

Task Experience as a Boundary Condition for the Negative Effects of Irrelevant Information on Learning

Gertjan Rop¹ · Margot van Wermeskerken^{1,2} ·
Jacqueline A. de Nooijer¹ · Peter P. J. L. Verkoeijen^{1,3} ·
Tamara van Gog^{1,2}

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Research on multimedia learning has shown that learning is hampered when a multimedia message includes extraneous information that is not relevant for the task, because processing the extraneous information uses up scarce attention and working memory resources. However, eye-tracking research suggests that task experience might be a boundary condition for this negative effect of extraneous information on learning, because people seem to learn to ignore task-irrelevant information over time. We therefore hypothesised that extraneous information might no longer hamper learning when it is present over a series of tasks, giving learners the chance to adapt their study strategy. This hypothesis was tested in three experiments. In experiments 1a/1b, participants learned the definitions of new words (from an artificial language) that denoted actions, with matching pictures (same action), mismatching pictures (another action), or without pictures. Mismatching pictures hampered learning compared with matching pictures. Experiment 2 showed that task experience may indeed be a boundary condition to this negative effect on learning: the initial negative effect was no longer present when learners gained experience with the task. This suggests that learners adapted their study strategy, ignoring the mismatching pictures. That hypothesis was tested in experiment 3, using eye tracking. Results showed that attention to the pictures waned with task experience, and that this decrease was stronger for mismatching than for matching pictures. Our findings demonstrate the importance of investigating multimedia effects over time and in relation to study strategies.

✉ Gertjan Rop
rop@fsw.eur.nl

¹ Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands

² Department of Education, Utrecht University, Utrecht, The Netherlands

³ Learning and Innovation Center, Avans University of Applied Sciences, Breda, The Netherlands

Keywords Multimedia learning · Coherence effect · Cognitive load · Eye tracking · Study strategies

Multimedia learning, which can be defined as learning with a combination of words (written or spoken) and pictures (static or dynamic), has been widely investigated in research inspired by the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2014) and Cognitive Load Theory (CLT; Sweller et al., 2011). This has led to the establishment of several principles for designing effective multimedia instructions. The present study is concerned with the coherence principle, which states that presenting extraneous information that is not relevant for the learning task should be avoided, because it hinders rather than helps learning (Mayer & Fiorella, 2014). Because eye-tracking research has shown that with increasing task experience, people learn to ignore irrelevant information during task performance, we hypothesised that task experience might be a boundary condition for the negative effect of extraneous information on learning. That is, the negative effect that the presentation of extraneous information initially has on learning might no longer occur when this information is present (in the same location) over a series of tasks, because learners might adapt their study strategy (i.e. learn to ignore the extraneous information). This hypothesis was tested in a series of three experiments, which will be introduced after discussing the relevant literature in more detail.

Cognitive Load in Multimedia Learning

Decades of multimedia research have shown that learning often improves when study tasks or materials combine pictorial and verbal representations of the content (i.e. the multimedia effect; Butcher, 2014). However, it soon became apparent that there are circumstances under which this hampers rather than aids learning (e.g. Chandler & Sweller, 1991; Harp & Mayer, 1998). These circumstances are related to the limitations of working memory, which can be defined as ‘a limited capacity [brain] system allowing the temporary storage and manipulation of information necessary for such complex tasks as comprehension, learning and reasoning’ (Baddeley, 2000, p. 418; text in square brackets added). Working memory is limited in both duration and capacity (e.g. Baddeley, 2000; Barouillet & Camos, 2007; Cowan, 1995; Miller, 1956). For instance, on average, our memory span is ‘seven plus or minus two’ chunks, where a chunk is one piece of information (Miller, 1956). Barouillet and Camos (2007) propose that the limited working memory resources have to be shared by rapidly switching attention between maintenance of ‘old’ information (prior knowledge from long-term memory or previously processed information during task performance), and processing of new, incoming information. Consequently, the higher the number of old information elements that have to be maintained active, and the faster new elements need to be processed, the higher the working memory load.

Learning (i.e. schema construction/elaboration in long-term memory; Sweller, 1994) requires that old information is maintained active in working memory and successfully integrated with the new information presented in the learning materials. When these processes are disrupted, learning is hampered. Moreover, learning may be hindered when scarce working memory resources are devoted to processing extraneous information that is not necessary for the learning task. Such extraneous processing may not be detrimental for learning when capacity limits are not exceeded, for instance with simple materials (that contain few interacting information elements), or when there is sufficient time available to compensate for the extraneous processing. However, it will

start to hamper learning when materials are complex (with many interacting information elements) or when time is constrained (Barouillet & Camos, 2007; Sweller et al., 2011). Consequently, it is important to avoid the presentation of extraneous information that does not contribute to learning as much as possible.

Avoiding the Presentation of Extraneous Information in Multimedia Learning

That the presentation of extraneous information can have a negative effect on learning has been established in many experiments. However, there are two different types of extraneous information presentation effects. The first, which is generally called the redundancy effect (Kalyuga & Sweller, 2014; Mayer & Fiorella, 2014; Sweller et al., 2011), concerns the negative effect of the presentation of *identical* extraneous information to learners in two modalities compared with a single modality. For example, it has been shown that when the text accompanying pictures or animations is presented simultaneously in both spoken and written form, this hampers learning compared with spoken text only (e.g. Craig et al., 2002; Mayer et al., 2001; but see Mayer & Johnson, 2008; Yue et al., 2013).

The second, which is called the coherence effect in CTML (Mayer & Fiorella, 2014; Mayer & Moreno, 2003), concerns the negative effect on learning of the presentation of extraneous information that is not relevant or necessary for learning, but is added to enrich or elaborate learning materials, compared with when this is left out. For instance, learning is hampered when interesting and entertaining information that is related to the topic but irrelevant for the learning task at hand is added to enrich materials (i.e. seductive details, e.g. ‘fun’ facts, pictures, videos or sounds; Harp & Mayer, 1998; Mayer et al., 2001; Moreno & Mayer, 2000); when learning materials are unnecessarily elaborate, presenting textual explanations with self-explanatory diagrams (Bobis et al., 1993; Chandler & Sweller, 1991) or presenting details and examples whereas a concise and coherent summary would suffice (e.g. Mayer et al., 1996; Reder & Anderson, 1982); or when information on related systems is presented when learning about a specific system (Mayer et al., 2007).

The negative effects of extraneous information presentation on learning presumably arise because learners attend to, process and attempt to integrate the extraneous information with the essential information, which unnecessarily depletes valuable working memory resources. Moreover, in some cases of the coherence effect, the content of the additional information that is presented may actively interfere with learning the essential information. For instance, Mayer et al. (2007) showed that adding explanations about calliper brakes and air brakes interfered with learning the working mechanisms of hydraulic brakes. Participants who learned about the calliper and air brakes made more intrusion errors (i.e. including information about calliper and air brakes in their answers) than participants who only learned about hydraulic brakes. The present study addresses the coherence effect, by investigating the effects of extraneous pictorial information on word learning in a ‘second’ (artificial) language.

Multimedia and Coherence Effects in Word Learning

As mentioned above, the multimedia effect refers to the finding that learning often improves when study tasks or materials combine pictorial and verbal representations of the content

(Butcher, 2014). In this case, additional information is also presented, but it is not extraneous to (i.e. does not hamper) learning and may even facilitate learning.

With regard to word learning, some studies have shown that adding pictures of the word to be learned (and example sentences in which the word is used), does not hamper and—in accordance with the multimedia effect—can even foster word learning (for a review, see Sadoski, 2005). For example, Smith et al. (1987) taught undergraduate students novel words in their first language (English) in one of three conditions: definition only, definition and a sentence using the word and definition, a sentence and a picture. Results on the immediate retention test favoured the condition with pictures, although the differences were not significant. On a delayed retention test, the condition with pictures performed significantly better than the definitions only condition (but not than the definitions plus sentences condition), demonstrating a multimedia effect (i.e. pictures facilitating word learning; Butcher, 2014).

Whether a multimedia (i.e. facilitative) effect of pictures on word learning is found, however, may depend on how easily the words can be mentally simulated or visualized. For instance, in a second language vocabulary learning study, Farley et al. (2014) taught Spanish vocabulary to English-speaking university students and found that abstract words (i.e. words without a physical referent) were learned better with pictures, while there was no such effect for concrete words (i.e. words with a physical referent). Concrete words may be easy to mentally visualize and learn because of the physical referent (e.g. Altarriba & Bauer, 2004), in which case pictures do not have much added value for learning. Shen (2010) found a comparable effect in learning Chinese as a foreign language: Pictures improved word learning, but only for abstract words. For non-abstract words, even though presenting pictures did not help, it did not hinder learning either.

While additional presentation of pictures of the words to be learned, might not hamper and could even foster learning, presentation of pictures that do not match the words might interfere with learning. For instance, a recent study on effects of animations on action word learning (e.g. to chisel, to hoe) in the first language (Hald et al., 2015), included a mismatched animation condition, to control for effects of movements shown in the animations. Results showed that word learning was significantly hampered in the animation condition in which the actions depicted in the animation mismatched the word to be learned compared with when it matched the word to be learned.

Another study investigated the effects of pictures on action word learning in a ‘second’ artificial language, which matched or mismatched the learners’ handedness (De Nooijer et al., 2013). Participants first saw the artificial language word (e.g. ‘luko’) on the screen, and heard a verbal definition of the word (e.g. ‘luko’ means ‘to dispense from a container’). Hearing actions described tends to automatically result in a body-specific (Casasanto, 2009) mental simulation/visualization of the action (e.g. Hauk et al., 2004; Pulvermüller et al., 2005; Tettamanti et al., 2005). Then, they heard this definition a second time, but now a picture was shown along with the artificial language word. This picture always showed the defined action, but either matched (i.e. right handed for right handers) or mismatched (i.e. left handed for right handers) participants’ mental simulation. The mismatching pictures hampered right-handers’ learning (this was a small but consistent effect across multiple experiments),¹ presumably because attending to and processing the picture that mismatched with their mental simulation of the action, interfered with learning the verbal definition.

¹ Left handers’ learning, in contrast, was not affected by right-handed pictures, presumably because they have visual and actual experience with the right hand being used for these actions.

In sum, additional presentation of pictures that match the word to be learned might not hamper, and may even help word learning. In contrast, pictures (whether static or dynamic) that mismatch with (the mental visualization of) the word to be learned have been shown to hinder learning. It is assumed that mismatching pictures hamper word learning because they capture learners' attention, and lead them to engage in extraneous—and conflicting—information processing. However, it is unclear whether extraneous information would continue to hamper learning when it is present (in the same location) over a series of tasks. This would give learners the chance to adapt their study strategy and ignore the extraneous information. Indeed, eye-tracking research suggests that with increasing task experience, people learn to focus their attention more on task-relevant information and to ignore task-irrelevant information.

Learning to Ignore Task-Irrelevant Information

Several studies have shown that with increasing task experience, people learn to focus on task-relevant and ignore task-irrelevant information. For instance, expertise research showed that chess experts had different viewing patterns compared with intermediate chess players, which included making fewer fixations and fixating more on relevant pieces (Charness et al., 2001). Likewise, Van Gog et al. (2005) demonstrated that participants with more expertise in an electrical circuit-troubleshooting task had shorter mean fixation duration and fixated more on task-relevant components of the electrical circuit in the first phase of troubleshooting, compared with participants with less expertise. Furthermore, Jarodzka et al. (2010) showed that, in a visually complex dynamic task, experts attended more to the relevant parts of a stimulus compared with novices. While these studies suggest that people are able to ignore irrelevant information with increasing expertise, they did not show a causal relationship between task experience and viewing behaviour, as they compared existing groups of experts and novices or intermediates.

Such evidence also exists, however, and comes from studies that investigated how viewing behaviour changed as expertise developed. These studies confirm that, even after relatively little practice, participants start to ignore task-irrelevant information and focus more on task-relevant information. For example, Haider and Frensch (1999) used an alphabetic string verification task and showed that participants implicitly learned to ignore the task-irrelevant information when they became more experienced with the task. Furthermore, Canham and Hegarty (2010; see also Hegarty et al., 2010) used a task in which participants had to learn to make inferences from weather maps, and showed that participants fixated more on task relevant information after a short training (10–15 min) than before a training about relevant meteorological principles.

Taken together, these findings strongly suggest that people may learn to ignore irrelevant information during task *performance*, as a result of increasing experience with a task. However, these studies did not investigate effects on *learning*. When learners are able to start ignoring extraneous information with increasing experience during learning, then it would no longer capture attention and working memory resources, and therefore, the negative effect of extraneous information should decrease or no longer occur with increasing task experience. In other words, these findings suggest that task experience might be a boundary condition to the negative effect of extraneous information on learning. This is not only theoretically relevant to establish, as it provides more insight into the underlying cognitive mechanisms of multimedia

effects and the role that study strategies play over time, but it is also relevant for instructional designers, because some multimedia principles are hard to implement generally (e.g. what may be essential information for someone with low prior knowledge, may be extraneous for a more advanced learner). The present study addressed this hypothesis in a series of three experiments, using a word-learning task.

The Present Study

Experiments 1a and 1b were conducted to first establish whether mismatching pictures would negatively affect word learning compared with no pictures and matching pictures. We used a similar experimental design as De Nooijer et al. (2013), but with right-handed pictures and right-handed participants only. Participants learned words from an artificial language called Vimmi (Macedonia & Knösche, 2011) that we coupled with action verb definitions. We used an artificial language to exclude any influences of prior knowledge of the words to be learned and associations on idiomatic level between two languages. Participants saw the artificial language word on the screen and heard the verbal definition of the word (e.g. 'ifra' means 'to polish or scrape with sandpaper'), after which they heard this definition a second time, either without a picture (control condition), with a matching picture (showing the action) or a mismatching picture (showing another action). Note that matching pictures, even though they showed the action that was being defined, are not entirely redundant to the definition and may in fact provide useful additional information. For instance, in case of known actions, these pictures may automatically prompt participants to think of the English verb that was not part of the definition (e.g. in the case of ifra, the picture might clarify that they are hearing the definition of 'to sand'), and in case of unknown actions, the pictures might help understand the meaning of the word.

We expected that the mismatching pictures would capture attention and interfere with processing the verbal definition the second time it was presented, which would hamper learning compared with the no-pictures condition (i.e. a coherence effect) and the matching picture condition (as we expected this condition to do as well or better than the no-picture condition, see below). Learning could be hampered by mismatching pictures via two (not mutually exclusive) routes: (1) less attention could be devoted to processing the materials relevant for learning and (2) processing the content of the mismatching action pictures could actively interfere with learning the correct action word definition. With respect to the matching pictures, we expected based on the studies described in the introduction that these would either improve learning the meaning of new words (Farley et al., 2014; Shen, 2010; Smith et al., 1987), providing evidence of a multimedia effect, or not affect word learning. If they do not significantly contribute to word learning, one could argue that matching pictures also constitute extraneous information, in the sense that they are redundant and could be left out at no expense to learning. As such one could argue that this null finding would also be a kind of coherence effect, though not in the strict definition (Mayer & Fiorella, 2014), as excluding the pictures would not have a positive effect on learning compared with including them.

To foreshadow, experiment 2 addressed the main question of whether the negative effect of mismatching pictures on word learning would decrease or no longer occur with increasing task experience, which would suggest that learners adapted their study strategy. Experiment 3

subsequently investigated study strategies directly, by means of eye tracking (to measure attention allocation), to determine whether participants indeed learned to ignore the mismatching pictures over time, with task experience.

Experiments 1a and 1b

As described above, experiments 1a and 1b were conducted to establish whether mismatching pictures would negatively affect word learning compared with no pictures and matching pictures. Experiment 1b was a direct replication of experiment 1a to test the reliability of our results.

Method

Participants and Design Participants (experiment 1a: $n = 85$, experiment 1b: $n = 144$) were recruited via Amazon's Mechanical Turk (Paolacci et al., 2010) and were paid 0.75 dollar for their participation.² A priori-defined criteria for post hoc exclusion were the following: being a non-native English speaker ($n = 2$ and $n = 1$, respectively), being left handed (i.e. the pictures were right handed and even though left handers seem less hampered by right-handed pictures than right handers are by left-handed pictures according to findings by De Nooijer et al., 2013, we wanted to rule out any potential handedness effects; $n = 9$ and $n = 21$, respectively), participating in the experiment twice or having participated in a similar earlier experiment ($n = 2$ and $n = 3$, respectively) and being in a noisy environment (i.e. self-reported noise of seven or higher on a scale of one to nine; $n = 4$ and $n = 7$, respectively). Thus, for experiment 1a, our final sample comprised 68 participants ($M_{\text{age}} = 39.13$ years, $SD = 14.12$ years, range 20–74; 47 females) and for experiment 1b, our final sample comprised 112 participants ($M_{\text{age}} = 34.58$ years, $SD = 12.66$ years, range 19–89; 78 females). Both experiments employed a within-subjects design, so participants learned words under all three conditions.

Materials Participants learned 18 Vimmi words presented in Qualtrics software (Qualtrics, Provo, UT). Each word was randomly coupled to the definition of an action verb (e.g. ifra means to polish or scrape with sandpaper). In every condition, the participants saw the Vimmi word and heard the definition (spoken by a female voice; $M_{\text{length}} = 3.44$ s, $SD = 1.01$ s) of the word they had to learn twice, each presentation lasted 11 s and the program automatically progressed. In the two picture conditions, a matching picture (showing the action) or a mismatching picture (showing another action) accompanied the word (see Fig. 1) the second time participants heard the definition. Participants' knowledge of the definition was tested with a cued recall test, in which they were presented with the written Vimmi word and had to type in the definition as literally as possible.

Procedure Participants learned the words in three blocks of six, and after each block, the cued recall test for those words was administered. During the cued recall test, no time constraint was imposed. To avoid confusion, the six words in the control (no-picture) condition were presented in one block (either the first or the last) because participants might think a

² Note that this was a common level of payment at the time the experiments were conducted; we are aware of the recent discussions of and increases in MTurk wages.

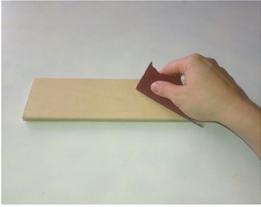
Word	Definition	Condition	Picture
<i>Ifra</i> (to sand)	To polish or scrape with sandpaper	Matched	
		Mismatched	

Fig. 1 Example materials. The spoken definition is presented twice; the second time accompanied by a picture in the matched condition (*top row*) and mismatched condition (*bottom row*)

technical error occurred when no picture would be present at random words during the experiment. The other two blocks consisted of three words with matching and three words with mismatching pictures. Order within block was randomized, so participants did not know beforehand if the next picture would be matching or mismatching the definition, and, thus, could not anticipate on the usefulness of the picture. Furthermore, the artificial words were rotated across definitions to control for possible item effects, resulting in the use of eight lists. In total, the experiment lasted about 20 min and was administered without breaks.

Scoring In all experiments, participants were awarded 1 point if the complete definition was given on the cued recall test. When a part of the definition was missing, they received 0.5 point. If they did not provide a definition, or if it was completely wrong, 0 points were awarded. So, every participant could score a maximum of 6 points on each test. We adopted the scoring scheme of De Nooijer et al. (2013) who used the same materials (they found an interrater reliability of $\kappa = .82$), but established interrater reliability again for the present study, by having part of the data (10.6 %) scored by a second independent rater. Both raters saw only the definitions and words during scoring, so they were blind to the experimental condition under which each word was learned. Because interrater reliability was sufficient ($\kappa = .76$; ‘substantial agreement’ according to Landis & Koch, 1977), the scores from the first rater (first author) were used in this and all subsequent experiments.

Results

Throughout all experiments, we maintained an alpha level of .05, and when the sphericity assumption was violated, we reported the Greenhouse-Geisser correction. Effect size measures used were partial eta-squared and Cohen’s d . Both can be interpreted in terms of small ($\eta_p^2 \sim .01$, $d \sim 0.2$), medium ($\eta_p^2 \sim .06$, $d \sim 0.5$), and large ($\eta_p^2 \sim .14$, $d \sim 0.8$) effect sizes (Cohen, 1988). In addition, effect sizes can also be interpreted with respect to median effects sizes found for these effects (in other studies with different materials). For example, Mayer and Fiorella (2014) report an average effect size of $d = 0.86$ for the coherence effect (based on

$n = 23$ comparisons). In experiments 1a and 1b, data were analysed with a repeated-measure ANOVA with condition (matched, mismatched, control) as a within-subjects variable and the scores on the cued recall test as dependent variable.

Table 1 shows the means and standard deviations for the cued recall tests of experiments 1a and 1b for the different conditions. In experiment 1a, there was a main effect of condition ($F(1.83, 122.26) = 3.37, p = .042, \eta_p^2 = .05$). Bonferroni-corrected post hoc tests showed that the mismatched condition performed worse than the matched condition ($p = .028, d = 0.32$). There were no significant differences between the control and matched condition ($p > .999, d < 0.01$) nor between the control and mismatched condition ($p = .125, d = 0.25$).

Experiment 1b replicated the results of experiment 1a, again showing a main effect of condition ($F(1.82, 202.19) = 4.15, p = .017, \eta_p^2 = .04$) with Bonferroni-corrected post hoc tests, indicating that the mismatched condition performed worse than the matched condition on the recall test ($p = .002, d = 0.32$), and that there were no significant differences between the control and matched condition ($p = .769, d = 0.11$) nor between the control and mismatched condition ($p = .371, d = 0.15$).

Discussion

The results of experiments 1a and 1b provided evidence that adding mismatching pictures to word learning decreased recall performance compared with matching pictures, but there was no significant negative effect compared with no pictures (i.e. no coherence effect). Processing the matching pictures did not significantly improve word learning compared with the no-picture control condition (i.e. there was no multimedia effect) but did not hurt learning either (which could be regarded as a kind of coherence effect, though not in the strictest definition, as exclusion did not lead to better learning than inclusion of the matching pictures). This may be due to the rather concrete object-manipulation verbs we used in our experiment, as previous research has suggested that pictures benefit learning abstract words more than learning concrete words (Altarriba & Bauer, 2004; Farley et al., 2014; Shen, 2010).

The fact that we did not find a coherence effect, indicating that including mismatching pictures would lead to poorer learning outcomes than excluding mismatching pictures (control condition) limits our conclusions somewhat. However, we did find a negative effect of mismatching compared with matching pictures on word learning. This effect of mismatching vs. matching pictures was small, and although it is smaller than the median effect size from other studies on the coherence effect, using different materials ($d = 0.86$ as reported by Mayer & Fiorella, 2014), it was comparable with the effects found by De Nooijer et al. (2013) who used the same materials but with mismatches in terms of handedness. Furthermore, the effect sizes were consistent across experiments 1a and 1b. As this difference between mismatching and matching picture conditions is still interesting for our purpose of investigating whether task experience affects processing of mismatching pictures and thereby, learning outcomes, we

Table 1 Mean (and SD) recall performance (max. = 6) as a function of condition in experiments 1a and 1b

	Experiment 1a	Experiment 1b
Matched	3.65 (1.83)	3.56 (1.75)
Mismatched	3.24 (1.83)	3.15 (1.81)
Control	3.65 (1.78)	3.38 (1.80)

conducted experiment 2 to address this question concerning the effects of task experience. We did not find significant differences compared with the control condition for either picture condition in experiment 1a/1b. Performance was somewhat lower in experiment 1b than in experiment 1a, but the pattern of results was the same. We opted to again include a no-picture control condition in experiment 2, because we could not exclude the possibility that differences between the picture conditions and the control condition might arise over time.

Experiment 2

In experiment 2, we used the same materials as in experiments 1a and 1b, but now in a between-subject design. Participants learned the words in three blocks of five, with matching, mismatching or no pictures. We hypothesised that the negative effect of mismatching compared with matching pictures would occur initially (after the first block) but would decrease or no longer occur with increasing task experience (blocks 2 and 3). That is, when participants would learn that the mismatching pictures are unnecessary for learning and adapt their study strategy accordingly, ignoring these pictures in the later blocks, their recall performance would increase in blocks 2 and 3. For the matched and control conditions, we had no reason to expect that recall performance would change with increasing task experience.

We also explored how participants experienced the task.³ To get an indication of potential differences in processing demands between conditions and over time, participants were asked to rate how much mental effort they invested in learning the words (which is an indicator of how much cognitive load participants experienced: Paas et al., 2003).

Method

Participants and Design All participants ($n = 232$) were recruited via Amazon's Mechanical Turk and were paid 0.75 dollar for their participation (see footnote 2). They were randomly assigned to one of the three conditions. The same exclusion criteria as in the former experiments were used: being a non-native English speaker ($n = 4$), being left handed ($n = 36$), participating in the experiment twice or having participated in a similar earlier experiment ($n = 14$) and being in a noisy environment ($n = 6$). Finally, there were participants in the mismatched condition who did not follow the instructions,⁴ who were also excluded ($n = 5$). Thus, our final sample comprised 166 participants ($M_{\text{age}} = 36.50$ years, $SD = 12.53$ years, range 18–69; 105 females), distributed across the control ($n = 62$), matched ($n = 55$) and mismatched ($n = 49$) conditions.

Materials and Procedure The materials and procedure were similar to experiments 1a and 1b, except that: (1) participants learned 15 word definitions in three blocks of five words under the same condition, (2) after each block of five words and prior to the cued recall test of that

³ We asked participants to rate invested effort, enjoyment, and whether they preferred to learn words without, with matching, or with decorative pictures in the future (cf. Yue et al., 2013) in experiments 2 and 3 (this was not possible in experiments 1a and 1b, where the blocks comprised words presented under different conditions). Only results on effort are reported; data on enjoyment (only null effects) and preferences can be obtained from the first author.

⁴ These participants wrote down one word that described the mismatching picture (i.e. wrench when the artificial word *Ifra* was tested), for each and every definition they had to provide.

block, participants were instructed to indicate how much effort they invested in learning the words, on a 9-point rating scale ranging from 1, *very, very low effort*, to 9, *very, very high effort* (Paas, 1992) and (3) at the end of the experiment, enjoyment and instruction preference was measured (not reported, see footnote 3). The order of the blocks was alternated between participants using a Latin-square design, resulting in three lists per condition. The experiment lasted around 20 min and was administered without breaks.

Results

Unless otherwise specified, all data were analysed with a mixed ANOVA with condition (matched, mismatched, control) as a between-subject factor and word block (first, second, third) as a within-subject factor.

Test Performance Table 2 shows the scores on the cued recall test. The analysis showed no main effect of condition ($F(2, 163) = 1.44, p = .239, \eta_p^2 = .02$), but there was a main effect of word block ($F(1.85, 300.90) = 10.34, p < .001, \eta_p^2 = .06$), indicating that the scores on the cued recall test improved over the course of the experiment. Importantly, this main effect was qualified by an interaction effect ($F(3.69, 300.90) = 2.46, p = .050, \eta_p^2 = .03$).

To follow-up on the interaction effect, we first analysed differences in test performance between conditions per block, using one-tailed independent-sample *t* tests with a Bonferroni-corrected *p* value (i.e. multiplying the *p* value by 3, the number of tests that were performed). In line with our hypothesis, performance was significantly lower in the mismatched condition than in the matched condition in block 1 ($t(102) = 2.18, p = .047, d = 0.43$), but only numerically, not significantly lower than in the control condition ($t(109) = 0.91, p = .545, d = 0.17$; cf. experiment 1). Finally, there was no difference in performance between the matched and the control condition in block 1 ($t(115) = 1.42, p = .239, d = 0.26$; cf. experiment 1). In blocks 2 and 3, there were no significant differences between the conditions. Mismatched vs. matched: block 2, $t(102) = 0.12, p > .999, d = 0.02$; block 3, $t(102) = 0.01, p > .999, d = 0.02$. Mismatched vs. control: block 2, $t(109) = 1.73, p = .128, d = 0.34$; block 3, $t(109) = 1.37, p = .260, d = 0.26$. Matched vs. control: block 2, $t(115) = 1.62, p = .161, d = 0.30$; block 3, $t(115) = 1.32, p = .287, d = 0.24$.

Secondly, we probed the interaction effect by analysing the change in test performance over blocks within each condition. Repeated-measures ANOVAs per condition, with word block (first, second, third) as a within-subject factor, showed no effect of word block in the matched

Table 2 Mean (and SD) recall performance (max. = 5) and mental effort rating (max. = 9) as a function of condition and word block in experiment 2

	Performance			Mental effort		
	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3
Matched	3.15 (1.46)	3.21 (1.51)	3.14 (1.63)	7.20 (1.63)	7.33 (1.63)	7.33 (1.75)
Mismatched	2.51 (1.51)	3.17 (1.45)	3.13 (1.44)	7.73 (1.15)	7.94 (1.25)	7.86 (1.47)
Control	2.77 (1.43)	3.65 (1.39)	3.51 (1.42)	7.52 (1.29)	7.85 (1.13)	7.69 (1.43)
Overall	2.82 (1.48)	3.36 (1.46)	3.27 (1.50)	7.48 (1.38)	7.70 (1.37)	7.62 (1.56)

condition ($F(2, 108) = 0.08, p = .926, \eta_p^2 < .01$), a main effect of word block in the mismatched condition ($F(2, 96) = 4.75, p = .011, \eta_p^2 = .09$) and a main effect of word block in the control condition ($F(2, 122) = 9.86, p < .001, \eta_p^2 = .14$). Repeated contrasts showed that, in line with our expectations, participants' performance in the mismatched condition improved from blocks 1 to 2 ($F(1, 48) = 7.63, p = .008, \eta_p^2 = .14$) and remained stable from blocks 2 to 3 ($F(1, 48) = 0.04, p = .848, \eta_p^2 < .01$). Similarly, though rather surprisingly, participants' performance in the control condition also improved from blocks 1 to 2 ($F(1, 61) = 18.34, p < .001, \eta_p^2 = .23$) but not from blocks 2 to 3 ($F(1, 61) = 0.53, p = .469, \eta_p^2 = .01$).

Mental Effort Table 2 shows the descriptive statistics of the mental effort ratings. Despite a trend in the means suggesting that participants in the matched condition invested less mental effort in the task, the analysis revealed no significant main effect of condition ($F(2, 163) = 2.68, p = .072, \eta_p^2 = .03$).⁵ There was a main effect of word block ($F(1.90, 309.30) = 3.59, p = .029, \eta_p^2 = .02$), but no interaction effect ($F(3.80, 309.30) = 0.32, p = .859, \eta_p^2 < 0.01$). To follow-up on the main effect of word block, repeated contrasts were conducted, which showed that invested mental effort increased from blocks 1 to 2 ($F(1, 163) = 7.60, p = .006, \eta_p^2 = .05$) and remained stable from blocks 2 to 3 ($F(1, 163) = 1.08, p = .299, \eta_p^2 = .01$).

Discussion

Consistent with our results from experiment 1a/1b, the cued recall performance in the first block was lower in the mismatched condition than in the matched condition. Moreover, and in line with our hypothesis, this negative effect of mismatching pictures compared with matching pictures disappeared when participants gained more task experience. The finding that performance in the mismatched condition improved from blocks 1 to 2 (and remained constant from blocks 2 to 3) suggests that participants adapted their study strategy and learned to ignore the mismatching pictures.

Although this should be interpreted with caution, as the effect of condition did not reach statistical significance, an explorative analysis (see footnote 5) suggested that mental effort invested in the mismatched condition was higher than in the matched condition. Furthermore, participants in all conditions started to invest somewhat more effort in block 2 compared with block 1, which then remained at the same level in block 3. Perhaps, the initial cued recall task in the first word block might have been more difficult than expected, leading participants to adjust their effort investment in the following blocks (cf. Brehm & Self, 1989; Kahneman, 1973). However, this increased effort investment only resulted into better test performance in the mismatched and control conditions, not in the matched condition (perhaps because the first block baseline score was already relatively high in this condition).

⁵ Although the effect of condition failed to reach statistical significance, we conducted an explorative analysis based on a remark by one of the reviewers, which showed that average mental effort invested in the mismatched condition ($M = 7.84, SD = 1.10$) was higher than in the matched condition ($M = 7.28, SD = 1.59$) when compared directly in a t test, $t(96.18) = 2.11, p = .038$ (two-sided), $d = 0.41$.

Experiment 3

The most important finding from experiment 2 was that task experience indeed seems to be a boundary condition for the negative effect of mismatching compared with matching pictures on learning. We assume that this finding arose because participants adapted their study strategy and started to ignore the mismatching pictures in blocks 2 and 3. Experiment 3 was set up as a direct test of this assumption. Using eye-tracking methodology, we measured participants' attention allocation to the pictures in the matched and mismatched conditions over time. We hypothesised that with increasing task experience, participants in the mismatched condition would allocate less attention to the pictures. We had no specific hypothesis about attention distribution in the matched condition. Because of the smaller sample size in experiment 3 (which was sufficiently large to address our hypotheses regarding attention allocation) and the small effect size in experiments 1 and 2, we expected to replicate the trends in performance scores from experiment 2 (i.e. mismatched < matched in block 1; increase from blocks 1 to 2 in the mismatched condition), but we did not necessarily expect these to be significant.

Method

Participants and Design Participants were 96 Dutch undergraduate university students ($M_{\text{age}} = 20.35$ years, $SD = 2.12$ years; 80 female) who participated for course credit. All participants were native Dutch speakers with normal or corrected-to-normal vision. Participants were randomly assigned to either the matched or mismatched condition. Within each condition, the picture location was counterbalanced, for half of the participants, the picture was presented above the word, and for the other half, the picture was presented underneath the word, to rule out the possibility that the hypothesised effects would be due to a particular location.

Thirteen participants (nine in the matched and four in the mismatched condition) turned out to be left handed and were therefore excluded from the analyses. The data of the remaining 83 participants (39 in the matched and 44 in the mismatched condition) was scored and analysed.

Apparatus and Materials The words and pictures were the same as those used in experiment 2, but the definitions were now presented in Dutch. The Vimmi words and pictures were presented in SMI Experiment Center (version 3.3; SensoMotoric Instruments) on a monitor with a resolution of 1680×1050 pixels (see Fig. 2). Pictures were 500×400 pixels; the centre

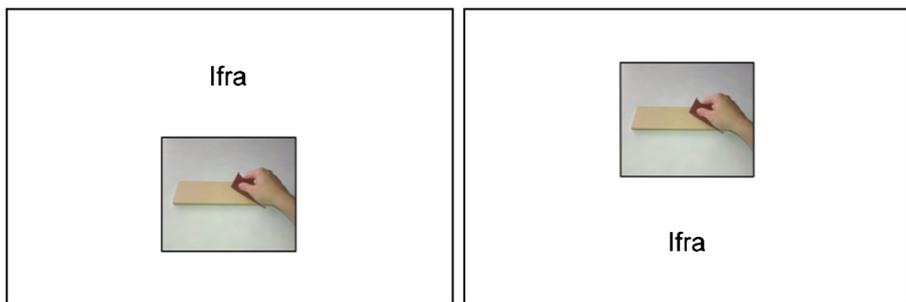


Fig. 2 Example materials. Vimmi word 'Ifra' (to sand) accompanied by the matching picture underneath (*left*) or above (*right*) the word

of the pictures was either at 656 or at 393 pixels (above or underneath the word, respectively) on the vertical axis and at 825 pixels on the horizontal axis. Participants' eye movements were recorded using an SMI RED 250 eye tracker (SensoMotoric Instruments) that recorded binocularly at 250 Hz using iView software (version 2.8; SensoMotoric Instruments) and subsequently analysed using BeGaze software (version 3.3; SensoMotoric Instruments).

Procedure The procedure was similar to that of experiment 2; participants learned the words in three blocks of five words and had to fill out the mental effort scale and complete a cued recall test after each block. At the start of the experiment, participants were seated in front of the monitor with their head positioned in a chin and forehead rest. The distance to the monitor was approximately 70 cm. After a short introduction, the eye tracker was calibrated using a 5-point calibration plus 4-point validation procedure, and participants were instructed to move as little as possible. After studying the first block of words, participants were allowed to move freely as they had to complete the recall test. After completion of the test, participants were calibrated again, after which the second block started. This procedure was repeated for block 3. The order of the blocks was alternated between participants using a Latin-square design, resulting in three lists per condition. The experiment lasted around 20 min and was administered without breaks.

Scoring and Data Analysis Due to a programming error, five participants in the mismatched condition were presented with one word in the final block twice, while five participants in the matched condition were not presented with this word at all. Two of these ten participants were excluded for being left handed, for the others, this was handled as follows. In the mismatched condition, the eye-tracking data for the second presentation of the word were discarded, and the recall score for the word was replaced with the average score in this condition to eliminate any advantages from the double presentation. In the matched condition, the eye-tracking data in block 3 were based on four words instead of five (as these participants were only presented with four words in block 3), and the recall score for the fifth word (which was missing) was replaced with the average score in the matched condition (see footnote 6).

For the eye-tracking analyses, we first checked the accuracy of calibration. We had to exclude six participants (all from the matched condition) because of inaccurate calibration (i.e. deviation from the four validation points exceeded 1° in one or more word blocks), leaving 77 participants. We then checked the tracking ratio (i.e. the percentage of time for which the eye tracker actually measured the eye movements) for each trial, and trials were excluded when this ratio was more than two standard deviations below the mean. As a result, the data of two other participants in the matched condition were excluded from further analyses as most of their trials (i.e. 13 and 14 out of the 15 trials) had a poor tracking ratio, leaving 75 participants for the eye-tracking analyses (33 in the matched and 42 in the mismatched condition). Eighteen individual trials (six in the matched and twelve in the mismatched condition), divided over 11 participants, were excluded, due to low tracking ratios in these respective trials. Taken together (i.e. including the participants whose trials were entirely excluded), 48 of 1155 trials (4.15 %) were excluded due to too low tracking ratios (see footnote 6). For the remaining 75 participants, mean calibration accuracy for blocks 1, 2 and 3 was 0.40° (SD = 0.12°), 0.40° (SD = 0.13°) and 0.40° (SD = 0.10°), respectively. Average tracking ratio based on the remaining 1107 trials was 88.82 % (SD = 10.75 %).

For the eye-tracking analyses, we defined fixations using a $40^\circ/s$ velocity threshold and a minimal duration of 100 ms (cf. Holmqvist et al. 2011). We created two areas of interest

(AoIs), one for the word (437×184 pixels) and one for the picture (536×442 pixels). The part of the screen not covered by either word or picture AoI was labelled as ‘white space’. We calculated the percentage of fixation time on the word and picture AoIs by dividing the total fixation duration (i.e. the sum of fixation duration on the word AoI, picture AoI and white space) by the fixation duration on the word or picture AoI, respectively. Finally, to explore whether the time spent looking at the mismatching pictures was indeed associated with lower test performance, we computed the correlation between the test performance and fixation duration on the picture AoI for each condition, per block.

Results

Unless otherwise specified, all data were analysed with a repeated-measure ANOVA with condition (matched or mismatched) as a between-subject factor and word block (first, second, third) as a within-subject factor.

Test Performance Table 3 shows the scores on the cued recall test per condition per block. Numerically, the mismatched condition performed worse than the matched condition, but there was no statistically significant main effect of condition ($F(1, 81) = 3.30, p = .073, \eta_p^2 = .04$). There was a main effect of word block ($F(2, 162) = 13.06, p < .001, \eta_p^2 = .14$). Repeated contrasts showed that for all participants, recall performance improved from blocks 1 to 2 ($F(1, 81) = 9.67, p = .003, \eta_p^2 = .11$) but not from blocks 2 to 3 ($F(1, 81) = 3.62, p = .061, \eta_p^2 = .04$) (see footnote 6). However, we did not find an interaction effect ($F(2, 162) = 0.23, p = .847, \eta_p^2 < .01$), probably because—in contrast to experiment 2—performance of participants in the matched condition also improved with task experience (see Table 3).

Mental Effort Table 3 shows the descriptive statistics of the mental effort scores in experiment 3. The analysis revealed no significant main effect of condition ($F(1, 81) = 0.02, p = .894, \eta_p^2 < .01$). There was a main effect of word block ($F(1.63, 131.76) = 7.39, p = .002, \eta_p^2 = .08$) but no interaction effect ($F(1.63, 131.76) = 2.20, p = .113, \eta_p^2 = .03$). To follow-up on the main effect of word block, repeated contrasts were conducted, which showed that, as in experiment 2, invested mental effort increased from blocks 1 to 2 ($F(1, 81) = 12.52, p = .001, \eta_p^2 = .13$) and remained stable from blocks 2 to 3 ($F(1, 81) = 0.08, p = .780, \eta_p^2 < .01$).

Table 3 Mean (and SD) recall performance (max. = 5) and mental effort rating (max. = 9) as a function of condition and word block in experiment 3

	Performance			Mental effort		
	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3
Matched	2.56 (1.27)	3.04 (1.18)	3.45 (1.37)	5.95 (1.38)	6.69 (1.15)	6.72 (1.32)
Mismatched	2.12 (1.28)	2.82 (1.39)	3.04 (1.61)	6.27 (1.30)	6.55 (0.98)	6.45 (1.36)
Overall	2.31 (1.29)	2.89 (1.31)	3.23 (1.51)	6.12 (1.34)	6.61 (1.06)	6.58 (1.34)

Table 4 Mean (and SD) percentage of fixation time on the picture and word AoI as a function of word block and condition

	Block 1	Block 2	Block 3
Picture			
Matched	76.79 (13.68)	67.34 (19.87)	59.97 (21.59)
Mismatched	56.00 (26.24)	35.73 (28.30)	27.58 (24.52)
Word			
Matched	22.28 (13.56)	30.93 (19.49)	37.77 (21.31)
Mismatched	42.14 (26.30)	59.15 (28.31)	69.32 (26.88)

Eye Movement Data As stated earlier, participants were presented with each definition twice, first without and then with a picture present. We analysed the eye movement data for the second part of the trials, in which a picture was present. As we only postulated hypotheses about the picture AoI, we did not analyse the data on the word AoI, but for completeness, we do provide the descriptive statistics in Table 4.

The analysis of the percentage of fixation time on the picture revealed a main effect of condition ($F(1, 73) = 41.80, p < .001, \eta_p^2 = .36$), showing that participants in the mismatched condition fixated less on the pictures than participants in the matched condition. Furthermore, we found a main effect of word block ($F(1.83, 133.72) = 32.22, p < .001, \eta_p^2 = .31$), indicating that participants allocated less attention to the picture AoI over the course of the experiment. The interaction effect did not reach statistical significance ($F(1.83, 133.72) = 2.93, p = .062, \eta_p^2 = .04$). However, the pattern of the mean fixation times across conditions seems to suggest that the decrease in fixation time on the picture AoI was larger in the mismatched condition.⁶ The descriptive statistics for the word AoI suggest that the attention that was no longer allocated to the picture was now dedicated to the word.

Exploratory Analysis: Correlation Between Picture Fixation and Test Performance To explore whether more attention to the mismatching pictures was indeed associated with lower recall performance, we computed the correlation between picture fixation time and test performance for each condition, per block. In the matched condition, we found no significant correlations between picture fixation time and test performance in any of the blocks: block 1, $r(31) = -.104$ ($p = .566$); block 2, $r(31) = .031$ ($p = .866$); and block 3, $r(31) = -.174$ ($p = .333$). In the mismatched condition, there was a negative correlation between picture fixation time and test performance, which became stronger with increasing task experience: block 1, $r(40) = -.120$ ($p = .448$); block 2, $r(40) = -.310$ ($p = .046$); and block 3, $r(40) = -.523$ ($p < .001$).

⁶ Exclusion of the eight participants who were affected by the programming error would have a minor impact on our results. Performance scores: The difference between blocks 2 and 3 is now significant ($F(1, 73) = 6.26, p = .015, \eta_p^2 = .08$); eye movement data: the interaction between condition and word block is now significant ($F(2, 130) = 4.10, p = .019, \eta_p^2 = .06$).

The exclusion of the 48 trials for which the tracking ratio was to low also had a minor impact on our results for the eye movement data: without the exclusion, the condition \times word block interaction was significant ($F(1.82, 136.27) = 3.23, p = .042, \eta_p^2 = 0.04$).

Exploratory Analysis: Picture Fixation During/After the Definition Was Spoken The pattern of results in the main analysis, indicating that fixation time on the picture AoI decreased over time, and seemed to decrease more strongly in the mismatched condition, is in line with our hypothesis that participants allocate less attention to the mismatching pictures with increasing task experience. However, due to our experimental setup (i.e. all trials lasted 11 s, while the audio was shorter, $M_{\text{length}} = 4.81$ s, $SD = 0.88$ s.), it is unclear to what extent the picture is being ignored during encoding of the verbal definition. Therefore, an exploratory analysis was performed, analysing the fixation data during the time the audio was playing, and after the audio had stopped (see Appendix Table 6 for audio length per Vimmi word definition). Because audio duration differed among words (with some definitions being longer than others), we could not compare absolute fixation times during or after the audio. Instead, we calculated the relative fixation duration by dividing the fixation duration on each AoI by the audio length (for during the audio) or non-audio length (for after the audio ended). Because we assumed that the pictures would hinder the encoding of the spoken definitions, attention allocation *during* the audio was of most interest.

We performed a 3 (word block: first, second, third) \times 2 (audio: during or after) \times 2 (condition: matched or mismatched) ANOVA on picture fixation duration. The analysis revealed a main effect of condition ($F(1, 73) = 39.45, p < .001, \eta_p^2 = .36$), showing that participants in the mismatched condition looked less at the picture AoI than participants in the matched condition (see Table 5). Furthermore, we found a main effect of word block ($F(2, 146) = 40.50, p < .001, \eta_p^2 = .36$), indicating that the relative fixation duration on the picture AoI decreased with increasing task experience. This main effect was qualified by an interaction effect between word block and condition ($F(2, 146) = 3.87, p = .023, \eta_p^2 = .05$), showing that the decrease in relative fixation duration was strongest in the mismatched condition. The main effect of audio was also significant ($F(1, 73) = 374.06, p < .001, \eta_p^2 = .84$), indicating that after the audio ended, relative fixation duration on the picture AoI decreased. The audio \times condition and word block \times audio interaction effects were not significant, respectively ($F(1, 73) = 0.85, p = .360, \eta_p^2 = .01$ and $F(2, 146) = 2.29, p = .105, \eta_p^2 = .03$). However, the word block \times audio \times condition interaction effect was significant ($F(2, 146) = 3.73, p = .026, \eta_p^2 = .05$). This suggests that the relative picture fixation difference between the matched and mismatched condition decreased more strongly over time *during* the audio than *after* the audio.

Table 5 Mean (and SD) relative fixation duration on the picture and word AoI as a function of word block and condition, split out for during audio and after audio

	Block 1		Block 2		Block 3	
	During	After	During	After	During	After
Picture						
Matched	0.70 (0.09)	0.47 (0.16)	0.62 (0.14)	0.40 (0.17)	0.58 (0.17)	0.34 (0.18)
Mismatched	0.58 (0.17)	0.32 (0.21)	0.40 (0.25)	0.19 (0.20)	0.30 (0.21)	0.14 (0.15)
Word						
Matched	0.11 (0.08)	0.22 (0.14)	0.16 (0.11)	0.27 (0.17)	0.19 (0.14)	0.32 (0.17)
Mismatched	0.22 (0.18)	0.32 (0.17)	0.36 (0.24)	0.46 (0.21)	0.46 (0.25)	0.51 (0.20)

Discussion

The eye movement data of experiment 3 are in line with our hypothesis that with increasing task experience, participants in the mismatched condition allocated less attention to the pictures. Interestingly, they seemed to pay less attention to the pictures overall (i.e. from the start) than participants in the matched condition. Furthermore, participants in the matched condition also looked less at the pictures with increasing task experience. However, this decrease was stronger in the mismatched condition (although the interaction in the main analysis did not reach statistical significance, see footnote 6). Moreover, only in the mismatched condition did we find an increasingly negative correlation between fixation time on the picture AoI and test performance: In blocks 2 and 3, less looking at the mismatching pictures was associated with higher test performance. Thus, it seems that students are indeed capable of adapting their study strategy and start to ignore extraneous information that is not relevant for—and may even interfere with—the learning task.

Similar to experiment 2, invested mental effort increased from blocks 1 to 2 and remained stable from blocks 2 to 3. Again, this is likely a result of participants' experience with the first recall test. In contrast to experiment 2, however, the increased effort investment seems to have translated into better test performance in blocks 2 and 3 in *both* conditions. Whereas we had expected this for participants in the mismatched condition, we did not expect recall performance of participants in the matched condition to improve based on experiment 2. It seems that participants in experiment 3 scored somewhat lower on the cued recall test after block 1 than participants in experiment 2 (Tables 2 and 3), leaving more room for improvement in the matched condition in experiment 3. We cannot rule out that awareness of being eye tracked had something to do with this lower performance in the first block, although very few studies have addressed effects of eye tracking on viewing behaviour (see Nasiopoulos et al., 2015; Risko & Kingstone, 2011), and we do not know of studies addressing effects on learning or performance, so this remains speculative. It should also be kept in mind that this might just be chance variation due to the smaller sample size in experiment 3.

The exploratory analysis on viewing behaviour during and after the spoken definition suggests that attending to the mismatching pictures negatively affects word learning in the first block by disrupting the encoding of the verbal definition. That is, overall, more attention was paid to the pictures while listening to the verbal definition than after, but the difference in attention to the pictures decreased more strongly over time in the mismatched than matched condition while hearing the verbal definition than after the audio had ended. In other words, the increase in recall performance in the mismatched condition over time seems to result mainly from being able to ignore the pictures while listening to the verbal definition.

General Discussion

According to the coherence principle in multimedia learning, presenting extraneous information that is not relevant for the learning task should be avoided, because it hinders learning (Mayer & Fiorella, 2014). However, based on eye-tracking research (e.g. Canham & Hegarty, 2010; Haider & Frensch, 1999), we hypothesised that task experience might be a boundary condition for the negative effect of extraneous information on learning. With increasing task experience, learners might adapt their study strategy and ignore extraneous information, which would reduce or lift the negative effects on learning. We assessed this hypothesis in a series of

three experiments. Although we did not find evidence for a coherence effect, we did establish that being presented with pictures that mismatched the action words to be learned, had a negative effect on learning outcomes compared with being presented with pictures that matched the action to be learned (experiment 1). We then confirmed that task experience nullified this negative effect (experiment 2), and finally, established that participants indeed adapted their study strategy, allocating less attention to the pictures over time, especially to the mismatching pictures (experiment 3).

Theoretical Relevance

Although we did not find a multimedia effect of matching pictures, or a coherence effect of mismatching pictures, in the sense that neither picture condition differed significantly from the control condition, we did find a negative effect of mismatching compared with matching pictures on word learning. This difference was interesting for our purpose of investigating whether task experience affects processing of mismatching pictures and thereby, learning outcomes. Our findings—as summarized at the beginning of the ‘[General Discussion](#)’—are relevant for theories of (multimedia) learning and instructional design.

First, they confirm directly (using eye tracking) the mechanism through which the presentation of extraneous information that is irrelevant for and conflicts with the task at hand initially hinders learning. That is, our findings show that such information initially captures attention (i.e. is being processed). This presumably hinders learning by drawing on valuable working memory resources that can no longer be used for processes relevant for learning (Kalyuga & Sweller, 2014, Sweller et al., 2011) and/or by actively conflicting with the to be learned information (cf. Mayer et al., 2007). The additional exploratory analyses of the eye-tracking data provide some more insight into the potential mechanisms through which this occurs. First, the reduction in attention to the extraneous information (i.e. mismatching pictures) was associated with an increase in test performance (i.e. negative correlation that became stronger over time). Second, participants seemed to suppress attention to the extraneous information over time particularly while the verbal definition was spoken (i.e. during encoding). Since the reduction in attention was associated with improved performance, this suggests that the extraneous information particularly interferes with encoding. Such direct tests of assumptions about the (attentional) mechanisms underlying multimedia principles are important, because they can support existing ideas about how and why effects on learning occur, or may generate new insights and explanations for these principles (Van Gog & Scheiter, 2010). Thus far, eye-tracking research on the coherence principle is scarce and focussed mostly on effects of presenting ‘seductive details’ (e.g. Lehman et al., 2007; Rey, 2014; Sanchez & Wiley, 2006). Seductive details are additional but irrelevant pieces of information (pictures, text) that are usually added in an attempt to increase learners’ interest or motivation, but have the (unintended) side effect that they often hamper learning. Our findings are in line with those from the eye-tracking studies on seductive details. These studies showed that seductive details hampered test performance by attracting attention, reducing the time learners spent on the relevant learning materials (Lehman et al., 2007). Furthermore, this effect was stronger for participants with lower attention control (Rey, 2014) or working memory capacity (Sanchez & Wiley, 2006).

Second, our findings not only reinforce but also extend prior multimedia research by showing that initial negative effects of extraneous information on learning disappear because

people learn to adapt their study strategy. Furthermore, this finding also extends prior eye-tracking research. It had been shown that people learn to ignore task-irrelevant information during task *performance* as a result of task experience (Haider & Frensch, 1999) or increased prior knowledge (Canham & Hegarty, 2010; Hegarty et al., 2010). Our findings demonstrate that people can also learn to ignore irrelevant information during *learning*, and that this is associated with better learning outcomes. Thus, our findings suggest that task experience may be a boundary condition for the negative effect of extraneous information on learning, because participants stop allocating attention to this information.

Third, the finding that learners seem able to adapt their study strategy to cope with the extraneous information (at least when this information always appears in the same location), without explicit instruction to do so, demonstrates the importance of studying multimedia learning principles over time, taking into account potential changes in study strategy. This focus on spontaneous adaptations of study strategies in multimedia learning is scarce, but fits well with a relatively new line of research in which participants are successfully instructed to self-manage their cognitive load during multimedia learning by changing their study strategy (e.g. Agostinho et al., 2013; Gordon et al., 2016; Roodenrys et al., 2012), and with research on training multimedia learning strategies (e.g. Bodemer et al., 2004; Mason et al., 2016; Stalbovs et al., 2015). These studies on strategy training, combined with those of the present study, show the importance of studying multimedia learning principles over time, as participants might be able to overcome suboptimal instructional design, with or without explicit instruction to do so.

Although these findings on effects of task experience may bring to mind an ‘expertise reversal effect’ (Kalyuga, 2014), they are rather different. The expertise reversal effect states that learning materials that are essential and non-redundant for novices become redundant when learners gain or have more prior knowledge, at which point they will no longer aid, and might even hinder learning. Thus, an expertise reversal effect would imply that information becomes extraneous and *starts* to hamper learning as task experience increases, whereas in our study the extraneous information *stops* hampering learning as task experience increases. Moreover, in contrast to the expertise reversal effect, participants in our study did not gain experience with (or knowledge of) the task *content* (i.e. the word definitions) over time, but with the task *presentation*.

Limitations and Future Research

A limitation of the present study lies in the materials used. In experiments 1a and 1b, we failed to establish a coherence effect, as there was no negative effect of mismatching pictures compared with a no-picture control condition. We did, however, find a small but significant and consistent negative effect on learning of mismatching compared with matching pictures, which was suitable for addressing our main hypothesis that such a negative effect might diminish with task experience. Furthermore, the learning materials used in the present study were specifically designed to test our hypothesis and as such have rather low ecological validity. Yet, pictures that are irrelevant for the learning task, and might even conflict with it, are ubiquitous in textbooks and e-learning materials, and our study provides a first indication that students can learn to ignore this extraneous information, and that when they do, it no longer hampers their learning.

Future research should investigate the process by which this occurs in more detail. For instance, it is possible that it was relatively easy to suppress attention to the pictures in the

present study, because they always appeared in the same location. An interesting question, therefore, is whether participants have truly learned that the content mismatched the definitions. In this case, they would continue to ignore the pictures even when the location would suddenly change (which would indicate a relatively conscious process of attention inhibition).

Another potential limitation of our materials is that our mismatching pictures were not only irrelevant for learning the action word definition, but might also have actively interfered with learning, by depicting another action. Therefore, it would be interesting for future research to further disentangle effects of various types of extraneous information. For instance, one could use decorative pictures that are superficially related to the action but do not conflict with the to-be-learned definition (e.g. a picture of a toolbox with a wrench, hammer and sanding paper inside when learning that *ifra* means to polish or scrape with sandpaper). An interesting question is whether it would be more difficult for participants to learn to ignore such pictures, as the lack of conflict might make it less obvious that the pictures can/should be ignored.

Moreover, it would be interesting to examine whether our findings would generalize to more ecologically valid and more complex materials that test meaningful learning instead of rote memory (as the present materials required). This would also allow for determining whether our findings would also extend to other types of coherence effects, such as the presentation of unnecessary elaborations or details when learning from illustrated texts. When the learning materials and the type of learning that is required become more complex, it may become less obvious what information is and is not relevant for learning, which may make it harder for learners to adapt their study strategy. Relatedly, the way in which extraneous information is presented may also affect how easily students can learn to ignore it. For instance, it may be harder for learners to start to ignore irrelevant text when it is integrated in pictures (cf. Bobis et al., 1993; Chandler & Sweller, 1991) or to ignore incoherent additional explanations about mechanical systems that interfere with learning the relevant materials (Mayer et al., 2007) than it is to ignore a picture (that always appears in the same location) in its entirety. At the same time, it is arguably even more important to be able to ignore extraneous information in complex materials such as illustrated texts, as such materials already pose a high demand on limited working memory resources.

Finally, it would also be relevant to address the role of individual differences in attention control or working memory (cf. Rey, 2014; Sanchez & Wiley, 2006) in future research. Such individual differences, might affect whether people learn to ignore task-irrelevant information or the rate with which they learn to do so.

Practical Relevance

Despite these limitations, our findings may prove to be relevant for educational practice. Although our findings require replication with more complex learning materials and more complex types of learning, our study provides a first indication that students can learn to ignore extraneous information, and that when they do, it no longer hampers their learning. Moreover, although learning was ‘merely’ defined in terms of rote memory in the present study, this also has its place in educational practice. When learning a new language, for example, it is important to first acquire a sufficient vocabulary before learning grammar. As Wilkins (1972, pp. 111–112) put it: ‘... while without grammar very little can be conveyed, without vocabulary nothing can be conveyed’.

Knowing that learners might be able to adapt their study strategies and ignore extraneous information with task experience, is relevant for instructional designers. That is, it is very hard for instructional designers to take into account all multimedia principles at once, because

individual learner characteristics may interact with some of the principles. For example, the split attention principle states that information from two mutually referring sources (e.g. text and picture) should be integrated rather than presented separately (Ayres & Sweller, 2014). However, whereas the integrated text may be crucial for novices' understanding, it can become redundant, and start to hamper learning for more advanced learners (Kalyuga & Sweller, 2014). Therefore, it is important to investigate whether and to what extent students themselves are able to adapt their study strategy spontaneously or after training, and we took a first step in that direction.

Conclusion

Concluding, this study suggests that task experience may be a boundary condition for the negative effect of extraneous information on learning, because experience allows learners to change their study strategies to cope with (i.e. ignore) information that interferes with their learning. Future research should establish whether this boundary condition generalizes to more complex learning and other types of extraneous information. If so, this is relevant knowledge for multimedia learning theories as well as for instructional designers.

Acknowledgements This research was funded by a Research Excellence Initiative grant from the Erasmus University Rotterdam awarded to the Educational Psychology section. The authors would like to thank Susan Ravensbergen and Kirsten Versijde for their help during the data collection.

Appendix

Table 6 Vimmi words used in experiments 2 and 3 and audio duration of the definitions for experiment 3

	Vimmi word	Definition of	Duration (s)
1	Repo	To write	4.14
2	Lapo	To iron	4.56
3	Rifa	To saw	6.54
4	Dawu	To inject	4.52
5	Dupi	To hammer	5.28
6	Kune	To paint	3.22
7	Luko	To scrub	4.56
8	Bepa	To screw	5.88
9	Redu	To stir	6.07
10	Buto	To cut	4.82
11	Lozu	To pour	3.89
12	Tari	To erase	5.22
13	Kori	To stamp	3.79
14	Lefa	To beat	5.24
15	Ifra	To sand	4.43

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agostinho, S., Tindall-Ford, S., & Roodenrys, K. (2013). Adaptive diagrams: handing control over to the learner to manage split-attention online. *Computers and Education*, *64*, 52–62. doi:10.1016/j.compedu.2013.01.007.
- Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: a comparison between emotion, abstract, and concrete words. *American Journal of Psychology*, *117*, 389–410. doi:10.2307/4149007.
- Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 206–226). New York: Cambridge University Press. doi:10.1017/CBO9781139547369.011.
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. doi:10.1016/S1364-6613(00)01538-2.
- Barouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 59–80). Oxford: Oxford University Press.
- Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary-school geometry task. *Learning and Instruction*, *3*, 1–21. doi:10.1016/S0959-4752(09)80002-9.
- Bodemer, D., Plöetznner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction*, *14*, 325–341. doi:10.1016/j.learninstruc.2004.06.006.
- Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*, 109–131. doi:10.1146/annurev.ps.40.020189.000545.
- Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 174–206). New York: Cambridge University Press. doi:10.1017/CBO9781139547369.010.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*, 155–166. doi:10.1016/j.learninstruc.2009.02.014.
- Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*, 351–367. doi:10.1037/a0015854.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332. doi:10.1207/s1532690xci0804_2.
- Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: evidence from eye movements. *Memory and Cognition*, *29*, 1146–1152. doi:10.3758/BF03206384.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
- Cowan, N. (1995). *Attention and memory: an integrated framework*. New York: Oxford University Press.
- Craig, S., Gholson, B., & Driscoll, D. (2002). Animated pedagogical agents in multimedia educational environments: effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, *94*, 428–434. doi:10.1037/0022-0663.94.2.428.
- De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). When left is not right: handedness effects on learning object-manipulation words using pictures with left- or right-handed first-person perspectives. *Psychological Science*, *24*, 1–7. doi:10.1177/0956797613498908.
- Farley, A. P., Ramonda, K., & Liu, X. (2014). The concreteness effect and the bilingual lexicon: the impact of visual stimuli attachment on meaning recall of abstract L2 words. *Language Teaching Research*, *16*, 449–466. doi:10.1177/1362168812436910.
- Gordon, C., Tindall-Ford, S., Agostinho, S., & Paas, F. (2016). Learning from instructor-managed and self-managed split-attention materials. *Applied Cognitive Psychology*, *30*, 1–9. doi:10.1002/acp.3159.
- Haider, H., & Fensch, P. A. (1999). Eye movement during skill acquisition: more evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 172–190. doi:10.1037/0278-7393.25.1.172.
- Hald, L. A., Van Den Hurk, L., & Bekkering, H. (2015). Learning verbs more effectively through meaning congruent action animations. *Learning and Instruction*, *39*, 107–122. doi:10.1016/j.learninstruc.2015.05.010.

- 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. doi:10.1037/0022-0663.90.3.414.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301–307. doi:10.1016/S0896-6273(03)00838-9.
- Hegarty, M., Canham, M., & Fabrikant, S. I. (2010). Thinking about the weather: how display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 37–53. doi:10.1037/a0017683.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: a comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- 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. doi:10.1016/j.learninstruc.2009.02.019.
- Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice-Hall.
- Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 576–597). New York: Cambridge University Press. doi:10.1017/CBO9781139547369.028.
- Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 247–262). New York: Cambridge University Press. doi:10.1017/CBO9781139547369.10.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. doi:10.2307/2529310.
- Lehman, S., Schraw, G., McCrudden, M., & Hartly, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology, 32*, 568–587. doi:10.1016/j.cedpsych.2006.07.002.
- Macedonia, M., & Knösche, T. R. (2011). Body in mind: how gestures empower foreign language learning. *Mind, Brain, and Education, 5*, 196–211. doi:10.1111/j.1751-228X.2011.01129.x.
- Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology. Advance online publication* doi:10.1111/bjet.12271.
- Mayer, R. E. (Ed.) (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). New York: Cambridge University Press. doi:10.1017/CBO9781139547369.
- Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). New York: Cambridge University Press. doi:10.1017/CBO9781139547369.015.
- Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*, 380–386. doi:10.1037/0022-0663.100.2.380.
- Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52. doi:10.1207/S15326985EP3801_6.
- Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*, 64–73. doi:10.1037/0022-0663.88.1.64.
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: when presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187–198. doi:10.1037/0022-0663.93.1.187.
- Mayer, R. E., DeLeeuw, K. E., & Ayres, P. (2007). Creating retroactive and proactive interference in multimedia learning. *Applied Cognitive Psychology, 21*, 795–809. doi:10.1002/acp.1350.
- Miller, G. (1956). The magic number seven, plus or minus two: some limits to our capacity for processing information. *Psychological Review, 63*, 81–97.
- Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: the case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*, 117–125. doi:10.1037/0022-0663.92.1.117.
- Nasiopoulos, E., Risko, E. F., Foulsham, T., & Kingstone, A. (2015). Wearable computing: will it make people prosocial? *British Journal of Psychology, 106*, 209–216. doi:10.1111/bjop.12080.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive load approach. *Journal of Educational Psychology, 84*, 429–434. doi:10.1037/0022-0663.84.4.429.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71. doi:10.1207/S15326985EP3801_8.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical Turk. *Judgment and Decision making, 5*, 411–419.
- Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21*, 793–797. doi:10.1111/j.1460-9568.2005.03900.x.

- Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory and Cognition*, *10*, 97–102. doi:10.3758/BF03209210.
- Rey, G. D. (2014). Seductive details and attention distraction—an eye tracker experiment. *Computer in Human Behavior*, *32*, 133–144. doi:10.1016/j.chb.2013.11.017.
- Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Attention, Perception and Psychophysics*, *73*, 291–296. doi:10.3758/s13414-010-0042-1.
- Roodenrys, K., Agostinho, S., Roodenrys, S., & Chandler, P. (2012). Managing one's own cognitive load when evidence of split attention is present. *Applied Cognitive Psychology*, *26*, 878–886. doi:10.1002/acp.2889.
- Sadoski, M. (2005). A dual coding view of vocabulary learning. *Reading & Writing Quarterly*, *21*, 221–238. doi:10.1080/10573560590949359.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory and Cognition*, *34*, 344–355. doi:10.3758/BF03193412.
- Shen, H. H. (2010). Imagery and verbal coding approaches in Chinese vocabulary instruction. *Language Teaching Research*, *14*, 485–499. doi:10.1177/1362168810375370.
- Smith, B. D., Stahl, N., & Neil, J. (1987). The effect of imagery instruction on vocabulary development. *Journal of College Reading and Learning*, *20*, 131–137. doi:10.1080/10790195.1987.10849909.
- Stalbovs, K., Scheiter, K., & Gerjets, P. (2015). Implementation intentions during multimedia learning: using if-then plans to facilitate cognitive processing. *Learning and Instruction*, *35*, 1–15. doi:10.1016/j.learninstruc.2014.09.002.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295–312. doi:10.1016/0959-4752(94)90003-5.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer. doi:10.1007/978-1-4419-8126-4.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281. doi:10.1162/0898929053124965.
- Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, *20*, 95–99. doi:10.1016/j.learninstruc.2009.02.009.
- 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. doi:10.1002/acp.1112.
- Wilkins, D. (1972). *Linguistics in language teaching*. London: Arnold.
- Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: an undesired desirable difficulty? *Journal of Educational Psychology*, *105*, 266–277. doi:10.1037/a0031971.