cognitive theory of multimedia learning (CTML; Mayer, 2014), or the categorization of picture functions in prose (Carney & Levin, 2002; Levin et al., 1987), and fundamental findings on neurocognitive functioning (Hellige, 1990). The expository texts applied these theories to explain (1) why there is no use in tailoring teaching materials to students' supposed learning styles, (2) why pictures do not merely serve a motivating function (see Table A1, for the expository text), (3) why it is sometimes better to use fewer media and modalities to present information, and that (4) both of students' brain hemispheres are activated even when they learn with text alone. As such, the expository texts map onto four common misconceptions about multimedia of (1) learning styles, (2) motivation primacy, (3) naïve summation and (4) hemispheric isolation. We administered these four misconceptions in the pretest and the posttest of this study using the Misconceptions About Multimedia Learning Questionnaire (see Eitel et al., 2021; see Section 3.4 and 3.5).

3.2.2 | Refutation texts

Teachers in this condition read the same four expository texts in the same order as described before. The only difference being that there was a text segment added before each expository text stating "Many people assume that […]. However, this assumption does not align with scientific evidence." This sentence is typical of the design of refutation texts (e.g., Tippett, 2010). A complete example of a refutation text can be found in Table A1. Due to the addition of the text segment refuting the misconception, refutation texts were marginally longer than expository texts (text 1: 216 words, text 2: 193 words, text 3: 239 words and text 4: 202 words).

3.2.3 | Personalized refutation texts

Teachers in this condition read the same four expository texts and in the same order as described in the Section 3.2.1. Moreover, adaptive personalized refutation texts included one sentence stressing personal relevance saying "Our evaluation shows that you […]" The exact formulation of this sentence depended on each participant's

answers in the misconception questionnaire (for details, see Section 3.3).

If the teachers responded correctly to all four questions pertaining to a misconception in the pretest, they read the sentence: "Our evaluation shows that you answered the questions […] in accordance with current scientific findings – great! In the following you can familiarize yourself with the scientific findings once again and supplement your knowledge on the topic." Afterwards, they read the expository text (text 1: 253 words, text 2: 237 words, text 3: 284 words and text 4: 243 words).

If the teachers responded with uncertainty in the pretest, they read the sentence: "Our evaluation shows that you answered the questions […] with insecurity. Therefore, we would like to give you the opportunity to familiarize yourself with the current state of research in this field." After, they read the expository text (text 1: 200 words, text 2: 197 words, text 3: 238 words and text 4: 182 words).

If teachers' pretest results indicated that they held the misconception, they read the sentence "Our evaluation shows that you did not answer the questions […] in accordance with current scientific findings. Therefore, we would like to give you the opportunity to familiarize yourself with the current state of research on this field." Further teachers with misconceptions read an additional sentence before the expository text stating "Many people assume that […]" (see section: *Refutation Texts*). The length of the texts was similar to the refutation texts, but a bit further increased (text 1: 205 words, text 2: 200 words, text 3: 242 words and text 4: 188 words).

3.3 | Functioning of the computer algorithm

We developed and ran the computer algorithm within the software unipark (Questback). The program reacted adaptively only in the condition with personalized refutation texts. In this condition, teachers' answers in the pretest affected how the subsequently presented texts addressed them (see Figure 1, for a graphical overview).

There were three types of addressing texts, which could vary both within and between teachers: Teachers received the feedback of *holding a correct conception* if they had answered all four items per misconception-scale correctly, regardless of the certainty of their answers.

Teachers received the feedback of *having answered with uncertainty* if they had indicated a medium to high uncertainty (*a little certain*, *uncertain*, or *absolutely uncertain* on Likert scale) for all of their four answers per misconception-scale, regardless of correctness of their answers.

Teachers received the feedback that *their answers were incorrect*, if they answered incorrectly with high certainty (*absolutely certain* or *very certain* on Likert scale) to at least one of four items per misconception-scale. So even though teachers might have answered correctly to the other three items of this scale, their answers overall suggested a misconception, which still requires refutation on its path to the correct conception (Vosniadou, 1994).

Because of this sensitive algorithm to detect misconceptions, teachers received the feedback about the presence of a misconception for approximately three of the four texts on average. Moreover, exactly half of the participants in the personalized refutation condition (n = 20) received feedback about the presence of a misconception for all of the four misconception-scales. So, feedback was variable/balanced for only the other half of participants in the condition with personalized refutation texts. No participant from this condition answered always correctly or always with high uncertainty. See Table A1 for an example of how the intervention looked like in all three conditions.

3.4 | Measures

3.4.1 | Control variables

Teachers filled in their age, gender (male, female, diverse), mother tongue (German or else), school subjects, school type and grade level. Moreover, we asked for teachers' intrinsic cognitive load (ICL; 2 items) and extraneous cognitive load (ECL; 3 items) while studying the instructional texts. We used the two respective subscales from the cognitive load questionnaire by Klepsch et al. (2017). Items had to be answered via seven-point Likert-scales from *completely wrong* to *absolutely right*. An item for intrinsic cognitive load was "For this task, many things needed to be kept in mind simultaneously" (correlation of the two ICL items: $r = 0.68$), for extraneous cognitive load "during this task, it was exhausting to find the important information" (McDonald's omegas [ω] = 0.88)*. In addition, the time to read expository and refutation texts was recorded by the system and analyzed.

3.4.2 | Affect

We assessed affect with the positive and negative affect schedule (PANAS; Watson et al., 1988). The PANAS assesses the current affect via responses to 20 adjectives—10 adjectives for positive affect (e.g. happy) and 10 adjectives for negative affect (e.g. angry). Participants rate on a five-point Likert-scale from *little or not at all* to *extremely* how strongly they currently endorse each affect. In the analysis, we focused on specific negative affect 1) resulting from internal blame being related with guilt, shame, or surprise about the feedback to one's answers (guilty, ashamed, startled; $\omega = 0.82$) and 2) resulting from external blame being related with anger (upset, irritable, hostile; $\omega = 0.73$). The remaining negative affect adjectives were rather unrelated to blame (afraid, nervous, confused, distressed; $\omega = 0.76$), and therefore not considered for analyses.

3.4.3 | Pretest

We used the Misconceptions about Multimedia Learning Questionnaire in order to assess participants' misconceptions (MMLQ; see Eitel

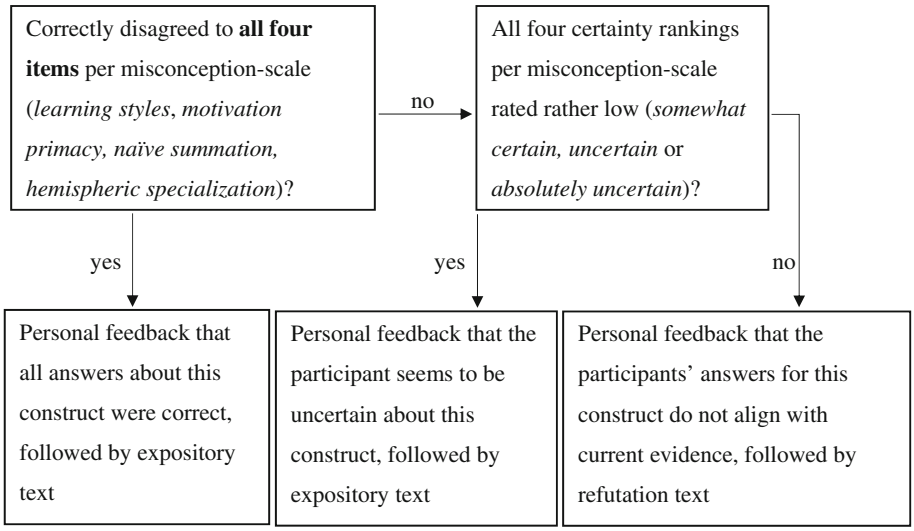

FIGURE 1 Flow-chart for the algorithms' decision process of providing feedback in the personalized refutation condition

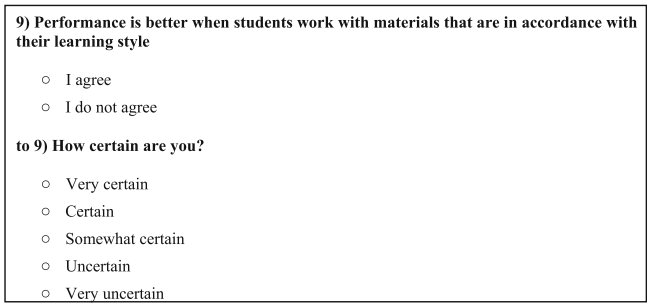

FIGURE 2 Example of an item presentation in the Misconceptions about Multimedia Learning Questionnaire (used both as pretest and posttest). There was always a true-false statement above, with the certainty rating beneath it

et al., 2021, for the development process). Eitel et al. (2021) found the MMLQ answers to be consistent with a four-factor structure of misconceptions about multimedia learning. These four MMLQ-scales showed modest to good reliabilities. The MMLQ consists of 32 items and is designed to assess misconceptions, not just missing conceptions, by asking both for agreement *and* for response certainty on a five-point Likert scale (see Figure 2). One of these items, for example, is "whether students learn better with visual or verbal materials depends on their learning style." Participants both indicate whether they agree or not, and provide a certainty rating on a five-point Likert scale ("absolutely certain"–"certain"–"somewhat certain"–"uncertain"–"absolutely uncertain"). Ratings of agreement and certainty were combined per item. Specifically, a misconception was only counted as such, if there was agreement to a wrong statement made with *sufficiently high certainty*. Wrong answers made with uncertainty indicate missing conceptions, while wrong answers with medium to high certainty indicate misconceptions (Eitel et al., 2021). Accordingly, MMLQ items are scored by multiplying the values for correct and incorrect responses (i.e., +1 and −1) by the respective certainty rating, coded from 0 = *absolutely uncertain* to 4 = *absolutely certain*. Hence, when a teacher indicated a very low certainty to an item response, they received a score of 0 for this item, regardless of the response correctness (i.e., $\pm 1 \times 0 = 0$). The more certain a teacher was about their false response (e.g., 4 vs. 2), the stronger the belief in a misconception, and, accordingly, the higher was the negative value in the misconception score (i.e., −4 vs. −2). By contrast, a higher certainty in correct responses likewise yields higher positive values (+4 vs. +2). Thus, average scores for the pretest ranged between the theoretical minimum of −4.0 points and the theoretical maximum of +4.0 points. The lower the score the stronger misconceptions are present.

All 32 items of the MMLQ are statements about multimedia learning: Half of the items (16) describe common misconceptions about multimedia learning (e.g., learning styles, naïve summation; see below). These 16 misconception-related items were in focus here. Eitel et al. (2021) showed that these 16 items can be categorized into four misconception-subscales, with four items per misconception-scale (see Table A2, for the full list of misconception items). The four misconception-subscales (translated from German) are (1) *Learning styles*: Multimedia instruction needs to be adapted to students' learning styles to be effective (e.g., "Performance is better when students work with materials that are in accordance with their learning style"), (2) *Motivation primacy*: Pictures' main function in multimedia learning is motivating learners (e.g. "The main reason for why pictures are beneficial is that they motivate"), (3) *Naïve summation*: Multimedia instruction is more helpful the more sensory channels are used (e.g., "To what degree information is remembered particularly depends on the number of sensory channels used to perceive it [e.g., seeing, hearing, touching]") and (4) *Hemispheric isolation*: Text *and* graphics activate both brain hemispheres and enable hemispheric communication (e.g., "The two brain hemispheres do not work together automatically so this needs to be stimulated by appropriate study materials"). See Table A2, for all misconception-related items of the MMLQ.

Eitel et al. (2021) found that in confirmatory factor analyses these four subscales can be subsumed under the umbrella of one misconception about multimedia learning factor, which is also treated as the main dependent variable here. Also, aggregating items from the four misconception scales to one single misconception factor increased reliability. The internal consistency of the 16-item

misconception-scale, administered as pretest, was $\omega = 0.75$, which is acceptable (Field, 2009).

The other half of the items (16) of the MMLQ align with current research on multimedia learning (e.g., pictures are beneficial to learning especially when they visualize the contents to-be-learned). These were used as filler items to balance correct and incorrect statements in the questionnaire, which is why they were not part of the analysis.

3.4.4 | Posttest

The posttest was identical to the pretest. Again, we administered and scored the MMLQ as in the case of our pretest. This repeated measurement allowed to compare pretest to posttest scores and calculate improvement scores reflecting the degree of conceptual change (posttest–pretest score). The larger the positive value of this difference score, the stronger the conceptual change. Again, aggregating items from the four misconception scales to one single misconception factor increased reliability. The internal consistency of the 16-item posttest misconception-scale was good ($\omega = 0.84$, Field, 2009).

3.4.5 | Transfer test

This test required teachers to select one of two alternatives of multimedia instructional materials (consisting of multiple representations) and to provide an explanation for why they chose the one material over the other; one of the two alternatives was always the better choice according to cognitive theories of multimedia learning (Ainsworth, 2006; Mayer, 2014). Overall, there were four such tasks. For instance, participants should choose among two textbook pages to teach predator–prey relations in biology class, one comprising text predator–prey relations and a decorative image of a fox eating a mouse and the other one comprising text about predator–prey relations and a line graph about population sizes of predators and prey over time. Participants scored one point only if they both chose the better answer according to cognitive theories of multimedia learning (here: the line graph) and provided an explanation for it based on what they had read in the instructional texts (e.g., the line graph made abstract relations easier to grasp, or the line graph constrained the range of [false] interpretations based on the text etc.). Participants scored zero points if they either chose the worse option, and/or did not provide a reasonable explanation (e.g., one that contradicts information presented in the expository/refutation texts). See Table A3, for correct and incorrect example answers for all four transfer test tasks.

Teachers could score a maximum of 4 points, and a minimum of 0 points. Two independent raters scored participants' explanations. They reached excellent agreement, ICC = 0.92 (two-way random effects, consistency). Internal consistency of this scale was rather low ($\omega = 0.46$), which is partially due to the shortness of the scale as indicated by the higher asymptotic $\omega = 0.59$. Because for heterogeneous constructs such as transfer performance, lower scores of internal consistencies are to be expected (Schmitt, 1996), we nevertheless analyzed results for the transfer test although they should be considered with caution.

3.5 | Procedure

Upon clicking on the recruiting link, participants saw a welcome page stating that this was a study by the University of Freiburg and that participation was voluntary. Participants then needed to provide informed consent to their anonymized data being stored and processed only for the purpose of this study. The experiment started only for the participants who provided informed consent. Participants were then instructed that there was no option of going back to previous pages in the experiment, and that they should answer all questions to proceed to the next page. Then, participants provided demographic data, and filled in the MMLQ (pretest). Afterwards, participants read four texts always in the same order (learning styles, motivation primacy, naïve summation, hemispheric isolation). The contents of these texts depended on the experimental condition. Reading times were self-paced in all conditions. However, participants needed to remain on the text page for at least 30–45 s (depending on the text length) before they could proceed. After, participants filled in the cognitive load questionnaire and the PANAS. Then, participants filled in the MMLQ again (posttest), followed by the transfer test. In the end, participants were thanked and read a debriefing. We additionally provided them with the literature about multimedia learning we had used for the four texts. The experiment took on average 40 min. For the timeline of the experiment see Figure 3.

4 | RESULTS

We first analyzed the control variables, followed by hypothesis testing. Mediation analyses were based on the procedure by Hayes (2018) using 10,000 bootstrapping samples. Partial eta-squared and R-squared were reported as effect sizes, where small values are 0.01, medium values are around 0.06, large values are 0.14 and larger.

4.1 | Control variables

We checked for whether the relevant control variables were equally distributed across the three conditions. Descriptive values can be found in Table 1. Results revealed that neither gender, $\chi^2(2) = 3.79$, $p = 0.16$, German as mother tongue, $\chi^2(2) = 2.05$, $p = 0.36$, age, $F(2, 126) = 1.852$, $p = 0.16$, nor pretest scores, $F(2, 126) = 1.929$, $p = 0.15$, differed significantly across conditions. Neither ICL nor ECL differed between conditions, both $F_s < 1$, indicating that the manipulation did neither increase complexity nor difficulty of the task. Reading times differed significantly between the conditions after excluding extreme outliers, $F(2, 115) = 3.82$, $p = 0.03$, $\eta^2_p = 0.06$. However, differences in reading times were unrelated to the dependent

variables of conceptual change scores, $r = 0.14$, $p = 0.15$, and transfer test scores, $r = -0.004$, $p = 0.97$.

4.2 | Personalized refutation hypothesis

We expected that personalized refutation texts stimulate teachers' conceptual change more than both common refutation texts and expository texts. We tested this hypothesis by translating it into a contrast, in which both the expository text and the refutation text conditions were supposed to display lower conceptual change scores than the condition with personalized refutation texts (contrast #1: [−1 −1 +2]). Within a second contrast, we tested whether the expository text condition differed from the refutation text condition (contrast #2: [−1 +1 0]). These two contrasts served as independent variables in a linear regression analysis. Results revealed that contrast #1 significantly predicted conceptual change scores, $b = 0.21$, SE = 0.09, $t(126) = 2.43$, $p = 0.02$, $R^2 = 0.05$, whereas contrast #2 did not, $t < 1$. Our findings support the personalized refutation hypothesis. Descriptive values can be found in Table 2 and in Figure 4. Conceptual change scores were positive and different from zero in all three conditions (all $p_s < 0.001$). In addition, we applied the same contrast to the results of the transfer test, including pretest scores as covariate. Results of both the focal contrast, $b = 0.13$, SE = 0.02, $t(126) = 1.47$, $p = 0.14$, $R^2 = 0.02$ and the residual contrast, $t < 1$, did not reach statistical significance.

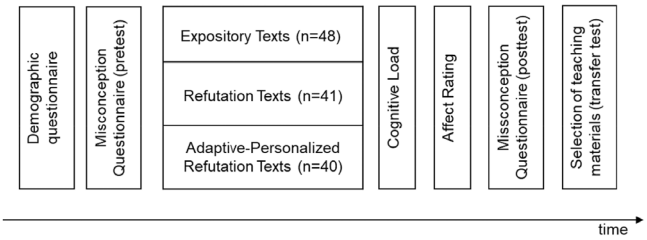

FIGURE 3 Graphic overview of the experimental procedure

4.3 | Shame-drives-change hypothesis

We expected that stronger conceptual change due to personalized refutation texts is driven by a moderate increase in the specific negative affect of shame and guilt, which leads to stronger conceptual change. We tested this hypothesis by means of a mediation analysis with condition as independent variable (contrast coding: [−1 −1 +2]), reported feelings of shame and guilt as mediator, and conceptual change scores as dependent variable. Results revealed a significant mediation, with an indirect effect of ab = 0.06, SE = 0.04, BCa CI 95 [0.01, 0.15], $R^2 = 0.02$. More specifically, personalized refutation texts led to moderately higher feelings of guilt and shame, $a = 0.10$, SE = 0.03, $p < 0.001$ (descriptive increase: $\Delta = 0.33$ on 5-point Likert scale), that in turn led to marginally stronger conceptual change, $b = 0.54$, SE = 0.29, $p = 0.06$. The total effect of $c = 0.22$, SE = 0.09, $p = 0.02$, was not significant anymore with the mediator in the model (direct effect: $c' = 0.16$, SE = 0.09, $p = 0.09$), indicating a full mediation effect (see Figure 5). This pattern of results supports the shame-drives-change hypothesis in a way that moderately higher shame drove stronger conceptual change. For transfer, the mediation missed significance, ab = 0.006, SE = 0.01, BCa CI 95 [−0.0009, 0.02].

4.4 | Anger-hinders-change hypothesis

We expected stronger conceptual change due to personalized compared to common refutation texts to be accompanied by reduced anger, resulting from the more balanced feedback that teachers received—at least when teachers neither endorsed all nor no misconception as identified by the pretest. We tested the anger-hinders-change hypothesis by means of a moderated mediation analysis with experimental condition as independent variable (refutation texts [−1] versus personalized refutation texts [+1]), feelings of anger as mediator, and the conceptual change score as dependent variable. Moreover, we included the number of misconceptions in the pretest (none vs. some) as moderator for the link between the independent variable and feelings of anger in a way that personalized refutation texts should reduce feelings of anger

TABLE 1 Means (with standard deviations) and distributions for the control variables as a function of experimental condition

	Expository texts	Refutation texts	Personalized refutation texts	All teachers
Gender (female/male)	39/9	28/13	34/6	101/28
German as mother tongue (yes/no)	47/1	40/1	37/3	124/5
Age in years	41.67 (8.78)	45.90 (11.10)	44.03 (11.44)	43.74 (10.48)
ICL (min. = 1, max. = 7)	3.67 (1.36)	3.81 (1.33)	3.88 (1.66)	3.78 (1.44)
ECL (min. = 1, max. = 7)	2.60 (1.23)	2.76 (1.46)	2.78 (1.38)	2.71 (1.35)
Reading times (in s)[a]	61.35 (15.17)	57.31 (10.70)	66.52 (15.83)	61.70 (14.52)
Pretest scores	−0.64 (1.21)	−0.55 (1.30)	−0.97 (1.08)	−0.71 (1.21)

Abbreviations: ECL, extraneous cognitive load; ICL, intrinsic cognitive load.

[a]These were the reading times that teachers invested in addition to what was pre-set by the system (30–45 s depending on text type).

TABLE 2 Means (with standard deviations) for the dependent variables as a function of experimental condition

	Expository texts	Refutation texts	Personalized refutation texts	All teachers
Conceptual change scores (min. = −8, max. = 8)	2.23 (1.34)	2.32 (1.40)	2.92 (1.41)	2.47 (1.40)
Posttest scores (min. = −4, max. = 4)	1.59 (1.26)	1.77 (1.52)	1.94 (1.42)	1.76 (1.39)
Transfer test scores (min. = 0, max. = 4)	2.50 (0.80)	2.44 (1.12)	2.65 (1.10)	2.88 (0.92)
Guilt-Shame (min. = 1, max. = 5)	1.08 (0.33)	1.11 (0.29)	1.41 (0.60)	1.19 (0.44)
Anger (min. = 1, max. = 5)	1.08 (0.20)	1.14 (0.34)	1.28 (0.48)	1.16 (0.36)

compared to normal refutation texts especially for teachers who received variable feedback by the system (i.e., those who endorsed no misconceptions in the personalized refutation condition). This reduced anger should in turn foster conceptual change. Results did not reveal a moderated mediation: index = 0.03, SE = 0.09 BCa CI 95 [−0.07, 0.35]. More specifically, regardless of the number of misconceptions, feelings of anger did not mediate the relation between the experimental condition and the dependent variable (mediation for some misconceptions: ab = 0.03, SE = 0.05, BCa CI 95 [−0.03, 0.20]; mediation for all or none misconception: ab = 0.05, SE = 0.09, BCa CI 95 [−0.06, 0.38]). For transfer, the moderated mediation also missed statistical significance: index = −0.0005, SE = 0.01, BCa CI 95 [−0.03, 0.02].

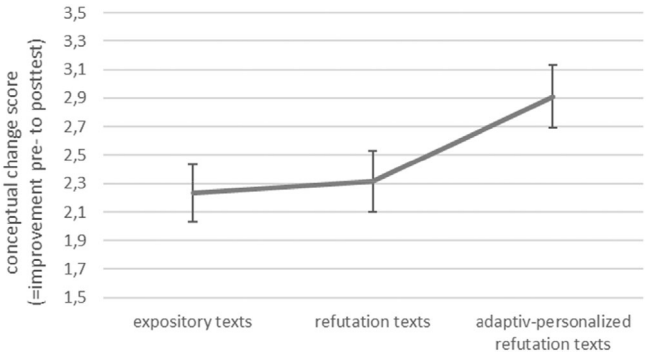

FIGURE 4 Improvement from pretest to posttest (conceptual change) scores as a function of the experimental condition

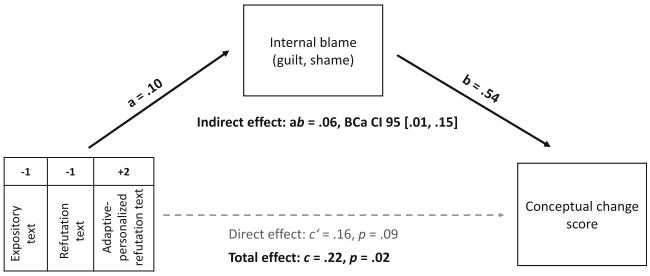

FIGURE 5 Graphic mediation model with the hypothesis contrast as independent variable, ratings of guilt-shame as mediator, and conceptual change scores as dependent variable

5 | DISCUSSION

We investigated whether and how teachers' conceptual change about multimedia learning can be best supported. We compared affective processing and conceptual change scores between teachers studying expository texts, refutation texts, or personalized refutation texts that were provided by an adaptive online instruction. Results indicate online instructions that aim to refute teacher misconceptions about multimedia learning are most effective when personalized. When teachers are directly challenged in their previous ideas, their emotional involvement (higher feelings of guilt-shame) drives conceptual change.

5.1 | Personalized refutation amplifies conceptual change

We expected personalized refutation texts to stimulate teachers' conceptual change more than both common refutation texts and expository texts, because only the former directly addresses misconceptions and increases personal involvement (personalized refutation hypothesis). In line with our hypothesis, teachers showed stronger conceptual change with personalized refutation texts compared to both common refutation texts and expository texts. There was no difference between the two latter conditions (see Figure 4). Thus, unlike in previous research (see Tippett, 2010, for a meta-analysis), we did not find common refutation texts fostering conceptual change more than expository texts. Nevertheless, teachers showed a strong degree of conceptual change (from incorrect to correct conceptions) in all three conditions including expository texts (see Table 2). Therefore, the common refutation texts did not fail to stimulate conceptual change at all, but rather, expository texts stimulated conceptual change to a similar degree. One potential reason why teachers did not show stronger conceptual change with common refutation texts is that the teachers, who are experienced with teaching, might not have felt sufficiently addressed by the common refutation text starting with "most people believe that." Unlike novices, teachers as experienced practitioners, might not have counted themselves as "most people" as described at the beginning of the refutation texts about multimedia learning. Therefore, refutation texts might not have worked as well as for novices (see also results by Poehnl & Bogner, 2013).

The present results suggest that refutation texts can be made more effective for teachers' conceptual change when they are designed in a personalized manner. When teachers read "your answers indicate that,"

they potentially felt more addressed, which is known to increase the likelihood for conceptual change (Gregoire, 2003; Sinatra, 2005). In addition, the present results indicate that personalized refutation of misconceptions can be effectively implemented with a parsimonious computer algorithm (Okoye, 2015). Specifically, we applied a conservative algorithm, deciding for a misconception if only one misconception-item per scale was agreed to with certainty. Nevertheless, we had to provide half the teachers the feedback that all their previous answers reflected misconceptions about multimedia learning. Only the other half of the teachers (in this condition) received balanced feedback. Still, teachers showed conceptual change to a similar degree, regardless of whether feedback for them was balanced or not, pointing towards the importance of personalization. It is still an open question whether teachers in the condition with unbalanced negative feedback actually believed the personalized feedback to being adaptive (in fact, it was). This question is an issue for further research.

It is noteworthy that there were no significant differences in transfer performance, even though we found descriptively the highest scores for personalized refutation (see Table 2). Further research might extend the transfer test by developing additional items to assess whether personalized refutation texts also benefit transfer performance.

5.2 | Specific negative affect is related to conceptual change

We expected stronger conceptual change due to personalized refutation texts to be driven by a moderate increase in the specific negative affect of shame and guilt, resulting from blaming oneself for incorrect prior beliefs (shame-drives-change hypothesis). Results support this hypothesis. Teachers reported stronger feelings of shame and guilt that led to the stronger conceptual change with personalized refutation texts (see Figure 5). These finding fits the assumption of the CAMCC (Gregoire, 2003) that a certain degree of discomfort is required for teachers' conceptual change to occur. Teachers were directly challenged in their pre-existing conceptions about multimedia learning in this condition (by receiving personalized corrective feedback) and then presented with alternative, scientifically-based conceptions. On average, teachers reacted to this discrepancy with internal blame being associated with somewhat stronger feelings of shame and guilt, which presumably made them process subsequent information more deeply leading to increased conceptual change (see mediation model; Figure 5). Concerning this mediated effect, however, it is to note that ratings of shame and guilt were generally low, albeit somewhat higher in the personalized refutation condition (see Table 2). This finding confirms that *up to a certain degree*, negative affect such as shame and guilt drives attitude changes because learners process subsequent information more systematically (Nabi, 1999). Nevertheless, our mediation model showed this path only indirectly, because we could not assess learning process data on a fine-grained level. Assessing such data (e.g., by eye-tracking) would be worthwhile in further research to examine these links more directly.

Another potential reason underlying more pronounced conceptual change due to personalized refutation texts is that anger might be reduced. More specifically, we expected that teachers as experienced in (multimedia) learning might react with anger or repulse being lessoned on their topic (here: learning), as typically the case in refutation texts. Such specific negative affect might hinder conceptual change. Personalized refutation texts hold the potential to provide more balanced feedback. When correct previous conceptions are identified, they are reflected as such. The recognition of performance may prevent from an increase in anger or repulse, leading to stronger conceptual change in this condition (anger-hinders-change hypothesis). Our findings did not, however, support this hypothesis. Personalized refutation did not reduce anger, and anger itself did not affect the strength of conceptual change, regardless of how balanced the feedback was according to pre-test scores. However, it cannot be concluded that anger does not hinder conceptual change in general, since the descriptive values for anger were very low in this study. Further studies producing stronger anger would be necessary to fully examine this relationship.

5.3 | Limitations and further research

We conducted this experiment online with volunteers. Therefore, we cannot exclude the possibility that participants with stronger negative affect dropped out of the study so that we could not analyze their data. Hence, selective drop-out might have led to the generally low scores of negative affect. Nevertheless, there are significant differences between conditions in the specific negative affect of guilt-shame, and conditions did not differ with respect to the number of dropouts. Also, we found significant differences in conceptual change scores despite the online situation. These significant differences speak of the robustness of this manipulation.

The high dropout of 45% indicates that only motivated teachers willing to learn about multimedia learning finished the experiment. However, a selection of motivated participants is the case for any psychological experiment and increases with duration of the assessment, especially in an online setting (Hoerger, 2010). Further, conducting the present study online, allowed us collecting data of teachers from very different regions, increasing the external validity of our data.

We found conceptual change to be most effective with personalized refutation instruction compared to non-personalized refutation texts as well as non-personalized expository texts. Further, we found specific negative affect (guilt, shame) to mediate conceptual change. However, the transfer test did not reveal these same effects. One reason may be, that while the MMLQ overlapped with the instructional texts content-wise, the items of the transfer test did not. However, as the contents of the instructional texts did not differ between the three conditions, our between-group-effects are not confounded. Apart from that, the low reliabilities of the transfer test may have obscured potential between-group-effects in our sample. We invite future research to develop a more valid and reliable transfer test.

Even though results were clear for our conceptual change scores, we measured conceptual change just on a short temporal scale.

Teachers responded to the misconception questionnaire directly before and directly after the (refutation) instruction. We thus showed immediate benefits on conceptual change due to personalized refutation instruction. The stability or further development of this conceptual change remains an issue for further research. Such research may administer the misconception questionnaire and a transfer test after a significant temporal delay. Future research is further invited to include learners' individual traits such as attributional styles (e.g., Hewitt et al., 2004) as well as beliefs on teaching (Buehl & Fives, 2009) and teaching experience (Eitel et al., 2019). These traits may have influenced learning as well as affect ratings in our study (Trevors et al., 2016). It would also be important to test the present effects with different learning contents that are also important for teaching practice (e.g., gender stereotypes or stigmata about psychological disorders).

5.4 | Conclusions and implications

The present experiment studied conceptual change about multimedia learning of in-service teachers, a professional group with a higher (perceived) expertise status compared to the student groups tested in most previous research. Here, common refutation texts did not foster conceptual change to a stronger degree than expository texts. The beneficial effects of common refutation texts, as mostly found with novices (Tippett, 2010), did not apply to in-service teachers. The present findings thus suggest that the experience status might moderate effects of common refutation texts. Further studies should assess teachers' experience status and systematically test for its moderating role. Nevertheless, the present results already indicate that common refutation texts may not be the best way to stimulate teachers' conceptual change. Rather, the present results suggest that refutation texts can be made more effective for teachers' conceptual change by personalizing them with the help of technology. Exploiting the power of technology to provide adaptive feedback to one's previous responses (here by presenting refutation texts, if needed) can benefit conceptual change, and may thus make trainings in the profession of teaching, and others, more effective in the future.

ENDNOTE

* We calculated McDonald's omega (ω) for reliability estimation of scales comprising at least three items. McDonald's omega is the model-based analogue of Cronbach's alpha that is also appropriate when item-scale correlations vary in their magnitude (non tau-equivalence). We further calculated *asymptotic omega* for short scales (e.g. four items). Asymptotic omega is the omega that would be obtained for a test of infinite length with a structure similar to the observed test. Values of McDonald's omega are interpreted similarly to Cronbach's alpha.

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict in interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12671.

DATA AVAILABILITY STATEMENT

Data will be made available upon request via email to the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ETHICS STATEMENT

The authors state that the appropriate ethical approval has been received from the internal review board of the psychology department of Justus-Liebig-University Gießen, and that the study conforms to APA standards.

ORCID

Anna-Sophia Dersch https://orcid.org/0000-0002-6012-3411

REFERENCES

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

Alvermann, D. E., & Hynd, C. R. (2015). Effects of prior knowledge activation modes and text structure on nonscience majors' comprehension of physics. *The Journal of Educational Research*, *83*(2), 97–102. https://doi.org/10.1080/00220671.1989.10885937

Ariasi, N., & Mason, L. (2011). Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional Science*, *39*, 581–601. https://doi.org/10.1007/s11251-010-9142-5

Asterhan, C. S., & Dotan, A. (2018). Feedback that corrects and contrasts students' erroneous solutions with expert ones improves expository instruction for conceptual change. *Instructional Science*, *1-19*(46), 337–355. https://doi.org/10.1007/s11251-017-9441-1

Beijaard, D., Verloop, N., & Vermunt, J. D. (2000). Teachers' professional identity: an ex-ploratory study from a personal knowledge perspective. *Teaching and Teacher Education*, *16*, 749–764. https://doi.org/10.1016/S0742-051X(00)00023-8

Buehl, M. M., & Fives, H. (2009). Exploring teachers' beliefs about teaching knowledge: Where does it come from? Does it change? *The Journal of Experimental Education*, *77*(4), 367–408. https://doi.org/10.3200/JEXE.77.4.367-408

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, *14*, 5–26. https://doi.org/10.1023/A:1013176309260

Chi, M. T., & Roscoe, R. D. (2002). The processes and challenges of conceptual change. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: issues in theory and practice* (pp. 3–27). Springer. https://doi.org/10.1007/0-306-47637-1_1

Chin, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1–49. 10.3102%2F00346543063001001

Dekker, S., Lee, N. C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, *3*, 429. https://doi.org/10.3389/fpsyg.2012.00429

Dole, J. A., & Sinatra, G. M. (1998). Reconceptalizing change in the cognitive construction of knowledge. *Educational psychologist*, *33*(2–3), 109–128. https://doi.org/10.1080/00461520.1998.9653294

Eitel, A., Dersch, A.-S., & Renkl, A. (2019). Scientific assumptions and misconceptions of prospective teachers about learning with multimedia.

Educational Science, *47*, 451–474. https://doi.org/10.1007/s42010-019-00049-4

Eitel, A., Prinz, A., Kollmer, J., Niessen, L., Russow, J., Ludäscher, M., Renkl, A., & Lindner, M. A. (2021). The misconceptions about multimedia learning questionnaire. An empirical evaluation study with teachers and student teachers. *Psychology Learning and Teaching*, *20*(3), 420–444. 10.1177%2F14757257211028723

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

Ferrero, M., Hardwicke, T. E., Konstantinidis, E., & Vadillo, M. A. (2020). The effectiveness of refutation texts to correct misconceptions among educators. *Journal of Experimental Psychology: Applied*, *26*(3), 411–421. https://doi.org/10.1037/xap0000258

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford University Press. https://doi.org/10.1038/scientificamerican1062-93

Field, A. (2009). *Discovering statistics using IBM SPSS statistics* (3rd ed.). SAGE.

Friedman, I. A., & Farber, B. A. (1992). Professional self-concept as a predictor of teacher burnout. *The Journal of Educational Research*, *86*, 23–35. https://doi.org/10.1080/00220671.1992.9941824

Gill, M. G., Trevors, G., Greene, J. A., & Algina, J. (2020). Don't take it personally? The role of personal relevance in conceptual change. *The Journal of Experimental Education*, *90*, 1–22. https://doi.org/10.1080/00220973.2020.1754152

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, *15*, 147–179. https://doi.org/10.1023/A:1023477131081

Guzzetti, B. J., Snyder, T. E., & Glass, G. V. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quaterly*, *28*, 116–159. https://doi.org/10.2307/747886

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

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

Hellige, J. B. (1990). Hemispheric asymmetry. *Annual review of psychology*, *41*(1), 55–80. https://doi.org/10.1146/annurev.ps.41.020190.000415

Hewitt, A. K., Foxcroft, D. R., & MacDonald, J. (2004). Multitrait-multimethod confirmatory factor analysis of the attributional style questionnaire. *Personality and Individual Differences*, *37*(7), 1483–1491. https://doi.org/10.1016/j.paid.2004.02.005

Hoerger, M. (2010). Participant dropout as a function of survey length in internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking*, *13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about psychological science: A Review. *Psychology Learning & Teaching*, *12*, 20–31. https://doi.org/10.2304/plat.2013.12.1.20

Jacob, R., Hill, H., & Corey, D. (2017). The impact of a professional development program on teachers' mathematical knowledge for teaching, instruction, and student achievement. *Journal of Research on Educational Effectiveness*, *10*, 379–407. https://doi.org/10.1080/19345747.2016.1273411

Kendeou, P., & O'Brien, E. J. (2014). The knowledge revision components (KReC) framework: Processes and mechanisms. D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (p. 353–377). MIT Press.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education*, *106*, 166–171. https://doi.org/10.1016/j.compedu.2016.12.006

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Krammer, G., Vogel, S. E., Yardimci, T., & Grabner, R. H. (2019). Neuromythen sind zu Beginn des Lehramtsstudiums prävalent und unabhängig vom Wissen über das menschliche Gehirn [Neuromyths are prevalent and independent of the knowledge of the human brain at the beginning of teacher education]. *Zeitschrift für Bildungsforschung*, *9*, 221–246. https://doi.org/10.1007/s35834-019-00238-2

Larkin, D. (2012). Misconceptions about "misconceptions": Preservice secondary science teachers' views on the value and role of student ideas. *Science Education*, *96*(5), 927–959. https://doi.org/10.1002/sce.21022

Lassonde, K. A., Kendeou, P., & O'Brien, E. J. (2016). Refutation texts: Overcoming psychology misconceptions that are resistant to change. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 62–74 10.1037/stl0000054

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration: I. Basic research* (pp. 51–85). Springer. https://doi.org/10.1007/978-1-4612-4674-9_2

Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of Personality and Social Psychology*, *61*(1), 13. https://doi.org/10.1037/0022-3514.61.1.13

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

McCrudden, M. T., & Kendeou, P. (2012). Exploring the link between cognitive processes and learning from refutational text. *Journal of Research in Reading*, *37*, 116–140. https://doi.org/10.1111/j.1467-9817.2011.01527.x

McElvany, N., Schroeder, S., Baumert, J., Schnotz, W., Horz, H., & Ullrich, M. (2012). Cognitively demanding learning materials with texts and instructional pictures: teachers' diagnostic skills, pedagogical beliefs and motivation. *European Journal of Psychology of Education*, *27*, 403–420. https://doi.org/10.1007/s10212-011-0078-1

Menz, C., Spinath, B., & Seifried, E. (2021). Misconceptions die hard: Prevalence and reduction of wrong beliefs in topics from educational psychology among preservice teachers. *European Journal of Psychology of Education*, *36*, 477–494. https://doi.org/10.1007/s10212-020-00474-5

Molloy, E., Borrell-Carrió, F., & Epstein, R. (2012). The impact of emotions in feedback. In D. Bound & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 60–81). Routledge.

Nabi, R. L. (1999). A cognitive-functional model for the effects of discrete negative emotions on information processing, attitude change, and recall. *Communication Theory*, *9*(3), 292–320. https://doi.org/10.1111/j.1468-2885.1999.tb00172.x

Okoye, I. U. (2015). *Building an educational recommender system based on conceptual change learning theory to improve students' understanding of science Concepts*. University of Colorado.

Poehnl, S., & Bogner, F. X. (2013). A modified refutation text design: Effects on instructional efficiency for experts and novices. *Educational Research and Evaluation*, *19*(5), 402–425. https://doi.org/10.1080/13803611.2013.789229

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of sci-entific conception: Toward a theory of conceptual change. *Science Edcuation*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

Prinz, A., Golke, S., & Wittwer, J. (2019). Refutation texts compensate for detrimental effects of misconceptions on comprehension and metacomprehension accuracy and support transfer. *Journal of Educational Psychology*, *111*(6), 957–981.

Prinz, A., Kollmer, J., Flick, L., Renkl, A., & Eitel, A. (2021). Refuting student teachers' misconceptions about multimedia learning. *Instructional Science*, 1–22.

Renkl, A. (2015). Different roads lead to Rome: The case of principle-based cognitive skills. *Learning: Research & Practice*, *1*, 79–90. https://doi.org/10.1080/23735082.2015.994255

Rohrer, D., & Pashler, H. (2012). Learning. *Styles: Where's the evidence?*., *46*(7), 634–635. https://doi.org/10.1111/j.1365-2923.2012.04273.x

Sánchez, E., García-Rodicio, H., & Acuña, S. R. (2009). Are instructional explanations more effective in the context of an impasse? *Instructional Science*, *37*, 537–563. https://doi.org/10.1007/s11251-008-9074-5

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self–affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, *11*(4), 119–123. 10.1111%2F1467-8721.00182

Sinatra, G. M. (2005). The "warming trend" in conceptual change research: The legacy of Paul R. Pintrich. *Educational Psychologist*, *40*(2), 107–115. https://doi.org/10.1207/s15326985ep4002_5

Sinatra, G. M., & Broughton, S. H. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly*, *46*(4), 374–393. https://doi.org/10.1002/RRQ.005

Smith, J. P., III, DiSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*(2), 115–163.

Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, *8*(6), 951–970. https://doi.org/10.1007/s10763-010-9203-x

Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P. H. (2016). Identity and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes*, *53*(5–6), 339–370. https://doi.org/10.1080/0163853X.2015.1136507

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69. https://doi.org/10.1016/0959-4752(94)90018-3

Vosniadou, S., Lawson, M. J., Wyra, M., Van Deur, P., Jeffries, D., & Ngurah, D. I. G. (2020). Pre-service teachers' beliefs about learning and teaching and about the self-regulation of learning: A conceptual change perspective. *International Journal of Educational Research*, *99*, 101495. https://doi.org/10.1016/j.ijer.2019.101495

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

How to cite this article: Dersch, A.-S., Renkl, A., & Eitel, A. (2022). Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning. *Journal of Computer Assisted Learning*, *38*(4), 977–992. https://doi.org/10.1111/jcal.12671

APPENDIX A

TABLE A1 Overview of the different variations of the text on motivation primacy (text 2 out of 4)

	Expository text (=scientific explanation for correct conception)	**Refutation text (=naming common misconception, disproving it + expository text)**	**Personalized Refutation Text (=Personal feedback + refutation text [*only if necessary*] + expository text)**
Correct conception (refutation not necessary	What function do images fulfill in learning? Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus, can prevent false interpretations of the text. Images can also make difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)	What function do images fulfill in learning? Many people assume that images are conductive to learning because they motivate the learners and increase their interest. However, this assumption does not align with scientific evidence. Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus, can prevent false interpretations of the text. Images can also make difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)	What functions do images fulfill in learning? Our evaluation shows that you answered the questions about the motivating functions of images during learning in accordance with current scientific findings—great! In the following you can familiarize yourself with the scientific findings once again and supplement your knowledge on the topic. Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus, can prevent false interpretations of the text. Images can also make difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)
Insecurity (refutation not necessary	What function do images fulfill in learning? Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus, can prevent false interpretations of the text. Images can also make difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use	What function do images fulfill in learning? Many people assume that images are conductive to learning because they motivate the learners and increase their interest. However, this assumption does not align with scientific evidence. Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus, can prevent false interpretations of the text. Images can also make	What function do images fulfill in learning? Our evaluation shows that you answered the questions about the motivation functions of images during learning with insecurity. Therefore, we would like to give you the opportunity to familiarize yourself with the current state of research in this field. Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus,

TABLE A1 (Continued)

	Expository text (=scientific explanation for correct conception)	Refutation text (=naming common misconception, disproving it + expository text)	Personalized Refutation Text (=Personal feedback + refutation text [*only if necessary*] + expository text)
	because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)	difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)	can prevent false interpretations of the text. Images can also make difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)
Misconception (refutation necessary)	What function do images fulfill in learning? Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus, can prevent false interpretations of the text. Images can also make difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)	What function do images fulfill in learning? Many people assume that images are conductive to learning because they motivate the learners and increase their interest. However, this assumption does not align with scientific evidence. Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus, can prevent false interpretations of the text. Images can also make difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)	What function do images fulfill in learning? Our evaluation shows that you did not answer the questions about the motivating functions of images during learning in accordance with current scientific findings. Therefore, we would like to give you the opportunity to familiarize yourself with current state of research in this field. Many people assume that images are conductive to learning because they motivate the learners and increase their interest. However, this assumption does not align with scientific evidence. Images are beneficial for learning because they can contribute to a better understanding of the learning material. For example, text-based statements about spatial relationships such as "The aorta is next to the superior vena cava." are ambiguous. In contrast, an image specifies whether the aorta is to the left or right of the superior vena cava. If an image is presented in addition to a text, the image can reduce ambiguity in the text; thus, can prevent false interpretations of the text. Images can also make difficult or abstract verbal descriptions easier to understand (Ainsworth, 2006). However, not all types of images support learning equally. For example, decorative pictures are of limited use because they do not provide any explanatory reference to the learning material. In contrast, logical images are usually more conducive to learning, especially when the text refers to essential elements and contents of the image (Carney & Levin, 2002)

Notes: Texts were translated from German. Participants in the *ETC* and the *RTC* received the same texts regardless of their performance in the pretest, participants in the *FRC* received personal feedback and a refutation text only if necessary. Expository texts were a component of every intervention text regardless of condition.

TABLE A2 Misconceptions about Multimedia Learning Questionnaire (MMLQ) ordered by subscales, without filler items

Learning styles (LS)	
LS1	Performance is decreased when visual learners study with text or when verbal learners study with animations or diagrams
LS2	Performance is better when students work with materials that are in accordance with their learning style
LS3	A necessary condition for good teaching is to know the students' learning style (visual, verbal, kinesiological).
LS4	Whether students learn better with visual or verbal materials depends on their learning style
Motivation primacy (MP)	
MP1	The main reason for why pictures are beneficial is that they motivate
MP2	Students have less fun learning with an unillustrated textbook so that this is never as effective as learning with an illustrated textbook
MP3	The picture contents are not so decisive as long as the picture serves the purpose to motivate
MP4	Text without illustrations is not helpful to learning because students do not become motivated
Naïve Summation (NS)	
NS1	The more sensory channels are addressed simultaneously, the better information will be remembered
NS2	Information that is simultaneously read and listened to will be better remembered than just read information
NS3	Videos are more effective for learning because they address multiple sensory channels
NS4	To what degree information is remembered particularly depends on the number of sensory channels used to perceive it (e.g., seeing, hearing, touching)
Hemispheric isolation (HI)	
MP1	Whether students learn better with text or pictures depends on which brain hemisphere is the better trained
MP2	Most students use only one of both brain hemispheres for learning
MP3	The two brain hemispheres do not work together automatically so this needs to be stimulated by appropriate study materials
MP4	Differences in hemispheric dominance (left vs. right brain) explain differences learning outcomes

TABLE A3 Examples of correct and incorrect answers for all four learning materials of the transfer test translated from German

	Example of a correctly explained answer	Example of an incorrectly explained answer
Material 1	There is no depiction of the explained painting technique to be found on the attached illustration of Material B, so it is nice to look at, but unnecessary and does not help the understanding of the text. Thus, the material without the picture seems fitting	It's art class, so I expect a picture. It serves as an eye-catcher
Material 2	Lecture and simultaneous reading of slides with the exact same and thus redundant content leads to listeners or readers focusing on matching content and trying to tune out one of both channels [reading or listening]. Hence they focus less on the relevant content	Structure is better
Material 3	In material A, the information in the text is additionally explained in terms of content by the diagram. The fox in material B rather distracts from the content of the text and does not add value to the information read	Pictures help make the topic more vivid, necessary to get motivated for the topic
Material 4	The constellation of the three celestial bodies can be understood in a structured and correct way with the help of a (good) animation, while the focus on one of the other media—text, illustration or model—represents a static situation in each case. The static model further is irrelevant as learning types supposedly do not exist	The selection appeals to as many learning types as possible