Effects of Observing and Creating Video Modeling Examples on Cognitive and Motivational Aspects of Learning

Vincent Hoogerheide

Cover design by Nadia Chien
Layout by Vincent Hoogerheide
Printed by Ridderprint BV

ISBN: 978-94-6299-406-5

The research presented in this dissertation was funded by Kennisnet.

Effects of Observing and Creating Video Modeling Examples on Cognitive and Motivational Aspects of Learning

Effecten van het observeren en creëren van videomodelvoorbeelden op cognitieve en motivationale aspecten van het leren

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam

op gezag van de rector magnificus
Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
Donderdag 20 oktober 2016 om 11:30 uur

door
Vincent Hoogerheide
geboren te Rotterdam

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotoren
Prof.dr. T. van Gog
Prof.dr. S. M. M. Loyens

Overige leden
Prof.dr. L. Kester
Prof.dr. F. Paas
Prof.dr. A. Renkl

Contents

Chapter 1

Introduction

Imagine a young adult male whose car has a flat tire that needs to be replaced. If he has little if any knowledge of how to tackle this problem, then trying to replace his car's tire without any form of guidance from others would likely be a very effortful, time consuming, and potentially dangerous activity. After a while, that may cause him to stop trying and give up. Even if he would figure out the necessary steps to solve the problem on his own, the chances are high that he would have spent a lot less time and effort if he had been able to watch a video of an expert who demonstrated the necessary steps first.

This everyday example illustrates what decades of educational research has shown (e.g., Sweller & Cooper, 1985), namely that for novice learners, instruction consisting of studying examples of how to solve a problem or complete a task, is very effective and efficient for learning new skills (for reviews, see Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Sweller, Ayres, & Kalyuga, 2011; Sweller, Van Merriënboer, & Paas, 1998; Van Gog & Rummel, 2010). Research inspired by cognitive theories, such as cognitive load theory (Sweller, 1988; Sweller et al., 1998; Sweller et al., 2011), has demonstrated the effectiveness and efficiency of studying text-based *worked examples* that provide a problem statement and a step-by-step explanation of how to solve that problem for learners with low prior knowledge. Research inspired by social-cognitive theories such as social learning theory (Bandura, 1977, 1986) has shown that observational learning from *modeling examples* in which a person (the so-called 'model') explains and/or demonstrates how to solve a problem is an effective way of acquiring new problem-solving skills and to enhance the belief that learners have in their own capabilities to perform the modeled task (i.e., self-efficacy; Bandura, 1997).

For many centuries, modeling examples could only be observed in person, for instance in master-apprenticeship learning situations (Collins, Brown, & Newman, 1989), and these so-called *live* modeling examples are still being used (e.g., Bjerrum, Hilberg, Van Gog, Charles, & Eika, 2013; Ryalls, Gul, & Ryalls, 2000). The development of the video tape recorder enabled observational learning from modeling examples on *video* (e.g., Koran, Snow, & McDonald, 1971; Schunk, Hanson, & Cox, 1987). Nowadays, video modeling examples are often digital (e.g., Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013a, 2013b; Schwan & Riempp, 2004) and accessible via online learning environments or websites, usually allowing for viewing on demand. So the use of video modeling examples is not a new development, but what is quite new is the scale on which this nowadays occurs. These days, learners in many parts of the world have access to countless video modeling examples created by experts from all over the globe (e.g., on websites such as Google Videos, Vimeo, and

YouTube) and they can view these on mobile devices at a time and location of their preference (Van der Meij & Van der Meij, 2014).

Moreover, with modern technology, the distinction between worked examples and modeling examples is blurring (Van Gog & Rummel, 2010). For instance, Khan Academy video examples (www.khanacademy.org) show a model writing out an example of how to solve a math, science, or economics problem step by step while explaining it in a voice over. The model is not visible in these videos, only what s/he is writing. Other video examples show the model's actions in an online environment (mouse clicks, typing) with or without a voice over (e.g., McLaren et al, 2010; Van Gog, 2011; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009a). Again, the model is not visible in such videos, only his or her actions. A third example are animated modeling examples, which do tend to show the 'model', but in animated form (i.e., animated pedagogical agents), often with a human voice over (e.g., Atkinson, 2002; Wouters, Paas, & Van Merriënboer, 2009). Animated pedagogical agents are frequently embedded in online learning environments (e.g., Frechette & Moreno, 2010; Liew, Tan, & Jayothisa, 2013; Rosenberg-Kima, Baylor, Plant, & Doerr, 2008) in which they can simulate human instructional roles such as an expert or mentor (Baylor & Kim, 2005; Johnson, Rickel, & Lester, 2000; Kim & Baylor, 2015).

Considering the effectiveness of video modeling examples and the fact that they are nowadays so easily created and distributed, it is no surprise that they are also increasingly being used for formal learning purposes. Students and teachers have also started creating video modeling examples themselves. For instance, in the context of the Dutch project Leerlingen voor Leerlingen (Students for Students; http://lvoorl.wikiwijs.nl), an online database was developed that is filled with student-created video modeling examples that their peers can use in class or at home to prepare for class or while making homework. The website www.wiskundeacademie.nl provides an example of a database filled with teacher-created video modeling examples that secondary education students of all educational levels can use to learn math. The most extensive online database comprised of video modeling examples is Khan Academy (www.khanacademy.org), a website that was founded in 2006 and although it originally focused mainly on mathematics, the website now provides over 5.500 video modeling examples covering a wide range of domains and gets over 10 million unique viewers per month.

Despite this widespread use and creation of video modeling examples in recent years, there is a lack of specific guidelines that would help students and teachers decide how to design their videos so that motivation and learning outcomes of viewers are optimized (Van Gog, 2013; of course there are well-established multimedia design guidelines that provide more general guidance, see Mayer, 2014). Therefore, the aim of the studies in the first part of this dissertation was to further our empirical knowledge of *whether and how the design of video modeling examples would affect students' motivation and learning.* Another open question, which is relevant for teachers who use the creation of video examples as a learning activity for their students, is whether the act of explaining a learning task to others on video would have a benefit for learning. In that case, the creation of video examples might have a double benefit, not only for the observer, but also for the model. Therefore, the aim of the studies in the second part of this dissertation was to investigate *whether providing explanations on video would yield potential benefits for learning and motivation*.

Part I: Effects of Example Design on Motivation and Learning

As already mentioned, example-based instruction has proven to be very effective for novice learners (see reviews by Atkinson et al., 2000; Renkl, 2014; Sweller et al., 1998; Sweller et al., 2011; Van Gog & Rummel, 2010) and has been studied from both a cognitive perspective and a social cognitive perspective. The main foci and findings from both perspectives are discussed below, resulting in the research questions addressed in Part I of this dissertation.

Cognitive Research: Worked Examples

Research conducted from a cognitive perspective has predominantly focused on the effects of studying worked examples in comparison to problem solving, and has shown that studying worked examples (or studying worked examples alternated with practicing problem solving) is more effective and efficient than practice problem solving (e.g., Van Gog, Kester, & Paas, 2011). By effectiveness, it is meant that learners who study worked examples achieve higher learning outcomes than students solving practice problems (e.g., Cooper & Sweller, 1987; Nievelstein, Van Gog, Van Dijck, & Boshuizen, 2013; Paas, 1992; Paas & Van Merriënboer, 1994). With efficiency, it is meant that learners who study worked examples often achieve higher (or equal) learning outcomes with less investment of study time (e.g., Cooper & Sweller, 1987; Paas & Van Merriënboer, 1994) and/or effort during study than students solving practice problems (e.g., Paas & Van Merriënboer, 1994; Van Gog et al., 2011). While most studies compared studying worked examples to practicing problem solving without any guidance (which, some have argued, seems to be a weak control condition; Koedinger & Aleven, 2007), more recent studies have demonstrated that studying worked examples is more effective and/or efficient compared to *tutored* problem solving with feedback and hints (e.g., McLaren, Van Gog, Ganoe, Karabinos, & Yaron, 2016; Schwonke, Renkl, Krieg, Wittwer, Aleven, & Salden, 2009).

Cognitive load theory (CLT) explains the effectiveness of example study in terms of the working memory load imposed by practice problem solving vs. example study. The issue with practice problem solving when learners lack knowledge of how to proceed, is that it forces learners to using weak problem-solving strategies that contribute very little to learning, that is, to building a cognitive schema of how to solve the problem (Sweller, 1988). Learners often engage in a trial-and-error strategy (randomly trying out steps), or continuously search for ways to reduce the gap between the problem state and goal state (i.e., means-end analysis), which imposes a heavy cognitive load and is very time consuming, and consequently inefficient for learning (Sweller, 1988; Sweller & Levine, 1982). Worked example study prevents learners from using weak problem-solving strategies like trial-and-error or means-ends analysis because no problem solving is required. Instead, all available cognitive capacity can be devoted to studying the worked-out procedure. This ensures that learners are enabled to, as Renkl (2014) describes: 'acquire knowledge about problem states, operators, and the consequences of the application of operators integrated into schemas that are applicable for later problem solving' (p. 4). Learners may also abstract general rules from worked examples (e.g., Anderson & Fincham, 1994) and consequently may be able to transfer their newly acquired skills to solving similar problems and/or novel problems for which adaptation of the problem-solving procedure is necessary (e.g., Paas, 1992; Paas & Van Merriënboer, 1994).

After early work by Sweller and Cooper (1985) and Cooper and Sweller (1987) it soon became clear, however, that studying worked examples is only better for novices' learning than practicing problem solving when the worked examples are well designed (Tarmizi & Sweller, 1988). Tarmizi and Sweller showed in their initial experiments that cognitive load increased and test performance decreased when learners were required to split their attention between a diagram and written solution steps that had to be integrated with one another in order to understand the solution procedure (this has become known as the 'split-attention effect'; for reviews, see Ayres & Sweller, 2014; Sweller et al., 2011). In the final two experiments of their study, Tarmizi and Sweller showed that the negative effects of split attention –increased cognitive load and decreased test performance– were remedied when students were presented with an integrated diagram (i.e., the written solution steps integrated into the diagram). Another way to avoid split attention is to use spoken text rather than written text (i.e., the modality effect; Ginns, 2005; Kühl, Scheiter, Gerjets, & Edelmann, 2011; Mayer & Moreno, 1998; Mousavi, Low & Sweller, 1995).

Social-cognitive Research: Modeling Examples

Research conducted from a social-cognitive perspective has not typically been concerned with the effectiveness of modeling examples in comparison to other instructional strategies. It has mainly investigated how the design characteristics of examples, in particular the type of model used, affect students' motivation (especially self-efficacy, i.e., people's beliefs about their capabilities to attain certain levels of performance; Bandura, 1994, 1997; Schunk, 1987). While it is well-established that studying modeling examples can enhance novice learners' self-efficacy (Bandura, 1981, Schunk, 1984), effects of worked example study on motivational variables are often ignored in worked examples research (Van Gog & Rummel, 2010). Effects of example study on motivational variables such as self-efficacy, is important for learning though: Self-efficacy affects how people feel, think, and act and as such plays a crucial role in learning (see Bandura, 1986, 1997; Honicke & Broadbent, 2016; Pajares, 2006; Schwarzer, 1992). Compared to low self-efficacy learners, those higher in self-efficacy tend to experience less anxiety and stress, have enhanced motivation, perform more challenging tasks, use more cognitive and metacognitive strategies, and set higher goals and stick to those goals more (Schwarzer, 1992). Moreover, the high efficacious invest more effort and persist more in a task, and in case of setbacks, recover sooner. Although self-efficacy is correlated with students' actual ability, the effects of self-efficacy do not seem to be moderated by students' ability (e.g., Collins, 1982).

When developing video modeling examples, various design choices have to be made, the most salient of which are: 1) whether or not to show the model in the example, and 2) who the model should be. According to the model-observer similarity hypothesis (MOS; Bandura, 1994; Schunk, 1987), these choices matter greatly, as students' learning and motivation may differ depending on characteristics of the model, and it is possible that these effects will be stronger when the model is seen in the video. In his social learning theory, Bandura (1977, 1986) postulated that cognitive symbolic representations (cf. cognitive schemas) can last beyond the modeling example situation when four qualifications are met. That is, for learning to occur, one firstly has to pay attention to the relevant behavior of the model. Secondly, this relevant behavior has to be encoded and stored into memory to ensure that, thirdly, the behavior can be reproduced. Finally, a learner has to be motivated to emulate the modeled behavior, that is, to try and produce or even surpass the same modeled behavior.

The MOS hypothesis (Bandura, 1984; Schunk, 1987; Schunk & Zimmerman, 2007) postulated that the extent to which learners perceive themselves to be similar to the model moderates the effectiveness of learning from (video) modeling examples, because modeling enables social comparison (Berger, 1977; Johnson & Lammers, 2012). Or as Bandura (1994) stated:

> 'The impact of modeling on perceived self-efficacy is strongly influenced by perceived similarity to the models. The greater the assumed similarity the more persuasive are the models' successes and failures. If people see the models as very different from themselves their perceived self-efficacy is not much influenced by the models' behavior and the results its produces.' (p.72)

Moreover, learners may also allocate more attention and/or be more attracted to a more similar model (Berscheid & Walster, 1969), and similarity can help learners to determine whether behavior is appropriate and to form outcome expectations (Schunk, 1987). MOS should especially be important when learners are still novices because novice learners are more inclined to engage in social comparison (Buunk, Zurriaga, Gonzalez-Roma, & Subirats, 2003).

In sum, the degree to which modeling examples enhance cognitive and affective aspects of learning may depend on characteristics of the model and the degree to which the model is present in the example.

Effects of Model Presence

There are several differences between worked examples, modeling examples, and examples that combine features of both that may affect motivation and learning. First, the degree to which a model is present in examples varies across the different example forms, and it has been argued that hearing and/or seeing the model provide learners with social cues that evoke them to connect the video content to their own personal self, thereby increasing their self-efficacy and potentially their learning outcomes (Mayer, 2014). Concerning the traditional example forms, video modeling examples contain a much stronger model presence than worked examples because learners can both see and hear the model. As already mentioned, the traditional distinction between worked examples and modeling examples is fading (Van Gog & Rummel, 2010). In many of these emergent example forms, the model still narrates the instructional text, but is no longer visible in the video; only (the consequences of) the model's actions are. For instance, examples may present a model narrating an instructional text while a computer screen is shown on which s/he is writing out the problem-solving steps (e.g., www.khanacademy.org), illustrating these steps using pictures and/or slides (e.g., Leahy & Sweller, 2011), or performing the steps by mouse-clicking or typing (e.g., McLaren et al., 2008; McLaren et al., 2016; Van Gog, 2011; Van Gog et al., 2009a). If the effectiveness of examples depends on the degree to which learners can identify with the model, then the degree to which the model is present in examples might also affect learning and motivation.

Note, though, that whether full presence of the model would help students' learning or hinder it, is currently being investigated. Even though observing others is a natural process, it has been hypothesized that seeing the model's face might perhaps attract students' attention at the detriment of attention to what the model is doing. Research has shown that we tend to look at the face of those who provide a spoken explanation as much as 95% of the time in conversation situations (Gullberg & Holmqvist, 2006). Recent findings of Van Gog, Verveer, and Verveer (2014), however, indicate that in learning situations with demonstration videos, the model's face still draws a substantial amount of attention (although much less than in a conversation situation, ca. 20% on average), but can actually foster performance compared to not seeing the model's face, possibly because seeing the model's gaze direction might be helpful. Findings of Kizilcec, Bailenson, and Gomez (2015) and Koran, Snow, and McDonald (1971) suggest that for instructional videos that focus more on providing explanations rather than a demonstration, the presence of the model's face draws attention but in such a way that learning is unaffected. This is consistent with Mayer's image principle, which states that people do not necessary learn more from multimedia when a speaker is visible on the screen (see Mayer, 2014). With regard to self-efficacy though, Rosenberg-Kima et al. (2008) showed that adding an animated pedagogical agent who lip-synched and gestured enhanced students' self-efficacy more compared to a voice-only condition. Research on this question with human modeling examples is scarce, however.

Second, differences among examples in the degree of model presence also connect to other multimedia design principles. For instance, the model's narration of the instructional text could be an advantage over written worked examples, not only because it provides a higher degree of model presence, but also because the modality effect shows that spoken text combined with visual information, enables learners to spread the cognitive load over multiple modalities (i.e., Ginns, 2005; Leahy & Sweller, 2011; Sweller et al., 2011). When written text is combined with visual information, in contrast, learners may be forced to split their attention between the visual information and written text (i.e., the split-attention effect; Ayres & Sweller, 2014; Sweller et al., 2011). Potentially, there are also disadvantages to using spoken text, however. For example, the transient information effect has shown that, relative to written text, spoken text can increase cognitive load and hamper learning because of transience (Singh et al., 2012; Sweller et al., 2011).

Given the paucity of comparisons of different example types, it is as yet unclear whether the degree of model presence, and the design consequences this has, would affect self-efficacy and learning outcomes. Therefore, the question is addressed in this dissertation whether and how the presence of the model in examples affects cognitive and motivational aspects of learning, by comparing written worked examples in which the model is absent, narrated examples in which the model can only be heard, and modeling examples in which the model is heard and seen (Chapter 2).

Effects of Model-observer Similarity

The model-observer similarity (MOS) hypothesis has inspired researchers to examine whether the effectiveness of modeling examples depends on how similar learners perceive themselves to be relative to the model in terms of factors such as age (e.g., Hicks, 1965; Rodriguez Buritica, Eppinger, Schuck, Heekeren, & Shu-Chen Li, 2015; Weeks et al., 2005; Zmyj, Aschersleben, Prinz, & Daum, 2012), expertise (e.g., Braaksma et al., 2002; Sonnenschein & Whitehurst, 1980; Schunk & Hanson, 1985), gender (e.g., Linek, Gerjets, & Scheiter, 2010; Schunk, Hanson, & Cox, 1987), and to a lesser extent also background (e.g., Rosekrans, 1967). Overall, however, research investigating MOS effects has led to mixed results.

One possible explanation for the mixed findings is that similarity may be most important when it is a cue for how appropriate the modeled behavior is (Schunk, 1987). This could explain why, for instance, Bandura, Ross, and Ross (1963) found that compared to observing a female model, observing a male model display aggressive behavior towards a doll led to more imitation for boys, whereas Schunk et al. (1987) found no differences for children learning how to solve fraction problems from a male vs. female model; the children were likely too young to evaluate mathematics as more appropriate for males than females (Ceci, Ginther, Kahn, & Williams, 2014; Steffens, Jelenec, & Noack, 2010).

Another possible explanation is that in most MOS studies, manipulations of the similarity between learners and the model across conditions also affected the content of the example, that is, what the model's said or did, which makes it difficult to determine whether beneficial effects on learning outcomes and self-efficacy were due to the difference in content of the examples or the (perceived) similarity with the model. For example, in quite a few studies learning from a coping model who first makes errors but corrects these later on, was contrasted to a mastery model who performs the task faultlessly from the beginning (e.g., Huang, in press; Kitsantas, Zimmerman, & Cleary, 2000; Schunk et al., 1987). In another line of research, learning from a low expertise model was contrasted to learning from a high expertise model (e.g., Becker & Gliden, 1979; Sonnenschein & Whitehurst, 1980), or learning from text-based examples created by advanced peer student models was compared to learning from text-based examples created by experts (Boekhout, Van Gog, Van de Wiel, Gerards-Last, & Geraets, 2010; Lachner & Nückles, 2015). Thus, because the performance of the model differed across conditions in these studies, any effects on learning and self-efficacy might have been due to the content of the examples, making it difficult to draw conclusions regarding effects of MOS.

In this dissertation, therefore, the question is addressed whether cognitive and motivational aspects of learning are enhanced when learners study video modeling examples with a more similar model relative to a more dissimilar model, in terms of gender (Chapter 3) or age and expertise (Chapter 4), *when the content of the examples is kept identical across conditions*. In addition, in the study presented in Chapter 5, the open question was addressed whether MOS would affect learning from text-based worked examples. That is, would perceived similarity with the one who created the example in terms of gender or age and expertise matter? If perceived similarity indeed positively affects learning from worked and/or video modeling examples when the content is identical, then online learning environments could adaptively present examples with a model that matches student characteristics to optimize learning and self-efficacy.

Part II: Effects of Explaining to Fictitious Others on Video

While it has been established that observational learning from video modeling examples that display a student model can be effective (e.g., Braaksma et al., 2002; Schunk & Hanson, 1985), it is still an open question whether asking students to create video modeling examples for each other might be an effective learning activity. When students are instructed to act as a model in a video example as a learning activity, they first have to study the learning materials with the intention of being able to explain the content later on, and then they have to explain the learned materials to (non-present) others in front of a video camera. Both processes could be hypothesized to benefit learning outcomes.

Effects of Explanation Intention / Teaching Expectancy

Studying with an explanation intention is different than studying with the intention of completing a test, which is how students normally study. It has been hypothesized that compared to studying for a test, studying with the intention to explain the learning materials to others (cf. teaching expectancy) stimulates study processes such as active studying (Benware & Deci, 1984), and self-explaining (Chi, De Leeuw, Chiu, & LaVancher, 1994; Renkl, 1997), as well the use of metacognitive strategies such as comprehension monitoring and planning (Muis, Psaradellis, Chevrier, Leo, & Lajoie, 2015; Roscoe, 2014). It may also evoke learners to focus more on interpreting and integrating new knowledge with old knowledge rather than on memorizing facts (Benware & Deci, 1984). Research on the effectiveness of studying with an explanation expectancy has, however, produced a mixed pattern of results: some studies found positive effects on learning outcomes compared to studying for a test (e.g., Fiorella & Mayer, 2013, 2014; Nestojko, Bui, Kornell, & Bjork, 2014), while others did not (e.g., Ehly, Keith, & Bratton, 1987; Renkl, 1995). One possible explanation for the mixed findings could be that students benefit more from studying with an explanation intention when they have a certain degree of prior knowledge and/or experience with studying in such a manner (Nestojko et al., 2014). If learners do not have sufficient prior knowledge or experience, they may experience more anxiety (Ross & DiVesta, 1976) and/or a decrease in intrinsic motivation (Renkl, 1995), which could negatively affect the learning process and/or learning outcomes. Another reason could be that an explanation intention leads to immediate benefits that decay or disappear after a delay (Fiorella & Mayer, 2013, 2014).

Effects of Explaining

Actually providing explanations on video can be expected to be beneficial for learning because research has shown that generating explanations can be a powerful instructional strategy (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Fiorella & Mayer, 2015; Leinhardt, 2001; Lombrozo, 2012; Ploetzner, Dillenbourg, Preier, & Traim, 1999; Richey & Nokes-Malach, 2015; Wilie & Chi, 2014). However, most explanation-studies have examined effects of self-explaining or explaining in interactive contexts (Ploetzner et al., 1999; Richey & Nokes-Malach, 2015).

Self-explaining can be defined as self-generated explanations of instructional materials. Students may spontaneously generate self-explanations (e.g., Renkl, 1997) or be prompted to do so (e.g., Chi, de Leeuw, Chiu, & LaVancher, 1994). According to Richey and Nokes-Malach (2015), self-explaining can:

> '… encourage learners to identify and elaborate on the critical features of problems, including the underlying principles (Atkinson, Derry, Renkl, & Wortham, 2003; Chi & VanLehn, 1991), the conditions for applying those principles (Chi et al., 1989), and

the logic and subgoals for applying them (Catrambone, 1998; Crowley & Siegler, 1999).' (p. 23)

Because learners recognize and understand these features, they are more likely to transfer their knowledge to new problems (Atkinson, Renkl, & Merrill, 2003).

Generating explanations in interactive situations can also be conducive to learning, for instance during small group discussions (Cohen, 1994; Johnson, Johnson, & Smith, 2007) or when tutoring (Cohen, Kulik, & Kulik, 1982). The finding that not only those who receive tutoring benefit greatly, but also the tutors themselves, is especially interesting because tutors also prepare learning materials with the intention of explaining followed by actually explaining to others. Explaining in interactive situations could also be beneficial because other students may, for example, provide explanations themselves, ask questions, or point out faults (Ploetzner et al., 1999; Webb, 1989). Because these studies typically do not isolate the benefits of providing explanations, it is still an open question to what degree the interactive elements contribute to the effectiveness of tutoring. Findings of Coleman, Brown, and Rivkin (1997) suggest, however, that explaining in an interactive situation is beneficial even when the interactive elements are controlled for. They showed that relative to self-explaining, providing instructional explanations to someone in the same room fostered measures of deep learning.

Explaining to non-present, fictitious others on video entails a different situation than explaining to oneself or explaining in interactive situations. Because the explanations are aimed at an imagined audience, the explainer receives no verbal or non-verbal feedback regarding the quality of the explanation (e.g., the extent to which the listeners understand it). While explaining to non-present, fictitious others may seem like an artificial situation, it is actually common in online (learning) environments nowadays: people generate explanations for others in various asynchronous situations, such as discussion forums (Andresen, 2009), weblectures (e.g., Chen & Wu, 2015; Day & Foley, 2006; Traphagan, Kucsera, & Kishi, 2010; Zhang, Zhou, Briggs, & Nunamaker, 2006), and demonstration videos that are typically recorded without the presence of an audience, using a webcam or a digital camera (e.g., Ouwehand, Van Gog, & Paas, 2015; Orús, Barlés, Belache, Casaló, Fraj, & Gurrea, 2016; Spires, Hervey, Morris, & Stelpflug, 2012).

Spires et al. (2012) examined effects of creating explanation videos. The authors argued that the process of preparing to explain and of explaining during video creation fostered secondary education students' motivation and learning outcomes; however, the lack of a control condition makes it difficult to draw any definite conclusions from that study. Fiorella and Mayer (2013, 2014) did experimentally control for study intention and explaining. They consistently found that for university students learning about the Doppler effect, studying with an explanation expectancy led to immediate benefits on a comprehension test relative to studying for a test. However, this benefit decayed on a delayed comprehension test unless it was coupled with actually explaining to non-present fictitious other students on video for 5 minutes. In some of those experiments, the effects of explaining might have resulted from the additional time students were given to do so, but in one experiment, the effect was still found even when time was controlled for by giving the other condition additional study time (Fiorella & Mayer, 2014; Experiment 2).

In this dissertation (Chapters 6 and 7), the question is addressed whether the two processes that would be involved in acting as a peer model (i.e., studying with an explanation intention and providing explanations to non-present others on video) would foster learning and transfer. In the experiments reported in Chapter 7 it was also investigated whether beneficial effects would be found when students explain to non-present, fictitious others in writing, which would be easier to implement as a learning activity in the classroom.

Overview of the Studies in this Dissertation

This dissertation is divided into two parts. **Part I**, *Effects of Example Design on Motivation and Learning*, contains four Chapters (2, 3, 4, and 5) describing experimental studies in which it was investigated whether and how the design of examples affects cognitive and motivational aspects of learning. **Part II**, *Effects of Explaining To Fictitious Others on Video*, contains two Chapters (6 and 7) in which it was experimentally examined whether learning outcomes are fostered by the two processes involved in acting as a model, namely studying learning materials with the intention of explaining them to someone else and then actually providing explanations to fictitious others on video.

Part I: Effects of Example Design on Motivation and Learning

In **Chapter 2**, two experiments are described in which the effects of video modeling examples with a visible model, video modeling examples without a visible model, and worked examples were compared. Students learned how to solve probability calculation problems by observing either two examples (Experiment 1) or one example (Experiment 2) under one of the three conditions. Effects on their self-efficacy and perceived competence were assessed, as well as effects on their learning outcomes in relation to invested mental effort (higher/equal performance reached with equal/less effort investment in one condition relative to the others, would indicate higher efficiency of the learning process; Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008). Effects on learning enjoyment and the willingness to receive similar instruction in the future were explored.

In **Chapter 3**, an experiment is described in which it was examined whether model-observer similarity (MOS) in terms of gender would affect learning from video modeling examples that were otherwise identical in content. Male and female students learned how to solve probability calculation problems by observing a video modeling example with either a male or female model. Effects on learning outcomes, effort investment, self-efficacy, and perceived competence were assessed. Based on the MOS hypothesis, it would be expected that effects on learning and motivation would be more enhanced when studying a same-gender model than an opposite-gender model. Effects on students' learning enjoyment and the willingness to receive similar instruction in the future were explored.

The aim of the experiment presented in **Chapter 4** was to investigate whether (perceived) similarity with the model in terms of age and perceived expertise would affect students' learning and motivation when the example content is otherwise kept identical across conditions. Students first observed a short video in which an adult or a peer model introduced herself as having low or high expertise. Students then learned how to solve electrical circuit problems by observing two video modeling examples. Effects on learning and effort investment were examined, as well as on self-efficacy and perceived competence. Based on the MOS hypothesis, it could be expected that students would benefit most from a low expertise peer model. Effects on learning enjoyment and students' assessment of the quality of explanations in the video modeling examples were explored.

In **Chapter 5**, two experiments are described in which the open question was addressed whether MOS in terms of gender, or in terms of age and expertise, would affect learning from text-based worked examples. Male and female students were led to believe via a short story and pictures that the worked examples were created by a male or female peer student (Experiment 1) or a peer student or teacher (Experiment 2). Notably, the content of the examples was kept identical across conditions. In both experiments, students learned how to solve electrical circuit problems with four worked examples. Effects on learning, mental effort, self-efficacy, and perceived competence were assessed. Effects on students' impression of the quality of the examples were explored in both experiments, and effects on learning enjoyment were explored in Experiment 2.

Part II: Effects of Explaining to Fictitious Others on Video

In **Chapter 6**, two experiments are described in which it was examined whether the processes that would be involved in acting as a peer model (i.e., studying with an explanation intention and providing explanations to non-present others on video) would foster learning outcomes for secondary education students (Experiment 1) and university students (Experiment 2). Students read a text on syllogistic reasoning with the intention of completing a test (one group) or with the intention of explaining the content to others (two groups). One of the groups of students who had studied with an explanation intention, subsequently explained the learning materials to (non-present) other students via a webcam. The other explanation intention group and the test intention group restudied the text. Effects on learning and transfer, and effort investment were investigated, as well as effects on self-efficacy and perceived competence explored.

The two experiments reported in **Chapter 7** extended the findings from the study reported in Chapter 6, by examining whether providing explanations to fictitious others in *writing* would foster learning outcomes for students learning about syllogistic reasoning. In Experiment 1, students first read a text on syllogistic reasoning with a test intention or an explanation intention, and then restudied the text or provided explanations to non-present others in writing. In Experiment 2, the effects of explaining in writing, explaining on video, and studying with a test intention were compared. Effects on learning and transfer and effort investment were assessed. Effects on perceived competence were explored (Experiment 2).

The final chapter (**Chapter 8**) presents a summary of the main findings and a discussion of both theoretical and practical implications.

Part I

Effects of Example Design on Motivation and Learning

Chapter 2

Comparing the effects of worked examples and modeling examples on learning

This chapter has been published as:

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior, 41*, 80-91. doi:10.1016/j.chb.2014.09.013

Abstract

Example-based learning is an effective instructional strategy for students with low prior knowledge, and is increasingly being used in online learning environments. However, examples can take many different forms and little is known about whether and how form affects learning outcomes. Therefore, this study investigated whether worked examples and modeling examples with and without a visible model would be equally effective in fostering learning of a problem-solving task. In Experiment 1, secondary education students (N = 78) learned how to solve a probability calculation problem by watching two videos that, depending on the assigned condition, provided worked examples (written text, pictures of problem states), modeling examples with a visible model (spoken text, a demonstration of the task), or modeling examples without a visible model (spoken text, pictures of problem states). Results showed that all three conditions were equally effective at fostering learning, near transfer, effort reduction, self-efficacy, and perceived competence. Experiment 2 (N = 134) replicated these results with a younger student population that only studied one example. These findings suggest that the format of examples does not affect learning outcomes for this task; future research should investigate whether this would generalize to other problem-solving tasks.

Introduction

Example-based learning is an effective and efficient instructional strategy for teaching novice learners new problem-solving skills (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Sweller, Ayres, & Kalyuga, 2011; Sweller, Van Merriënboer, & Paas, 1998; Van Gog & Rummel, 2010), and for enhancing learners' self-efficacy (Bandura, 1997; Crippen & Earl, 2007). Research has traditionally focused on two forms of example-based learning (for reviews: Renkl 2014; Van Gog & Rummel, 2010). Studies conducted from a cognitive perspective (e.g., cognitive load theory; Sweller, 1988; Sweller et al., 1998; Sweller et al, 2011) have mostly examined text-based *worked examples* that explain how to complete a task. Studies conducted from a social-cognitive perspective (e.g., social learning theory; Bandura, 1977, 1986; cognitive apprenticeship: Collins, Brown, & Newman, 1989) have mostly examined *modeling examples*, in which an expert, teacher, or peer student –the model– demonstrates and (often) explains how to complete a task. Modeling examples can be both live (e.g., Bjerrum, Hilberg, Van Gog, Charles, & Eika, 2013; Krautter et al., 2011) or on video (e.g., Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013a; Schunk, Hanson, & Cox, 1987; Schwan & Riempp, 2004; Van Gog, Verveer, & Verveer, 2014). In many video modeling examples, the model is (partly) visible in the video (e.g., Atienza, Balaguer, & Garcia-Merita, 1998; Schunk et al., 1987; Van Gog et al., 2014; Xeroulis, Park, Moulton, Reznick, LeBlanc, & Dubrowski, 2007). These are henceforth referred to as 'modeling examples with a visible model'.

With technological advances, new forms of video examples are being developed that combine features of worked and modeling examples and are widely used in online learning environments. For instance, one may hear the model explaining the task while seeing the model's computer screen on which he or she is performing the problem-solving steps (e.g., McLaren, Lim, & Koedinger, 2008; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009a; Van Gog, 2011), writing out the steps (e.g., www.khanacademy.org), or illustrating the steps with pictures or slides (e.g., Leahy & Sweller, 2011). Consequently, learners observing such modeling examples can only hear the models but do not see them. These examples are henceforth referred to as 'modeling examples without a visible model'.

Not only do worked examples, modeling examples with a visible model, and modeling examples without a visible model vary in terms of social aspects, but as we will explain below, their design may also have potential advantages and disadvantages in terms of cognitive aspects of learning. However, despite the substantial amount of research on example-based instruction, and the fact that worked and modeling examples are increasingly being used in online learning environments or on computers, tablets, or smart boards in the classroom, little is known about how the social and cognitive consequences of the various example types would affect the learning process and learning outcomes. The present study therefore compares the effects of worked examples and modeling examples with and without a visible model on cognitive factors such as learning (i.e., applying what has been learned to new tasks of the same type that differ in surface features but have the same structural features), near transfer (i.e., applying what has been learned to new tasks of the same type that are slightly more complex; i.e., differ in surface features and partly in structural features) and far transfer (i.e., applying what has been learned to new tasks of a different type; i.e., different structural features/solution procedures), and motivational factors of learning such as self-efficacy when learning a problem-solving task.

Effects of Example Design on Cognitive and Motivational Factors

Worked examples and modeling examples with and without a visible model have different design characteristics that may be more or less conducive to learning when seen from a cognitive load perspective. Cognitive load theory (Sweller, 1988; Sweller et al., 1998; Sweller et al., 2011) argues that it is important to design instruction in such a way that ineffective working memory load (i.e., extraneous load) is reduced so that more working memory resources are available for processes that are effective for learning (i.e., germane load).

There are several known causes of extraneous load. For instance, the split attention effect (for reviews, see Ayres & Sweller, 2005; Sweller et al., 2011) shows that separately presenting mutually referring written text and graphical information requires learners to split their attention between both information sources and to mentally integrate them, which can hinder learning. This split in attention can be (partially) remedied by providing learners with an integrated format (i.e., text mapped onto corresponding parts of the picture; e.g., Chandler & Sweller, 1992), by visually guiding learners' attention to the corresponding elements (i.e., color-coding cues; e.g., Ozcelik, Karakus, Kursun, & Cagiltay, 2009), or by replacing written text with spoken text (i.e., making use of the modality effect; Kühl, Scheiter, Gerjets, & Edelmann, 2011; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995). A potential benefit of using multiple modalities is that information processing is divided over different working memory systems (Baddeley, 1986), which helps reduce working memory load, because learners can devote their visual attention to the pictures while listening to the spoken text (Ginns, 2005; Leahy & Sweller, 2011; Sweller et al., 2011). Note that, with regard to example design, replacing written text with spoken text could result in a modeling example in which the model is not visible, but learners only hear the model's explanation of how to perform the task, supported by pictures of the problem states or slides.

The use of spoken text can, however, also have a potential drawback, because it leads to information transience. The transient information effect shows that, because of the temporal limits of working memory, spoken text can lead to an increase of extraneous load and can therefore hamper learning outcomes compared to written text, which is often permanently available (Singh, Marcus, & Ayres, 2012; Sweller et al., 2011). Yet transient information is not always detrimental to learning. For instance, when learners already have a substantial amount of prior knowledge, or when the instructional text is short in length or low in complexity, then learners should be able to easily process the learning materials regardless of information transience (Leahy & Sweller, 2011). Consequently, the negative effects of transient information on cognitive load and learning can be remedied by dividing lengthy sections into smaller segments (Mayer & Chandler, 2001; Spanjers, Wouters, Van Gog, & Van Merriënboer, 2011; Wong, Leahy, Marcus, & Sweller, 2012).

With regard to modeling examples with a visible model, the *model* might possibly present a source of split attention. Although it is a natural process to observe a demonstration by another person and we learn human-movement tasks (e.g., assembly, origami, knot-tying) more effectively by observing dynamic visualizations compared to static ones (Höffler & Leutner, 2007; Van Gog, Paas, Marcus, Ayres, & Sweller, 2009b), it could be argued that in tasks to which human movement is not inherent, seeing the model might draw attention away from the task because we tend to automatically focus our attention on other people's –and even animated pedagogical agents'- faces (see Van Gog et al., 2014). In addition, just like certain features of animated pedagogical agents (i.e., virtual characters that simulate human instructional roles) have been hypothesized to be able to draw attention away from the learning materials (Moreno & Flowerday, 2006; Walker, Sproull, & Subramani, 1994), a model may do the same because he or she provides learners with additional, redundant information compared to written text, such as the model's tone of voice, clothes, and task-irrelevant movements.

The social cues that are more strongly available in modeling examples (regardless of whether or not the model is visible) than in written worked examples may also affect motivational aspects of learning, such as self-efficacy or perceived competence. Because modeling allows for social comparison (Berger, 1977; Johnson & Lammers, 2012), observing another person successfully explain and perform a task increases the likelihood that learners believe that they can perform the task as well (Bandura, 1981, Schunk, 1984). According to the cognitive theory of multimedia learning, social cues such as the model's voice and visibility allow for this social comparison to take place because they stimulate learners to link the presented content to their own personal self (Mayer, 2005). A study with animated pedagogical agents by Rosenberg-Kima, Baylor, Plant, and Doer (2008) showed that adding an animated pedagogical agent that lip-synchronized the instructional text and occasionally gestured, enhanced self-efficacy compared to a voice-only condition. If this would also apply to video modeling examples, then seeing the model might positively affect self-efficacy.

Research on worked examples has largely ignored motivational effects of example-based learning (Van Gog & Rummel, 2010), presumably because social cues are less prominent in written examples. However, when comparing the effectiveness of worked examples and modeling examples, it is important to not only focus on cognitive effects, but to also take into account effects on self-efficacy and perceived competence. That is, self-efficacy seems to have significant bearing on factors such as academic motivation, study behaviour, and learning outcomes (Bandura, 1997; Bong & Skaalvik, 2003; Schunk, 2001). The closely related construct of perceived competence, which refers to more broad perceptions and knowledge that are consequently also more stable and enduring (Bong & Skaalvik, 2003; Hughes, Galbraith, & White, 2011; Klassen & Usher, 2010), has also been shown to have significant influence on academic motivation and learning outcomes (Bong & Skaalvik, 2003; Harter, 1990; Ma & Kishor, 1997).

To conclude, it is an open question whether and how the different design characteristics of worked examples and modeling examples with and without a visible model would affect the learning process (e.g., effort investment) and learning outcomes. Regarding self-efficacy and perceived competence, the literature reviewed above suggests that modeling examples can be expected to lead to higher self-efficacy and perceived competence gains because they have a stronger social component.

The Present Study

The present study investigated whether it is more effective to study worked examples, modeling examples with a visible model, or modeling examples without a visible model in terms of learning outcomes, cognitive load, and the confidence in one's own capabilities to perform a task (i.e., self-efficacy and perceived competence). In two experiments, secondary education students were taught how to solve probability calculation problems (with replacement and order important) with the help of two videos (See Figure 1 for snapshots taken from the videos) that showed either a modeling example with a visible model (spoken text, demonstration of the task), a modeling example without a visible model (spoken text, pictures of the problem-states), or a worked example (written text, pictures of the problem-states). Several design factors were kept constant to ensure that potential effects on social and cognitive aspects of learning could be explained by the difference between written/spoken text and by presence/absence of the model: (1) all examples were *videos* that presented the same, *didactical* procedure[1] on how to successfully solve a probability calculation problem without replacement in which order is important, (2) to keep time constant, worked examples were divided into 5 segments that were presented for the same length of time as in the modeling examples (see Figure 1 for an example of a segment). Effects were measured both on an immediate posttest and after a delay of one week (i.e., delayed posttest), because different patterns of results might be found over time (cf. Töpper, Glaser, & Schwan, 2014).

Hypotheses

The effectiveness (higher posttest performance) and efficiency (higher posttest performance attained with less investment of effort) of example-based learning has been well documented (for reviews, see Atkinson et al., 2000; Renkl, 2014; Sweller et al., 1998; Van Gog & Rummel, 2010). We therefore expect that the examples presented in all three conditions would be beneficial for learning, as measured by an increase in performance, reduction in mental effort, and increase in self-efficacy and perceived competence from pretest to posttest.

The more interesting, yet open question, however, is whether the design differences among these three example forms would differentially affect test performance and investment of mental effort (an indicator of cognitive load) during learning. With regard to self-efficacy and perceived competence, it can be expected given the degrees of social presence that observing modeling examples with a visible model would be more effective than those without a visible model, which would be more effective than worked examples.

However, increases in students' confidence in their own abilities might not necessarily be a positive outcome when they would become overconfident, because overconfidence negatively affects students' regulation of the learning process (Dunlosky & Rawson, 2012; Rhodes & Tauber, 2011; Thiede, Anderson, & Therriault, 2003). Therefore, we explored how accurate participants' beliefs about their own abilities were, by asking them to provide a judgment of learning of the predicted performance on the posttest and

[1] Note that a didactical procedure refers to how a student should learn to solve a problem, not necessarily how an expert would solve it, as experts tend to skip steps (Blessing & Anderson, 1996; Kalyuga & Sweller, 2004).

relating that to their actual performance (i.e., judgment of learning accuracy). The more accurate students' judgment of learning, the better they are able to regulate their study time and restudy choices (Kornell & Metcalfe, 2006; Thiede et al., 2003). We also explored whether differences would occur in how enjoyable participants found the three forms of example-based instruction to be, and in their preference to receive similar instruction in the future, which would be important for learners' use of online examples during self-study (Yi & Hwang, 2003). Lastly, to determine whether example-based learning is more effective for some learners than for others, we explored whether secondary education students' processing and regulation strategies can explain a part of this variance in learning outcomes.

Figure 1. A snapshot from the first video example watched in the modeling example with a visible model condition (left), the modeling example without a visible model condition (middle), and the worked example condition (right).

Experiment 1

Method

Participants and Design. Participants were 78 Dutch students (M^{age} = 14.29, *SD* = 0.79; 39 boys, 39 girls) in their second or third year of pre-university education. This is the highest level of secondary education in the Netherlands and has a total duration of six years. They were from two schools located in middle class neighbourhoods and although we do not have the exact details, participants were predominantly of Caucasian origin. The experiment consisted of 4 phases: pretest, learning phase, immediate posttest, and delayed posttest. Participants were randomly allocated to the Worked Example Condition (WE; *n* = 26), the Modeling Example With Visible Model Condition (ME+; *n* = 27), or the Modeling Example Without Visible Model Condition (ME-; *n* = 25). The experiment was conducted at a point in the curriculum when probability calculation had not yet been taught.

Materials

All materials were presented using the Qualtrics platform (www.qualtrics.com).

Pretest and Posttests. Three parallel test versions were created, all of which presented eight probability calculation problems. The eight probability calculation problems across the three parallel tests differed in surface characteristics (cover stories) but were structurally equivalent. Each test version consisted of 6 problems of the same problem category as was demonstrated in the learning phase, to be able to measure learning (i.e., applying what has been learned to same-domain tasks that differ in surface features but have the same structural features): four probability calculation problems without replacement and with order important (two with low numbers: 1/4*1/3=1/12 and 1/6*1/5=1/30; two with high numbers: 1/11*1/10=1/110 and 1/14*1/13=1/182), and two probability calculation problems without replacement and with order important and more events than shown in the learning phase (one with three events: 1/6*1/5*1/4=1/120; one with four events: 1/8*1/7*1/6*1/5=1/1680). Furthermore, to be able to measure near transfer, two probability calculation problems *with* replacement and order important were used (both low numbers: 1/6*1/6=1/36, 1/5*1/5=1/25). As such, the near transfer problems asked participants to apply what they had learned to more complex tasks that were both partly different in surface features (i.e., a different cover story) and in one of the structural features (i.e., probability calculation with replacement instead of without replacement, which was briefly addressed in the video examples). All eight problems required participants to both write down the calculation and the correct answer.

Learning Phase. Two examples were created that described how to solve a probability calculation problem (without replacement and with order important). The problem-state of the first example was as follows: “The scouting staff brings 4 coloured balls for the cub scouts to play with. There is a red ball, a blue ball, a yellow ball, and a green ball. The cub scouts get to choose a ball one by one and prefer every colour equally. What is the chance that the red ball gets picked first and the green ball second?” Subsequently, the remainder of the example explained how to solve the problem step-by-step. The problem-state of the second example was: “The tutor of a class brings candy bars to hand out to his students. There is one Mars, one Twix, one Snickers, one Kit-Kat, and one Bounty. The students get to choose a candy bar one by one and prefer each candy bar equally. What is the chance that the Mars gets picked first, and the Bounty gets picked second?” The remainder of the example then described stepwise how to solve the problem correctly. Both examples addressed that it was an example of a probability calculation task without replacement in which order mattered, and briefly mentioned what would happen in case of an example of a probability problem with replacement.

Condition specific videos were created for both the balls and candy bars examples. To ensure that participants in all three conditions spent an equal amount of time studying the example, time was kept constant across conditions (balls example: 176 s.; candy bars example: 169 s.). The video examples for the ME+ Condition presented a young Caucasian woman in her early twenties wearing a neutral (black) outfit while sitting behind a desk with the example-specific materials (as such, the model resembled the majority of students in terms of ethnicity; see e.g., Baylor & Kim, 2004). An example is shown in Figure 1. In the begin state of the problem, these materials lay together on a platter. The model then used these materials to illustrate the problem-solving steps while explaining how to solve the problem. For instance, while explaining the first event (i.e., chance that the red ball or Mars bar is picked first) she picked up the first item (i.e., red ball or Mars bar), held it in the air, and then placed it on one side of the platter. The video examples for the ME- Condition consisted of the audio (i.e., narration by the model) taken from the ME+ Condition and a full-screen picture that depicted the change that occurred at the end of the problem step (see Figure 1). These pictures resembled screenshots taken of the most crucial moments in the ME+ Condition, however, without a visible model. Three pictures were used: one picture depicted the begin state of the problem (all items grouped), one the first event (one item placed to the side), and another the second and final event (two items placed to the side). Near the end of the model's narration, the calculation she mentioned that was needed to attain the final answer was overlaid as subtitling on the video (ME+) or the picture (ME-). The video examples for the WE Condition presented the same text as the model narrated in the ME+ and ME- Condition in written form (see Figure 1). The text was split into 5 segments, each supported by the same pictures that were used in the ME- Condition (only the desk and objects), which was placed at the bottom of the screen. The first segment presented the begin state of the problem (first picture), the second segment the first event (second picture), and the third to fifth segment the second and final event (third picture). While segmenting introduced a degree of transience compared to written worked examples presented on a single page, it ensured that study time of the problem steps was identical to the ME+ and ME- Condition.

Mental Effort. After each probability calculation problem on the pretest, immediate, and delayed posttest, and after both video examples during the learning phase, participants were asked to indicate how much effort they invested in studying the example or solving the problem, on the 9-point rating scale developed by Paas (1992), which ranges from (1) very, very low effort to (9) very, very high effort. Average effort invested in the learning and near transfer items was computed separately, as well as average effort invested in the examples.

Self-efficacy and Perceived Competence. Self-efficacy was measured by asking participants to rate on a 9-point rating scale (ranging from (1) very, very unconfident to (9) very, very confident) to what extent participants believed that they had mastered the skill of probability calculation. Perceived competence was measured with an adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996). This is a four item questionnaire, which, on a scale of 1 (not at all true) to 7 (very true), asks people to rate the questions: “I am capable of learning the material in this course”, “I feel confident in my ability to learn this material”, “I feel able to meet the challenge of performing well in this course”, and “I am able to achieve my goals in this course”. We altered the scale by removing the last question on the topic of goals because this question did not apply to the present study. For the remaining questions, we rephrased the wording of ‘course’ to ‘probability calculation problems’. For each measurement, the final score of self-efficacy and perceived competence was computed by averaging the ratings of the three questions.

Judgment of Learning. Participants were also asked to assess on a scale of 0 to 8 how many of the 8 probability calculation problems they believed they would answer correctly if presented with a test.

Instruction Evaluation. To assess how participants felt about learning with the video examples, they were asked after observing the two WE, ME+, or ME- video examples to indicate on a scale of 0 to 10 how enjoyable watching the videos was for them, and to what extent they would prefer receiving instruction this way more often.

Self-regulation and Processing. To measure students’ regulation and processing strategies, Part A of Vermunt’s Inventory of Learning Patterns (ILS; Vermunt, 1994) was administered. This part distinguishes between three processing strategies: A) ‘deep processing’ (11 items), which is a combination of the learning activities relating, restructuring, and critical processing, B) ‘stepwise processing’ (11 items), which combines the learning activities analyzing and memorizing, and C) ‘concrete processing’ (5 items), which combines the learning activities concretizing and applying. Furthermore, part A of the ILS measures three regulatory processes: A) ‘self-regulation’ (11 items), which measures the degree that participants regulate their own study behavior, B) ‘external regulation’ (11 items), which measures the degree that participants let their study behavior be regulated by external sources such as teachers, and C) ‘lack of regulation’ (5 items), which measures students’ inability to regulate their learning as well as the support they receive from their learning environment. Each item asks participants to indicate to what extent participants believe the question applies to them. This is measured on a scale of 1 (“I never use/do this”) to 5 (“I almost always use/do this”). Because the scale was designed for use in higher education, the phrasing of some questions was simplified to ensure that each item was comprehensible for the secondary education students.

Procedure

The study was run in two sessions. Both sessions took place in the computer lab of the participants’ school with an entire class of students present. Participants were randomly distributed among the three conditions prior to the experiment.

Prior to the first session (ca. 50 min.), A4-papers were randomly distributed over the available seats in the computer lab, containing participants’ name as well as a link to the condition specific Qualtrics questionnaire. The Qualtrics questionnaire presented 4 ‘blocks’ of questions. Prior to each block, the experimenter provided participants with a plenary verbal instruction after which they individually completed all the items in that particular block. Block 1 presented a general demographic questionnaire for which participants received 90 seconds. Next, participants received 10 minutes for block 2, which contained the pretest (eight probability calculation problems and mental effort ratings), followed by self-efficacy and perceived competence measurements. The experimenter verbally instructed participants to write down both the calculation and the correct answer to the pretest problems. The third block presented the two video examples and mental effort ratings after each example, followed by the instruction evaluation questions. The fourth block first presented self-efficacy and perceived competence, and judgment of learning measurements, followed by the immediate posttest (eight probability calculation problems and mental effort ratings). Half of the participants within each condition received posttest version A and the other half posttest version B as immediate test. It was re-emphasized that participants should write down both the calculation and the correct answer. Participants received 20 minutes to complete the immediate posttest.

The second session took place exactly 7 days later and lasted circa 50 minutes. This time the Qualtrics questionnaire presented two blocks of questions, block 5 and 6. The experimenter first gave a plenary introduction, after which participants could individually work through those blocks of questions at their own pace. The fifth block presented self-efficacy and perceived competence, and judgment of learning measurements, followed by the delayed posttest (eight probability calculation problems and mental effort ratings). Participants who had received version A on the immediate posttest, now received version B, and vice versa. The sixth block asked participants to complete part A of the ILS by

Vermunt (1994). Afterwards, the experimenter explained the nature of the experiment and provided participants with the correct answers to the test items.

Due to technical difficulties, part A of the ILS measuring self-regulation and processing strategies could only be administered for half of the participants. Because it would not be meaningful to analyze the results for only these participants, the ILS results were omitted from this experiment.

Data Analysis

A research assistant scored all test answers based on a standard developed by the authors. To measure the reliability of the ratings, two raters independently scored 10% of the tests. The intra-class correlation coefficient was .985. Because this was high, the assistant continued to score all data and his scores were used for the analyses. Participants were assigned max. 2 points for each probability calculation problem, 1 point per answer (1 point for the correct answer, 0 points for an incorrect or missing answer) and 1 point per calculation (1 point for the correct calculation, 0.5 points for a partially correct calculation, and 0 points for an incorrect or missing calculation). There was one exception to the scoring system: because it was assumed that participants who answered a probability calculation problem correctly but did not write down the calculation would know how to correctly calculate the answer, both points were granted in case of a correct answer. A maximum of 12 points could be earned on the six probability calculation problems that measured learning, and a maximum of 4 points on the two probability calculation problems that measured near transfer.

A measure of judgment of learning accuracy was then computed by matching the judgment of learning measurement (how many out of 8 problems similar to the pretest participants expected to solve correctly) to their actual test performance (how many they did solve correctly), by multiplying participants' judgment of learning by 2 and then subtracting their actual test performance.

Results

Sixteen participants were removed from all analyses because their pretest scores indicated too much prior knowledge (scores equal to or higher than 50%), leaving 62 participants in total (WE: n = 21; ME+: n = 22; ME-: n = 19). Six participants were absent during the Delayed Posttest (two in each condition). Their missing scores on the Delayed Posttest were replaced with the condition average. All analyses were conducted using a Repeated Measures ANOVA (RM-ANOVA) with Test Moment (Pretest, Immediate Posttest, Delayed Posttest) as within-subject factor and Instruction Condition (WE, ME+, ME-) as between-subject factor unless otherwise specified. In case of violations of the sphericity assumption, a Huynh Feldt correction was applied for estimates of sphericity greater than 0.75, and a Greenhouse-Geisser correction for estimates equal to or lower than 0.75 (Field, 2009). All follow-up tests were conducted with the Bonferroni correction.

Test performance and mental effort scores on the test are shown in Table 1, and self-efficacy, perceived competence, judgment of learning, and judgment of learning accuracy ratings in Table 2. Because the analyses of performance on the learning and near transfer items showed exactly the same pattern of results, we only report the analyses of the total test performance and invested mental effort scores here for the sake of brevity.

Table 1.
Mean (SD) of Learning, Near Transfer, and Total Test Scores and Mental Effort per Condition in Experiment 1

	Pretest			Immediate Posttest			Delayed Posttest		
	WE	ME+	ME-	WE	ME+	ME-	WE	ME+	ME-
Test Scores									
Learning	0.43	0.46	0.55	11.81	10.61	10.71	11.39	10.60	11.59
	(1.02)	(0.92)	(1.13)	(0.60)	(3.02)	(2.09)	(1.08)	(3.03)	(1.00)
Near	1.19	1.18	1.18	2.60	2.93	2.97	2.42	3.03	3.41
Transfer	(1.62)	(1.67)	(1.58)	(1.66)	(1.29)	(1.64)	(1.49)	(1.43)	(1.11)
Total	1.62	1.63	1.73	14.40	13.55	13.68	13.81	13.63	15.00
	(1.92)	(2.22)	(2.29)	(1.93)	(3.86)	(3.37)	(1.94)	(4.21)	(1.33)
Mental Effort									
Learning	6.14	5.48	6.09	2.75	2.79	3.29	3.23	2.84	3.07
	(1.22)	(1.28)	(1.32)	(0.80)	(1.08)	(1.13)	(1.38)	(1.05)	(0.97)
Near	5.95	5.77	6.24	3.02	2.71	3.08	3.10	2.80	2.97
Transfer	(1.80)	(1.55)	(0.98)	(1.28)	(1.14)	(1.26)	(1.29)	(1.06)	(1.25)
Test	6.11	5.56	6.31	2.84	2.77	3.21	3.20	2.82	3.04
Total	(1.28)	(1.27)	(1.08)	(0.88)	(1.07)	(1.10)	(1.41)	(1.04)	(1.06)

Test Performance

The RM-ANOVA on the total test scores showed no main effect of Instruction Condition, $F < 1$, but there was a significant main effect of Test Moment, $F(1.76, 104.02) = 758.89$, $p < .001$, $\eta_p^2 = .928$. Participants performed significantly better on the Immediate Posttest and Delayed Posttest than on the Pretest, *p*s < .001. Moreover, participants' scores on the Immediate and Delayed Posttest did not significantly differ, $p = 1.000$, which shows that performance stayed about the same after a one-week interval. There was no interaction effect between Test Moment and Instruction Condition, $F(3.53, 104.02) = 1.33$, $p = .268$.

Mental Effort

A RM-ANOVA on invested mental effort during example study was conducted. Participants in the WE (*M* = 2.00, *SE* = 0.23), ME+ (*M* = 2.57, *SE* = 0.23), and ME- Condition (*M* = 1.95, *SD* = 0.25) did not differ in the amount of effort invested in learning the content of the videos, $F(2, 59) = 2.13$, $p = .121$.

A RM-ANOVA on average invested mental effort across all 8 probability calculation problems showed no main effect of Instruction Condition, $F(2, 59) = 1.19$, $p = .310$. However, a main effect of Test Moment was found, $F(1.38, 81.67) = 246.27$, $p < .001$, $\eta_p^2 = .807$. Significantly less mental effort was invested in solving the probability calculation problems on the Immediate Posttest and Delayed Posttest than on the Pretest, *p*s < .001. No statistically significant difference was found between the Immediate and Delayed Posttest, $p = 1.000$, nor an interaction between Test Moment and Instruction Condition, $F < 1$.

Self-efficacy and Perceived Competence

No main effect of Instruction Condition on the self-efficacy scores was found, $F < 1$. There was, however, a main effect of Test Moment, $F(1.55, 91.58) = 177.48$, $p < .001$, $\eta_p^2 = .747$. Participants estimated their self-efficacy to be significantly higher on the Immediate and Delayed Posttest than on the Pretest, *p*s < .001. Moreover, participants rated their self-efficacy to be significantly higher at the Immediate Posttest measurement than one week later at the Delayed Posttest measurement, $p = .001$. Lastly, no interaction was found between Test Moment and Instruction Condition, $F < 1$.

For perceived competence, the RM-ANOVA showed no main effect of Instruction Condition, $F < 1$. There was, however, a main effect of Test Moment, $F(1.89, 111.63) = 156.31$, $p < .001$, $\eta_p^2 = .726$. Participants' perceived competence significantly increased from Pretest to Immediate Posttest as a result of the learning phase, $p < .001$. It decreased again from Immediate to Delayed Posttest, $p < .001$, although on the Delayed Posttest participants' perceived competence was still significantly higher than on the Pretest measurement, $p < .001$. There was no significant interaction between Test Moment and Instruction Condition, $F(3.78, 111.63) = 1.57$, $p = .191$.

Judgment of Learning

For judgment of learning, there was no main effect of Instruction Condition, $F < 1$, nor a main effect of Test Moment $F(1, 59) = 2.71$, $p = .105$. There was no significant interaction effect between Test Moment and Instruction Condition, $F < 1$.

With respect to accuracy of the judgments of learning, the RM-ANOVA with Test Moment (Immediate Posttest, Delayed Posttest) as between-subject factor showed no main effect of Instruction Condition, nor of Test Moment, *F*s < 1. One sample *t*-tests showed that the predicted test performance accuracy at both measurements were significantly different from zero, *t*s < .001. Thus, participants significantly underestimated (see Table 2) their skills on the Immediate and Delayed Posttest. In addition, there was no significant interaction effect, $F < 1$.

Table 2.
Mean (SD) of Self-efficacy, Perceived Competence, Judgment of Learning (JOL), and Judgment of Learning Accuracy Scores per Condition in Experiment 1

	Pretest			Immediate Posttest			Delayed Posttest		
	WE	ME+	ME-	WE	ME+	ME-	WE	ME+	ME-
Self-efficacy	3.00 (1.55)	3.41 (1.50)	2.84 (1.61)	6.62 (0.97)	6.55 (1.06)	6.79 (0.98)	6.00 (1.22)	6.15 (0.94)	6.24 (1.23)
Perceived Competence	3.57 (1.35)	3.47 (1.24)	3.19 (1.19)	5.76 (0.73)	5.47 (1.00)	5.91 (0.67)	5.14 (1.09)	5.05 (0.81)	5.37 (1.05)
JOL				5.86 (1.20)	6.00 (1.23)	5.68 (1.06)	6.05 (1.47)	6.30 (1.31)	6.00 (1.29)
JOL Accuracy				-2.69 (3.08)	-1.55 (4.24)	-2.32 (4.42)	-1.71 (3.54)	-1.03 (4.39)	-3.00 (2.58)

Instruction Evaluation

To assess how participants experienced watching the video examples, an ANOVA was conducted. With respect to how enjoyable watching the video examples was for them, participants seemed to rate the videos in the WE (M = 5.76, SE = 0.50) and the ME- Condition (M = 5.37, SE = 0.53) as more enjoyable than the ME+ Condition (M = 4.32, SE = 0.49); however, this difference was not statistically significant ($F(2, 59) = 2.29$, p = .110). Regarding the extent to which participants preferred to receive instruction in a similar manner in the future, there was a significant difference among conditions though, $F(2,59) = 3.78$, p = .029, η_p^2 = .114. The WE Condition had a significantly higher preference to receive similar instruction again (M = 7.14, SE = 0.45) compared to the ME+ Condition (M = 5.50, SE = 0.44), p = .032. There was no significant difference between the WE Condition and the ME- Condition (M = 6.74, SE = 0.47), p = 1.000, nor between both ME Conditions, p = .173.

Discussion

The results of this experiment showed that worked examples and modeling examples with and without a visible model were equally effective at fostering learning and near transfer when learning how to solve probability calculation problems. The delayed posttest results show that performance stayed high one week later. The amount of effort participants invested in studying the examples, as well as in completing the posttests, did not differ among conditions. In other words, all three instructional conditions were (approximately) equally efficient in terms of the learning process (combination of test scores and effort invested during learning) and in terms of the quality of the acquired cognitive schemas (combination of test scores with effort invested to complete the test problems; see Van Gog & Paas, 2008, for a discussion of both types of efficiency).

As expected, all three forms of example-based instruction helped participants to significantly enhance their self-efficacy and perceived competence. From immediate to delayed posttest, participants' self-efficacy and perceived competence decreased somewhat, but both remained substantially higher than on the pretest. In contrast to our expectation, there were no differences between the conditions in the degree to which these perceived capabilities were enhanced. Furthermore, after watching the two video examples, all three conditions showed that they underestimated how they would actually perform on the tests.

The instruction evaluation results showed that participants who studied worked examples were significantly more positive about receiving similar instruction in the future than those who studied modeling examples with a visible model. There was no significant difference in how enjoyable participants found the examples to be.

The results of Experiment 1 suggest that the form of the examples does not seem to affect cognitive and motivational aspects of learning. However, the results of experiment 1 should be interpreted with caution. The near-ceiling effect on the immediate and delayed posttest suggests that the learning materials may have been too easy for participants despite the fact that both video examples presented new information that had not been taught in the curriculum yet. Possibly, the second example may have been redundant, which could perhaps also explain why participants who watched the modeling example with visible model were less appreciative of receiving similar instruction in the future, as a demonstration by a model is more difficult to ignore than written text or listening to an explanation with static pictures. To address these issues, as well as to be able to explore whether secondary education students' processing and regulation strategies can explain variance in learning outcomes (which was not possible in Experiment 1 due to a technical error), a second experiment was conducted.

Experiment 2

Experiment 2 replicated Experiment 1 but with a few adjustments. To prevent participants from having too much prior knowledge as well as to prevent a ceiling effect on the posttests, participants were younger and partly lower in educational level, and they only received one video example. Also, the tests were adjusted to measure far transfer (i.e., applying what has been learned to new tasks of a different type; i.e., different structural features/solution procedures) in addition to learning and near transfer.

Method

Participants and Design. The design of Experiment 2 was the same as that of Experiment 1. Participants were 134 Dutch students (M_{age} = 13.77, *SD* = 0.53; 76 boys, 58 girls) in the second year of pre-university education (the highest level of secondary education in the Netherlands with a six year duration) or the second year of general secondary education (i.e., the second highest level of secondary education in the Netherlands with a five year duration). As in Experiment 1, participants were from two schools located in middle class neighbourhoods and although we do not have the exact details, participants were predominantly of Caucasian origin. Participants were randomly allocated to either the WE (*n* = 43), ME+ (*n* = 46), or ME- Condition (*n* = 45); because differences among conditions, if existent at all, might be very small, we also increased the number of participants in each condition in Experiment 2. The experiment was conducted before probability calculation was covered in the curriculum.

Materials and Procedure

The same materials and procedure were used as in Experiment 1, with a few exceptions. Firstly, two new probability calculation problems that measured far transfer were added to the pretest and both posttests: one probability calculation problem without replacement with *order unimportant* (i.e., 2/6*1/5) and one probability calculation problem without replacement with order important in which the *numerator was higher than one* (4/7*3/6). Thus, the far transfer problems asked participants to apply what they had learned to more complex tasks that differed in surface features (i.e., cover stories) and in structural features (i.e., probability calculation problems with order unimportant or a higher numerator instead of without order important or a numerator of one, both of which were not explained in the video examples). To keep the total number of problems on every test the same as in Experiment 1, all three of the test versions had two probability calculation problems without replacement and with order important (i.e., that measured learning) removed (1/6*1/5=1/30 and 1/14*1/13=1/182). Secondly, participants in each condition only observed one video example (i.e., the balls example) as opposed to two.

Data Analysis

The same research assistant scored all test answers and again, two raters independently scored 10% of the tests. The intra-class coefficient was .996. The scoring method was identical to that of Experiment 1, but this time a maximum of 8 points could be earned on the probability calculation problems that measured learning, a maximum of 4 points on problems that measured near transfer, and a maximum of 4 points on problems that measured far transfer.

Results

One participant from the ME+ Condition was removed from all analyses because the pretest score indicated too much prior knowledge (50% or more correct on the learning items). Two participants from the ME+ and one from the WE Condition were removed because of technical difficulties while watching the videos during the learning phase. Therefore, data of 130 participants in total remained (WE: n = 45; ME+: n = 43; ME-: n = 42). Eight participants were absent during the delayed posttest (three from both ME Conditions, two from the WE Condition). Their missing scores on the Delayed Posttest were replaced with the condition average.

The test performance and mental effort scores on the test are shown in Table 3, the self-efficacy, perceived competence scores, judgment of learning, and judgment of learning accuracy scores in Table 4, and the ILS scores in Table 5.

Test Performance

A RM-ANOVA on the scores obtained on the learning test items showed no main effect of Instruction Condition, $F(2, 127) = 1.62$, $p = .201$. There was a significant main effect of Test Moment, $F(2, 254) = 187.69$, $p < .001$, $\eta_p^2 = .60$. Participants' test scores were significantly better on the Immediate Posttest and Delayed Posttest than on the Pretest, ps < .001. Furthermore, participants' performance significantly decreased from the Immediate to the Delayed Posttest, $p < .001$. There was no significant interaction effect between Test Moment and Instruction Condition, $F(4, 254) = 2.26$, $p = .063$.

For near transfer, there was no main effect of Instruction Condition, $F < 1$. A main effect of Test Moment was found, $F(2, 254) = 52.16$, $p < .001$, $\eta_p^2 = .291$. Performance was significantly better on the Immediate and Delayed Posttest compared to the Pretest, ps < .001. Moreover, performance on the Delayed Test seemed to be somewhat lower than on the Immediate Posttest, but this difference was not statistically significant, $p = .073$. There was no significant interaction between Test Moment and Instruction Condition, $F < 1$.

For far transfer, a RM-ANOVA showed no main effect of Instruction Condition, $F(2,127) = 1.24$, $p = .292$. There was, however, a main effect of Test Moment, F(2.15, 38.02) = 7.19, p = .002, η_p^2 = .054. Moreover, a significant interaction was found between Test Moment and Instruction Condition, $F(1.59, 38.02) = 2.66$, p = .044, η_p^2 = .040. A closer look at the data suggested that the interaction effect was a result of how the scores developed from the Immediate to the Delayed Posttest, with the WE Condition showing a decline and the ME+ and ME- Condition an increase in performance. It should be noted, however, that because the far transfer scores on all tests were extremely low (see Table 3), it is questionable how meaningful this difference is.

Table 3.
Mean (SD) of Learning, Near Transfer, and Far Transfer Test Scores and Mental Effort per Condition in Experiment 2

	Pretest			Immediate Posttest			Delayed Posttest		
	WE	ME+	ME-	WE	ME+	ME-	WE	ME+	ME-
Test scores									
Learning	0.19	0.29	0.47	5.39	4.80	4.38	4.16	4.24	3.05
	(0.43)	(0.58)	(0.79)	(2.45)	(2.82)	(2.67)	(3.25)	(2.87)	(2.74)
Near	0.16	0.30	0.24	1.61	1.42	1.51	1.50	1.05	1.18
Transfer	(0.28)	(0.72)	(0.47)	(1.53)	(1.49)	(1.66)	(1.68)	(1.41)	(1.58)
Far	0.01	0.00	0.06	0.21	0.04	0.14	0.07	0.24	0.30
Transfer	(0.07)	(0.00)	(0.16)	(0.58)	(0.13)	(0.44)	(0.25)	(0.59)	(0.65)
Mental									
Effort	5.13	5.02	5.23	3.90	3.71	3.75	4.08	4.39	4.71
Learning	(1.60)	(1.53)	(1.77)	(1.39)	(1.31)	(1.64)	(1.84)	(1.47)	(2.00)
Near	5.26	5.19	5.55	4.16	3.79	3.90	4.05	4.19	5.00
Transfer	(2.06)	(1.93)	(2.03)	(1.70)	(1.58)	(2.15)	(1.95)	(1.88)	(2.34)
Far	5.20	5.45	5.70	4.19	4.15	3.86	4.59	4.68	4.99
Transfer	(1.82)	(1.68)	(1.71)	(1.68)	(1.51)	(1.86)	(1.66)	(1.37)	(1.93)

Mental Effort

An ANOVA showed that there was no significant difference between the conditions on the mental effort invested in the learning phase, $F < 1$. Participants in the WE (M = 2.64, SD = 1.73), ME+ (M = 2.57, SD = 1.45), and ME- Condition (M = 2.44, SD = 1.44) invested an equal amount of mental effort on learning their condition-specific video example.

A RM-ANOVA on mental effort invested in completing the learning test items showed no main effect of Instruction Condition, $F < 1$. There was, however, a main effect of Test Moment, $F(2, 254) = 30.65$, $p < .001$, η_p^2 = .194. Significantly less mental effort was invested on the probability calculation tasks that measured learning on the Immediate and Delayed Posttest than on the Pretest, $p < .001$. Moreover, participants invested significantly more effort on the Delayed Posttest than on the Immediate Posttest, $p = .001$. There was no significant interaction between Test Moment and Instruction Condition, $F < 1$.

A RM-ANOVA on invested mental effort on the near transfer test items showed no main effect of Instruction Condition, $F < 1$. There was, however, a main effect of Test Moment, $F(2, 254) = 29.16$, $p < .001$, $\eta_p^2 = .187$. Compared to the Pretest, participants invested significantly less effort on the near transfer items on the Immediate Posttest, $p < .001$, and the Delayed Posttest, $p < .001$. Moreover, participants' invested mental effort significantly increased from Immediate to Delayed Posttest, $p = .027$. No significant interaction was found between Test Moment and Instruction Condition, $F(4, 254) = 1.86$, $p = .119$.

There was no main effect of Instruction Condition on the mental effort invested on the far transfer test items, $F < 1$. A main effect of Test Moment was found, $F(2, 254) = 32.62$, $p < .001$, $\eta_p^2 = .204$. The effort that participants invested was significantly higher on the Pretest tasks than on the Immediate ($p < .001$) and Delayed Posttest ($p = .001$). Moreover, participants invested more effort in solving the far transfer problems on the Delayed Posttest than on the Immediate Posttest, $p < .001$. Lastly, there was no significant interaction between Test Moment and Instruction Condition, $F(4, 254) = 1.27$, $p = .282$.

Self-efficacy and Perceived Competence

For self-efficacy, no main effect of Instruction Condition was found, $F(2, 127) = 1.55$, $p = .217$. There was, however a main effect of Test Moment, $F(1.93, 244.79) = 76.93$, $p < .001$, $\eta_p^2 = .377$. Participants' self-efficacy significantly increased from Pretest to Immediate Posttest, $p < .001$, and then significantly decreased again from Immediate to Delayed Posttest, $p < .001$. However, participants still rated their self-efficacy higher after one week than on the Pretest measurement, $p < .001$. The interaction between Test Moment and Instruction Condition was not significant, $F < 1$.

The RM-ANOVA on perceived competence ratings also showed no main effect of Instruction Condition, $F < 1$. Again, a main effect of Test Moment was found, $F(2, 254) = 85.98$, $p < .001$, $\eta_p^2 = .404$, with perceived competence being significantly higher on the Immediate test than on the Pretest, $p < .001$. Then, participants' perceived competence ratings significantly dropped from Immediate to Delayed Posttest, $p < .001$, but at the Delayed Posttest their perceived competence was still higher than on the Pretest, $p < .001$. No interaction effect was found, $F < 1$.

Judgment of Learning

For judgment of learning, there was no main effect of Instruction Condition, $F(2, 127) = 1.65$, $p = .197$. There was, however, a main effect of Test Moment, $F(2, 254) = 32.11$, $p < .001$, $\eta_p^2 = .202$. Significantly more answers were estimated to be correct on the Immediate and Delayed Posttest than on the Pretest, $p < .001$. No significant difference between the Immediate and Delayed Posttest was found, $p = 1.000$. Moreover, no significant interaction effect was found, $F < 1$.

On accuracy of the predicted test performance (i.e., judgment of learning accuracy), the RM-ANOVA showed no main effect of Instruction Condition, $F(1, 127) = 1.05$, $p = .353$. A main effect of Test Moment was found, $F(1, 127) = 7.39$, $p = .007$, $\eta_p^2 = .055$. Participants were significantly more accurate in assessing their own performance on the Immediate Posttest ($M = 1.55$, $SD = 4.56$) than on the Delayed Posttest ($M = 2.72$, $SD = 5.03$). One sample *t*-tests showed that performance accuracy estimates differed significantly from zero, *p*s < .001, which means that participants significantly overestimated (Table 4) their performance on both measurements. There was no significant interaction between Test Moment and Instruction Condition, $F < 1$.

Table 4.
Mean (SD) of Self-efficacy, Perceived Competence, Judgment of Learning (JOL), and Judgment of Learning Accuracy Scores per Condition in Experiment 2

	Pretest			Immediate Posttest			Delayed Posttest		
	WE	ME+	ME-	WE	ME+	ME-	WE	ME+	ME-
Self-efficacy	3.58	3.74	3.53	5.69	5.90	5.58	4.77	5.38	4.50
	(2.04)	(1.81)	(1.65)	(1.38)	(1.59)	(1.26)	(2.04)	(1.62)	(1.66)
Perceived	3.66	3.45	3.43	5.04	5.10	4.86	4.35	4.34	3.90
Competence	(1.57)	(1.48)	(1.29)	(1.38)	(1.32)	(1.33)	(1.34)	(1.48)	(1.24)
JOL				4.11	4.17	3.81	3.86	4.51	3.62
				(1.42)	(1.36)	(1.52)	(1.93)	(1.59)	(1.67)
JOL				1.01	2.08	1.59	1.99	3.49	2.72
Accuracy				(4.07)	(4.52)	(5.09)	(6.06)	(4.20)	(4.58)

Instruction Evaluation

An ANOVA showed no significant differences among the ME+ ($M = 4.67$, $SE = 0.39$), WE ($M = 4.67$, $SE = 0.38$), and ME- Condition ($M = 4.30$, $SE = 0.39$) in how enjoyable observing the video example was for participants, $F < 1$. Moreover, although numerically participants in the ME+ Condition ($M = 7.40$, $SE = 0.34$) had a significantly higher preference to receive similar instruction again than the WE ($M = 6.60$, $SE = 0.33$) and ME- Condition ($M = 6.30$, $SE = 0.34$), this difference did not reach statistical significance, $F(2, 127) = 2.79$, $p = .065$.

Self-regulation and Processing

Correlations were calculated of participants' scores on the three subscales that measured processing strategies (deep processing, stepwise processing, and concrete processing) and the three subscales that measured self-regulatory processes (self-regulation, external regulation, and lack of regulation) with the performance measures. None of the scales significantly correlated with participants' performance on test items that measured learning, near transfer, or far transfer, except for 'lack of regulation' which was significantly negatively correlated with test performance on the items that measured learning ($r = -.234$, $p = .009$).

Discussion

Experiment 2 replicated the results of Experiment 1 with a younger student population that only studied one video example. The findings from Experiment 2 again showed that it is effective and efficient to study worked examples and modeling examples with and without a model when learning how to solve probability calculation problems, and that the form of the examples does not seem to affect learning and transfer performance. Even though the far transfer results suggest that modeling examples may be more effective than worked examples for far transfer, these test scores were extremely low, so caution is warranted when interpreting these results –even though the analysis was significant.

Possibly as a result of redundancy of the second example combined with that example being harder to ignore in the 'model visible' condition, participants in Experiment 1 were significantly less appreciative of receiving instruction consisting of modeling examples with a visible model in the future. The instruction evaluation results of Experiment 2, where this redundancy was avoided by presenting only one example, showed that participants who watched a modeling example with a visible model seemed to be more appreciative about receiving similar instruction in the future, although this difference did not reach statistical significance ($p = .065$).

Exploration of secondary education students' processing and regulation strategies showed that these could not explain variance in learning outcomes as they did not correlate with performance. The lack of explained variance may be a result of the experiment being experimenter-paced as opposed to user-paced and the learning phase being relatively short. When given more freedom, students can be expected to employ more of the skills that they regularly use during self-study and thus reap a greater benefit from effective processing or regulation skills.

General Discussion

Two experiments investigated whether worked examples, modeling examples with a visible model, and modeling examples without a visible model would differentially affect cognitive and motivational aspects of learning for novices learning a problem-solving task. Across the two experiments with secondary education students, we found a consistent pattern of results. In line with our hypothesis, all three forms of example-based instruction were effective at enhancing students' test performance and reducing invested mental effort from pretest to posttest. The fact that high performance was attained with relatively low effort investment on not only the immediate but also the delayed posttest indicates a high quality of the acquired cognitive schemas (Van Gog & Paas, 2008). With regard to our open question, the three different types of examples did not differ in the extent to which they fostered test performance and effort reduction. Thus, the extraneous load that participants may have experienced as a result of a split in attention when studying worked examples and transient information when studying modeling examples did not differentially affect cognitive aspects of learning, nor did the model's task-relevant and task-irrelevant movements.

Similar results were found with regards to self-efficacy and perceived competence, which are important to take into account because they have significant bearing on factors such as academic motivation and learning outcomes (Bandura, 1997; Bong & Skaalvik, 2003; Hughes et al., 2011; Klassen & Usher, 2010; Schunk, 2001). Research on worked examples has largely neglected motivational effects of example-based learning (Van Gog & Rummel, 2010). Again, all three forms of example-based learning helped students to improve confidence in their own capabilities. In contrast to our expectations, however, the conditions did not differ in the extent to which they fostered self-efficacy and perceived competence. While it was expected that studying a modeling example with or without a visible model would lead to higher self-efficacy and perceived competence gains than worked examples because modeling examples have a stronger social component (e.g., the model's voice and visibility; Mayer, 2005), these findings suggest that worked examples can be equally beneficial for motivational aspects of learning,

The conditions also did not differ in their judgment of learning accuracy. Overall, students were moderately accurate in assessing their own learning, which is important because the more accurate students' judgment of learning, the better they are able to regulate their study time and restudy choices (Kornell & Metcalfe, 2006; Thiede et al., 2003). It is interesting to note that while participants in Experiment 1 *under*estimated their performance, participants in Experiment 2 *over*estimated their performance. This could possibly be a result of differences in participants' experience with math (participants in Experiment 2 were younger), and the number of examples across the experiments (participants in Experiment 1 received two examples, those in Experiment 2 only one example), because –at least for adults– repeated exposure to materials has been shown to result in underconfidence (i.e., "underconfidence with practice effect"; Koriat, Sheffer, & Ma'ayan, 2002) and because harder items have been shown to result in overconfidence, whereas medium or easy items result in underconfidence (Baars, Visser, Van Gog, De Bruin, & Paas, 2013; Pulford & Colman, 1997). With less experience and less examples, it is likely that participants in Experiment 2 found the materials harder to learn.

These results seem to suggest that the form examples take does not matter. However, form may matter for cognitive aspects of learning if characteristics of the to-be learned task are different. For example, when object-manipulation is inherent to the task and thus requires human movement to successfully complete (e.g., writing, sports), then it may be more natural and effective to observe a model's demonstration. For tasks that do not inherently require object manipulation, which is the case for learning how to solve probability calculation problems, there may not be any additive effect to observing a model's demonstration compared to pictures of problem states. This would be similar to findings that show that human-movement tasks (e.g., assembly, knot-tying, origami tasks) are more effectively taught by dynamic visualizations compared to static ones; while for other types of tasks there is typically a smaller advantage in favour of dynamic visualizations, if such an advantage exists at all (Höffler & Leutner, 2007; Van Gog et al., 2009b).

With regard to self-efficacy and perceived competence, different results may also be obtained when a different task is used or when information is presented in a different manner. Firstly, if the to-be learned task is social by nature and/or requires multiple models to interact with one another (e.g., negotiation skills; Gentner, Loewenstein, & Thompson, 2003), then a modeling example may lead to higher pre- to posttest gains because more socially-relevant information is available to make a social comparison. A modeling example could perhaps also lead to higher pretest to posttest gains when the instructional text is specifically designed to follow a more conversational style and to entail making and correcting errors, because, compared to a more formal and ideal style, this would allow novices to identify themselves more with a human model (Mayer, 2005; Töpper et al., 2014).

In summary, the present study is to our knowledge the first to explore whether form of example-based instruction affects cognitive and motivational aspects when learning a problem-solving skill, which is important because worked examples and modeling examples with and without a visible model are increasingly being used in (online) learning environments. The findings suggest that form has little impact when learning how to solve probability calculation tasks. However, future research is recommended to further investigate whether these findings generalize to different tasks, and different manners of presentation.

Chapter 3

Learning from video modeling examples: Does gender matter?

This chapter has been published as:

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Does gender matter? *Instructional Science, 44*, 69-86. doi:10.1007/s11251-015-9360-y

Abstract

Online learning from video modeling examples, in which a human model demonstrates and explains how to perform a learning task, is an effective instructional method that is increasingly used nowadays. However, model characteristics such as gender tend to differ across videos, and the model-observer similarity hypothesis suggests that such characteristics may affect learning. Therefore, this study investigated whether the effectiveness of learning how to solve a probability calculation problem from video modeling examples would vary as a function of the model's and observer's gender. In a 2 (Model: Female/Male) X 2 (Observer: Female/Male) between-subject design, 167 secondary education students learned how to solve probability calculation problems by observing video modeling examples. Results showed no effects of Model or Observer gender on learning and near transfer. Male students reported higher self-efficacy than female students. Compared to a female model, observing a male model enhanced perceived competence more from pretest to posttest, irrespective of observers' gender. Furthermore, learning from a male model was less effortful and more enjoyable for male students than for female students. These results suggest that gender of both model and observer can matter in terms of affective variables experienced during learning, and that instructional designers may want to consider this when creating (online) learning environments with video modeling examples.

Introduction

Students of all ages and educational levels increasingly watch instructional videos for informal learning purposes online on websites such as YouTube and Google Videos, but such videos are also increasingly used in formal learning (Lenhart, 2012; Spires, Hervey, Morris, & Stelpflug, 2012). In formal learning, online instructional videos can be consulted while making homework, or can replace activities that normally took place face to face. For instance, some educators have even argued in favor of a “flipped classroom”, which entails having learners study videos at home to free up time in school for practice and teacher support (Bergmann & Sams, 2012). Various types of videos are used for both informal and formal learning purposes, such as web lectures (e.g., Day & Foley, 2006; Traphagan, Kucsera, & Kishi, 2010), short knowledge clips (e.g., Day, 2008), and how-to demonstration videos (e.g., Ayres, Marcus, Chan, & Qian, 2009). Regarding the latter, research inspired by social-cognitive theories such as social learning theory (Bandura, 1977, 1986) and cognitive apprenticeship (Collins, Brown, & Newman, 1989) has demonstrated the effectiveness of acquiring problem-solving skills from these so-called *video modeling examples* in which a (human) model explains and/or demonstrates how to perform a task on video (e.g., Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013a, 2013b; Hoogerheide, Loyens, & Van Gog, 2014a; Van Gog, Verveer, & Verveer, 2014). In addition to being effective for acquiring cognitive skills, observing video modeling examples has also been shown to enhance affective variables, such as students’ belief in their own ability to perform the modeled task at a certain level (i.e., self-efficacy; Bandura, 1997; Schunk, 1987).

When creating a video modeling example, an instructional designer is confronted with various design choices, which might affect learning, both cognitively as well as affectively. For instance, should the video present a natural task performance procedure, which might entail making and correcting errors (e.g., Groenendijk et al., 2013a, 2013b), or a more didactical procedure that reflects how a student should ideally learn the skill (e.g., Hoogerheide et al., 2014a; Simon & Werner, 1996; Van Gog et al., 2014)? Another design consideration is whether the model should be (partly) visible in the video while explaining the task (e.g., Hoogerheide et al., 2014a; Van Gog et al., 2014; Xeroulis, Park, Moulton, Reznick, LeBlanc, & Dubrowski, 2007), or whether only the model’s computer screen should be shown (e.g., McLaren, Lim, & Koedinger, 2008; Van Gog, 2011; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009a). If a form is chosen in which the model is visible, the question arises who the model should be in terms of expertise, age, background, and gender.

Because the widespread use of online video modeling examples is relatively recent, there is as of yet little empirical knowledge available to guide design choices. Recent studies have started to uncover effects of different ways of presenting the content in video modeling examples (e.g., to which degree the model should be visible; Hoogerheide et al., 2014a; Van Gog et al., 2014). Potential effects of model characteristics that are unrelated to how the learning task is presented, such as gender, on the learning process and learning outcomes, have received little attention in recent research on video modeling examples. However, earlier research inspired by the model-observer similarity hypothesis (Schunk, 1987, 1991a), as well as recent research on pedagogical agents (e.g., Baylor and Kim, 2004; Ozogul, Johnson, Atkinson, & Reisslein, 2013), suggests that similarity in factors such as gender may matter. Building on these findings, which will be reviewed below, the present study examined whether the effectiveness and efficiency of video modeling examples can vary as a function of the observer's and model's gender.

Model-observer Similarity

The model-observer similarity hypothesis (Schunk, 1987, 1991a; see also the similarity-attraction hypothesis; Moreno & Flowerday, 2006) states that model characteristics can matter when learning from modeling examples because the effectiveness of modeling is at least partly moderated by the degree to which observers perceive a model to be similar to them. Modeling evokes social comparison (Berger, 1977; Johnson & Lammers, 2012) and observing a model that successfully performs a task may lead observers to believe that they can perform the task as well, if they identify with the model (Bandura, 1981, Schunk, 1984). Moreover, an observer may be more attracted to and pay more attention to a model that is perceived as similar (Berscheid & Walster, 1969).

As Schunk (1987) noted, "similarity serves as an important source of information for gauging behavioural appropriateness, formulating outcome expectations, and assessing one's self-efficacy for learning or performing tasks" (p. 149). It is likely that particularly novice learners whose prior knowledge as well as self-efficacy and perceived competence are still low, are affected by model-observer similarity, as they are especially likely to engage in social comparison (Buunk, Zurriaga, Gonzalez-Roma, & Subirats, 2003). In other words, the higher the degree of similarity between observer and model, particularly when the observer is novice to the task at hand, the more cognitive outcomes of learning (e.g., performing the same or novel tasks) and affective aspects of the learning process (e.g., self-efficacy, perceived competence) may be enhanced.

With respect to those affective variables, self-efficacy is important because it influences factors such as academic motivation, study behaviour, and learning outcomes (Bandura, 1997; Bong & Skaalvik, 2003; Schunk, 2001). Similarly, perceived competence, which is a related construct that reflects broader perceptions and knowledge (Bong & Skaalvik, 2003; Hughes, Galbraith, & White, 2011; Klassen & Usher, 2010), also affects academic motivation and learning outcomes (Bong & Skaalvik, 2003; Harter, 1990; Ma & Kishor, 1997). Moreover, when students' confidence in their own capabilities increases, they tend to use more cognitive and metacognitive strategies irrespective of previous achievement or ability (Pajares, 2006) and the willingness to invest mental effort in a task changes as well (Bandura, 1977; Salomon, 1983, 1984).

Gender can perhaps be expected to be the most important factor of model-observer similarity because gender is among the first things being noticed when interacting with others (Contreras, Banaji, & Mitchell, 2013). Schunk (1987), however, reported mixed results on both learning outcomes and self-efficacy in his review, and suggested that one possible explanation for these mixed findings might lie in the appropriateness of the modelled behaviour: students' beliefs that a skill or behaviour is more appropriate for one of the genders may moderate effects of gender similarity. This might explain why Bandura, Ross, and Ross (1963) and Hicks (1965) found that for boys, observing a male model displaying aggressive behaviour towards a doll led to more imitative aggression than observing a female model. In contrast, no such effects were found for grade 4 to 6 students who observed a male or female model solving fraction problems (Schunk, Hanson, & Cox, 1987). Although mathematical tasks are typically more associated with males than females (Forgasz, Leder, & Klosterman, 2004; Stewart-Williams, 2002), young children do not yet seem to hold this association, which becomes stronger during adolescence (Steffens, Jelenec, & Noack, 2010; see also Ceci, Ginther, Kahn, & Williams, 2014). In other words, the 10 year olds in the study by Schunk et al. (1987) may have been too young to associate a mathematical task with gender.

More recent studies also suggest mixed findings, however. Surprisingly in light of the above, a study with university students learning probability calculation with dynamic visualizations accompanied by a male or female model's narration showed that a female model was preferred and led to better learning outcomes than a male model (Linek, Gerjets, & Scheiter, 2010). However, findings of Rodicio (2012) and Lee, Liao, and Ryu (2007) suggest the opposite, namely that male narrations should be preferred. More specifically, Rodicio (2012) found that university students learned more about geology from dynamic visualizations with a male voice over than a female voice over, and Lee et al.

(2007) found that for male students, a male computer-generated voice was more positively evaluated, trusted, and led to higher confidence levels than a female computer-generated voice. Note though, that in these studies, the model was not visible and therefore the cues available to make a social comparison may have been less strong compared to video modeling examples with a visible model (Hoogerheide et al., 2014a).

Several animated pedagogical agent studies, in which a cartoon-like (humanoid) agent functions as a model or teacher, did show a preference for male agents, particularly for tasks that may be believed to be more appropriate for men. For instance, Moreno, Person, Adcock, Eck, Jackson, and Marineau (2002) found that university students' knowledge about blood pressure was enhanced more after interacting with a male agent than a female agent. Arroyo, Woolf, Royer, and Thai (2009) found that for secondary education and university students, a male agent led to more positive attitudes about mathematics and better learning outcomes. Furthermore, a study in educational technology found that male agents were evaluated as more interesting, intelligent, useful, and satisfactory than female agents (Baylor & Kim, 2004). However, other research has shown that when learning an engineering task, often considered a stereotypically male domain in Western countries, interacting with a female model decreased women's beliefs about engineering stereotypes compared to interacting with a male agent (Rosenberg-Kima, Baylor, Plant, & Doerr, 2008). Moreover, when given the choice, students tend to select an agent of the same gender (Ozogul et al., 2013).

In sum, the model-observer similarity hypothesis suggests that if one observes a same-gender model, affective and cognitive aspects of learning are more enhanced. More recent studies, particularly those with animated pedagogical agents, seem to suggest however, that for tasks that are more appropriate for males, male agents are preferred over female ones. Therefore, when it comes to video modeling examples, it is still an open question how gender affects learning.

The Present Study

The present study investigated whether it is more effective for male and female secondary education students to study video modeling examples depicting a same-gender model explaining and demonstrating a math task in terms of cognitive aspects of learning (i.e., learning and near transfer) and motivational aspects of learning (i.e., self-efficacy and perceived competence). In addition, the study measured cognitive load (i.e., effort investment) during the learning and test phase to investigate effects on the learning process and explored effects on judgment of learning accuracy and instruction evaluation. Female and male secondary education students learned how to solve probability calculation problems with replacement and order important by watching a video modeling example in which either a male (see Figure 1) or a female (see Figure 2) model explained and demonstrated the task. Both were instructed to wear a neutral, black t-shirt, and participated in an extensive practice training session to ensure that they showed the same behaviour throughout the video (e.g., identical movements and gestures). An autocue was used to guarantee that the models gave the same explanation and spent the same amount of time on the steps shown in the video (and consequently on the video as a whole). After sufficient practice (as judged by the first author who was present at all times), the definitive recordings were created. Moreover, other factors that might affect perceived similarity were kept constant across conditions by selecting a young adult male and female Caucasian model (the majority of our participant population was Caucasian), who had a comparable educational background and were both in their early 20s. Therefore, we could be confident that effects (if any) would not be caused by differences in the content that was being presented."

We firstly hypothesized that for male and female secondary education students who have little if any knowledge of solving probability calculation problems, it would be effective to study video modeling examples with both a male and female model, because research has consistently shown that example-based learning is very effective and efficient for novice learners (Atkinson et al., 2000; Renkl, 2014; Sweller, Ayres, & Kalyuga, 2011; Van Gog & Rummel, 2010)[1]. Thus, we expected high pretest to posttest performance gains (Hypothesis 1a) attained with a low to medium amount of effort investment during example study (Hypothesis 1b), while the amount of mental effort required to solve the test problems would decrease (Hypothesis 1c). Students' self-efficacy and perceived competence were also expected to increase from pretest to posttest (Hypothesis 1d), since observing a model successfully explain and demonstrate a task has been shown to positively affect novices' confidence in their own abilities (Bandura, 1981, Hoogerheide et al., 2014a; Schunk, 1984).

The more interesting and open question was whether model-observer similarity would have an effect on cognitive and affective variables. In other words, would male and female students differ in the degree to which learning and transfer (Question 2a) and self-efficacy and perceived competence (Question 2d) would be enhanced, mental effort

[1] Note that for students who have some prior knowledge of solving probability calculation problems, examples would lose their effectiveness or may even start to hamper learning compared to practice problem solving (Kalyuga, Chandler, Tuovinen, & Sweller, 2001; this is an example of the expertise-reversal effect; see Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga & Renkl, 2010).

invested in the test reduced (Question 2c), and in the degree that students invest mental effort during example study (Question 2b), depending on whether they observed a video modeling example that presented a male or a female model? Based on the model-observer similarity hypothesis, we could expect novice learners to identify more with a same-gender model relative to an opposite-gender one and therefore show cognitive and affective benefits when learning from a same-gender model (Schunk, 1987). However, based on research with animated pedagogical agents (e.g., Arroyo et al., 2009, Moreno et al., 2002) and dynamic visualizations with a voice over (Lee et al., 2007; Rodicio, 2012), we might expect that novices benefit more from a male model than a female model because mathematical tasks are associated more with males than females (Forgasz et al., 2004; Stewart-Williams, 2002). Moreover, because the confidence that learners have in their own capabilities is associated with how much effort they invest (Bandura, 1977; Salomon, 1983, 1984), differences in perceived capabilities across conditions could affect how much mental effort students invest during example study.

Because enhanced confidence can also be a negative outcome if it leads to overconfidence, which can be detrimental to students' regulation of their learning process (Dunlosky & Rawson, 2012; Rhodes & Tauber, 2011; Thiede, Anderson, & Therriault, 2003), we instructed participants to predict their performance on the posttest. This judgment of learning was then matched to their actual performance to explore whether students' judgment of learning accuracy would depend on the gender of the model (Question 3). Because an increase in confidence leads to using more cognitive and metacognitive strategies (Pajares, 2006), differences might especially arise if students differ in their self-efficacy and perceived competence depending on the model's gender.

Previous research has shown that instruction evaluation measures such as learning enjoyment may vary depending on the form of example-based instruction (Hoogerheide et al., 2014a; see also Liew, Tan, & Jayothisa, 2013), and therefore we also explored effects on learning enjoyment and willingness to receive similar instruction in the future (Question 4) because these can be important indicators for the use of online examples during future self-study (Yi & Hwang, 2003). Differences on these instruction evaluation measures might especially be dependent on whether there are differences in effort investment during example study because when practice effort decreases, enjoyment of practice may increase (Hyllegard & Bories, 2009).

Figure 1. Female model

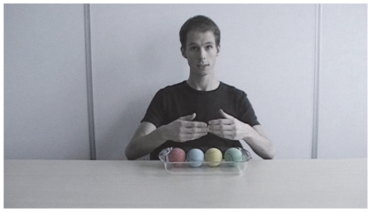


Figure 2. Male model

Method

Participants and Design

The experiment had a 2 X 2 design, with Gender Model (Male vs. Female) and Gender Observer (Male vs. Female) as between-subject factors. Participants were 167 predominantly Caucasian secondary education students (M^{age} = 13.50, *SD* = 0.59; 80 male, 87 female) in their second year of general secondary education, which is the second highest level of secondary education in The Netherlands and has a total duration of five years. The students were randomly allocated to a female model (38 girls, 43 boys) or a male model (42 boys, 44 girls) condition. The experiment was conducted at a point in time at which probability calculation had not yet been taught in the curriculum.

Materials

All materials were presented using Qualtrics, which is a web-based survey software tool platform (http://www.qualtrics.com).

Video Modeling Example. Two video modeling examples were created, one with a female model (see Figure 1) and one with a male model (see Figure 2). Both models used the same example to address how one would ideally solve a probability calculation problem without replacement and with order important (i.e., an ideal procedure). The problem-state of this example was as follows: “The scouting staff brings 4 coloured balls for the cub scouts to play with. There is a red ball, a blue ball, a yellow ball, and a green ball. The cub scouts get to choose a ball one by one and prefer every colour equally. What is the chance that the red ball gets picked first and the green ball second?” The example then explained step-by-step how to solve this problem and briefly addressed what would happen in case it was an example of a probability calculation *with* replacement.

Both models were in their 20s, Caucasian, and wore a black neutral outfit while sitting behind a desk with the learning materials placed on the desk (i.e., the 4 different coloured balls and a platter; see Figure 1 and 2). An autocue was used to guarantee that the models gave the same explanation and spent the same amount of time on the steps shown in the video (and consequently on the video as a whole). After sufficient practice, the definitive recordings were created. At the beginning of the video, all four items rested inside a platter. While explaining the models interacted with the learning materials to illustrate the problem-solving steps. For example, while explaining the first event —the chance that the red ball is picked first— both models picked up the red ball and held it in the air, after which they placed the red ball at the side of the platter.

Pretest and Posttest. Two test versions were created that both consisted of six probability calculation problems. Within each test, four items measured learning (i.e., applying what has been learned to new tasks of the same type that have the same structural features but differ in surface features; solution procedures: 1/4 * 1/3 = 1/12, 1/11 * 1/10 = 1/110, 1/6 * 1/5 * 1/4 = 1/120, and 1/8 * 1/7 * 1/6 * 1/5 = 1/1680) and two near transfer (i.e., applying what has been learned to new tasks of the same type that differ partly in structural features and differ in surface features; solution procedures: 1/6 * 1/6 = 1/36, 1/5 * 1/5 = 1/25). All problems required participants to fill in the correct answer and calculation. For example, one problem provided the following problem-state: “On a cold Sunday, a fisherman catches all the fish at from a small lake, one at a time. There are four fish swimming in the lake: a perch, a bream, a pike, and an eel. What is the chance that the bream is caught first, and the pike caught second?” The correct answer would be 1/4 * 1/3 = 1/12. The two test versions were parallel to each other, that is, the problems were structurally equivalent across both tests, but they differed in surface features (i.e., cover stories). The internal consistency (Cronbach’s alpha) of the pretest was .775 and of the posttest it was .741.

Mental Effort. Effort investment was measured after every test item on the pretest and posttest and after watching the video modeling example using the subjective rating scale of Paas (1992), which asks participants to indicate the effort they invested on a 9-point scale that ranges from (1) very, very low effort to (9) very, very high effort.

Self-efficacy and Perceived Competence. Self-efficacy was measured by asking participants to indicate on a 9-point scale, ranging from (1) very, very unconfident to (9) very, very confident, to what degree they believed that they mastered the skill probability calculation. This measure was adopted from Hoogerheide et al. (2014a) and the phrasing of the questions is similar to Bandura (2006). To measure perceived competence, an adapted version of the scale by Williams and Deci (1996) was used. This scale consists of four items and asks participants to indicate to what degree the item applies to them, on a scale of 1 (not at all true) to 7 (very true). The item “I am able to achieve my goals in this course” was removed because this question did not apply to the present experiment, leaving the following three items: “I feel confident in my ability to learn this material”, “I am capable of learning the material in this course", and "I feel able to meet the challenge of performing well in this course". The word "course" was rephrased to "probability calculation problems".

Judgment of Learning. To measure judgment of learning, participants were asked on a scale of 0 to 6 to indicate how many probability calculation problems they expected to answer correctly if presented with a test.

Instruction Evaluation. To investigate how participants experienced the video modeling example, they were asked after observing the video modeling example to indicate how enjoyable watching the video was and to what degree they would prefer to receive similar instruction in the future on a scale of 0 (lowest) to 10 (highest).

Procedure

The session took place at the computer lab of participants' school (ca. 45 min.). Before participants arrived, A4-papers were distributed over the computer lab containing the name of participants and a link to the Qualtrics questionnaire. This questionnaire presented 4 'question blocks'. Prior to each question block, participants received a plenary verbal instruction, after which they completed that specific question block. Question block 1 asked participants to fill in a general demographic questionnaire for which they received 90 seconds. Question block 2 contained the pretest (6 probability calculation problems and mental effort ratings), for which participants were instructed to not only write down their answer, but also the calculation. The remainder of question block 2 presented questions to measure self-efficacy and perceived competence. Question block 3 presented the video example (a YouTube video embedded in Qualtrics) followed by a mental effort rating and the instruction evaluation questions (i.e., learning enjoyment and willingness to receive similar instruction). Lastly, question block 4 first presented self-efficacy, perceived competence, and judgment of learning questions, followed by the posttest, which consisted of six probability calculation problems and mental effort ratings. Those that received version A as the pretest now received version B as the posttest, and those that received B as the pretest now received version A.

Data Analysis

A maximum of 8 points could be earned on both tests for the problems that measured learning, and a maximum of 4 for the problems that measured near transfer. Participants could earn 2 points per probability calculation problem: 1 point for a correct answer (0.5 for a partially correct answer; 0 for an incorrect or missing answer) and 1 point for the correct calculation (0 for an incorrect calculation). Both points were granted if participants wrote down the correct answer.

Averages were computed for participants invested mental effort in completing the learning and near transfer test items, as well as the three items that measured perceived competence, on the pretest and posttest separately. We then computed a measure of judgment of learning accuracy by multiplying participants' judgment of learning (i.e., how many of the 6 problems participants predicted to correctly solve) by two and subsequently subtracting their actual test performance (range -12 to +12).

Four participants were removed from all analyses because of technical issues during the experiment (one participant) or too high prior knowledge as indicated by a total attained score greater than 50% on the pretest (three participants). This left 163 participants in total, of which 87 observed a female model (43 female students, 44 male students) and 76 a male model (38 female students, 38 male students). One male student who observed a male model was removed from all test performance analyses and mental effort analysis (excluding invested mental effort in learning the video content) because he had to leave the experiment shortly after he started working on the posttest.

Results

The test performance and invested mental effort scores can be found in Table 1, the self-efficacy, perceived competence, and judgment of learning (accuracy) scores in Table 2, and the instruction evaluation ratings in Table 3.

Test Performance

We tested Hypothesis 1a and Question 2a using a mixed ANOVA, with Test Moment (Pretest, Posttest) as within-subject factor and Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors. The scores obtained on the test items that measured learning showed a significant main effect of Test Moment, $F(1, 158) = 658.79$, $p < .001$, $\eta_p^2 = .807$. Participants performed significantly better on the Posttest ($M = 5.56$, $SD = 2.55$) than on the Pretest ($M = 0.30$, $SD = 0.64$). There was no main effect of Gender Model, $F(1, 158) = 1.71$, $p = .192$, nor of Gender Observer, $F < 1$. No other interaction effects were significant, Fs < 1. With regard to near transfer, the main effect of Test Moment was significant, $F(1, 158) = 154.96$, $p < .001$, $\eta_p^2 = .495$. Performance was significantly higher on the Posttest ($M = 1.85$, $SD = 1.66$) than on the Pretest ($M = 0.28$, $SD = 0.54$). The main effects of Gender Model and Gender Observer were not significant, Fs < 1. Furthermore, no other interaction effects were found (Fs < 1; Test Moment and Gender Observer, $F(1, 158) = 3.15$, $p = .116$).

Table 1.
Mean (SD) of Learning (range 0 to 12) and Near Transfer (range 0 to 4) Scores and Mental Effort (range 1 to 9) per Condition.

	Male Observer		Female Observer	
	Male Model (n = 42)	Female Model (n = 44)	Male Model (n = 43)	Female Model (n = 38)
Test Scores				
Learning Pretest	0.74 (1.52)	0.41 (0.68)	0.31 (0.57)	0.12 (0.45)
Learning Posttest	5.65 (2.69)	5.28 (2.59)	5.89 (2.35)	5.38 (2.67)
Near Transfer Pretest	0.51 (1.00)	0.43 (0.47)	0.14 (0.25)	0.26 (0.73)
Near Transfer Posttest	1.80 (1.70)	1.70 (1.63)	2.03 (1.77)	1.87 (1.59)
Mental Effort				
Learning Pretest	4.82 (1.92)	5.08 (2.03)	4.95 (1.93)	5.05 (1.85)
Learning Posttest	3.27 (1.51)	3.64 (1.78)	3.77 (1.26)	3.95 (1.45)
Near Transfer Pretest	5.28 (2.02)	5.24 (2.27)	4.97 (2.08)	4.92 (2.14)
Near Transfer Posttest	3.25 (1.79)	3.67 (1.90)	3.68 (1.46)	3.67 (1.68)
Effort during Example Study	2.12 (1.27)	2.97 (1.72)	3.11 (1.69)	2.77 (1.41)

Mental Effort

We tested Hypothesis 1b and Question 2b via a 2 x 2 ANOVA with Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors. There was no significant main effect of Gender Model on the invested mental effort *during example study*, $F < 1$, nor of Gender Observer, $F(1, 159) = 1.93$, $p = .167$. The interaction effect between Gender Model and Gender Observer was significant, $F(1, 159) = 5.03$, $p = .026$, $\eta_p^2 = .031$. To explore this interaction effect, we firstly compared the effects of Model Gender for each Observer Gender condition separately. There was only an effect of Model Gender for male students: it was less effortful for them to study an example by a male ($M = 2.24$, $SD = 1.28$) than a female model ($M = 2.97$, $SD = 1.72$), $t(74) = 2.12$, $p = .037$, $d = 0.486$ (medium effect; Cohen, 1988). Secondly, we compared the effects of Observer Gender for each Model Gender separately. There was only an effect of Observer Gender for male model: observing a male model was more effortful for female students ($M = 3.11$, $SD = 1.69$) than male students ($M = 2.24$, $SD = 1.28$), $t(80) = 2.62$, $p = .011$, $d = 0.585$ (medium effect; Cohen, 1988).

A mixed ANOVA with Test Moment (Pretest, Posttest) as within-subject factor and Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors was used to test Hypothesis 1c and Question 2c. The results showed a main effect of Test Moment on invested mental effort in completing the probability calculation problems that measured *learning*, $F(1, 158) = 75.90$, $p < .001$, $\eta_p^2 = .325$. Participants invested less effort to complete the problems that measured learning on the Posttest (M =

3.71, *SD* = 1.49) than on the Pretest (*M* = 5.04, *SD* = 1.90). There were no main effects of Gender Model and Gender Observer, *F*s < 1. None of the interaction effects were significant (*F*s < 1; Test Moment and Gender Observer, *F*(1, 158) = 1.65, *p* = .201).

For the average mental effort invested in completing the *near transfer* problems on the tests, a main effect of Test Moment was found, *F*(1, 158) = 84.24, *p* < .001, η_p^2 = .348. Again, participants invested less effort to complete the near transfer problems on the Posttest (*M* = 3.60, *SD* = 1.69) than on the Pretest (*M* = 5.18, *SD* = 2.10). There were no main effects of Gender Model and Gender Observer, nor were there significant interaction effects, *F*s < 1.

Self-efficacy and Perceived Competence

Hypothesis 1d and Question 2d were tested using a mixed ANOVA with Test Moment (Pretest, Posttest) as within-subject factor and Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors. There was a main effect of Test Moment, *F*(1, 159) = 113.26, *p* < .001, η_p^2 = .416. Participants showed higher confidence in their abilities on the posttest (*M* = 5.60, *SD* = 1.35) than on the pretest (*M* = 3.96, *SD* = 1.96). There was no main effect of Gender Model, *F* < 1, but there was a main effect of Gender Observer, *F*(1, 159) = 10.16, *p* = .002, η_p^2 = .060, showing that male students (*M* = 5.14, *SE* = 0.15) were significantly more confident in their own abilities than female students (*M* = 4.47, *SE* = 0.14). None of the interaction effects were significant, *F*s < 1. With regards to perceived competence, a main effect was found of Test Moment, *F*(1, 159) = 191.72, *p* < .001, η_p^2 = .547. Participants' perceived their competence to be higher on the posttest (*M* = 4.82, *SD* = 1.27) than on the pretest (*M* = 3.37, *SD* = 1.46). There was no main effect of Gender Model, *F* < 1, nor of Gender Observer, *F*(1, 159) = 3.14, *p* = .078. There was no interaction between Gender Model and Gender Observer, *F* < 1. The interaction between Test Moment and Gender Model was significant, *F*(1, 159) = 4.81, *p* = .030, η_p^2 = .029. A closer look at the data showed that observing a male model enhanced perceived competence more from pretest (*M* = 3.32, *SE* = 0.16) to posttest (*M* = 4.98, *SE* = 0.14), than observing a female model enhanced perceived competence improvement from pretest (*M* = 3.46, *SE* = 0.16) to posttest (*M* = 4.67, *SE* = 0.14). No other interaction was significant, *F*s < 1.

Judgment of Learning

We tested Question 3 via a 2 x 2 ANOVA with Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors. On the judgment of learning scores, there was no main effect of Gender Model, *F* < 1, nor of Gender Observer, *F*(1, 159) = 1.90, *p* = .170. The interaction between Gender Model and Gender Observer was not significant either, $F < 1$. With respect to the accuracy of the judgments of learning, no main effect of Gender Model was found, $F(1, 159) = 1.47$, $p = .227$, nor of Gender Observer, $F(1, 159) = 2.21$, $p = 1.39$. There was no significant interaction either, $F < 1$. One sample *t*-tests showed that for all four combinations of the 2 X 2 design, judgment of learning accuracy was not statistically different from zero, *t*s > .10, indicating that male and female students were highly accurate in predicting their performance.

Table 2.
Mean (SD) of Self-Efficacy (range 1 to 9) and Perceived Competence (range 1 to 7) Scores and Judgment of Learning (range 0 to 8) and Judgment of Learning Accuracy (range -12 to 12) per Condition.

	Male Observer		Female Observer	
	Male Model	Female Model	Male Model	Female Model
Self-efficacy Pretest	4.57 (2.06)	4.32 (2.03)	3.73 (1.88)	3.49 (1.75)
Self-efficacy Posttest	6.12 (1.37)	5.87 (1.28)	5.39 (1.19)	5.28 (1.52)
Perceived Competence Pretest	3.67 (1.33)	3.63 (1.52)	3.06 (1.43)	3.29 (1.54)
Perceived Competence Posttest	5.30 (1.06)	4.72 (1.21)	4.79 (1.25)	4.63 (1.49)
Judgment of Learning	4.12 (1.21)	4.05 (1.06)	3.70 (0.98)	3.86 (1.25)
Judgment of Learning Accuracy	0.79 (3.81)	1.13 (4.51)	-0.51 (3.56)	0.47 (3.75)

Instruction Evaluation

The 2 (Gender Model: male, female) X 2 (Gender Observer: male, female) ANOVA on how enjoyable watching the video examples was (Question 4) showed no main effects of Gender Model and Gender Observer, *F*s < 1. There was, however, a significant interaction effect between Gender Model and Gender Observer, $F(1, 159) = 4.27$, $p = .040$, $\eta_p^2 = .026$. To explore this interaction effect, we firstly examined the effects of Model Gender for each Observer Gender condition separately, but none of the effects were significant. However, when the effects of Observer Gender were compared for each Model Gender separately, it was found that learning from a male model was significantly more enjoyable for male students ($M = 5.47$, $SD = 2.45$) than for female students ($M = 4.46$, $SD = 2.27$), $t(80) = 2.13$, $p = .036$, $d = 0.428$ (medium effect; Cohen, 1988).

With respect to the degree to which participants preferred to receive instruction in a similar manner in the future, the same pattern of results was found as on the learning enjoyment question. Again, we found no main effect of Gender Model, $F < 1$, nor of Gender Observer, $F(1, 159) = 1.45$, $p = .230$, but there was a significant interaction between Gender Model and Gender Observer, $F(1, 159) = 4.02$, $p = .047$, $\eta_p^2 = .025$. When investigating the effects of Model Gender for each Observer Gender condition separately, no effects were found, but when we compared the effects of Observer Gender for each Model Gender separately, it was found that observing a male model caused male students (M = 7.13, SD = 2.51) to be significantly more positive about receiving similar instruction in the future than female students (M = 5.82, SD = 2.45), $t(80)$ = 2.39, p = .019, d = 0.528 (medium effect; Cohen, 1988).

Correlations were computed between the invested mental effort ratings for learning the video content and the two instruction evaluation questions because these constructs showed a very similar pattern of results (i.e., significant interaction effects showing a very similar pattern). Surprisingly, effort invested in learning did not significantly correlate with how enjoyable watching the videos was, r = -0.17, p = .831, nor with the degree to which participants preferred to receive similar instruction in the future, r = -0.10, p = .204.

Table 3.
Mean (SD) of Learning Enjoyment and Willingness to Receive Similar Instruction (ranges 0 to 10) Scores per Condition.

	Male Observer		Female Observer	
	Male Model	Female Model	Male Model	Female Model
Learning Enjoyment	5.36 (2.50)	4.47 (2.46)	4.36 (2.27)	4.98 (2.74)
Willingness to Receive Similar Instruction	6.98 (2.63)	6.18 (2.74)	5.82 (2.45)	6.51 (2.72)

Discussion

This experiment investigated whether it would be more effective for secondary education students to study a video modeling example in which it was demonstrated how a math problem should be solved, with a same-gender model than an opposite gender model, as the model-observer similarity hypothesis would predict (Schunk, 1987, 1991a). With respect to cognitive aspects of learning, the results showed that, as expected, example study was effective for fostering learning and near transfer (i.e., high gains from pretest to posttest; Hypothesis 1a), regardless of the model’s or the observer’s gender. That is, gender did not affect the degree to which students improved their performance (Question 2a).

As one would expect, given the knowledge gains, the amount of mental effort students had to invest in solving the probability calculation problems decreased from pretest to posttest (Hypothesis 1c), and this effort reduction was not affected by gender either (Question 2c). In accordance with Hypothesis 1b, students invested a low/medium degree of effort during example study. There were, however, differences in the effort invested during example study as a function of model/observer gender (Question 2b). For male students it was less effortful to study a male model than a female model and observing a male model was less effortful for male students than female students (both medium effect sizes). This indicates that the learning process was more efficient for male students who observed a male model compared to female students and compared to male students observing a female model (see Van Gog & Paas, 2008, for a discussion of efficiency in terms of the relation between mental effort and performance).

The affective variables of self-efficacy and perceived competence, both of which have been associated with better learning outcomes (Bandura, 1997; Bong & Skaalvik, 2003; Harter, 1990; Ma & Kishor, 1997; Schunk, 2001), were also enhanced from pretest to posttest (Hypothesis 1d), although no effect of model-observer similarity was found (Question 2d). Male students did show higher self-efficacy than female students, which was, however, not associated with higher learning outcomes. This may be a consequence of the stereotypical perception that males are more competent in math than female students (Steffens et al., 2010), particularly among older students (Ceci et al., 2014), although typically very few, if any, actual performance differences are found between the genders (Hyde, Fennema, & Lamon, 1990; Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Although the findings on perceived confidence combined with performance suggest that male students may have overestimated their performance, the judgment of learning accuracy results show that gender did not affect how accurate participants were at judging their own skills (Question 3). The stereotype that males are better than females at math could also explain why observing a male model enhanced perceived competence more from pretest to posttest than observing a female model, for both male and female students. Perhaps all students saw the male model as more of an expert than the female model (despite the fact that the content of the examples was fully identical) at this stereotypically male task. This is in line with findings on the effectiveness of animated pedagogical agents (e.g., Arroyo et al., 2009; Moreno et al., 2002).

We also found gender effects on both learning enjoyment and willingness to receive similar instruction in the future (Question 4), which may be indicators of how students might use such examples during future self-study online (Yi & Hwang, 2003). Results showed that studying a male model was more enjoyable for male students than female students and caused male students to be more positive about receiving similar instruction in the future than female students (both medium effects). While at first sight the pattern on the instruction evaluation questions and invested mental effort during learning appear identical, these measures did not correlate, indicating separate effects.

In sum, our results suggest that the gender of the model in video examples does not affect learning outcomes, but may influence affective aspects of learning. Notably, our study kept the content of the example videos entirely equal across conditions, so these effects only result from the differences in models. Effects on affective variables are important as these might influence students' self-study behaviour. So with video modeling examples being increasingly used in online learning environments, as they have become much easier to create and share, instructional designers creating such environments may want to consider the effects of model gender on male and female students' affect. Given that learning outcomes did not differ, but perceived competence was higher for students who studied a male video model, educational practitioners could give preference to designing and using video modeling examples with a male model when students learn a task that is associated more with males than females. However, given that students' gender interacted with the gender of the models on the evaluation of the instruction and on invested mental effort during example study, it is likely more advisable to create both a male and a female model version with identical content. These videos could be distributed to the learners via either an adaptive system that allocates students a male or female model depending on their own gender, or by allowing students to choose the model they want to learn from. The latter would have the added benefit of giving students an extra opportunity of regulating their own study behaviour, which should increase feelings of autonomy and thereby possibly raise their motivation and self-efficacy (Bandura, 2001; Behrend & Thompson, 2012; Clark & Mayer, 2011; Ryan & Deci, 2000). A similar argument has previously been made in the animated pedagogical agent literature (Ozogul et al., 2013). Because the gender of the model in a video modeling example does not seem to affect students' test performance, there seems to be no harm in providing students with the opportunity to choose the gender of their model, although future research should first examine whether our findings are replicated using tasks from other domains and over longer study periods.

Given that we used a single example, future research should also explore effects of the model's gender in relation to students' gender with multiple models to investigate whether the effects on affective variables would become stronger or weaker over time and if they would become stronger, whether they start to influence learning outcomes over time. It would also be interesting to compare effects of a set of examples by multiple male or female models to a mixed set of examples by male and female models.

Chapter 4

Learning from video modeling examples: Content kept equal, adults are more effective models than peers

This chapter has been published as:

Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction, 44*, 22-30. doi:10.1016/j.learninstruc.2016.02.004

Abstract

Learning from (video) modeling examples in which a model demonstrates how to perform a task is an effective instructional strategy. The model-observer similarity (MOS) hypothesis postulates that (perceived) similarity between learners and the model in terms of age or expertise moderates the effectiveness of modeling examples. Findings have been mixed, however, possibly because manipulations of MOS were often associated with differences in example content and manipulations of (perceived) expertise confounded with age. Therefore, we investigated whether similarity with the model in terms of age and putative expertise would affect cognitive and motivational aspects of learning when the example content is kept equal across conditions. Adolescents (*N*=157) watched a short video in which a peer or adult model was introduced as having low or high expertise, followed by two video modeling examples in which the model demonstrated how to troubleshoot electrical circuit problems. Results showed no effects of putative expertise. In contrast to the MOS hypothesis, adult models were more effective and efficient to learn from than peer models.

Introduction

Instructional videos are rapidly gaining popularity in education. They form the backbone of massive open online courses (MOOCs) and blended courses, and support students during self-study at home or at school. Next to web lectures (e.g., Chen & Wu, 2015; Korving, Hernández, & De Groot, 2016; Traphagan, Kucsera, & Kishi, 2010) and short knowledge clips (e.g., Day, 2008), demonstration (i.e., "how-to") videos (e.g., Ayres, Marcus, Chan, & Qian, 2009; Van der Meij & Van der Meij, 2013) make up an important part of the instructional videos on offer. Such demonstration videos are also known as *video modeling examples*. Research inspired by Bandura's (1977, 1986) social learning theory has shown the effectiveness of observational learning from human models, and this dovetails nicely with findings from cognitive psychology and instructional design research (e.g., Anderson, 1993; Sweller, Ayres, & Kalyuga, 2011) that has shown the effectiveness of example-based learning (for reviews: Renkl, 2014; Sweller et al., 2011; Van Gog & Rummel, 2010).[1]

Video modeling examples in which a model demonstrates and explains how to solve a problem are effective for acquiring new skills (e.g., Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Schunk, Hanson, & Cox, 1987; Schwan & Riempp, 2004; Van Gog, Verveer, & Verveer, 2014) and may enhance the confidence learners have in their own capabilities to perform the modeled task (i.e., self-efficacy and perceived competence; Bandura, 1997; Hoogerheide, Loyens, & Van Gog, 2014a, 2016; Schunk & Hanson, 1985). Yet, when developing video modeling examples, several design choices have to be made that may influence their effectiveness, the most salient of which is the choice of model. The present study investigates whether similarity between the learner and the model in terms of age and (putative) expertise would affect self-efficacy and learning outcomes, as predicted by the model-observer similarity hypothesis.

The Model-observer Similarity Hypothesis

The model-observer similarity (MOS) hypothesis (Bandura, 1994; Schunk, 1987; see also the similarity-attraction hypothesis, Montoya & Horton, 2013; Moreno & Flowerday, 2006; Reeves & Nass, 1996) postulates that, because modeling enables social comparison (Berger, 1977; Johnson & Lammers, 2012), the effectiveness of observational learning from (video) modeling examples depends in part on how similar to the model learners perceive themselves to be. Or in Bandura's (1994) words:

[1] Note that examples can lose their effectiveness or may even hamper learning when students have some prior knowledge of the problem (Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Kalyuga & Renkl, 2010).

> 'The impact of modeling on perceived self-efficacy is strongly influenced by perceived similarity to the models. The greater the assumed similarity, the more persuasive are the models' successes and failures. If people see the models as very different from themselves their perceived self-efficacy is not much influenced by the models' behavior and the results it produces.' (p.72)

Self-efficacy and the closely related construct of perceived competence are important, as they have been linked to factors such as academic motivation (Self-efficacy: Bandura, 1994; Schunk, 1991b, 2001; Schwarzer, 1992; Perceived competence: Bong & Skaalvik, 2003; Harter, 1990) and learning outcomes (Self-efficacy: Bandura, 1994; Schwarzer, 1992; Perceived competence: Bong & Skaalvik, 2003; Harter, 1990; Ma & Kishor, 1997). Learners who perceive themselves as more similar to the model may also feel more attracted to the model and pay more attention to the model (Berscheid & Walster, 1969), and a high degree of similarity can help them form outcome expectations (Schunk, 1987). Similarity factors may be particularly important for novice learners whose self-efficacy and prior knowledge are still low, as they are especially prone to engaging in social comparison (Buunk, Zurriaga, Gonzalez-Roma, & Subirats, 2003). The present study focuses on MOS in terms of age and putative expertise.

Model-observer Similarity in Age and (Perceived) Expertise

With regard to the age of a model, the MOS-hypothesis predicts that primary or secondary education students would benefit more from a model that is perceived as similar in age, such as a peer model, than dissimilar in age, such as an adult model. Findings have been mixed however, with some studies showing stronger effects of observing a peer model compared to an adult model (e.g., Davidson & Smith, 1982; Rodriguez Buritica, Eppinger, Schuck, Heekeren, & Shu-Chen Li, 2015; Schunk & Hanson, 1985; Zmyj, Aschersleben, Prinz, & Daum, 2012), some showing no differences (Robert, 1983; Strauss, 1978), and others showing stronger effects of an adult model (e.g., Hicks, 1965; Jakubczak & Walters, 1959). A possible explanation for these mixed findings may be that peer models are especially beneficial for learners who have encountered difficulties in learning or for learners of low ability (Schunk, 1987). Schunk and Hanson (1985), for instance, examined whether children who previously showed difficulties learning fractions benefited more from a peer model, a teacher model, or no model, and found that peer modeling was more conducive to both self-efficacy and learning than teacher modeling, while both models were more effective than no modeling. Another possible explanation is that age only becomes a salient cue when coupled with (perceived) expertise. That is, students may particularly imitate peer models when they believe them to be high in expertise, and age may become an informative cue especially for tasks in which peers are generally (perceived as) less of an expert than adults (Bandura, 1986; Schunk, 1987).

Research on the MOS-hypothesis in terms of expertise has used different approaches. One line of research contrasted learning from a mastery model (i.e., a model who displays faultless performance from the start) to learning from a coping model (i.e., a model who shows performance errors that he or she corrects later on), and this has led to mixed results. For instance, in math, no differences in the effectiveness of both model types were found for low ability students who had had prior successful experiences with the task (e.g., Schunk & Hanson, 1985) or for average ability students (Schunk & Hanson, 1989). However, for low ability students without prior success with the task, coping models were more effective for learning (Schunk et al., 1987).

Another line of research has compared the effects of learning from a high expertise (e.g., expert) model to a lower expertise (e.g., advanced student) model, the latter being closer in knowledge and skill to novice learners. Contrary to the model-observer similarity hypothesis, older findings indicate that for primary school children, a more expert model was more beneficial for a wide range of measures such as learning communication skills or paired-associates relative to a low expertise model (e.g., Simon, Ditrichs, & Speckhart, 1975; Sonnenschein & Whitehurst, 1980). In line with the MOS-hypothesis, however, Braaksma, Rijlaarsdam, and Van den Bergh (2002) showed more recently that secondary education students who had weak writing skills benefitted more from being instructed to focus on weak models who explained and demonstrated how to write an argumentative text (on video) than from focusing on strong models, whereas the reversed effect was found for more competent students. Studies in higher professional education, however, showed no benefit of (advanced) peer models: written examples created by experts fostered transfer (i.e., applying the acquired knowledge to novel tasks) more than examples created by advanced peer students, possibly because experts' explanations contain a higher degree of abstraction (Boekhout, Van Gog, Van de Wiel, Gerards-Last, & Geraets, 2010; Lachner & Nückles, 2015).

Clearly, findings regarding both age and expertise have been mixed. There are two important things to note, however. First, in many of those studies, there were actual differences in how the models behaved across conditions or in other words, in the content of the examples. This applies, for instance, to studies that contrasted learning from coping models and mastery models because only coping models' behaviour contains expressions of uncertainty and/or errors (e.g., Kitsantas, Zimmerman, & Cleary, 2000; Schunk & Hanson, 1985; Zimmerman & Kitsantas, 2002), and to studies that compared high and lower expertise models because their explanations differ in quality (e.g., Lachner & Nückles, 2015; Simon et al., 1975; Sonnenschein & Whitehurst, 1980). This makes it hard to evaluate whether any differences in motivational or learning outcomes were due to (perceived) similarity or to differences in content. Some evidence indicating that perceived similarity may still influence cognitive, affective, or motivational aspects of learning when all else is equal, comes from studies with *animated* models (i.e., animated pedagogical agents) in which the content was kept equal. For instance, Rosenberg-Kima, Baylor, Plant, and Doerr (2008) found that self-efficacy was enhanced more for students who learned about engineering from a 'young and cool' agent than a 'young and uncool' and an 'older and (un)cool' agent. Liew, Tan, and Jayothisa (2013) found that for female university students, a peer-like agent was more enjoyable to learn programming skills from than an expert-like agent, although the expert-like agents were more credible and led to less anxiety during learning, possibly because people are more easily persuaded by those whom they perceive as experts (Chaiken & Maheswaran, 1994; Debono & Harnish, 1988). Lastly, Kim, Baylor, and Reed (2003) found that a mentor-like agent was as beneficial for learning compared to an expert-like agent, but was considered more motivating to interact with and learn from.

Secondly, age and expertise manipulations were often confounded. For example, Davidson and Smith (1982) investigated the relationship between model expertise and children's self-evaluation skills, and, instead of keeping model age constant across conditions, children observed a peer of equal skill, an adult of superior skill, or a child of inferior skill. Animated agent studies have also confounded age and expertise manipulations. In the studies of Kim et al. (2003) and Baylor and Kim (2004), for instance, the expert-like agent looked much older than the mentor-like agent. Some early video modeling example studies have tried to disentangle the effects of age and expertise. For instance, Sonnenschein and Whitehurst (1980) showed that observing high ('informative') and low expertise ('uninformative') peer models and high expertise adult models enhanced children's communicative skills more than watching low expertise adult models. Children did, however, evaluate the low expertise adult as more knowledgeable than the low expertise peer. Becker and Glidden (1979) found that expertise and age interacted for low ability children. Children observed a low or high expertise peer or adult model performing a motor task while displaying certain social behaviors. The behavior of high expertise models and peer models was imitated more than that of low expertise models and adult models, presumably because the social behavior of the peers was evaluated as more appropriate. Note though, that in these studies, the example content again varied across conditions, and it is therefore uncertain whether these effects were caused by differences in perceived similarity to the model.

In sum, it is unclear whether similarity to a model in terms of age and (perceived) expertise would play a role in learning when the content of the examples would be kept equal and whether age and perceived expertise of the model contribute independently to effects on motivation and learning outcomes or only in interaction. Therefore, the present study examined whether, when the content of the example is controlled for, the effectiveness of studying video modeling examples for novice students' perceptions of their own capabilities to perform the modelled task (i.e., perceived competence and self-efficacy) and learning outcomes (i.e., posttest performance) depends on whether the model is of similar or dissimilar age and whether the model is introduced as having low or high expertise.

The Present Study

We addressed the question of whether model-observer similarity in age, putative expertise, or both would affect novice secondary education students' learning (i.e., adolescents of about 15 years of age who did not have prior knowledge of the task). They studied two video modeling examples on how to solve a science problem (troubleshooting electrical circuits). The models were either peers (17 years old) or adults (42 years old) who were introduced prior to example study as being enrolled in a tutor-training (peers) or teacher-training (adults) program and as having low expertise or high expertise in science. We kept all else equal, both with respect to model characteristics (i.e., all models were Caucasian females from the same region of the country, wearing a black t-shirt and blue jeans) and the content of the videos (i.e., all models narrated the exact same text, spent an equal amount of time on the parts of the video and the video as a whole, and were trained to show the same movements and gestures). Moreover, to ensure that any effects of condition were not associated with one particular model, the two adult and the two peer models featured in both the high and low expertise conditions (i.e., half of the participants in the low expertise adult condition saw "adult model 1" the other half "adult model 2").

The primary research question was whether students would perform better on the posttest and show greater self-efficacy and perceived competence when they were more similar to the model in age, expertise, or both. Given that students are novices with regard to the modeled task, the MOS-hypothesis (Schunk, 1987) would predict that students' self-efficacy, perceived competence, and learning outcomes would benefit most from studying a peer model with low putative expertise. Because of the fact that prior research has produced mixed findings, often confounding age and expertise or expertise and example content, however, we are hesitant to adopt the MOS-hypothesis for the present study and rather approach this as an open question. We also measured mental effort invested during example study and the posttest to obtain more information on the cognitive efficiency of the instructional conditions (Van Gog & Paas, 2008). Effects on learning enjoyment were also explored because previous studies have shown influences on affect (e.g., Kim et al., 2003; Liew et al., 2013) and enjoyment may be an important cue for whether students would use examples during self-study (Yi & Hwang, 2003). Lastly, students evaluated the quality of the model's explanation.

Method

Participants and Design

Participants were 157 Dutch secondary education students (82 male; M^{age} = 14.99 years, *SD* = 0.64) in their third or fourth year of pre-university education. The experiment used a 2 x 2 design, with Model Age (Peer vs. Adult) and Model Expertise (Low vs. High) as between-subject factors. Students were quasi-randomly (i.e., matched for gender) allocated to the Low Expertise Peer (*n* = 39, 21 males), High Expertise Peer (*n* = 40, 21 males), Low Expertise Adult (*n* = 38, 21 males), or High Expertise Adult (*n* = 40, 21 males) Model conditions. There were two adult and two peer models, featuring in both the High and Low Expertise conditions. Within each condition, half of the students received one model, the other half the other model (e.g., half of the participants in the Low Expertise Adult Model condition saw "adult model 1" the other half "adult model 2"). At the time of the experiment, students had taken basic science classes but were novices with regard to the modelled task (troubleshooting electrical circuits) as this had not yet been covered in their curriculum according to the teachers.

Materials

The materials for this study were based on the pen-and-paper materials on troubleshooting parallel electrical circuits from prior studies on example-based learning (e.g., Hoogerheide, Loyens, Jadi, Vrins, & Van Gog, 2015; Van Gog & Kester, 2012; Van Gog, Kester, Dirkx, Hoogerheide, Boerboom, & Verkoeijen, 2015; Van Gog, Kester, & Paas, 2011), but were presented online in the web-based Qualtrics platform (http://www.qualtrics.com).

Conceptual Prior Knowledge Test. The prior knowledge test consisted of seven conceptual open-ended questions on troubleshooting and parallel circuits principles. This test was used as a check that students indeed had little if any prior knowledge of the principles required for troubleshooting parallel electrical circuits (e.g., relations between voltage, current, and resistance in parallel circuits) and to rule out differences among conditions in prior knowledge.

Introductory Text. A short introductory text explained what the abbreviations and components in a circuit drawing stand for and described Ohm's law and the three different forms of the formula (i.e., R=U/I; I=U/R; U=I*R).

Video Modeling Examples. Two video modeling examples were created by all four models (i.e., the two Peer and the two Adult Models) under both High and Low Expertise conditions. In the first example (240 s), the fault was that the measured current in one of the parallel branches was higher than one would expect, meaning that the resistance in that branch was lower; in the second example (244 s), the current was lower, meaning that resistance was higher. Each example showed the model standing to the right of a large screen displaying PowerPoint slides (see Figure 1). Each example began with a circuit drawing containing three parallel branches that was presented on the screen; the circuit indicated how much resistance each resistor provided as well as how much voltage the power source delivered. The model explained based on this circuit drawing that the information on voltage and resistance presented in the drawing can be used to calculate what current should be measured in all three parallel branches and overall if the circuit were functioning correctly. The model then provided a step-by-step demonstration of how to calculate the current in each branch as well as the total current (sum of the currents in the branches) using Ohm's law; this explanation was supported by a slide that showed the same circuit drawing (only smaller), Ohm's law, and the worked-out problem-solving steps. The next slide presented measured current at each ammeter below the currents that should be measured if the circuit were functioning correctly and the model pointed out the discrepancy in one of the branches (i.e., either higher in example 1 or lower in example 2) and explained that this meant the resistance was lower (example 1) or higher (example 2) than indicated in the drawing and demonstrated how to calculate the actual resistance, supported by a slide displaying the measured currents, Ohm's law, and the calculation.

Age Manipulation. Two adolescents (real ages 16 and 17) served as a peer model (introduced as being 17 years old) in both Peer Model conditions, and two adults (real ages 42 and 43) as an adult model (introduced as being 42 years old) in both Adult Model conditions.

Expertise Manipulation. A short video was created in which the models introduced themselves, to be presented prior to the examples. Each model created an introduction for both the Low and High Expertise condition (both 40 s). This video showed the model standing next to the screen (i.e., the setup of the video modeling examples), but the screen was empty. The introduction of the peer models started as follows: ‘My name is Natasja/Denise. I am 17 years old and I am enrolled in a homework tutor-training program. For a course within that program, I was instructed to create instructional videos that can be used for homework purposes.’ The introduction of the adult models started with: ‘My name is Natasja/Denise. I am 42 years old and I am enrolled in a lateral-entry teacher-training program. For a course within that program, I was instructed to create instructional videos that can be used for homework purposes.’ All introduction videos then continued: ‘Therefore, you will be shown two examples that demonstrate how to detect and solve a problem in an electrical circuit. Afterwards, you will be asked to solve similar problems yourself to see how much you have learned from my explanation.’ Next, the low vs. high expertise manipulation followed, with the Low Expertise introduction stating ‘I hope that I can explain this clearly, as I am not so proficient in physics and I am not taking [peer] / did not take [adult] physics as a final examination subject in secondary education[2]. But I will do my best’ while the High Expertise introduction was ‘I expect that I can explain this clearly, as I am very proficient in physics and I am taking [peer] / did take [adult] physics as a final examination subject in secondary education. So I will do my best’.

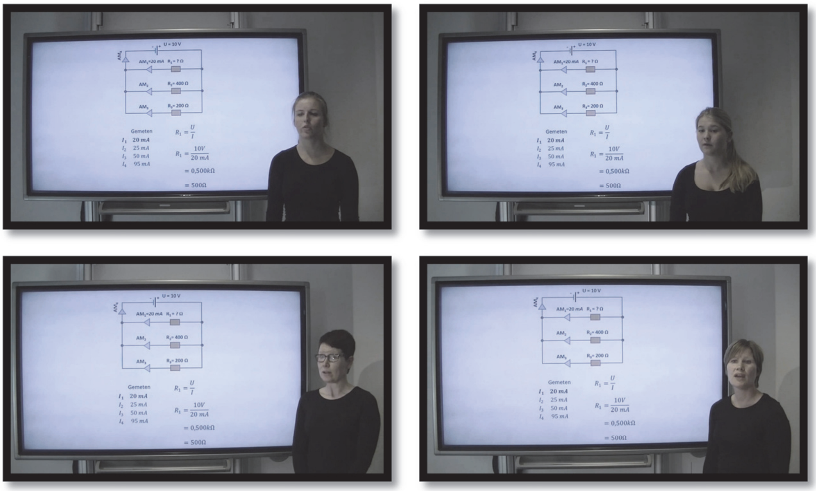


Figure 1. Peer models (top row) and adult models (bottom row)

[2] In the Netherlands, after some general years with a common curriculum, secondary education students can chose ‘profiles’ of subjects in which they will take the final examination, and not all of those include physics.

Posttest. The posttest presented two troubleshooting problems. The first one was isomorphic to the problems used in the video modeling examples (i.e., one fault), but the second problem was slightly different in the sense that it contained both faults that had been encountered in the training. Both problems reminded students that: 'The current (U) is expressed in volt (V), resistance (R) is expressed in Ohm (Ω), and power (I) is expressed in amperes (A).'

Mental Effort. After each video modeling example and each posttest task, invested mental effort was measured using the rating scale developed by Paas (1992), which ranges from (1) very, very low effort to (9) very, very high effort.

Self-efficacy and Perceived Competence. Self-efficacy was measured by asking students how confident they were that they had mastered the skill of detecting and solving electrical circuit problems (cf. Bandura, 2006), on a scale of 1 (not at all confident) to 9 (very, very confident). Perceived competence was measured using an adapted version of the Perceived Competence Scale for Learning of Williams and Deci (1996). Participants were asked to rate on a scale of (1) not at all true to (7) very true to what degree the following items apply to them: 'I feel able to meet the challenge of performing well in detecting and solving electrical circuit problems', 'I feel confident in my ability to detect and solve electrical circuit problems', and 'I am capable of detecting and solving electrical circuit problems'. That is, the adaptation consisted of rephrasing the questions to focus on detecting and solving electrical circuit problems and of removing the item 'I am able to achieve my goals in this course' because it did not apply to our study context.

Learning Enjoyment. Participants were asked to give a 'school-grade' on a scale of 0 (lowest) to 10 (highest) for how enjoyable studying the video modeling examples was for them (cf. Hoogerheide et al., 2014a).

Explanation Quality. Participants were asked to rate the quality of explanations provided in the video modeling examples on a scale of (1) very, very bad quality to (9) very, very good quality.

Procedure

Prior to the experiment, participants were quasi-randomly (i.e., matched for gender) allocated to one of four conditions and, within each condition, to one of the two peer or adult models, based on a name list. This was done to ensure that all conditions and models contained an approximately equal number of students and ratio of male to female students. The experiment was run in 8 sessions of ca. 50 min. duration in a computer lab at participants' schools. Participants were told to sit at the computer that was marked with their name on a sheet of A4 paper; that sheet also contained the link to the Qualtrics questionnaire of their assigned condition. The experimenter first gave a brief plenary general introduction and instructed students how to access the Qualtrics questionnaire. This questionnaire presented 4 ‘blocks’ of questions. Block 1 contained demographic questions and the conceptual prior knowledge test followed by self-efficacy and perceived competence ratings. Students were given 6 min. to complete this block. Next, students were instructed to study the introductory text for 2 minutes in block 2, and the experimenter emphasized that students needed to study this information carefully to be able to comprehend the demonstration videos later on. Block 3 first presented the expertise manipulation video in which the model introduced herself, followed by the two video modeling examples (the videos were embedded in Qualtrics via YouTube). After each example, participants were asked to rate how much effort they invested in studying the example. At the end of block 3, participants rated their learning enjoyment, the perceived quality of the explanations provided in the examples, and their self-efficacy and perceived competence. Block 4 presented the posttest for which participants received 12 minutes; after each test problem participants were asked to rate how much effort they invested in solving it.

Data Analysis

Averages were computed for effort invested during example study, effort invested in the posttest, perceived competence after the prior knowledge test, and perceived competence prior to the posttest. Test performance was scored based on straightforward coding schemes that had been developed and used in prior research (e.g., Hoogerheide et al., 2015; Van Gog et al., 2011; Van Gog et al., 2015; Van Gog & Kester, 2012). Ten points could be earned for the prior knowledge test (and partial credit was given for partially correct answers). The maximum score to be earned for the posttest was eight points. For the first task that contained only one fault, one point could be earned for calculating the correct value of all ammeters, one for indicating which resistor was faulty, and one for indicating what the faulty resistor’s actual value was. For the second task that contained two faults, an extra point was granted for correctly indicating the second faulty resistor and for correctly calculating its resistance. Incomplete or partially correct answers were given half a point. Two raters scored 10% of the pretests and posttests, and Cohen's κ was run to determine if there was agreement between them (on item level). There were high levels of agreement between the two raters on the pretest scores, $\kappa = .909$, $p < .001$, and on the posttest scores, $\kappa = .922$, $p < .001$. Therefore, the remainder of the tests was scored by a single rater.

One student in the High Expertise Adult Model condition was removed from all the analyses due to non-compliance with the instructions, leaving 156 participants. One participant from the Low Expertise Adult Model condition had to leave early and was therefore excluded from the posttest and invested mental effort in the posttest analyses. Three participants had one missing value on invested mental effort in the posttest, which were replaced by the series mean.

Results

Two types of preliminary analyses were conducted. First, it was investigated whether there was a difference between the two peer models and the two adult models by means of independent samples *t*-tests on all outcome variables. Because there were no significant differences, we proceeded analysing the data at condition level. Second, we checked whether participants' prior knowledge was indeed low (which it was, as can be seen in Table 1) and did not differ among conditions. Indeed, a 2 x 2 ANOVA on the prior knowledge test scores with Model Age (Peer, Adult) and Model Expertise (Low, High) as between-subject factors revealed no main or interaction effects (all *F*s < 1).

Test performance, invested mental effort, learning enjoyment, and explanation quality results were analysed using 2 x 2 ANOVAs, with Model Age (Peer, Adult) and Model Expertise (Low, High) as between-subject factors. The self-efficacy and perceived competence results were analysed using repeated measures ANOVAs with Test Moment (Before and After Example Study) as within-subjects factor and Model Age (Peer, Adult) and Model Expertise (Low, High) as between-subjects factors. Mean (and SD) scores on all variables are shown in Table 1.

Posttest Performance

There was a main effect of Model Age on posttest performance, $F(1,151) = 5.92$, $p = .016$, $\eta_p^2 = .038$, indicating that participants who had observed an adult model ($M = 4.14$, $SD = 2.44$) outperformed those who had observed a peer model ($M = 3.20$, $SD = 2.34$). There was no main effect of Model Expertise, $F(1,151) = 1.45$, $p = .231$, $\eta_p^2 = .009$, nor an interaction effect, $F < 1$. [3]

[3] Upon a reviewer's request we explored whether there were differences among conditions in the degree to which participants correctly solved each problem as a whole. Chi-square tests showed that the number of students who managed to correctly solve the problem in its entirety (i.e., max score) did not differ among conditions (problem 1: $X^2(1, N = 156) = 0.01$, $p = .92$; problem 2: $X^2(1, N = 156) = 1.71$, $p = .19$).

Mental Effort

We found a main effect of Model Age on mental effort invested during example study, $F(1,152) = 4.84$, $p = .029$, $\eta_p^2 = .031$, indicating that participants who observed an adult model ($M = 3.43$, $SD = 1.56$) invested less effort than those who observed a peer model ($M = 4.03$, $SD = 1.82$). There was no main effect of Model Expertise, nor an interaction effect, both *F*s < 1. With regard to mental effort invested in completing the posttest tasks, there were no significant main or interaction effects (Model Age: $F < 1$; Model Expertise: $F(1,151) = 1.27$, $p = .263$, $\eta_p^2 = .008$; interaction: $F < 1$).

Table 1.
Mean (SD) of Test Performance, Invested Mental Effort, Self-efficacy, Perceived Competence, Explanation Quality, and Learning Enjoyment per Condition.

	Peer Model		Adult Model	
	Low Putative Expertise	High Putative Expertise	Low Putative Expertise	High Putative Expertise
Performance Pretest	1.59 (1.16)	1.61 (1.38)	1.79 (1.11)	1.49 (0.98)
Performance Posttest	2.90 (2.17)	3.50 (2.49)	3.97 (2.47)	4.29 (2.43)
Mental Effort Study Phase	3.96 (2.09)	4.10 (1.53)	3.55 (1.72)	3.31 (1.40)
Mental Effort Posttest	4.81 (2.01)	4.24 (1.72)	4.41 (1.97)	4.32 (1.51)
Self-efficacy Pretest	3.18 (1.73)	3.58 (1.68)	3.68 (1.65)	3.85 (2.05)
Self-efficacy Posttest	5.10 (1.93)	5.63 (1.17)	5.55 (1.39)	5.92 (1.36)
Perceived Competence Pretest	2.56 (1.30)	3.03 (1.08)	3.18 (1.16)	2.97 (1.39)
Perceived Competence Posttest	4.08 (1.47)	4.48 (0.99)	4.24 (1.34)	4.51 (1.32)
Explanation Quality	5.62 (1.60)	6.20 (1.29)	6.34 (1.17)	6.59 (1.14)
Learning Enjoyment	4.00 (2.24)	4.08 (2.31)	4.45 (2.33)	4.56 (2.28)

Self-efficacy and Perceived Competence

The analysis of self-efficacy showed a main effect of Test Moment, $F(1,152) = 233.53$, $p < .001$, $\eta_p^2 = .606$, indicating that self-efficacy improved from before ($M = 3.57$, $SD = 1.78$) to after example study ($M = 5.55$, $SD = 1.50$). Other than that, there were no significant main effects (Model Age: $F(1,152) = 2.77$, $p = .098$, $\eta_p^2 = .018$; Model Expertise: $F(1,152) = 2.51$, $p = .115$, $\eta_p^2 = .016$) or interaction effects (all *F*s < 1).

A similar pattern was found for students' perceptions of their own competence. There was a main effect of Test Moment, $F(1,152) = 244.36$, $p < .001$, $\eta_p^2 = .617$, indicating that perceived competence improved from before ($M = 2.94$, $SD = 1.25$) to after example study ($M = 4.33$, $SD = 1.29$). Other than that, there were neither main effects (Model Age: $F(1,152) = 1.06$, $p = .304$, $\eta_p^2 = .007$; Model Expertise: $F(1,152) = 1.74$, $p = .189$, $\eta_p^2 = .011$), nor interaction effects (Test Moment * Model Expertise, $F(1,152) = 1.30$, $p = .257$, $\eta_p^2 = .008$; Test Moment * Model Age, $F(1,152) = 1.09$, $p = .299$, $\eta_p^2 = .007$; Model Expertise *

Model Age, $F(1,152) = 1.24$, $p = .267$, $\eta_p^2 = .008$; Test Moment * Model Expertise * Model Age, $F(1,152) = 2.37$, $p = .126$, $\eta_p^2 = .015$).

Learning Enjoyment

There were no significant effects on learning enjoyment (Model Age: $F(1,152) = 1.63$, $p = .204$, $\eta_p^2 = .011$; Model Expertise: $F < 1$; interaction: $F < 1$).

Explanation Quality

Despite the fact that the models provided the exact same explanations, there was a main effect of Model Age on students' ratings of the quality of explanations provided in the examples, $F(1,152) = 7.06$, $p = .009$, $\eta_p^2 = .044$, indicating that students who had observed an adult model rated the explanations as being of higher quality ($M = 6.47$, $SD = 1.15$) than students who had observed a peer model ($M = 5.91$, $SD = 1.47$). Moreover, there was also a main effect of Model Expertise, $F(1,152) = 3.92$, $p = .049$, $\eta_p^2 = .025$, showing that students who observed low expertise models rated the explanations as being of lower quality ($M = 5.97$, $SD = 1.44$) than the explanations of high expertise models ($M = 6.39$, $SD = 1.22$). There was no interaction effect, $F < 1$.[4]

Discussion

This experiment examined whether similarity to a model in terms of age and (putative) expertise would affect secondary education students' learning from video modeling examples. Because prior research on the model-observer similarity (MOS) hypothesis led to mixed findings, and many studies confounded expertise manipulations with example content and age, we took care to keep the example content equal across conditions and to disentangle the age and perceived expertise factors. Moreover, we used two models in each condition to rule out that effects would be caused by incidental model characteristics.

Given that students were adolescents and novices with regard to the modeled task, the MOS-hypothesis (Schunk, 1987) would predict that students' self-efficacy, perceived competence, and learning outcomes would benefit most from studying a peer model with low (putative) expertise. Our results do not support this hypothesis. On the contrary, with regard to model age, we found the opposite of what the MOS-hypothesis would predict: learners who studied *adult* models invested less effort and attained better learning outcomes than those who studied peer models. Thus, an adult model was more effective

[4] Upon a reviewer's request we re-ran all analyses with students' gender and pretest scores as covariates. The ANCOVA led to the same outcomes as the ANOVA with regard to all dependent variables. Pretest scores were a significant predictor of posttest scores, effort invested during example study and the posttest, and self-efficacy and perceived competence (as one might expect); gender was not a significant predictor of any of the dependent variables.

to learn from and more efficient in the sense that higher test performance was achieved with less effort investment in example study (see Van Gog & Paas, 2008, for a discussion of efficiency in terms of the relation between mental effort and performance). Students also rated the adult models' explanations as being of higher quality. Note that these findings are quite remarkable, given that the content of the examples was exactly the same. They cannot be explained through increased self-efficacy or perceived competence, however, as all students' ratings increased after example study, but did not differ among conditions.

A possible explanation for the finding that adult models were more effective might lie in perceived age-appropriateness of the modeled task. It has been proposed (Bandura, 1986; Schunk, 1987; Zmyj & Seehagen, 2013) that adult models may be more beneficial than peer models for behaviours that are viewed as more appropriate for adults and in which adults are considered to be more of an expert. The tasks demonstrated by the model were in the domain of physics, and research has shown that students typically struggle with learning physics skills and often continue to experience difficulties after extensive instruction (Duit & Von Rhöneck, 1998; Fredette & Lockhead, 1980; McDermott & Shaffer, 1992; Shipstone, 1984). As such, they might have attributed more expertise to the adult models, which would explain why students who had observed adult models found the model's explanations to be of higher quality than those who observed peer models, even though –again– the peer and adult models provided the exact same explanations. This would also explain why we did not find effects on self-efficacy; while MOS-effects predominantly occur via enhanced self-efficacy, task-appropriateness effects may not (Schunk, 1987).

But how might task-appropriateness explain better learning outcomes? Although this is a question for future research to address definitively, the answer might lie in students' attention allocation during example study. Bandura (1977, 1986) postulated that paying attention to a model is an important prerequisite for being able to emulate the modeled behavior later on, and that, in addition to MOS, model characteristics can affect how much attention is paid to a model. Peer models might lead to focusing more on task-irrelevant aspects of the video such as the model's appearance rather than aspects of the video that contribute to building a cognitive schema as a result of increased interest in and attraction to the model due to higher levels of perceived similarity (Berscheid & Walster, 1969). Moreno and Flowerday (2006) made a similar argument in the animated pedagogical agent literature to explain why a similar-ethnicity agent hampered learning;

they suggested that students may have focused more on how the agent's appearance and behaviour represented them.

Next to peer models having 'negative' effects on attention, adults might have beneficial effects: students may find it easier and more natural to pay attention to adult models. Adolescents are used to learning from adults and to adults being more knowledgeable and therefore giving higher quality explanations when it comes to complex subjects such as physics, and the idea that this may enhance students' attention to what the model is saying resonates with findings from research on group interaction. It is well-established that group members are more influenced by those perceived as more knowledgeable (Bottger, 1984; Littlepage, Schmidt, Whisler, & Frost, 1995; Ridgeway, 1987), and individuals who observe group interactions on video have been shown to pay more visual attention to group members perceived as more knowledgeable (Cheng, Tracy, Kingstone, Foulsham, & Henrich, 2013). Although these findings cannot be directly translated to learning from modeling examples, as in group settings there are always multiple individuals that an observer or group member may pay attention to, they do suggest that attention processes are the key to the effect of adult models in our study. The hypothesis that students attribute more expertise to adult models and therefore pay more attention to them during example study also resonates well with the communication maxim's of Grice (1975), in which it is stated that it is a social rule in conversations to pay more attention to those expected to be more knowledgeable. Higher attention levels may prevent students' minds from wandering; especially for complex tasks, learners often have difficulties building an accurate cognitive schema of the task at hand because their mind wanders easily (Smallwood, Fishman, & Schooler, 2007; Smallwood & Schooler, 2015; Szpunar, Moulton, & Schacter, 2013). Of course, these attention explanations are tentative in nature and need to be examined in future research.

The model expertise manipulation, consisting of a brief introductory video in which the model stated she is (or is not) very proficient in physics and is (not) taking / has (not) taken physics classes, seems to have been effective in the sense that students evaluated the explanations provided by the low expertise models as being of lower quality (even though they were exactly the same as in the high expertise conditions). However, this manipulation did not result in differences in self-efficacy, perceived competence, mental effort, enjoyment, or posttest performance. It is possible that this manipulation was too subtle to affect attention processes; in contrast to age-related cues that are automatically processed upon seeing another person and were continuously available during example study, cues regarding expertise were not. Another possibility is that the expertise manipulation did not result in students perceiving the low expertise models as more similar to themselves than the high expertise models, but instead regarded them as lower in competence (i.e., having no credibility at all, due to the model's uncertainty about her own ability to explain the task), which could have affected students' willingness to listen to the explanation. Although they indeed rated the quality of the explanation as lower than students in the high expertise condition, there were no significant effects of model expertise on learning outcome. As such, it seems unlikely that the explanations provided by the low expertise models were discarded by students.

Note that we took care to use two different models in each condition, to decrease the likelihood that effects of model age or putative expertise would result from specific characteristics of the particular model in a condition. Nevertheless, we cannot definitively rule out that model characteristics could have affected our results. Because there were no significant differences between both peer models and between both adult models on any of the outcome measures, however, it seems unlikely that specific model characteristics would have a strong influence on the findings. Another limitation of this study is that we only used one type of learning task, so future research should investigate whether these results hold with different types of tasks and tasks from other domains. Moreover, we only asked students to rate the quality of the model's explanations, but not the model's expertise, the perceived similarity to the model, or the appropriateness of the task for the model. Such information would have been helpful in determining whether students actually viewed the low perceived expertise models as similar, and whether the effect of model age would indeed be due to attributions of task appropriateness or expertise. An interesting avenue for future research would be to compare the effects of learning from video modeling examples with peer models to adult models for students learning a task that they view as more appropriate for their own age. If students' views of the age appropriateness of troubleshooting electrical circuit problems indeed caused adults to be more effective models than peers, then peers can be expected to be more effective models than adults in this case. Considering that views of task-appropriateness and expertise may be different for students of different ages, it would be interesting to investigate whether these findings generalize to different age groups. Moreover, it would be interesting to investigate whether these findings extend to other kinds of instructional video, such as short knowledge clips (e.g., Day, 2008) or web lectures (e.g., Korving et al., 2016; Traphagan et al., 2010).

Despite these limitations, our findings are of interest for educational practice. With video modeling examples being increasingly used in online learning environments because they have become easier to create, instructional designers and educational practitioners may want to design and use video modeling examples with an adult model rather than a peer model when the skill to be learned is viewed as more appropriate for adults because they are perceived as more of an expert.

Chapter 5

Testing the model-observer similarity hypothesis with text-based worked examples

This chapter has been accepted for publication as:

Hoogerheide, V., Loyens, S. M. M., Jadi, F., Vrins, A., & Van Gog, T. (in press). Testing the model-observer similarity hypothesis with text-based worked examples. *Educational Psychology.* Advance online publication (2015), doi:10.1080/01443410.2015.1109609

Abstract

Example-based learning is a very effective and efficient instructional strategy for novices. It can be implemented using text-based worked examples that provide a written demonstration of how to perform a task, or (video) modeling examples in which an instructor (the 'model') provides a demonstration. The model-observer similarity (MOS) hypothesis predicts that the effectiveness of modeling examples partly depends on the degree to which learners perceive the models to be similar to them. It is an open question, however, whether perceived similarity with the person who created the example, would also affect learning from text-based worked examples. Therefore, two experiments were conducted to investigate whether MOS would also play a role in learning from worked examples. In Experiment 1 ($N = 147$), students were led to believe via pictures and a short story that the worked examples were created by a male or female peer student. Males showed higher performance and confidence, but no effects of MOS on learning were found. In Experiment 2 ($N = 130$), students were led to believe that a peer student or a teacher created the examples. Again, no effects of MOS were found. These findings suggest that the perceived origin of text-based worked examples is not important for learning.

Introduction

It is a well-established finding that example-based learning is a very effective and efficient instructional strategy for novices compared to practice problem solving (Renkl, 2014; Van Gog & Rummel, 2010) that may also enhance their self-efficacy (Bandura, 1997). Research conducted from a social learning theory perspective (Bandura, 1986) has shown the effectiveness of observational learning from *modeling examples* in which a person (the 'model') demonstrates and explains a procedure either live or on video. Research conducted from a cognitive load theory perspective (Sweller, Ayres, & Kalyuga, 2011) has shown the effectiveness of studying text-based *worked examples* in which the procedure is demonstrated in writing.

The design of the examples is a crucial factor in their effectiveness for learning and self-efficacy, however. Following early studies on learning from worked examples (e.g., Cooper & Sweller, 1987; Sweller and Cooper, 1985), it was soon discovered that this is not always more effective than problem solving and that the effectiveness depends on the design of the examples (e.g., examples that induce split-attention are less effective; Tarmizi & Sweller 1988). With regard to video modeling examples, design choices concern, for instance, which model should provide the demonstration. Research on the model-observer similarity (MOS) hypothesis has addressed the question of whether students' learning outcomes and self-efficacy are enhanced by studying a model who is similar to them (e.g., in terms of gender or expertise).

Model-observer Similarity

The MOS hypothesis states that the effectiveness of modeling depends in part on the degree to which observers perceive a model to be similar to them (Schunk, 1987). Modeling evokes an observer to engage in social comparison (Berger, 1977) and when a model successfully demonstrates a task, observers are likely to believe that they can perform the task as well, assuming they identify with the model (Bandura, 1981). Also, an observer may pay more attention and be more attracted to a model that is perceived as similar (Berscheid & Walster, 1969).

As Schunk (1987) postulated, 'similarity serves as an important source of information for gauging behavioural appropriateness, formulating outcome expectations, and assessing one's self-efficacy for learning or performing tasks' (p. 149). Novice learners are presumably most influenced by MOS, as social comparison is more likely when self-efficacy and perceived competence are still low (Buunk, Zurriaga, Gonzalez-Roma, & Subirats, 2003). Thus, the MOS hypothesis predicts that a higher degree of similarity between novice learners and their model positively affects both cognitive (i.e., test performance) and affective aspects of learning, such as self-efficacy and perceived competence. Self-efficacy and perceived competence are important factors in learning. Self-efficacy affects factors such as study behaviour, academic motivation, and learning outcomes (Bandura, 1997; Bong & Skaalvik, 2003). Additionally, the related construct of perceived competence, which reflects broader perceptions and knowledge of one's own abilities than self-efficacy (Bong & Skaalvik, 2003; Klassen & Usher, 2010), also affects learning outcomes and academic motivation (Bong & Skaalvik, 2003; Harter, 1990). Several MOS-factors may play a role in learning, such as gender and perceived expertise.

Gender

Gender is one of the most salient MOS-factors because another person's gender is among the first things that we notice in social interactions (Contreras, Banaji, & Mitchell, 2013). A substantial amount of research has investigated whether it is better to observe a same-gender model relative to an opposite-gender model. Schunk (1987) reviewed effects of gender similarity on learning outcomes and self-efficacy and reported mixed results. In his review, Schunk suggested that the appropriateness of the modelled behaviour might moderate effects of gender similarity; gender similarity might only play a role when a skill or behaviour is deemed more appropriate for males or females. This could explain why classic studies found that boys displayed more imitative aggression after observing a male model displaying aggression towards a doll relative to observing a female model (Bandura, Ross, & Ross, 1963; Hicks, 1965), while Schunk, Hanson, and Cox (1987) found no effects on self-efficacy nor learning for students (grade 4 to 6) who observed a female or male model solving fraction problems (participants were likely too young to associate a mathematical task with gender; Ceci, Ginther, Kahn, & Williams, 2014).

With technological advancements, new forms of modeling examples have emerged, such as video modeling examples in which one may hear models explaining the task while seeing the models' computer screens, without seeing the models themselves (e.g., McLaren, Lim, & Koedinger, 2008). More recent research has investigated the question how MOS affects learning from these learning types of examples. Several recent studies compared learning from dynamic visualizations (i.e., video, animation) accompanied by a male or female voice over. Surprisingly, in view of the idea that the gender-appropriateness of the behavior may be important for perceived similarity, Linek, Gerjets, and Scheiter (2010) found that for university students who learned about probability calculation, a female voice was more preferred and more conducive to learning outcomes. Rodicio (2012), however, did find effects in favour of male narration over female narration in terms of learning outcomes for university students learning about geography and Lee, Liao, and Ryu (2007) found that a male computer-generated voice relative to a female one led to a more positive evaluation and higher trust and confidence levels when learning about male topics such as football and knights.

Animated pedagogical agent studies have also shown benefits of interacting with male agents, especially for tasks that are more associated with men. For example, Arroyo, Woolf, Royer, and Thai (2009) found that students learned more about mathematics and showed a more positive attitude towards a male agent relative to a female agent. Baylor and Kim (2004) found that male agents were assessed as more intelligent, useful, interesting, and satisfactory relative to female agents when learning an educational technology task. However, both male and female students tended to select a same-gender agent when given the choice (Ozogul, Johnson, Atkinson, & Reisslein, 2013).

Expertise

Another MOS-factor that may play a role in learning is expertise of the model. The MOS hypothesis (Schunk, 1987) predicts that for novices, it is more conducive for self-efficacy and learning when a model is perceived as similar in expertise (e.g., a peer student) than when a model is perceived as dissimilar in expertise (e.g., a teacher). A substantial number of studies have compared learning from a coping model, who shows performance errors that are gradually corrected and/or expressions of uncertainty that are gradually reduced, to a mastery model, who displays faultless performance from the beginning. Findings have been mixed, however, and may depend on students' ability. For instance, in math, mastery and coping models seem to be equally effective for learning when students are of average ability (Schunk & Hanson, 1989) or when they are of weak ability but have already experienced success on the modeled task (e.g., Schunk & Hanson, 1985). When weaker students have no prior experience with the modeled task, however, coping models were more effective (Schunk et al., 1987). More recent studies have shown that for novice learners, studying coping models is not only more effective for learning, but also fosters affective aspects of learning such as self-efficacy and task interest more than mastery models do (Kitsantas, Zimmerman, & Cleary, 2000; Zimmerman & Kitsantas, 2002).

Another line of research has contrasted effects of a high expertise model to a low expertise model. Older studies indicate that on a wide range of measures, more expert models lead to more favorable outcomes for primary school children (e.g., Sonnenschein & Whitehurst, 1980). More recently, Braaksma, Rijlaarsdam, and Van den Bergh (2002) taught secondary education students how to write an argumentative text using video examples of both strong and weak peer models. As predicted, weaker students benefited more from focusing on weak models, whereas better students learned more from focusing on strong models. When looking at transfer, however, indications have been found in studies in professional education that studying worked examples created by expert models may be more conducive to transfer than studying worked examples created by advanced peer student models; presumably, the higher degree of abstraction of expert explanations fosters transfer (Boekhout, Van Gog, Van de Wiel, Gerards-Last, & Geraets, 2010; Lachner & Nückles, 2015).

In all of the studies discussed above, there were actual expertise differences in the models' performance, and therefore the effects on learning and self-efficacy might have been due to the content of the examples rather than the perceived expertise of the model. To the best of our knowledge, only animated pedagogical agent studies have examined effects of *perceived* similarity in expertise (sometimes in conjunction with related factors, such as model age), while keeping the learning content the same. For instance, Kim, Baylor, and Reed (2003) showed that a younger looking mentor-like agent was more motivating to learn from than an older looking expert agent, although no effects on test performance were found. Rosenberg-Kima, Baylor, Plant, and Doerr (2008) found that a 'young and cool' agent enhanced female university students' self-efficacy more than 'young and uncool' and 'older and (un)cool' agents when learning about engineering. Baylor and Kim (2004) compared learning about educational technology from agents that represented the roles of expert, motivator, or mentor. Their findings showed that learning from an expert was less conducive for self-efficacy relative to the other two roles, possibly because the expert was seen as more intelligent.

The Present Study

Most research on the MOS-hypothesis has been conducted on video (e.g., Braaksma et al., 2002; Schunk & Hanson, 1985, 1989) or animated (e.g., Baylor & Kim, 2004; Kim et al., 2003; Rosenberg-Kima et al., 2008) models. A few studies have used written text-based examples, although, as mentioned above, these written worked examples created by different models actually differed in content (Boekhout et al., 2010; Lachner & Nückles, 2015). It is, therefore, an open question whether perceived similarity with the 'model' who created the example, in terms of gender or expertise, would also affect learning from text-based worked examples of the exact same content. If so, information about the origin of examples could be adaptively changed in online learning environments to match student characteristics in order to optimize learning and/or self-efficacy. The present study addresses this question in two experiments.

It is investigated whether it is more effective for secondary education students to study text-based worked examples on a science topic that they believe to have been designed by a peer of the same or opposite sex (MOS: gender; Experiment 1) or a peer student or teacher (MOS: expertise; Experiment 2) in terms of learning outcomes, self-efficacy, and perceived (own) competence. To ensure that any potential MOS effects would not be due to the characteristics of one particular model, three models were used per condition in both experiments. Effort investment (i.e., an indicator of cognitive load) was also measured during learning and during the posttest to further examine effects on the learning process. Lastly, students evaluated the 'models' who they believed had 'created' the examples, in terms of attractiveness, friendliness, and intelligence, and evaluated the examples in terms of quality. The examples that were used have proven to be effective for learning in other studies (Van Gog, Kester, Dirkx, Hoogerheide, Boerboom, & Verkoeijen, 2015; Van Gog, Kester, & Paas, 2011) and the content of the examples was kept identical across conditions, so that only the perceived similarity between observer and model could affect the outcome variables.

Experiment 1

Method

Participants and Design. The experiment had a 2 x 2 design, with Gender Observer (Male vs. Female) and Gender Model (Male vs. Female) as between-subject factors. Participants were 147 secondary education students (M^{age} = 14.43, *SD* = 0.58; 74 male) in their third year of pre-university education[1]. Students were quasi-randomly (i.e., matched for gender, e.g., for each boy assigned to the male model condition, another was assigned to the female model condition) allocated to either the male peer model (37 girls, 37 boys) or female peer model (36 girls, 37 boys) condition. The students were novices to troubleshooting electrical circuit problems at the time of the experiment, as this content had not yet been covered in their curriculum.

Materials

The paper-based materials that focused on learning to solve electrical circuit troubleshooting problems had been used and proven effective in prior studies (Van Gog et al., 2011, 2015).

Conceptual Prior Knowledge Test. The conceptual prior knowledge test contained seven questions (open-ended) on troubleshooting and parallel circuits principles.

[1] Note that a power analysis (alpha = .05, power = .80) was conducted to determine how many participants we would need to be able to reliably detect medium-sized effects, and this analysis revealed that the minimum number of participants required in each experiment was 128.

Examples of pretest items are: 'What do you know about the total current in a parallel circuit? (answer: it is the sum of the currents in each of the parallel branches)' and 'If the total current in a parallel branch is lower than you would expect, what does that tell you about the resistance in the circuit? (answer: the resistance is higher than you would expect).'

Introduction and Formula Sheet. The abbreviations of the components of the circuit drawing were explained as well as a description of Ohm's law and the three different forms of the formula (i.e., U=R*I; R=U/I; I=U/R) on one A4-sized page.

Acquisition Phase Tasks. The worked examples presented a malfunctioning parallel electrical circuit and fully worked out solutions of how to identify the fault, using Ohm's law (see Appendix for an example). The top of the page in each example presented a circuit drawing that indicated how much voltage the power source delivered and how much resistance each resistor provided. The examples started by explaining how to determine the current that should be measured at each of the ammeters when the circuit would function correctly. Next, faulty ammeter measurements were presented and based on the information in the circuit and the formula sheet, the current that should be measured (i.e., if the system were functioning correctly) in each of the parallel branches as well as overall was calculated. It was then stated that if one compared the calculated measurements at step 1 to those given at step 2, it could be inferred in which branch the resistance differed from the resistance indicated in the diagram, and that the actual measurement at step 2 could be used to find the actual value of the resistor.

All conditions were presented with the same four worked examples, with the only difference being the picture and brief description (name, age, and class) of the model at the top left of the page. The first two and last two examples contained the same fault. That is, in the first two *lower* current was measured in a particular parallel branch, which is indicative of *higher* resistance in that branch, while in the last two *higher* current was measured in a particular parallel branch, which is indicative of *lower* resistance in that branch.

Posttest. The posttest presented two troubleshooting tasks in a problem-solving format. The first task was isomorphic to one pair of the training tasks and the second task contained the two faults encountered in the worked examples simultaneously. The test problems asked participants to answer the following questions: 'Determine how this circuit should function using Ohm's law, that is, determine what the current is that you should measure at each of the ammeters'; 'Suppose the ammeters indicate the following measurements: ….' (this was given); 'What is the fault and in which component is it located?'

Mental Effort. Invested mental effort was measured with the subjective rating scale developed by Paas (1992), which ranges from (1) very, very low effort to (9) very, very high effort.

Self-efficacy and Perceived Competence. The self-efficacy measure asked participants to indicate on a 9-point scale ranging from (1) very, very unconfident to (9) very, very confident to what extent participants believed that they had mastered the skill of troubleshooting electrical circuit problems. We adopted this measure from Hoogerheide et al. (2014a) in which the question was phrased in accordance with recommendations provided by Bandura (2006). Perceived competence was measured with an adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996). While this scale originally contains four questions, the question 'I am able to achieve my goals in this course' was removed because it did not apply in the context of the current study, leaving the following three questions: 'I am capable of learning the material in this course', 'I feel confident in my ability to learn this material', 'I feel able to meet the challenge of performing well in this course'. We changed 'course' into 'troubleshooting electrical circuit problems' in all three items. Note that for both self-efficacy and perceived competence, the phrasing of the questions differed before vs. after the acquisition phase. On the pretest, the questions asked to which degree participants were confident in *learning* how to troubleshoot electrical circuit problems, while after the acquisition phase the questions asked to which degree participants were confident in their acquired skills.

Example Quality and Model Evaluation. Participants' impression of the quality of the worked examples was measured on a scale of 1 (very, very bad quality) to 9 (very, very good quality). Moreover, participants indicated on a scale of 1 to 7 how attractive, friendly, and intelligent they found the model.

Procedure

Prior to the experiment, participants had been quasi-randomly assigned to the conditions (based on a name list). The experiment lasted ca. 50 min. and was run in a classroom in participants' schools with an entire class of students present. When entering the classroom, participants were instructed to sit down at the desk containing the envelope with their name on it. Each envelope contained three booklets. First, participants received seven minutes to fill in Booklet 1, which contained a brief demographic questionnaire and the conceptual prior knowledge test. At the end of

Booklet 1, participants indicated their self-efficacy and perceived competence. Participants were then instructed to put aside Booklet 1 and to take out Booklet 2, which presented the four examples and had the formula sheet inserted. Participants first received two minutes to study the formula sheet, after which they were instructed to read the short introduction on the first page for a minute. For the male peer model materials, the introduction stated: 'You are about to study four examples in which it is explained how to troubleshoot electrical circuit problems. These examples were created by Jan/Maarten/Peter van Zomeren (15 years old, 4th year pre-university education student). [A photo followed here (see Figure 1)]. Study these examples well, because later on you will be asked to solve similar problems yourself'. The introduction of the female models was identical, except that the names and photos differed (female names were chosen to closely resemble the male names: Janine/Maartje/Petra van Zomeren).

Participants then studied the four worked examples sequentially, each printed on a separate page. The top left corner of each worked example showed the same picture, age, and student status from the model's introduction. Participants were not allowed to turn over pages unless explicitly instructed. The experimenter indicated after three minutes that participants were allowed to move on to the next example. Each worked example was followed by a mental effort rating scale, and at the end of the booklet participants rated the quality of the examples, and their self-efficacy and perceived competence. Next, participants used the remaining time (max. 12 min.) for Booklet 3, which contained the two posttest tasks. Each test task was followed by a mental effort rating scale and at the end of the test the model evaluation questions were administered. Participants were allowed to use a calculator and a newly handed out formula sheet, and the model's picture was shown on the model evaluation question page.

Data Analysis

Participants could earn ten points on the conceptual prior knowledge test. For the posttest, the maximum score was three points for the first task with only one fault (one for calculating the correct value of all ammeters, one for indicating the faulty resistor, and one for indicating what the actual value of the faulty resistor was) and five points for the second task with two faults (an extra point for indicating the second faulty resistor and for calculating its resistance). For incomplete or partially correct answers, on both the prior knowledge and the posttest, half points were given. The scoring procedure was based on coding schemes that had been used before by Van Gog et al. (2011, 2015), which are very straightforward and leave little room for interpretation. Therefore, scoring was done by a single rater.

Figure 1. Male and *female* peer models used in Experiment 1

Results

A preliminary analysis was conducted to test whether the three male and three female models had differential effects on the outcome measures. Results showed a main effect of students' perceptions of the female model's attractiveness, $F(2,70) = 8.94$, $p < .001$, $\eta_p^2 = .203$. Follow-up tests with a Bonferroni-correction showed that one of the female models was rated as significantly more attractive than the other two ($p = .001$ and .002). No other significant effects were found, and therefore we proceeded analysing the data at condition level.

Data were analyzed using 2 x 2 ANOVAs, with Gender Observer (Female, Male) and Gender Model (Female, Male) as between-subject factors, unless otherwise specified. Interactions were further analyzed with *t*-tests. The test performance, invested mental effort, self-efficacy, and perceived competence scores can be found in Table 1, and ratings of the examples' quality and perceived model characteristics in Table 2.

Test Performance. The analysis of the conceptual prior knowledge test scores, revealed a main effect of Gender Observer, $F(1,143) = 5.97$, $p = .016$, $\eta_p^2 = .040$, indicating that male students ($M = 2.99$, $SD = 1.39$) had more prior knowledge than female students ($M = 2.43$, $SD = 1.40$). There was no main effect of Gender Model, $F(1,143) = 2.39$, $p = .142$, $\eta_p^2 = .015$, nor an interaction effect, $F(1,143) = 1.23$, $p = .269$, $\eta_p^2 = .009$.

The analysis of posttest performance showed a main effect of Gender Observer, $F(1,143) = 6.23$, $p = .014$, $\eta_p^2 = .042$, with male students ($M = 4.65$, $SD = 2.44$) outperforming female students ($M = 3.65$, $SD = 2.37$). There was no main effect of Gender Model, nor an interaction effect (both Fs < 1).

Mental Effort. The analysis of mental effort invested during example study revealed a main effect of Gender Observer, $F(1,143) = 5.35$, $p = .022$, $\eta_p^2 = .036$, indicating that male students ($M = 3.00$, $SD = 1.59$) invested less mental effort than female students ($M = 3.63$, $SD = 1.71$). There was no main effect of Gender Model nor an interaction effect, both Fs < 1. The same pattern of results was found with regard to effort invested in completing the posttest tasks: A main effect of Gender Observer, $F(1,143) = 11.13$, $p = .001$, $\eta_p^2 = .072$, with males ($M = 3.95$, $SD = 1.90$) investing less effort than females ($M = 4.99$, $SD = 1.91$), but no main effect of Gender Model and no interaction effect, both Fs < 1.

Self-efficacy and Perceived Competence. There were 54 missing values on the self-efficacy measurement after the pretest (leaving 24 participants in both female student conditions, and respectively 22 and 23 in the male student – male model and male student – female model condition) because students overlooked the self-efficacy question on the page. There were no significant differences on self-efficacy measured after the pretest and prior to the acquisition phase (Gender Observer: $F(1,89) = 3.12$, $p = .081$, $\eta_p^2 = .034$; Gender Model: $F < 1$; interaction: $F(1,89) = 1.14$, $p = .288$, $\eta_p^2 = .013$). On self-efficacy after the acquisition phase, there was a main effect of Gender Observer, $F(1,139) = 10.88$, $p = .001$, $\eta_p^2 = .073$. Male students ($M = 6.41$, $SD = 1.71$) showed higher confidence than female students ($M = 5.51$, $SD = 1.55$). There was no main effect of Gender Model, nor an interaction effect, both Fs < 1.

The analysis of students' perceived competence after the pretest (prior to the acquisition phase), showed a main effect of Gender Observer, $F(1,143) = 10.48$, $p = .001$, $\eta_p^2 = .068$. Male students ($M = 5.37$, $SD = 1.24$) showed higher perceived competence than female students ($M = 4.73$, $SD = 1.21$). There was no main effect of Gender Model, $F(1,143) = 1.61$, $p = .\ 207$, $\eta_p^2 = .011$, and no interaction effect, $F(1, 143) = 1.49$, $p = .224$, $\eta_p^2 = .010$. After the acquisition phase, male students ($M = 5.62$, $SD = 1.29$) still showed higher perceived competence than female students ($M = 4.85$, $SD = 1.39$), as indicated by a main effect of Gender Observer, $F(1,139) = 11.69$, $p = .001$, $\eta_p^2 = .078$. However, there was, no main effect of Gender Model nor an interaction effect, both Fs < 1.

Table 1.
Mean (SD) of Test Performance, Invested Mental Effort, Self-efficacy, and Perceived Competence per Condition in Experiment 1.

	Male Observer		Female Observer	
	Male Model	Female Model	Male Model	Female Model
Performance Pretest	3.28 (1.43)	2.69 (1.29)	2.47 (1.47)	2.39 (1.35)
Performance Posttest	4.77 (2.37)	4.53 (2.54)	3.45 (2.31)	3.86 (2.44)
Mental Effort Study Phase	3.16 (1.71)	2.85 (1.47)	3.55 (1.75)	3.72 (1.69)
Mental Effort Posttest	3.88 (1.89)	4.01 (1.92)	4.77 (2.15)	5.22 (1.62)
Self-efficacy Pretest	5.41 (1.79)	5.70 (1.66)	5.17 (1.49)	4.71 (1.76)
Self-efficacy Posttest	6.30 (1.79)	6.51 (1.64)	5.62 (1.42)	5.39 (1.68)
Perceived Competence Pretest	5.38 (1.28)	5.37 (1.22)	4.97 (1.15)	4.47 (1.23)
Perceived Competence Posttest	5.63 (1.42)	5.61 (1.18)	5.05 (1.23)	4.65 (1.52)

Example Quality and Model Evaluation. On the example quality and each of the model evaluation questions, there was one missing value, leaving 146 participants. The students rated the quality of the examples as high, and there were no significant differences (Gender Observer: $F(1,142) = 1.29$, $p = .259$, $\eta_p^2 = .009$; Gender Model: $F(1, 142) = 1.51$, $p = .221$, $\eta_p^2 = .011$; interaction: $F < 1$). The scores of the model's attractiveness showed no effect of Gender Observer, $F(1,142) = 2.96$, $p = .088$, $\eta_p^2 = .020$, or Gender Model, $F < 1$, but the interaction between Gender Observer and Gender Model was significant, $F(1,142) = 5.68$, $p = .018$, $\eta_p^2 = .038$. Follow up tests showed that this effect was caused by a difference in model evaluation by the female students: they rated the female models ($M = 4.56$, $SD = 1.50$) as more attractive than the male models ($M = 3.70$, $SD = 1.75$), $p = .029$, $\eta_p^2 = .066$, and they evaluated the female models as significantly more attractive than the male students did ($M = 3.41$, $SD = 1.91$). There were no significant effects on students' perceptions of model friendliness (Gender Observer: $F(1,142) = 2.29$, $p = .133$, $\eta_p^2 = .016$; Gender Model: $F < 1$; interaction: $F(1,142) = 1.17$, $p = .281$, $\eta_p^2 = .008$) or intelligence (Gender Observer: $F(1,142) = 3.41$, $p = .067$, $\eta_p^2 = .023$; Gender Model: $F < 1$; interaction: $F(1,142) = 2.45$, $p = .119$, $\eta_p^2 = .017$).

Table 2.
Mean (SD) of Example Quality and Model Evaluation scores per Condition in Experiment 1.

	Male Observer		Female Observer	
	Male Model	Female Model	Male Model	Female Model
Example Quality	6.73 (1.35)	6.78 (1.18)	6.28 (1.50)	6.75 (1.11)
Perceived Attractiveness	3.89 (1.58)	3.41 (1.91)	3.70 (1.75)	4.56 (1.50)
Perceived friendliness	5.32 (1.13)	5.42 (1.18)	5.00 (1.53)	5.56 (1.30)
Perceived Intelligence	5.22 (1.42)	5.28 (1.30)	4.95 (1.49)	5.69 (1.04)

Discussion

The results of this experiment showed no effects of the model's gender on learning outcomes; test performance after studying worked examples was not significantly different when students were told that the examples had been created by a male peer student or a female peer student. Male students showed higher confidence in their own capabilities and also showed better test performance than female students, with less effort investment during example study and test performance, but no MOS effects were found on test performance, self-efficacy, perceived competence, or effort investment.

Participants also did not differ in their estimation of the examples' quality. With respect to how the models were perceived, the results showed that overall the male and female peer models were perceived as equally intelligent and friendly. There was an interaction effect with regard to model attractiveness, however: female students evaluated the female models as being more attractive than male models, and female students also rated female models as being more attractive than male students did. Nevertheless, overall, based on Experiment 1, there seems to be no indication that MOS in terms of gender would affect learning how to troubleshoot electrical circuit problems from text-based worked examples. Experiment 2 investigated MOS in terms of perceived model expertise.

Experiment 2

Experiment 2 used the same materials to test the MOS hypothesis with regard to perceived expertise in text-based worked examples, by presenting those as having been created by a male peer model (i.e., more similar in expertise and age) or a male teacher model (i.e., dissimilar in expertise and age).

Participants and Design

The experiment had a 2 x 2 design, with Gender Observer (Male vs. Female) and Model Expertise (Peer vs. Teacher) as between-subject factors. Participants were 130 secondary education students (M^{age} = 14.63, *SD* = 0.68; 71 male) in their third year of pre-university education (the highest secondary education level in The Netherlands). Students were quasi-randomly allocated (i.e., matched for gender, e.g., for each boy assigned to the peer model condition, another was assigned to the teacher model condition) to the peer model (29 girls, 36 boys) or teacher model (30 girls, 35 boys) condition. The students were novices to troubleshooting electrical circuit problems at the time of the experiment.

Figure 2. Male *peer* (top row) and male *teacher* (bottom row) models used in Experiment 2

Materials, Procedure, and Data Analysis

The same materials and procedure were used as in Experiment 1, with two exceptions. Firstly, the secondary education students were either led to believe via a short story and pictures that the materials were created by one of three male peer models (the same peer models as in Experiment 1), or one of three male teacher models (see Figure 2). The same names used for the peer models (Jan/Maarten/Peter van Zomeren) were also used for the teacher models, who were further stated to be 42 years old and a science teacher.

Secondly, we also explored effects of learning enjoyment. Prior research has shown that with learning materials that present more social cues, learning enjoyment may differ depending on the design of the example-based instruction (Hoogerheide et al., 2014).

Moreover, learning with an animated peer agent has been shown to be more enjoyable to interact with than an animated teacher agent (Liew et al., 2013). Lesson enjoyment was measured on a scale of 0 (entirely unenjoyable) to 10 (very enjoyable). Scoring was done as in Experiment 1.

Results

Again, a preliminary analysis was conducted to test whether the three peer and three teacher models had differential effects on the outcome measures. Results showed a main effect of students' perceptions of the teacher model's intelligence, $F(2,62) = 4.15$, $p = .020$, $\eta_p^2 = .118$. Bonferroni-corrected post-hoc tests showed that one of the models was perceived as more intelligent than the other two ($p = .043$ and $.050$). No other effect was significant, and therefore we proceeded analysing the data at condition level.

Unless otherwise specified, the analyses were completed using 2 x 2 ANOVAs, with Gender Observer (Female, Male) and Model Expertise (Peer, Teacher) as between-subject factors. The test performance, invested mental effort, self-efficacy, and perceived competence scores can be found in Table 3, and lesson enjoyment, quality of examples, and perceived model characteristics in Table 4.

Test Performance. The analysis of the conceptual prior knowledge test scores, revealed a main effect of Gender Observer, $F(1,126) = 10.99$, $p = .001$, $\eta_p^2 = .080$. As in Experiment 1, male students ($M = 2.30$, $SD = 1.23$) scored significantly higher than female students ($M = 1.60$, $SD = 1.12$). There was no main effect of Model Expertise, nor an interaction, both *F*s < 1. There were no significant effects on posttest performance (Gender Observer: $F(1,126) = 2.88$, $p = .092$, $\eta_p^2 = .022$; Model Expertise and interaction: *F*s < 1).

Mental Effort. There were no significant effects on mental effort during example study (Gender Observer and Model Expertise: *F*s < 1; interaction, $F(1,126) = 3.32$, $p = .071$, $\eta_p^2 = .026$) or mental effort invested in completing the posttest tasks (Gender Observer, $F(1, 126) = 2.00$, $p = .159$, $\eta_p^2 = .016$; Model Expertise and interaction, *F*s < 1).

Self-efficacy and Perceived Competence. There were 35 missing values on the self-efficacy measurement on the pretest (leaving 22 and 23 participants in the female student – peer model and female student – teacher model condition, respectively, and 24 and 26 participants in the male student peer model and male student teacher model condition, respectively), and two missing values after the acquisition phase (both from the male student – peer model condition). There were no significant effects on the self-efficacy measurement after the pretest (Gender Observer: $F(1,91) = 1.86$, $p = .178$, $\eta_p^2 = .020$; Model Expertise: $F < 1$; interaction: $F(1, 91) = 1.91$, $p = .170$, $\eta_p^2 = .021$), or after the acquisition phase (Gender Observer: $F(1,124) = 2.62$, $p = .108$, $\eta_p^2 = .021$; Model Expertise: $F(1,124) = 1.47$, $p = .228$, $\eta_p^2 = .012$; interaction: $F(1, 124) = 2.60$, $p = .109$, $\eta_p^2 = .021$).

One participant from the male student peer model condition did not fill in the perceived competence measurement after the acquisition phase. With regard to students' perceived competence after the pretest, there was a main effect of Gender Observer as in Experiment 1, $F(1,126) = 11.74$, $p < .001$, $\eta_p^2 = .085$, with male students ($M = 5.27$, $SD = 1.32$) showing higher perceived competence than female students ($M = 4.50$, $SD = 1.29$). There was no main effect of Model Expertise, $F(1,126) = 1.46$, $p = .228$, $\eta_p^2 = .011$, nor an interaction effect, $F(1, 126) = 3.54$, $p = .062$, $\eta_p^2 = .027$. On perceived competence after the acquisition phase, there was no longer a main effect of Gender Observer, $F(1,125) = 3.14$, $p = .079$, $\eta_p^2 = .025$, and there was no effect of Gender Model, $F < 1$, nor an interaction, $F(1,125) = 1.16$, $p = .284$, $\eta_p^2 = .009$.

Table 3.
Mean (SD) of Test Performance, Invested Mental Effort, Self-efficacy, and Perceived Competence per Condition in Experiment 2.

	Male Observer		Female Observer	
	Peer Model	Teacher Model	Peer Model	Teacher Model
Performance Pretest	2.28 (1.24)	2.31 (1.27)	1.76 (1.07)	1.45 (1.15)
Performance Posttest	3.62 (2.54)	4.24 (2.39)	4.55 (2.42)	4.80 (2.57)
Mental Effort Study Phase	4.00 (1.94)	3.42 (1.66)	3.50 (1.73)	4.02 (1.43)
Mental Effort Posttest	4.46 (2.40)	4.29 (2.05)	5.02 (2.09)	4.77 (1.66)
Self-efficacy Pretest	5.08 (1.91)	5.81 (1.74)	5.09 (1.48)	4.91 (1.04)
Self-efficacy Posttest	5.59 (2.08)	6.43 (1.46)	5.59 (1.84)	5.47 (1.17)
Perceived Competence Pretest	4.92 (1.52)	5.63 (0.97)	4.57 (1.44)	4.42 (1.16)
Perceived Competence Posttest	5.03 (1.33)	5.47 (1.11)	4.87 (1.39)	4.83 (1.21)

Lesson Enjoyment, Example Quality, and Model Evaluation. The analysis of lesson enjoyment scores showed a main effect of Gender Observer, $F(1,125) = 5.91$, $p = .017$, $\eta_p^2 = .045$, with male students ($M = 5.27$, $SD = 2.62$) showing higher lesson enjoyment than female students ($M = 4.22$, $SD = 2.22$). There was, however, no main effect of Model Expertise, $F < 1$, nor an interaction effect, $F(1,125) = 3.01$, $p = .085$, $\eta_p^2 = .023$. There were no effects on students' perceptions of the quality of the examples (all *F*s < 1).

With respect to the model evaluation questions, there was a main effect of Gender Observer on perceived model's attractiveness, $F(1,126) = 4.09$, $p = .045$, $\eta_p^2 = .031$. Male students ($M = 3.97$, $SD = 1.93$) rated the models as more attractive than female students ($M = 3.37$, $SD = 1.35$). There was no main effect of Model Expertise, $F < 1$, but we did find a significant interaction between Gender Observer and Model Expertise, $F(1,126) = 4.92$, $p = .028$, $\eta_p^2 = .038$. When investigating the effects of Model Expertise for each Observer Gender condition separately, an effect was only found for the teacher models, $F(1,63) = 8.67$, $p = .005$, $\eta_p^2 = .121$, with male students rating the teacher model as significantly more attractive ($M = 4.29$, $SD = 1.95$) than female students ($M = 3.03$, $SD = 1.38$). When investigating the effects of Gender Observer for each Model Expertise condition separately, there was an effect only for the female students, $F(1,57) = 4.06$, $p = .049$, $\eta_p^2 = .067$, who evaluated the peer students ($M = 3.72$, $SD = 1.25$) as more attractive than the teacher models ($M = 3.03$, $SD = 1.38$). There were no effects on ratings of model friendliness (Gender Observer: $F(1,126) = 1.24$, $p = .268$, $\eta_p^2 = .010$; Model Expertise: $F(1,126) = 2.90$, $p = .091$, $\eta_p^2 = .023$; interaction: $F < 1$) or intelligence (all Fs < 1).

Table 4.
Mean (SD) of Lesson enjoyment, Example Quality, and Model Evaluation scores per Condition in Experiment 2.

	Male Observer		Female Observer	
	Peer Model	Teacher Model	Peer Model	Teacher Model
Lesson Enjoyment	4.89 (2.86)	5.66 (2.33)	4.59 (2.24)	3.87 (2.18)
Example Quality	6.33 (2.06)	6.60 (1.31)	6.72 (1.22)	6.57 (1.25)
Perceived Attractiveness	3.67 (1.90)	4.29 (1.95)	3.72 (1.25)	3.03 (1.38)
Perceived friendliness	5.53 (1.13)	5.17 (1.69)	5.31 (0.89)	4.87 (1.43)
Perceived Intelligence	5.26 (1.80)	5.46 (1.58)	5.41 (1.09)	5.60 (1.13)

Discussion

As in Experiment 1, there were no effects of MOS on posttest performance, self-efficacy, perceived competence, or effort investment. Male students, as in Experiment 1, overall showed somewhat higher conceptual knowledge and perceived competence prior to studying the examples, as well as higher lesson enjoyment than female students. However, this higher confidence and enjoyment shown by male students was –in contrast to Experiment 1- not accompanied by significantly higher posttest performance in Experiment 2. With respect to students' perceptions of the models, the results showed that overall, the peer and teacher models were perceived as equally intelligent and friendly. The only difference that was found, was on perceived attractiveness: male students rated the teacher models as more attractive than female students, and female students evaluated the peer students as more attractive than the teacher models.

General Discussion

Two experiments investigated the MOS-hypothesis (Schunk, 1987) with text-based worked examples. More specifically, it was investigated whether similarity to the 'model' who supposedly created the example, in terms of gender (Experiment 1) or expertise (Experiment 2), would affect learning from worked examples that were otherwise identical in content. Neither Experiment 1, nor Experiment 2 showed MOS-effects on learning outcomes, self-efficacy and perceived competence, or effort investment. Although there were interaction effects on ratings of the model's attractiveness, these were inconsequential with regard to the main outcome variables.

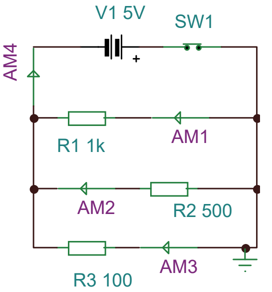
A likely reason for why learning outcomes did not differ as a function of MOS is that we, in contrast to prior studies (e.g., Boekhout et al., 2010; Braaksma et al., 2002; Lachner & Nückles, 2015), kept the content of the examples identical across conditions. This does not necessarily explain, however, why there were no MOS effects on self-efficacy or perceived competence. Possibly, our text-based examples with pictures of the 'models' did not provide sufficiently strong social cues. Such cues are more strongly available when learning with modeling examples or animated pedagogical agents, and they may stimulate learners to engage in a social comparison by linking the presented content to their own personal self (Mayer, 2014), which might affect self-efficacy or perceived competence. This could explain why, even with otherwise identical content, studies with animated agents did show MOS effects, whereas the present study did not. It would be interesting to address MOS effects in worked examples that contain more cues regarding the model, as was the case, for instance, in studies on learning from 'heuristic worked examples'. In such examples not only the solution procedure is presented, but students' attention is also focused on domain-specific principles as well as strategies that could help solve similar problems, by means of a worked out solution by a fictitious peer student (Kollar, Ufer, Reichersdorfer, Vogel, & Fischer, 2014) or a "discussion" among two fictitious students (e.g., Hilbert, Renkl, Kessler, & Reiss, 2008).

A strength of this study was the use of different models in each condition; as such, it can be ruled out that the effects on attractiveness or the lack of effects on other variables would be due to the characteristics of one particular model. A potential limitation of the current study is that only one type of learning task was used; we cannot exclude the possibility that effects would arise with a different type of task or a task from another domain. Prior research on perceived similarity effects has, however, used science tasks to great effect (e.g., Moreno & Flowerday, 2006; Rosenberg-Kima et al., 2008) and because science is typically more associated with males than females (Nosek et al., 2009). MOS effects could be expected, especially with regard to gender, because similarity functions as an important source of information for assessing behavioural appropriateness (Schunk, 1987). Another potential limitation is that, because several constructs were assessed by single item questions, we cannot rule out the possibility that the lower reliability associated with single item questions might have contributed to the lack of MOS effects. Finally, because we did not have a manipulation check of whether students remembered the model's name, age, gender, or expertise (which would have interfered with posttest performance when assessed straight after the learning phase, whereas posttest performance might have interfered with memory for this information when assessed after the test phase), we do not know to what extent the learners were aware of the model characteristics during example study. Note though, that we gave an elaborate description of the model on the first page, and gave photo, name and age on the top of the page with every example. As such, it is highly unlikely that learners entirely ignored this information. A related issue is that the present study cannot answer the question of whether the availability of information regarding the origin (i.e., the 'model') of text-based worked examples matters compared to not having such information available at all. That is, the examples used in this study do provide some –albeit limited– social cues compared to regular worked examples in which no information about their creator is provided.

Despite these limitations, our study adds to the current literature by showing that there is little to be gained from manipulating text-based worked examples to include information about a fictitious 'model' who supposedly created the example and matches characteristics of the learner. Had MOS played a role, then information about the origin of examples could have been adaptively changed in online learning environments to match student characteristics in order to optimize learning or self-efficacy. It is potentially useful to offer students a choice of 'models' even with text-based examples (cf. Ozogul et al., 2013), in the sense that male and female students found the models differentially attractive, and our findings suggest that when the content of the examples is kept the same, such a choice can be offered without negative effects on learning outcomes. Next to investigating what 'models' students would chose when it is up to them and how that would affect learning and self-efficacy, another interesting direction for future research would be to systematically vary the availability of social cues. This could be done, for instance, by first presenting a video in which the designer introduces oneself before studying the examples vs. merely a textual and pictorial description, which would help determine whether effects of MOS would arise when social cues become stronger.

Appendix

Example 1 (Maarten, 15, 4VWO)



1. Determine how this circuit should function using Ohm's law, that is, determine what the current is that you should measure at AM1 to AM4

In parallel circuits, the total current (I_t) equals the sum of the currents in the parallel branches (I_1, I_2, etc.).

The total current should be: $I_t = I_1 + I_2 + I_3$

or: $$I_t = \frac{U}{R_1} + \frac{U}{R_2} + \frac{U}{R_3} = \frac{5V}{1k\Omega} + \frac{5V}{500\Omega} + \frac{5V}{100\Omega} = 5mA + 10mA + 50mA = 65mA$$

This means you should measure:

AM1 = $5mA$ AM2 = $10mA$ AM3 = $50mA$ Am4 = $65mA$

2. Suppose the ammeters indicate the following measurements:
AM1 = $5mA$ AM2 = $7{,}14mA$ AM3 = $50mA$ AM4 = $62{,}14mA$

In this case, the calculation of what you should measure does not correspond to the actual measures, so something is wrong in this circuit.

3. What is the fault and in which component is it located?
If the current in a branch is lower than it should be, the resistance in that branch is higher (equal U divided by higher R results in lower I).

The current in the second branch is smaller than it should be: $I_2 = 7{,}14mA$ instead of $10mA$. Thus, R_2 has a higher resistance than the indicated 500Ω. The actual resistance of R_2 can be calculated using the measured current:

$$R_2 = \frac{U}{I_2} = \frac{5V}{7{,}14mA} = 0{,}7k\Omega = 700\Omega$$

Part II

Effects of Explaining To Fictitious Others on Video

Chapter 6

Effects of creating video-based modeling examples on learning and transfer

This chapter has been published as:

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction, 33*, 108-119. doi:10.1016/j.learninstruc.2014.04.005

Abstract

Two experiments investigated whether acting as a peer model for a video-based modeling example, which entails studying a text with the intention to explain it to others and then actually explaining it on video, would foster learning and transfer. In both experiments, novices were instructed to study a text, either with the intention of being able to complete a test (condition A), or being able to explain the content to others (condition B and C). Moreover, students in condition C actually had to explain the text by creating a webcam-video. In Experiment 1 (*N* = 76 secondary education students) there was no effect of study intention on learning (A=B), but explaining during video creation significantly fostered transfer performance (C>B; C>A). In Experiment 2 (*N* = 95 university students), study intention did have an effect on learning (C>A; B>A), but only actual video creation significantly fostered transfer performance (C>A).

Introduction

Example-based learning is an effective instructional strategy that has been studied from different perspectives. Research from a cognitive perspective (e.g., cognitive load theory; Sweller, 1988; Sweller, Van Merriënboer, & Paas, 1998) has mainly focused on observational learning from *worked examples*, which consist of a written, step-by-step worked-out procedure for completing the learning task. This is usually an “ideal” or “didactical” procedure, reflecting how a student should learn to complete a task, which may differ from how an expert would actually handle it, since experts sometimes can skip or chunk steps (Ericsson & Staszweksi, 1989). Research from a social-cognitive perspective (e.g., social learning theory; Bandura, 1977, 1986; cognitive apprenticeship; Collins, Brown, & Newman, 1989) has focused on observational learning from *modeling examples* in which a human model or humanoid agent demonstrates and explains how to complete a task (see Van Gog & Rummel, 2010). These models sometimes demonstrate an ideal, didactical procedure for the task, but they may also display “natural” behavior, which entails making and correcting errors (e.g., Braaksma, Rijlaarsdam, & Van den Bergh, 2002). In modeling examples, the model can be either an adult (e.g., Schunk, 1981; Simon & Werner, 1996) or a peer student (e.g., Braaksma, et al., 2002; Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013a, 2013b; Schunk & Hanson, 1985).

Research inspired by the cognitive perspective has demonstrated the effectiveness and efficiency of example-based learning. For novices, instruction consisting of example study (alternated with problem solving) leads to better learning outcomes with less investment of time and mental effort than instruction consisting of problem solving only (Atkinson, Derry, Renkl, & Wortham, 2000; Paas & Van Gog, 2006; Renkl, 2011; Sweller et al., 1998; Van Gog & Rummel, 2010) and instruction consisting of tutored problem solving (Salden, Koedinger, Renkl, Aleven, & McLaren, 2010). Research inspired by the social-cognitive perspective has not only demonstrated that example-based learning can be effective for learning, but also that it can increase learners’ self-efficacy, which is the perceived belief a learner has for learning, or performing a task at a certain level (Bandura, 1997; Schunk, 1987).

As mentioned above, peer students are known to be effective modeling examples, improving learning of students who observe them (Groenendijk et al., 2013a, 2013b; Schunk, 1987). For educators, an interesting question is whether there would also be potential benefits for learning, for the peer students who act as models in the examples (i.e., for the students who explain and/or demonstrate a task). However, despite the fact that a lot of research has investigated the effects of *observing* modeling examples, little is known about the effects on learning and transfer that *acting* as a peer model might have. Therefore, this study addresses that question.

Acting as a Peer Model for Video-based Modeling Examples

Nowadays, video-based modeling examples (e.g., Braaksma et al., 2002; Groenendijk et al., 2013a, 2013b; McLaren, Lim, & Koedinger, 2008; Van Gog, 2011; see also www.khanacademy.org) are increasingly used in education as they have become easier to create and store in online (learning) environments. It seems that video-based modeling examples are also increasingly being used for informal learning purposes. Research has shown that many students (age 12 to 17) watch videos on websites such as YouTube and Google Videos; moreover, an increasing number of students also indicate they create and share videos (Lenhart, 2012; Spires, Hervey, Morris, & Stelpflug, 2012). While not all of those would qualify as video-based modeling examples, it is likely that these form part of the videos watched and created.

If students have to act as a peer model for a video-based modeling example and are not yet experts on the topic themselves, they first have to study learning materials on the subject. These learning materials are studied with a different intention than the common intention of studying for a test. That is, the materials are *studied with the intention of being able to explain the task to others*. Secondly, the peer model *actually explains the task* during the creation of the video-based modeling example. Both steps may affect students' learning outcomes (with better outcomes being reflected by higher retention and transfer test performance reached with equal or less effort investment on those tests) and beliefs about their own capabilities.

Instructing learners to study with the intention of being able to successfully explain a task to others might invoke a more active study approach and cause learners to focus less on absorbing new facts and more on interpreting and integrating new knowledge (Benware & Deci, 1984). Some studies have shown that instructing learners to study with the expectation of teaching to another student (i.e., teaching expectancy) can invoke an active study approach, and enhance learning processes and/or outcomes when compared to the more passive approach of studying to complete a test (e.g., Bargh & Schul, 1980; Benware & Deci, 1984; Renkl, 1995). Moreover, the study intention of being able to explain to others could result in different comprehension monitoring processes (e.g., asking oneself: "Why is it that…?"; "Do I understand…?"; "Can I explain…?"), and could invoke self-explanation processes, both of which have been shown to foster deep learning and understanding (comprehension monitoring: Graesser, Baggett, & Williams, 1996; Sternberg, 1987; self-explaining: Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De

Leeuw, Chiu, & LaVancher, 1994; Renkl, 1997, 2002; deep questions and explanations: Craig, Gholson, Brittingham, Williams, & Shubeck, 2012; Craig, Sullins, Witherspoon, & Gholson, 2006). Enhanced understanding should be particularly beneficial for transfer performance (e.g., Van Gog, Paas, & Van Merriënboer, 2004).

Next to the effect an explanation study intention might have, *actually* explaining the learning materials to another (non-present) person during the creation of a video-based modeling example might further improve learning outcomes. It has been shown that generating explanations can foster learning more than rereading or receiving explanations (Lombrozo, 2012). For example, asking learners to generate explanations can help them to identify and then repair knowledge gaps (Chi, 2000), to integrate new knowledge with prior knowledge (Chi et al., 1994; Lombrozo, 2006), and to transform declarative knowledge into applicable procedures (Chi et al., 1989, 1994). Whereas these studies only prompted students to explain to themselves (i.e., self-explanation), explaining with the intention of providing instruction that can be shared with others (as one would do when creating a video) can be seen as what Leinhardt (2001) refers to as providing instructional explanations[1]. According to Leinhardt, providing instructional explanations differs from simply stating or describing a concept or procedure, by more carefully examining it. That is, providing a full explanation of the concept or procedure in which key features are identified, connections to prior knowledge are made, and effective and important examples are provided. In other words, providing such explanations would foster deeper processing and elaboration of the learning materials, which might foster the explainer's learning outcomes and especially, transfer performance.

Indeed, actively providing such instructional explanations to others during small group discussions (Van Blankenstein, Dolmans, Van der Vleuten, & Schmidt, 2011) and during tutoring (Cohen, Kulik, & Kulik, 1982) has been shown to aid learning. The finding that tutoring is not only effective in terms of the tutee's learning, but also in terms of the tutor's learning (Cohen, Kulik, & Kulik, 1982), is interesting because tutors also prepare by studying learning materials with the intention of being able to explain those to others, and subsequently explain what they have learned to the tutee. The tutor learning effect not only applies when the knowledge and age gap between tutor and tutee is large (e.g., Juel, 1996; Sharpley, Irvine, & Sharpley, 1983), but also when that gap is small (e.g., Coleman, Brown, & Rivkin, 1997; McMaster, Fuchs, & Fuchs, 2006; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003).

[1] Note that other authors have used the term 'instructional explanations' in a somewhat more restricted sense (e.g., Wittwer & Renkl, 2010).

These findings suggest that acting as a peer model may also have beneficial effects on learning. However, peer tutors' learning gains may stem from other factors than an explanation study intention and actual explaining: peer tutors' learning may also be affected by the interaction with the tutee, who may ask questions that stimulate the peer tutor's reflective knowledge-building in the process of formulating an answer to those questions (Graesser, Person, & Magliano, 1995; Roscoe & Chi, 2007). Peer models in video modeling examples, on the other hand, are explaining to fictitious peers who are not physically present.

To the best of our knowledge, the only study that has investigated the effects of acting as a peer model by a) preparing and studying learning materials and b) explaining what was learned by creating a video-based modeling example, was conducted by Spires, Hervey, Morris, and Stelpflug (2012). As part of a collaborative learning course, secondary education students were asked to create a 5 min. long video. The authors concluded, based on students' self-reports, that the video creation process fostered both motivation and learning. However, because of the lack of experimental control (e.g., no control group for study intention and actually explaining) and reliance on self-reports, no conclusions can be drawn from this study regarding the effects of study intention and video creation on learning and motivation.

The Present Study

The purpose of the current experiments was to investigate and disentangle the effects of acting as a peer model on learning and transfer and explore potential effects on self-efficacy and perceived competence. A study intention of being able to explain a task to others, and actually explaining that task during video creation, may not only affect learning and transfer, but also how peer models view and assess their own capabilities to perform that task. Again, it has been shown that observing models may enhance self-efficacy (Bandura, 1997; Schunk, 1987), but whether the process of *acting* as a peer model affects self-efficacy and perceived competence is not known. Exploring this question would be interesting, as self-efficacy beliefs have been hypothesized to underpin motivation, well-being, persistence, study behavior, and achievement (Bandura, 1997; Bong & Skaalvik, 2003; Schunk, 2001). Perceived competence is closely related to self-efficacy, but whereas self-efficacy represents specific expectations and convictions for a certain situation, perceived competence refers to more general knowledge and perceptions that are likely more stable and enduring (Bong & Skaalvik, 2003; Hughes, Galbraith, & White, 2011; Klassen & Usher, 2010). Like self-efficacy, perceived competence has also been shown to have a significant bearing on students' motivation and learning (Harter, 1990; Ma & Kishor, 1997).

In Experiment 1, secondary education students read a text on syllogistic reasoning (the content of which was new to them) with the study intention of being able to either successfully complete a test (condition A), or to successfully explain the learning materials to others (conditions B and C). Whereas students with an explanation study intention did not actually have to create a video in one condition (B), they did have to do so in the other condition (C). Students in condition C spent the last 5 min. of the allocated study time on creating a video-based modeling example. Because a teaching expectancy can influence learning and motivation (Bargh & Schul, 1980; Benware & Deci, 1984; Renkl, 1995), the students in the video creation condition were not informed beforehand that they would actually be creating the video. Experiment 2 replicated Experiment 1 with university students, but the instructions were slightly adapted to rule out potential alternative explanations for the findings of Experiment 1.

We hypothesized that compared to studying for a test (condition A), studying with the intention of being able to successfully explain learning materials to others (i.e., without actually doing so; condition B) would facilitate a more active study approach that would benefit learning (Hypothesis 1a) and transfer (Hypothesis 1b), because of the aforementioned comprehension monitoring and self-explanation processes that the explanation intention might elicit, which could enhance understanding. Enhanced understanding is particularly beneficial for transfer performance (e.g., Van Gog et al., 2004).

Secondly, we hypothesized that actually creating a video-based modeling example (condition C) would further benefit learning (Hypothesis 2a) and transfer (Hypothesis 2b) above and beyond effects of study intention, in line with findings that show that providing explanations during tutoring (Cohen et al., 1982) or during small group discussions (Van Blankenstein et al., 2011) can aid learning. In sum, we predict the following pattern of results: condition C > B > A.

Mental effort in answering test questions will be analyzed to get more insight into the quality of learning outcomes (Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008). Mental effort invested in the test provides an additional and more subtle indicator of the quality of cognitive schemas acquired under the different instructional conditions compared to performance measures alone (Van Gog & Paas, 2008). That is, when equal/higher performance is reached with lower/equal effort investment, cognitive schemas are more efficient (Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008). We expect no differences in effort investment among conditions prior to learning, that is, in completing the pretest (Hypothesis 3a). We do, however, expect that the hypothesized higher performance on the posttest in the explanation conditions (see Hypothesis 1 and 2) will be reached with equal or lower effort investment (Hypothesis 3b).

Finally, potential effects of study intention and video-creation on self-efficacy (Question 4a) and perceived competence (Question 4b) were explored. On the one hand, when students in the explanation intention conditions feel they were successful in answering the questions they may have posed themselves while monitoring their learning and in explaining while making the video, the hypotheses regarding learning and transfer may also apply to self-efficacy and perceived competence (i.e., condition C > B > A). On the other hand, the opposite may also occur when students in the explanation conditions would become uncertain because of the questions they ask themselves. For instance, for difficult tasks with the added pressure of understanding the learning materials, anxiety has been shown to increase (Ross & DiVesta, 1976) and intrinsic motivation has been shown to decline (Renkl, 1995).

Experiment 1

Method

Participants. Participants were 76 Dutch secondary education students (47 female) in the fourth year of pre-university education (age range 15 to 17 years), which is the highest level of secondary education in the Netherlands and has a six-year duration.

Design. The experiment consisted of four phases: pretest, study phase, immediate posttest, and delayed posttest. Participants were randomly assigned to one of three conditions: (1) Test Condition (study intention: be able to successfully complete a test; no video creation; *n* = 27), (2) Explanation Condition (study intention: be able to successfully explain to others; no video creation; *n* = 25), or (3) Explanation-Video Condition (study intention: be able to successfully explain to others; video creation; *n* = 24).

Materials

The tests and the text to be studied were paper-based.

Pretest. The pretest consisted of eight conditional syllogistic reasoning items. Each item (multiple choice) required participants to indicate whether the conclusion was ‘valid’ or ‘invalid’ (i.e., whether or not it logically followed from the two premises, which participants had to assume were true). For each of the four forms of syllogistic reasoning (i.e., 1) affirming the antecedent: valid; 2) affirming the consequent: invalid; 3) denying the antecedent: invalid; and 4) denying the consequent: valid), there were two test items: one with and one without belief-bias (i.e., the tendency to confirm based on prior beliefs of real world knowledge). The belief-bias makes it harder to judge whether a conclusion is valid or invalid (George, 1995; Newstead, Pollard, Evans, & Allen, 1992). To illustrate, the pretest item for affirming the antecedent without belief-bias was: “If a person works at the V&D [a Dutch department store], then that person is 16 years old or older. Elles works at the V&D. Conclusion: Elles is 16 years old or older.”, and the item for affirming the antecedent with belief-bias was: “If an event takes place in the 22nd century, then that event takes place in the future. The Second World War takes place in the 22nd century. Conclusion: The Second World War takes place in the future.” The maximum total score on the pretest was eight points (i.e., one point for each correct answer). For a description of the self-efficacy and perceived competence items used in the pretest, see below.

Study Text. The study text was 1930 words and six pages long. It described how one can judge for syllogism reasoning tasks whether or not a conclusion logically follows from the two premises. The text was specifically created for the present experiment and throughout the text, the first and second premises were respectively addressed to as ‘the rule’ and ‘the observation’, after which the ‘the conclusion’ followed. There were two versions of the text, one for the Test Condition and one for both Explanation Conditions. These two versions only differed with regard to the study intention specific prompts placed at the end of each page (in the footer). These prompts were: “Can you apply the information from this page to complete a test?” (Test Condition) or: “Can you explain the information on this page to a fellow student? (Explanation Conditions)”

The text started with a general introduction to deductive reasoning and an overview of the four forms of syllogistic reasoning. The different forms of syllogistic reasoning were explained using the same recurrent example (affirming the antecedent: “If John sees a clown, then he is afraid. John sees a clown. Conclusion: John is afraid”; denying the antecedent: “If John sees a clown, then he is afraid. John does *not* see a clown. Conclusion: John is *not* afraid”; et cetera). How to judge whether a conclusion is valid or invalid was explained both in terms of the example and in abstract terms (e.g., for denying the antecedent: “If p, then q, *not* p therefore *not* q”), and for each form of syllogistic reasoning belief-bias was explained, providing a specific example. The last page provided a table that summarized which forms of syllogistic reasoning were valid and invalid (see Appendix A), using another generic example (i.e., “If this is an apple, then it is fruit”).

Video Creation. Participants in the Explanation-Video Condition were instructed on paper to 1) explain the four forms of syllogistic reasoning in front of a webcam, as if explaining them to someone without any knowledge on the subject, with the help of an adapted table taken from the study text (see Appendix B), and 2) explain what errors people commonly make when judging whether a conclusion is valid or invalid, and why people make this mistake (this refers to the belief-bias, but belief-bias was not explicitly mentioned in the instructions).

Posttests. The immediate and delayed posttests consisted of eight conditional syllogistic reasoning items (two for each form; one with and one without belief-bias) to assess learning outcomes, that is, retention of information from the experimental text about syllogisms. These test items were structurally equivalent but different in surface features compared to the pretest and again required participants to decide whether a conclusion was valid or invalid, but now they were also asked to explain their answer.

To assess transfer (i.e., applying what had been learned from the text to new tasks), two Wason selection tasks were used (task 1: concrete context; task 2: abstract). The Wason selection tasks (Wason, 1966) ask people how they can test the validity of a rule such as “if a card has Y on one side, then it has 2 on the other side” by turning two cards from a set of four (e.g., showing X, Y, 2, 7). This task can be solved correctly if one understands the validity of the different reasoning forms as taught in the syllogism tasks and knows how to apply them. That is, to reach the correct solution, one needs to affirm the antecedent (turning Y) and deny the consequent (turning 7), but most people tend to affirm the antecedent (turning Y) as well as the consequent (turning 2). The Wason selection task items were introduced by stating that these were new tasks, but that what had been learned from the text could be used to successfully complete them. Again, participants were asked to select the right answer as well as to explain their answer.

Two parallel versions (A and B) of the posttest were created. Both test versions were structurally equivalent but different in surface features. On both posttests, the maximum score on the syllogistic reasoning items (i.e., learning outcomes) was 56 points. Each no belief-bias item was worth six points (one point for the correct choice on the multiple-choice question and five points for the explanation: one point for correctly recalling the form of syllogistic reasoning, one point for explaining correctly in abstract terms of p and q, two points for explaining correctly in concrete terms, and one point for correctly concluding in the explanation whether a conclusion was valid or invalid) and each belief-bias item was worth eight points (identical scoring to the no belief bias items, but with two points extra for correctly explaining the belief-bias). Note that scoring of the explanations was not dependent on correctness of the initial answer, so some points could still be gained for correct elements in the explanation even when the answer regarding validity of the conclusion was incorrect. Participants could earn 18 points in total for the Wason selection tasks (i.e., nine per question; one point for selecting the correct answer, and two points per correct explanation for each of the four forms of syllogistic reasoning as applied to the rule in the Wason selection task).

Two raters independently scored 10% of the tests. The intra-class correlation coefficient was .949. Because of the high inter-rater reliability the remainder of the tests was scored by one rater.

Mental Effort. Invested mental effort was measured after each test item on the pretest and both posttests using a subjective 9-point rating scale (Paas, 1992), which has a range from (1) very, very low effort to (9) very, very high effort. Average invested mental effort was computed separately for the syllogistic reasoning items and the Wason selection tasks.

Self-efficacy and Perceived Competence Scale. Self-efficacy was measured using an adapted version of Bandura's (2006) problem-solving self-efficacy scale, which asks participants to rate the degree of confidence (from 0% to 100%) in their ability to solve an incremental percentage of the total number of problems (from 10% to 100%, with increases of 10%). We altered this scale slightly by asking participants after the pretest to rate their degree of confidence (from 0% to 100%) in their ability to learn an in-depth explanation of the eight items (1 out of 8 to 8 out of 8, with increases of 1). Thus, participants firstly indicated their degree of confidence (0-100%) in their ability to learn an in-depth explanation of 1 out of 8 test items well, then the degree of confidence (0-100%) in their ability to learn an in-depth explanation of 2 out of 8 test items well, and so forth till 8 out of 8 test items. The wording of the self-efficacy measurements prior to each posttest was adjusted by asking participants to rate their confidence (from 0% to 100%) in their ability to answer an incremental amount of questions on a test (1 out of 8 to 8 out of 8, with increases of 1).

Perceived competence was measured using an adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996), which is a four item questionnaire that asks people to rate on a scale of 1 (not at all true) to 7 (very true): "I feel confident in my ability to learn this material", "I am capable of learning the material in this course", "I am able to achieve my goals in this course", and "I feel able to meet the challenge of performing well in this course". We altered this scale slightly by asking participants to rate the perceived competence in their capability to learn an in-depth explanation of the eight items (after the pretest) and in answering eight questions on a test (prior to each posttest). Both self-efficacy and perceived competence ratings were averaged for each measurement (i.e., after the pretest and before the immediate and delayed posttest).

Procedure

The study was run in two sessions. The first session (ca. 50 min.) took place in the university lab. In a room with an entire class of students present, students were randomly allocated to a condition, after which the class was split-up in small groups of max. 10 participants according to condition; each group was supervised by an experiment leader. Participants in the Explanation-Video Condition were seated in sound-proof cubicles with the doors open so that they could hear and see the experiment leader; the doors were only closed once they started the actual video creation so that they would not hear each other. Participants in the other two conditions were either tested in sound-proof cubicles or a larger room with several seats between them. Participants received an envelope with three booklets. The experiment leader gave a general introduction in which the study intention was mentioned. Then, participants were asked to take out the first booklet, containing the pretest and the first self-efficacy and perceived competence measures. They were given 12 min. to complete the pretest. When time was up, the experiment leader asked participants to return the first booklet to the envelope and take out the second booklet, which contained the study text. The experimenter repeated the study intention prompt and indicated the amount of time participants would get to study the text. The Test and Explanation Conditions received 17 min. study time, while participants in the Explanation-Video Condition received 12 min. study time. All participants were instructed several times to study for the full length of available study time. When time was up, participants were instructed to place booklet 2 back into the envelope. Participants in the Explanation-Video Condition were then told to turn on the webcam, and to create the video using the instructions that were handed out by the experiment leader. Note that participants in the Explanation-Video condition were not aware prior to this point that they would be asked to create a video. Participants could see themselves on the computer screen during video creation. After the learning phase was over, participants were instructed that they would have 25 min. for booklet 3. This booklet contained the self-efficacy and perceived competence scales, followed by the posttest. Half of the participants in each condition received version A as the immediate posttest while the other half received version B.

The second session took place 4 days later at participants' schools, and lasted 25 min during which the delayed posttest was completed. A fourth booklet was handed out that again started with the self-efficacy and perceived competence items, followed by the posttest items. Participants who had received version A as the immediate posttest, now received version B (and vice versa).

Results

Two participants from the Explanation-Video Condition were removed from all analyses because of non-compliance with the instructions when creating the video, leaving 74 participants in total.

Table 1.
Mean (SD) of Learning and Transfer Multiple Choice (MC) and Total Test Scores and Mental Effort per Condition in Experiment 1

	Learning			Transfer		
	Test	Explanation	Explanation-Video	Test	Explanation	Explanation-Video
Test Scores						
Pretest (MC)	5.00 (0.88)	4.80 (0.96)	4.86 (0.91)			
Immediate Posttest (MC)	7.07 (1.36)	6.88 (1.27)	7.19 (1.03)	0.63 (0.93)	0.56 (0.82)	1.14 (1.01)
Immediate Posttest Total	21.07 (7.72)	21.17 (4.87)	24.81 (6.59)	3.44 (3.11)	3.30 (3.12)	6.12 (4.09)
Delayed Posttest (MC)	7.32 (1.17)	7.00 (1.19)	7.42 (0.66)	0.88 (0.97)	0.96 (0.98)	1.26 (0.94)
Delayed Posttest Total	20.56 (6.33)	18.96 (5.76)	23.08 (5.25)	3.66 (3.25)	4.06 (3.49)	5.89 (3.92)
Effort						
Pretest	2.74 (0.92)	2.87 (0.91)	2.49 (1.06)			
Immediate Posttest	2.91 (1.35)	3.36 (1.07)	3.10 (1.34)	3.35 (1.55)	3.84 (1.30)	3.82 (1.51)
Delayed Posttest	2.43 (1.14)	3.21 (1.28)	2.78 (1.41)	2.27 (1.30)	3.36 (1.35)	3.05 (1.50)

Learning Outcomes and Transfer Test Performance. The test scores on the learning (i.e., syllogism) and transfer (i.e., Wason selection) tasks in Experiment 1 are provided in Table 1, both in terms of score on the 'multiple choice' part of the questions (mc) and in terms of a 'total' (tot) score, which consists of the multiple choice score plus the score on the open-ended part of the question (explain your answer) taken together.

Five participants were absent at the delayed test (two from the Test Condition, one from the Explanation Condition, and two from the Explanation-Video Condition). For the delayed posttest scores of the five absent participants, a missing value analysis using the expectation maximization (EM) method was performed in SPSS. For the ANOVAs, Cohen's *f* is reported as a measure of effect size, with values of 0.10, 0.25, and 0.40 representing a small, medium, and large effect size, respectively. For the post-hoc tests, Cohen's *d* is reported with values of 0.20, 0.50, and 0.80 representing a small, medium, and large effect size, respectively (Cohen, 1988).

As expected, there were no differences among conditions in pretest performance, $F(2, 70) = .33$, $p = .720$, $f = 0.097$. To test our hypotheses regarding the effects on learning (Hypothesis 1a and 2a), a repeated measures ANOVA on the total performance scores on the syllogism tasks, with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor showed a main effect of Test Moment, $F(1, 70) = 4.67$, $p = .034$, $f = 0.256$. This indicated that participants performed somewhat better on the Immediate ($M = 22.18$, $SD = 6.66$) than on the Delayed Posttest ($M = 20.74$, $SD = 5.99$). There was also a significant effect of Instruction Condition, $F(2, 70) = 3.17$, $p = .048$, $f = 0.301$. Follow-up Bonferroni corrected post-hoc tests showed that, although the Explanation-Video Condition ($M = 23.94$, $SE = 1.19$) scored higher on average than the Explanation Condition ($M = 20.06$, $SE = 1.09$), this difference was not statistically significant, $p = .057$, $d = 0.710$. There were no differences between the Explanation-Video Condition and Test Condition ($M = 20.82$, $SE = 1.05$), $p = .159$, $d = 0.572$, nor between the Explanation Condition and Test Condition, $p = 1.000$, $d = 0.138$. Furthermore, there was no interaction between Test Moment and Instruction Condition, $F(2, 70) = .58$, $p = .561$, $f = 0.125$.

To test our hypotheses regarding the effects on transfer (Hypothesis 1b and 2b), a similar repeated measures ANOVA on the total performance scores on the Wason selection tasks showed no main effect of Test Moment (Immediate and Delayed Posttest), $F(1, 70) = .46$, $p = .499$, $f = 0.081$. There was a main effect of Instruction Condition, $F(2, 70) = 4.45$, $p = .015$, $f = 0.357$. Bonferroni corrected post-hoc tests on participants' score over the Immediate and Delayed Posttest showed that performance in the Explanation-Video Condition ($M = 6.01$, $SE = 0.68$) was significantly better than in the Test Condition ($M = 3.55$, $SE = 0.60$), $p = .025$, $d = 0.791$, and the Explanation Condition ($M = 3.68$, $SE = 0.62$), $p = .041$, $d = 0.749$. No difference was found between the Test Condition and Explanation Condition, $p = 1.000$, $d = 0.042$, nor an interaction between Test Moment and Instruction Condition, $F(2, 70) = 0.57$, $p = .567$, $f = 0.127$.

Mental effort. Mental effort data are shown in Table 1. Six participants had one missing value on the syllogisms tasks (i.e., learning) and eight participants had a missing value on the Wason-selection tasks (i.e., transfer), which were replaced with the series mean. One participant with more than two missing values on the syllogism tasks and two participants with more than one missing value on the Wason-selection tasks were removed from the analyses. As expected (Hypothesis 3a), an ANOVA showed no significant differences among conditions in the mean invested mental effort during the pretest, $F(2, 70) = .86$, $p = .426$, $f = 0.157$. On the posttest (Hypothesis 3b), a repeated measures ANOVA on the average mental effort invested on the syllogism tasks with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor and Instruction Condition as between-subjects factor showed a main effect of Test Moment, $F(1, 63) = 13.63$, $p < .001$, $f = 0.456$, with participants investing more mental effort during the Immediate ($M = 3.12$, $SD = 1.25$) than during the Delayed Posttest ($M = 2.80$, $SD = 1.30$). There was no effect of Instruction Condition, $F(2, 63) = 1.54$, $p = .223$, $f = 0.221$, nor an interaction effect, $F(2, 63) = 1.31$, $p = .278$, $f = 0.185$. On the Wason selection tasks, there was also a main effect of Test Moment, $F(1, 62) = 13.67$, $p < .001$, $f = 0.468$. Again, participants invested more mental effort during the Immediate Posttest ($M = 3.65$, $SD = 1.45$) than during the Delayed Posttest ($M = 3.00$, $SD = 1.39$). There was no main effect of Instruction Condition, $F(2, 62) = 1.54$, $p = .223$, $f = 0.223$, nor an interaction effect, $F(2, 62) = .26$, $p = .773$, $f = 0.083$.

Table 2.
Mean (SD) of Self-efficacy and Perceived Competence Scores per Condition in Experiment 1

	Self-efficacy			Perceived Competence		
	Test	Explanation	Explanation-Video	Test	Explanation	Explanation-Video
Pretest	80.67 (12.55)	73.94 (13.33)	82.92 (12.71)	6.04 (0.67)	5.40 (1.10)	5.83 (0.68)
Immediate Posttest	83.45 (11.02)	79.20 (15.82)	78.49 (15.76)	6.39 (0.54)	6.15 (0.69)	5.97 (0.90)
Delayed Posttest	80.78 (13.62)	71.57 (16.04)	72.40 (16.97)	6.19 (0.71)	5.65 (0.72)	5.74 (0.86)

Self-efficacy and Perceived Competence. The self-efficacy and perceived competence data are provided in Table 2. Because the questionnaires that measured self-efficacy and perceived competence prior to the immediate and delayed posttests differed in phrasing compared to the pretest, we analyzed them separately to answer Questions 4a and 4b. For the self-efficacy measure conducted after the pretest, seven participants had to be excluded because of missing values (leaving $n = 25$ in the Test, $n = 23$ in the Explanation, and $n = 19$ in the Explanation-Video Condition). An ANOVA showed no significant differences among conditions on the self-efficacy measurement after the pretest, $F(1, 64) = 2.13$, $p = .127$, $f = 0.258$. A repeated measures ANOVA with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor on the self-efficacy ratings provided before the Immediate and Delayed Posttest showed a main effect of Test Moment, $F(1, 63) = 18.39$, $p < .001$, $f = 0.528$, indicating that self-efficacy was significantly lower at the Delayed Posttest ($M = 75.30$, $SD = 15.81$) than at the Immediate Posttest ($M = 80.60$, $SD = 14.13$). There was no main effect of Instruction Condition, $F(2, 63) = 1.80$, $p = .173$, $f = 0.239$, nor an interaction effect, $F(2, 63) = 1.44$, $p = .245$, $f = 0.188$.

The adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996) used in the present study showed high internal consistency on the Pretest (Cronbach's $\alpha = .925$), Immediate Posttest (Cronbach's $\alpha = .920$), and Delayed Posttest (Cronbach's $\alpha = .947$). Perceived competence after the pretest differed significantly among conditions, $F(1, 70) = 3.90$, $p = .025$, $f = 0.334$. Because the experimental procedure for both Explanation Conditions prior to the learning phase was still identical, a contrast test was conducted to investigate whether the Test Condition differed on perceived competence from both Explanation Conditions taken together after the pretest (i.e., after only one study intention prompt). Results showed a significant difference, $t(71) = 2.15$, $p = .035$, $d = 0.509$, with the Test Condition showing higher perceived competence ($M = 6.04$, $SE = 0.16$) than the Explanation Conditions ($M = 5.39$, $SE = 0.17$). A repeated measures ANOVA with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor on perceived competence ratings provided before the Immediate and Delayed Posttest, showed a main effect of Test Moment, $F(1, 64) = 18.99$, $p < .001$, $f = 0.530$, indicating a decrease in perceived competence from the Immediate ($M = 6.19$, $SD = 0.72$) to the Delayed Posttest ($M = 5.88$, $SD = 0.79$). Although the Test Condition seemed to have higher mean scores on perceived competence, there was no statistically significant effect of Instruction Condition, $F(2, 64) = 2.88$, $p = .063$, $f = 0.300$. There was no statistically significant interaction effect between Test Moment and Instruction Condition, $F(2, 64) = 1.76$, $p = .181$, $f = 0.206$.

Discussion

In contrast to our first hypothesis, studying with the intention of being able to successfully explain materials to others (i.e., Explanation Condition) was not more beneficial for learning (Hypothesis 1a) and transfer (Hypothesis 1b) than studying to be able to successfully complete a test (i.e., Test Condition). Regarding our second hypothesis, while actually explaining to non-present others by creating a video-based modeling example (i.e., Explanation-Video Condition) did not have an effect on learning (Hypothesis 2a), it did have a significant beneficial effect on transfer (Hypothesis 2b). This beneficial effect on transfer performance was reached with the same investment of mental effort on the posttest (Hypothesis 3b), indicating that the cognitive schemas acquired in the Explanation-Video Condition were not only more effective, but also more efficient (Van Gog & Paas, 2008).

The explorative analysis of potential effects that study intention and creating a video-based modeling example might have on self-efficacy and perceived competence (Questions 4a and 4b), showed no effects of instructional condition on self-efficacy. As for perceived competence, there was a significant difference between the Test Condition and both Explanation Conditions taken together after the pretest, that is, after just one study intention prompt. On the posttests, even though studying new materials with the intention of completing a test seemed to lead to a somewhat higher confidence in one's capabilities, there were no statistically significant differences among conditions.

In sum, the results of Experiment 1 show that the process of explaining during video creation is particularly effective for enhancing transfer performance, which suggests that acting as a peer model for a video-based modeling example can be an effective educational activity in and of itself.

Based on Experiment 1, however, it is not entirely clear whether this positive effect was 'simply' caused by recalling information from long-term memory (i.e., retrieval practice) or specifically by explaining during video creation. That is, only participants in the Explanation-Video Condition had to recall information from memory, and they were given a table that they could use to explain the four types of syllogisms. Engaging in (cued) retrieval practice has been shown to positively affect learning outcomes (Karpicke & Blunt, 2011). Furthermore, the Explanation-Video Condition also received the instruction to focus on what errors people commonly make when judging whether a conclusion is valid or invalid, and why people make this mistake (i.e., the belief-bias). Although the study text described the belief-bias repeatedly throughout the text, we cannot rule out the possibility that the emphasis on explaining this error has influenced the transfer results.

To address these issues, a second experiment was conducted. In Experiment 2 we investigated whether we could replicate the results from Experiment 1 with university students who were also novices on the topic of syllogistic reasoning, while ruling out retrieval practice and the focus on common errors as factors that could explain the positive findings of the process of acting as a peer model in Experiment 1.

Experiment 2

In Experiment 2, after reading the text for 12 min., participants in the Test and Explanation Condition also engaged in a short cued recall activity prior to continuing studying of the text. They were instructed to fill in the gaps of the table in Appendix B from memory (i.e., to apply the example 'If this is an apple, then it is fruit' to the four forms of syllogistic reasoning) to ensure that all three conditions engaged in retrieval practice after studying for 12 minutes. Furthermore, the Explanation-Video Condition no longer received the instruction to explain during video creation what errors people commonly make when judging whether a conclusion is valid or invalid, and why people make this mistake.

If the positive effects of the process of acting as a peer model for a video-based modeling example in Experiment 1 arose because of retrieval practice and/or focusing on the belief-bias during video creation, then no differences among the conditions should be found in Experiment 2. However, as in Experiment 1, we hypothesize that the key processes in the Explanation-Video Condition of a) studying with the intention of explaining to others later on and b) actually explaining during video creation, will have beneficial effects on learning outcomes, particularly on transfer. The reader is referred to the Introduction of this Chapter for a description of hypotheses and research questions.

Participants

Participants were 95 Dutch undergraduate students (age M = 20.41, SE = 0.19; 65 female; 90 of those participants studied Psychology in a Problem-Based Learning curriculum). Students who had studied or were studying philosophy or students with programming experience were not eligible for participation to ensure that the population consisted of novices with regard to the topic of syllogistic reasoning. Participants were rewarded for their participation with course credits or a monetary reward.

Design

The design was the same as in Experiment 1. Participants were randomly assigned to one of three conditions: (1) Test Condition (n = 32), (2) Explanation Condition (n = 31), or (3) Explanation-Video Condition (n = 32).

Materials and Procedure

The same materials and procedure were used as in Experiment 1, with a few exceptions.

Firstly, after the text (booklet 2) that was studied for 12 min., a booklet 3 was handed out that presented participants in all conditions with the table taken from the study text, although it was now no longer filled out as it was in the study text (see Appendix B). In the Test Condition and Explanation Condition, participants were instructed to fill in the gaps in the table from memory. If they were finished, they were instructed to proceed to turn the page and study the rest of booklet 3, which presented the text again for restudy. They were given max. 5 min. in total for recall and restudy. To investigate whether retrieval from long-term memory was successful (i.e., how well students had filled in the gaps in the table), recall performance was scored. Participants could earn a total of 8 points, 2 points per correctly applied form of syllogistic reasoning. They received 0 points for a wrong answer, 1 point for a partially correct answer, and 2 points for a fully correct answer. In the Explanation-Video Condition, the table was preceded by the instruction to use it while explaining the four forms of syllogistic reasoning in front of a webcam, as if explaining them to someone without any knowledge on the subject. Participants were given maximally 5 min. for creating the video.

Secondly, the study intention prompt during the general instruction at the start of the experiment was removed to ensure that the self-efficacy and perceived competence measures made after the pretest could act as control variables for the self-efficacy and perceived competence measures made prior to both posttests. Thirdly, all participants were tested in individual cubicles with maximally eight participants per session during both sessions. Finally, the delayed test took place 7 days later (instead of 4 days in Experiment 1).

Posttest performance was scored in the same manner as in Experiment 1. Two raters independently scored 10% of the posttests. The intra-class correlation coefficient was 0.898, and because of the high agreement the remainder of the tests was scored by one rater.

Results

One participant from the Explanation-Video Condition was removed from all analyses because of non-compliance with the instructions.

To assess the quality of recall (i.e., filling in the gaps in the table depicted in Appendix B) in the Test and Explanation Condition, an independent samples t-test was conducted. There was no statistically significant difference between the Test Condition (M = 7.84, SD = 0.63) and the Explanation Condition (M = 7.74, SD = 0.86), $t(61)$ = 0.54, p = .591, d = 0.136. Both were highly successful (the maximum score that could be obtained was 8) in filling in the gaps in the table from memory.

Learning Outcomes and Transfer Test Performance. The multiple choice and total (i.e., mc + explanation) test scores on the learning (i.e., syllogism) and transfer (i.e., Wason selection) tasks in Experiment 2 are displayed in Table 3. Three participants from the Test Condition were absent at the delayed test. An EM missing value analysis in SPSS was performed to replace their missing values.

Table 3.
Mean (SD) of Learning and Transfer Multiple Choice (MC) and Total Test Scores and Mental Effort per Condition in Experiment 2

	Learning			Transfer		
	Test	Explanation	Explanation-Video	Test	Explanation	Explanation-Video
Test Scores						
Pretest (MC)	5.47 (1.50)	5.52 (1.12)	5.58 (1.12)			
Immediate Posttest (MC)	7.44 (1.19)	7.58 (0.85)	7.52 (0.81)	1.31 (0.93)	1.55 (0.77)	1.58 (0.76)
Immediate Posttest Total	21.69 (6.39)	25.15 (5.78)	27.18 (6.09)	5.39 (3.33)	6.53 (3.00)	6.50 (3.04)
Delayed Posttest (MC)	6.88 (1.76)	7.26 (1.26)	7.07 (1.67)	1.14 (0.97)	1.71 (0.69)	1.74 (0.63)
Delayed Posttest Total	18.82 (6.51)	23.90 (5.68)	25.81 (7.99)	4.41 (2.86)	6.24 (2.65)	6.90 (2.98)
Effort						
Pretest	3.48 (1.09)	3.20 (1.15)	3.71 (0.85)			
Immediate Posttest	2.87 (1.33)	2.57 (1.09)	3.34 (1.36)	4.31 (1.78)	4.63 (1.65)	5.00 (2.19)
Delayed Posttest	2.91 (1.39)	2.43 (1.12)	3.41 (1.76)	3.88 (1.87)	3.48 (1.86)	4.29 (1.99)

An ANOVA showed no significant differences among conditions in pretest performance, $F(2, 91) = 0.06$, $p = .940$, $f = 0.037$. A repeated measures ANOVA on the total performance scores on the syllogism tasks (Hypothesis 1a and 2a), with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor showed a main effect of Test Moment, $F(1, 91) = 8.30$, $p = .005$, $f = 0.300$, indicating a decrease in learning performance from the Immediate ($M = 24.64$, $SD = 6.45$) to the Delayed Posttest ($M = 22.80$, $SD = 7.35$). There was also a significant main effect of Instruction Condition, $F(2, 91) = 9.99$, $p < .001$, $f = 0.469$. Bonferroni corrected post-hoc tests showed that the Test Condition ($M = 20.25$, $SE = 1.00$) was significantly outperformed by both the Explanation-Video Condition ($M = 26.49$, $SE = 1.02$), $p < .001$, $d = 1.100$, and the Explanation Condition ($M = 24.52$, $SE = 1.02$), $p = .011$, $d = 0.753$. There was no significant difference between the Explanation-Video and Explanation Condition, $p = .526$, $d = 0.347$, nor was there an interaction effect between Test Moment and Instruction Condition, $F(2, 91) = 0.68$, $p = .507$, $f = 0.117$.

To test our hypotheses regarding the effects on transfer (Hypothesis 1b and 2b), a similar repeated measures ANOVA on the total performance scores on the Wason selection tasks showed no main effect of Test Moment (Immediate and Delayed Posttest),

$F(1, 91) = 1.16$, $p = .283$, $f = 0.110$. There was a main effect of Instruction Condition, $F(2, 91) = 4.04$, $p = .021$, $f = 0.298$. Bonferroni post-hoc tests showed no difference between the Explanation-Video ($M = 6.70$, $SE = 0.48$) and Explanation Condition ($M = 6.39$, $SE = 0.48$), $p = 1.000$, $d = 0.117$, nor between the Explanation Condition and Test Condition ($M = 4.90$, $SE = 0.48$), $p = .093$, $d = 0.552$. However, as expected, performance in the Explanation-Video Condition was significantly higher than in the Test Condition, $p = .028$, $d = 0.700$. No interaction effect was found between Test Moment and Instruction Condition, $F(2, 91) = 2.24$, $p = .112$, $f = 0.221$.

Mental Effort. Mental effort data are shown in Table 3. For one participant, a missing value on the pretest was replaced with the series mean. On the posttests, five participants had one missing value on the syllogisms tasks (i.e., learning) and five participants had a missing value on the Wason-selection tasks (i.e., transfer), all of which were replaced with the series mean. One participant with more than two missing values on the syllogism tasks was removed from the analysis on learning. As expected (Hypothesis 3a), an ANOVA showed no significant differences among conditions on the mean invested mental effort during the pretest, $F(2, 91) = 1.87$, $p = .159$, $f = 0.203$. For the average mental effort invested on the syllogism tasks, a repeated measures ANOVA with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor and Instruction Condition as between-subjects factor showed no main effect of Test Moment, $F(1, 87) = 0.01$, $p = .913$, $f = 0.012$. There was, however, a significant main effect of Instruction Condition, $F(2, 87) = 3.471$, $p = .035$, $f = 0.282$. Bonferroni corrected post-hoc tests showed that participants in the Explanation-Video Condition ($M = 3.37$, $SE = 0.24$) invested significantly more mental effort than participants in the Explanation Condition ($M = 2.50$, $SE = 0.23$), $p = .030$, $d = 0.675$. No differences were found between the Explanation-Video Condition and the Test Condition ($M = 2.89$, $SE = 0.24$), $p = .469$, $d = 0.373$, nor between the Explanation Condition and Test Condition, $p = .740$, $d = 0.301$. No interaction effect was found between Test Moment and Instruction, $F(2, 87) = .68$, $p = .510$, $f = 0.125$.

For the average mental effort invested on the Wason selection tasks, a repeated measures ANOVA showed a main effect of Test Moment, $F(1, 88) = 19.49$, $p < .001$, $f = 0.463$. with invested mental effort being significantly lower on the Delayed Posttest ($M = 3.88$, $SD = 1.92$) than on the Immediate Posttest ($M = 4.65$, $SD = 1.89$). There was no main effect of Instruction Condition, $F(2, 88) = 1.14$, $p = .325$, $f = 0.161$, nor an interaction effect, $F(2, 88) = 1.44$, $p = .242$, $f = 0.164$.

Table 4.
Mean (SD) of Self-efficacy and Perceived Competence Scores per Condition in Experiment 2

	Self-efficacy			Perceived Competence		
	Test	Explanation	Explanation-Video	Test	Explanation	Explanation-Video
Pretest	71.60 (15.48)	74.20 (14.40)	70.75 (17.32)	5.73 (0.91)	5.92 (0.68)	5.49 (0.88)
Immediate Posttest	82.38 (13.55)	83.60 (13.12)	79.01 (16.97)	6.19 (0.72)	6.40 (0.60)	5.81 (0.99)
Delayed Posttest	80.44 (15.36)	82.19 (13.51)	73.67 (19.34)	6.01 (0.91)	6.13 (0.58)	5.66 (1.03)

Self-efficacy and Perceived Competence. The self-efficacy and perceived competence data are provided in Table 4. Note that in the present experiment the self-efficacy and perceived competence estimates made after the pretest acted as control variables as a result of the removal of the study prompt prior to the pretest. Because the questionnaires that measured self-efficacy and perceived competence prior to the immediate and delayed posttests differed in phrasing compared to the pretest, we analyzed them separately to answer Questions 4a and 4b. For self-efficacy, an ANOVA showed no significant differences among conditions on the measurement after the pretest, $F(2, 91) = 0.40$, $p = .669$, $f = 0.094$. A repeated measures ANOVA with Test Moment (Immediate Posttest and Delayed Posttest) as within-subjects factor, and Instruction Condition as between-subjects factor on the self-efficacy ratings provided before the Immediate and Delayed Posttest showed a main effect of Test Moment, $F(1, 88) = 6.00$, $p = .016$, $f = 0.258$, indicating that self-efficacy significantly decreased from the Immediate Posttest ($M = 81.65$, $SD = 14.64$) to the Delayed Posttest ($M = 78.73$, $SD = 16.52$). There was no main effect of Instruction Condition, $F(2, 88) = 1.75$, $p = .180$, $f = 0.199$, nor an interaction effect, $F(2, 88) = 1.10$, $p = .339$, $f = 0.153$.

The adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996) again showed high internal consistency on the Pretest (Cronbach's α = .842), Immediate Posttest (Cronbach's α = .945), and Delayed Posttest (Cronbach's α = .934). Perceived competence after the pretest did not differ significantly among conditions, $F(2, 91) = 2.07$, $p = .133$, $f = 0.213$. A repeated measures ANOVA showed a significant effect of Test Moment (Immediate Posttest and Delayed Posttest), $F(1, 88) = 4.30$, $p = .041$, $f = 0.220$. Participants' perceived competence was significantly lower on the Delayed Posttest ($M = 6.00$, $SD = 0.89$) than on the Immediate Posttest ($M = 6.13$, $SD = 0.82$). There was also a significant main effect of Instruction Condition, $F(2, 88) = 5.23$, $p = .007$, $f = 0.345$. Bonferroni corrected post-hoc tests showed that the Explanation Condition ($M = 6.36$, $SE =$

0.14) scored higher on perceived competence than the Explanation-Video Condition (*M* = 5.73, *SE* = 0.14), $p = .005$, $d = 0.818$. No difference was found between the Explanation Condition and Test Condition (*M* = 6.10, *SE* = 0.14), $p = .575$, $d = 0.340$, nor between the Explanation-Video Condition and the Test Condition, $p = .203$, $d = 0.477$. There was no interaction between Test Moment and Instruction Condition, $F(2, 88) = 0.25$, $p = .782$, $f = 0.073$.

Discussion

Congruent with our hypothesis, studying with the intention of being able to successfully explain materials to others without actually doing so (i.e., Explanation Condition) was more beneficial for learning (Hypothesis 1a) than studying to successfully complete a test (i.e., Test Condition). However, actually explaining to non-present others during video creation (i.e., Explanation-Video Condition) was not more effective for learning than studying to explain to others without actually doing so (Hypothesis 2a). Interestingly, actually having explained to others did result in higher mental effort investment on the posttest than only having had the intention to explain to others or the intention to study for a test. This suggests that with regard to learning, the Explanation Condition had acquired more efficient cognitive schemata than the Explanation-Video Condition (Van Gog & Paas, 2008).

This second experiment again showed positive effects on transfer of the processes involved in acting as a peer model relative to studying with the intention of successfully completing a test. No effect of study intention on transfer was found (Hypothesis 1b), but in line with our hypothesis, actual video creation significantly fostered transfer performance compared to studying for a test (Hypothesis 2b). This higher transfer performance was reached with equal investment of mental effort on the posttest as in the other conditions (Hypothesis 3b), suggesting that the cognitive schemas acquired in the Explanation-Video Condition were not only more effective, but also more efficient compared to the Test Condition (Van Gog & Paas, 2008).

Our exploration of self-efficacy (Hypothesis 4a) showed no differences among the conditions. However, studying with the intention of being able to successfully explain materials to others without actually doing so led to higher perceived competence than studying to explain to others followed by video creation (Hypothesis 4b).

In sum, Experiment 2 showed a positive effect of studying with the intention of explaining to others on learning. Moreover, Experiment 2 provided additional evidence for the notion that engaging in the two processes involved in acting as a peer model for a video-based modeling example (i.e., first studying with the intention of explaining later on and then actually explaining on webcam) are more effective for fostering transfer than studying with the intention of completing a test, while ruling out retrieval practice and the focus on common errors as factors for these positive effects.

General Discussion

The present experiments investigated the effects of study intention and creating a video-based modeling example on learning, transfer, and mental effort, and explored the effects on self-efficacy, and perceived competence. Experiment 1 showed no differences between studying to be able to successfully complete a test (i.e., Test Condition) and studying with the intention of being able to successfully explain materials to others without actually doing so (i.e., Explanation Condition) on learning (Hypothesis 1a) and transfer (Hypothesis 1b). Experiment 2, however, did show an effect of study intention on learning, but not for transfer. For university students, a study intention of explaining to others later on was more beneficial for learning than studying for a test. There are two potential explanations for this finding. First, the fact that the majority of participants in Experiment 2 were enrolled in a PBL curriculum means that they had a substantial amount of experience with explaining studied materials to others (a post-discussion of the literature that was studied at home is the last step of the PBL model used in this curriculum; Loyens, Kirschner, & Paas, 2012). Thus, these findings might suggest that the secondary education students may not have fully adopted the unfamiliar explanation study intention, while the university students for whom this was a more familiar approach, did adopt it. Second, in Experiment 2, participants in the Test and Explanation Condition also had to engage in memory retrieval, which was not the case in Experiment 1. Possibly, it was the explanation study intention combined with retrieval practice that fostered learning.

In line with our second hypothesis, actually explaining to non-present others by creating a video-based modeling example (i.e., Explanation-Video Condition) had an effect on learning (Hypothesis 2a) and on transfer (Hypothesis 2b). In Experiment 1, the process of video creation fostered transfer (Hypothesis 2b) relative to studying to successfully complete a test. In Experiment 2, video creation was more effective for both learning and transfer compared to studying for a test. Moreover, whereas the Explanation-Video Condition significantly outperformed the Explanation Condition on transfer performance in Experiment 1, this was not the case in Experiment 2. Similar to the findings with regard to learning outcomes, the better transfer performance of the Explanation Condition in Experiment 2 compared to Experiment 1, could be a result of participants' experience with a study intention of explaining to others later on and/or of engaging in memory retrieval before restudying. Nevertheless, the Explanation Condition did not perform significantly better on transfer than the Test Condition, only the Explanation-Video Condition did (d = 0.552). These results show that actually providing explanations to others not only has positive effects on learning in interactive situations such as tutoring (Cohen et al., 1982) and small group discussions (Van Blankenstein et al., 2011), but also in the non-interactive situation of explaining while creating a video-based modeling example. Moreover, the results showed that explaining mainly affects transfer, which was not measured in earlier studies.

In sum, two experiments showed that explaining to non-present others during video creation resulted in better transfer performance relative to studying for a test. To be able to solve and explain the Wason selection tasks correctly (i.e., transfer), students had to understand the validity of the different reasoning forms as taught in the syllogism tasks and had to know how to apply this information to a new task. Apparently, explaining during video creation helped participants in both experiments to see the structural analogy between the two types of tasks and realize how the syllogistic reasoning rules could be successfully applied to the Wason selection tasks (for a discussion of the importance of perceiving analogies for transfer, see Gick & Holyoak, 1983; Needham & Begg, 1991; Renkl, 2011). But what was it about explaining aloud during video creation that helped participants see this analogy? The results of Experiment 2 show that it is not simply an effect of recalling information from memory, nor a result of focusing on the belief-bias during video creation. It is likely that actively explaining to non-present others helped participants to make connections between different elements of information and to process information on a deeper level, resulting in improved understanding, which has been suggested to be particularly beneficial for transfer performance (Van Gog et al., 2004). Furthermore, speaking aloud while explaining may also have been a contributing beneficial factor. Research on the production effect has shown that producing a word aloud can increase its distinctiveness and therefore improve explicit memory compared to studying in silence (e.g., MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). Thus far, however, it is unclear whether the beneficial effect can be generalized from learning vocabulary to learning a text.

The results on invested mental effort show that participants who studied with the intention of being able to successfully explain materials to others without actually doing so invested significantly less mental effort in completing the posttest items that measured learning than participants who did explain to others during video creation, while both conditions performed equally on those items. This indicates that the Explanation

Condition acquired more efficient cognitive schemas (Hypothesis 3b; see Van Gog & Paas, 2008), a surprising finding that is hard to explain. Congruent with our expectations, however, both Experiment 1 and 2 showed that the beneficial effects of explaining during video creation on transfer performance were reached with the same investment of mental effort on the transfer items. This indicates that, in line with our hypothesis, the acquired cognitive schemas of students who actually explained to others during video creation were not only more effective but also more efficient for transfer performance than those of students in the other conditions.

It would be interesting for future research to also measure the invested mental effort during the learning phase, because it could be hypothesized that actually explaining on video might require more mental effort as a learning strategy than studying for a test or studying with the intention to explain but not actually doing so. In other words, it might be an activity that imposes germane, or effective, cognitive load (Sweller et al., 1998). Measuring invested effort during or shortly after the learning phase would ideally have to be in addition to measuring effort invested in the test though, because mental effort ratings during the learning phase can be difficult to interpret on their own. Not only are such ratings directly affected by the instructional conditions, but levels of mental effort invested during learning cannot tell us whether this effort is evoked by processes relevant (i.e., germane load) or irrelevant (i.e., extraneous) for learning without looking at the quality of the cognitive schemas that were acquired (Van Gog & Paas, 2008).

Our explorative analysis showed no effects of study intention and creating a video-based modeling example on self-efficacy (Question 4a). Self-efficacy did however significantly decrease from immediate to delayed posttest in both experiments. Effects on perceived competence (Question 4b) were found in both experiments. The results of Experiment 1 showed that, for secondary education students, receiving a prompt to study new materials with the intention of successfully completing a test led to more confidence in their own capabilities after completing a pretest (i.e., before the learning phase). After the learning phase, the differences among conditions were not statistically significant. For the university students in Experiment 2 –in which the study intention prompt that was given during the general instruction at the start of the experiment was removed– the results suggest that the process of video creation lowered confidence in one's own capabilities in comparison to studying with the intention of explaining later on without actually explaining, but not compared to studying with the intention to complete a test. A possible explanation for this finding might be that a more active approach to learning, such as explaining during video creation, can make novice learners more aware of the difficulty of the materials (see also findings that teaching expectancy may lead to increased anxiety, Ross & DiVesta, 1976, and decreased intrinsic motivation, Renkl, 1995), while they may not realize that these difficulties may be conducive to performance on a subsequent test. Another possible explanation is the degree to which the activities in the learning phase were new to the students; the majority of university students had experience both with test and explanation study intentions (due to their PBL curriculum), but not with video creation. However, these explanations are tentative at best, because we found no such effect in Experiment 1 and only found the effect compared to the Explanation Condition, not compared to the Test Condition in Experiment 2. Concerning the question what differences between conditions in perceived competence mean exactly, it would be interesting in future research to investigate whether students' beliefs about their abilities are accurate considering their level of test performance.

Our findings are also of interest for educational practice. The use of video-based instruction in general is increasing in formal education, with educators even arguing for "flipping the classroom", that is, having learners study video lessons at home, and using teaching time in school for practice and teacher support (Bergmann & Sams, 2012). Moreover, the use of video-based modeling examples is also increasing in both formal and informal learning (see e.g., www.khanacademy.org). Asking students to create video-based modeling examples for other students does not only provide a more cost-efficient alternative to using teachers as models, but our results show it might also be an effective learning activity in and of itself. However, before educators can start relying on modeling examples created by peer students, research should first verify whether the videos that students create when acting as a peer model constitute effective educational materials. Whereas previous research does indeed suggest that video modeling examples created by peers are effective for students to learn from (e.g., Groenendijk et al., 2013a, 2013b), the examples in those studies focused on different tasks and were created under very different circumstances from those in the current study. Before videos of the kind created in this study can be used in educational practice, students may have to be allowed to design, edit, or re-do the video as well.

Future research could investigate whether the positive effects of the processes involved in acting as a peer model on learning and transfer only apply to students learning reasoning tasks, or if they can be generalized to other learning materials and task types, such as problem solving tasks. Possibly, the learning effects might even increase when students would be given more time to think about how to present the material to others (allowing them to design and edit the video itself); research in other domains has shown that learning by designing hypermedia (Lehrer & Romberg, 1996; Penner, Lehrer, & Schauble, 1998) and by designing “slow animations” (Hoban, Loughran, & Nielsen, 2011) can have positive effects on learning outcomes. Future research could also vary the degree of support that peer models receive during video creation. For example, peer models could be instructed to follow a predetermined script, which has been shown to scaffold knowledge acquisition in the interactive domain of computer-supported collaborative learning (Fischer, Kollar, Stegmann & Wecker, 2013; Kollar, Fischer, & Slotta, 2007). Moreover, varying the degree of support by using scripts could help answer the question whether the process of explaining is already sufficient to produce the performance benefits. Another question that future research could explore is how learners’ conceptions of the posttest affect their study behavior and consequently the quality of acquired schemata. Possibly, if learners would be explicitly told that successful performance on the test relies on deep processing, they might study in a different manner and such a test study intention condition may become more similar to an explanation study intention. In addition, as mentioned above, future research should further investigate effects on self-efficacy and perceived competence, but should also try to address the question of whether students’ beliefs about their abilities are accurate considering their level of test performance. Finally, it remains an open question whether the presence of others would affect the quality of explanations and the confidence novice learners have in their own abilities. Whereas a peer tutor directly explains to the tutee, and consequently is fully aware of who the recipient of the explanation is, a peer model for a video-based modeling example may not be aware who the recipient is because the explanation is given indirectly. Further research on these issues would seem potentially fruitful for both educational theory and practice.

Appendix A.

Summary Table of Syllogistic Reasoning Using a Generic Example Presented in the Study Text

<table>
<tr><th colspan="3">If this is an apple, then it is fruit</th></tr>
<tr><th></th><th>Affirm</th><th>Deny</th></tr>
<tr><td>Antecedent</td><td>If it is an apple
Then it is fruit
Valid</td><td>If it is NOT an apple
Then it is NOT fruit</td></tr>
<tr><td>Consequent</td><td>If it is fruit
Then it is an apple</td><td>If it is NOT fruit
Then it is NOT an apple
Valid</td></tr>
</table>

Appendix B.

Adapted Summary Table of Syllogistic Reasoning with the Observation and Conclusion Removed, used in Video-Explanation Condition in Experiment 1, and in all Conditions in Experiment 2

If this is an apple, then it is fruit		
	Affirm	**Deny**
Antecedent	Observation …………………….. Conclusion: ……………………..	Observation …………. NOT……….. Conclusion: …………. NOT………..
Consequent	Observation …………………….. Conclusion: ……………………..	Observation …………. NOT……….. Conclusion: …………. NOT………..

Chapter 7

Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them

This chapter has been published as:

Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heiltjes, A., & Van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology, 44*, 95-106. doi:10.1016/j.cedpsych.2016.02.005

Abstract

Two experiments investigated whether studying a text with an explanation intention and then actually explaining it to (fictitious) other students in writing, would yield the same benefits as previously found for explaining on video. Experiment 1 had participants first studying a text either with the intention to explain it to others or to complete a test, and subsequently restudying or explaining in writing. Neither study intention nor explaining affected learning outcomes. Experiment 2 directly compared explaining in writing and on video. Participants studied a text with a test intention followed by restudy, or study with an explanation intention followed by either explaining in writing or on video. Explaining on video, but not in writing, enhanced learning more than restudy. These findings suggest that the benefits of explaining on video are not a result of engaging in explanation per se. Results are discussed in light of feelings of social presence.

Introduction

It is well established that explaining is a powerful learning strategy (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Fiorella & Mayer, 2015a, 2015b; Leinhardt, 2001; Lombrozo, 2012; Ploetzner, Dillenbourg, Preier, & Traim, 1999; Richey & Nokes-Malach, 2015; Wylie & Chi, 2014). Most research on the effects of explaining has focused on explaining instructional materials to oneself (i.e., self-explaining) or explaining to others in interactive tutoring situations (Ploetzner et al., 1999; Richey & Nokes-Malach, 2015). Recent studies, however, have shown that providing explanations of learned material to fictitious other students (i.e., not present, no interaction) is also effective for learning, and even more so than restudying that material (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014b).

Hoogerheide, Loyens, and Van Gog (2014b) provided students with a text on syllogistic reasoning problems. Students who were instructed to study with the intention to explain the learning material to someone else and then explained it to a fictitious other student by creating a webcam video, showed higher learning and transfer performance on an immediate and delayed posttest compared to students who were instructed to study with the intention of performing well on a test and engaged in restudying the material, which is how students normally study. The cognitive schemas acquired by those who explained on video were also more efficient in the sense that higher test performance attained with equal (perceived) effort investment on the posttest (for elaboration on instructional efficiency in terms of the relation between mental effort and performance, see Van Gog & Paas, 2008). This pattern of results was found across two experiments. In the second experiment, students in the restudy condition engaged in a recall activity prior to restudy, in order to rule out the possibility that the positive effects of explaining on video were simply caused by retrieval practice (inherent to explaining), which has been shown to positively affect learning outcomes (Roediger, Putnam, & Smith, 2011).

Fiorella and Mayer (2013, 2014) obtained similar results in two studies on the effects of studying with the expectation of teaching later on (i.e., a teaching expectancy) and actually teaching by creating a short five-minute video lecture. Their participants studied a text about the Doppler effect. Across both studies, those students who expected to have to teach later on, showed enhanced performance on an immediate but not on a delayed comprehension test compared to those studying for a test. Only the students who had actually created a video lesson showed better comprehension scores than those studying for a test on both the immediate and delayed comprehension test. Fiorella and Mayer also explored effects on (perceived) effort investment during learning. They found some tentative indications that studying with a teaching expectancy is more effortful than studying with a test expectancy. However, findings were mixed, possibly because effort investment was measured at the end of the experiment rather than directly after the learning phase.

Roscoe and Chi (2008) contrasted explaining learning materials to a fictitious peer student on video (i.e., creating a video lesson) to self-explaining and peer tutoring. In a first session, university students studied a text about the human eye (1025 words) for 30 minutes. One week later, in a second session, they generated explanations for 30 minutes with the materials still being available (at least in the peer tutoring and self-explaining conditions). Although all three strategies were beneficial for learning, explaining on video was less effective relative to peer tutoring and self-explaining. It is unclear how these findings relate to Fiorella and Mayer (2013, 2014) and Hoogerheide et al. (2014b), however. Next to self-explaining and peer tutoring being stronger control conditions than restudy, Roscoe and Chi's study had a very different design (i.e., a delay between sessions, materials available during explaining, the time spent on explaining), and the actual time spent explaining in the three conditions was not reported and therefore may have differed among conditions.

Regardless of what exactly caused explaining on video to be less effective than self-explaining and peer tutoring, the positive effect found by Hoogerheide et al. (2014b) and Fiorella and Mayer (2013, 2014) beg the question of whether there is something specific to the video creation process that promotes learning, or whether it is simply the fact that students engage in explaining that causes beneficial effects on learning outcomes. In case of the latter, one would expect no unique benefit from explaining on video compared to explaining in writing. Instructions to provide written explanations for others would also be easier to implement in the classroom. Therefore, Experiment 1 replicated and extended the study by Hoogerheide et al. (2014b) and Fiorella and Mayer (2013, 2014) by having students explain in writing instead of on video. Experiment 2 made a direct comparison between explaining on video versus explaining in writing. Before introducing the experiments in more detail, we will first review relevant literature on the effects of study intention and teaching expectancy, as well as on the effects of giving explanations on learning outcomes.

Effects of Studying with the Intention to Explain

Studying learning materials with the intention of explaining them to others later on, also referred to as 'teaching expectancy', can be expected to foster effective study processes. For example, studying with an explanation intention may stimulate more active processing (Benware & Deci, 1984), comprehension monitoring (e.g., asking oneself "why" questions; Roscoe, 2014), self-explaining (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De Leeuw, Chiu, & LaVancher, 1994; Renkl, 1997, 2002), metacognitive processing (Muis, Psaradellis, Chevrier, Leo, & Lajoie, 2015), and generating deep questions and explanations (Craig, Gholson, Brittingham, Williams, & Shubeck, 2012; Craig, Sullins, Witherspoon, & Gholson, 2006).

Research on studying with a teaching expectancy has led to mixed findings, however. Some studies found positive effects on learning outcomes. For example, in Bargh and Schul (1980), the university students who studied a passage with a teaching expectancy outperformed those who studied with a test expectancy on a subsequent recall and recognition test. Similarly, Nestojko, Bui, Kornell, and Bjork (2014) recently showed that university students recalled more information from a text and recalled more efficiently if they had studied the text with a teaching expectancy. This benefit was also found, albeit less consistently, on the short answer test. Muis et al. (2015) even found that for primary school children, studying with a teaching expectancy fostered the use of metacognitive strategies and learning outcomes. Other studies did not find such positive effects on learning outcomes, however. For example, Renkl (1995) showed that studying learning materials with a teaching expectancy evoked university students to study less superficially than those who studied with a test expectancy, but this did not result in higher learning outcomes. Those who studied with a teaching expectancy even showed less intrinsic motivation and increased levels of anxiety. Higher anxiety was also found by Ross and DiVesta (1976). Finally, Ehly, Keith, and Bratton (1987) found a detrimental effect of teaching expectancy in the sense that high school students performed worse on a test if they studied with a teaching expectancy than if they studied for a test.

Several explanations have been offered for the mixed findings. Regarding immediate vs. delayed tests, Fiorella and Mayer (2013, 2014) suggested that the effect of studying a text with the intention of explaining it later on might be short-lived. On a delayed posttest, this effect would have diminished unless the expectancy had been coupled with actually explaining (on video). However, other studies did not even find beneficial effects of teaching expectancy on an immediate posttest (e.g., Ehly et al., 1987; Renkl, 1995). A potential explanation for the differences in findings with regard to immediate test performance could be that learners might need a certain level of experience with studying with an explanation expectancy before it becomes beneficial for learning. In the study by Hoogerheide et al. (2014b), no effects of an explanation intention were apparent for secondary education students. For university students in a problem-based learning curriculum, who are used to explaining to other students, the explanation intention did positively affect learning both on the immediate and the delayed posttests. Note however that Muis et al (2015) showed that even primary school children could benefit from studying with a teaching expectancy, and it would seem unlikely that they would have had a lot of experience explaining to each other.

Generating Explanations

Generating explanations can be a powerful method for improving learning outcomes (Dunlosky et al., 2013; Fiorella & Mayer, 2015a, 2015b; Leinhardt, 2001; Lombrozo, 2012; Ploetzner et al., 1999; Richey & Nokes-Malach, 2015; Wylie & Chi, 2014). As mentioned above, research on generating explanations has mainly focused on the effects of self-explanations and the effects of explaining to others in tutoring or collaborative learning contexts (Ploetzner et al., 1999; Richey & Nokes-Malach, 2015). As Richey and Nokes-Malach (2015) describe, research on self-explaining has shown that:

> '… encourage learners to identify and elaborate on the critical features of problems, including the underlying principles (Atkinson, Renkl, & Merrill, 2003; Chi & VanLehn, 1991), the conditions for applying those principles (Chi et al., 1989), and the logic and subgoals for applying them (Catrambone, 1998; Crowley & Siegler, 1999). These critical features tend to apply across problems within a domain. By recognizing and understanding these features, a learner is more likely to successfully transfer knowledge to a novel problem (Atkinson et al., 2003).' (p. 203)

These cognitive benefits may in part arise because the process of self-explaining may stimulate metacognitive processes such as monitoring the quality of one's own understanding (i.e., comprehension monitoring; Roscoe & Chi, 2007). However, a caveat to self-explaining is that students may not always generate high quality self-explanations on their own (e.g., Renkl, 1997), and therefore may need self-explanation prompts (e.g., Nokes, Hausmann, VanLehn, & Gershman, 2011) or even an explicit training (e.g., Kurby et al., 2012) before generating self-explanations effectively.

Explaining to others has also been shown to enhance learning outcomes in interactive situations, for instance when tutoring (Cohen, Kulik, & Kulik, 1982) or during small group discussions (Cohen, 1994; Johnson, Johnson, & Smith, 2007). Several studies analyzed the quality of the explanations to identify the benefits of different discourse moves, and these studies typically show that explaining is most effective when the explanations are relevant, coherent, complete, and accurate (Coleman, Brown, & Rivkin, 1997; King, 1994; Roscoe & Chi, 2007; Webb, 1989). Moreover, learners benefit more from generating explanations when they engage in so-called 'reflective knowledge building activities' such as generating inferences, repairing knowledge gaps, elaborating, and comprehension-monitoring. In contrast, learners benefit less when they predominantly engage in 'knowledge-telling', which entails summarizing with little elaboration or monitoring of one's own understanding (King, Staffieri, & Adelgais, 1998; Roscoe, 2014; Roscoe & Chi, 2007, 2008). Studies on tutoring or small group discussions typically do not experimentally control for explaining as a contributing factor (for an exception, see Roscoe & Chi, 2008). Therefore beneficial effects could also, at least partly, be attributed to the fact that in these situations, explanations are aimed at others who are present and can be interacted with. The other students may, for instance, ask questions, point out inconsistencies, or provide explanations themselves, which might contribute to the effectiveness of tutoring or small group learning (Okita & Schwarz, 2013; Ploetzner et al., 1999; Webb, 1989). Indeed, King et al.'s (1998) concept of transactive peer tutoring postulates that the benefits of peer tutoring are for a large part a result of a cognitive partnership in which learning partners and their actions continuously depend on the others' level of understanding and their responses.

When interactive elements are controlled for, a situation remains in which explanations are aimed at instructing someone who is present but merely listens. Ploetzner et al. (1999) suggest that aiming explanations at someone else who is physically present may stimulate more elaborate explanations and monitoring of whether the recipient comprehends the explanations, compared to self-explaining, which may lead to skipping. In line with this view, Coleman et al. (1997) found that engaging in explanations with the aim of instructing another person in the room fostered measures of deep learning more so than self-explaining. These findings suggest that 'social presence', even without interaction, may foster the effectiveness of explaining for learning.[1]

Social presence was originally defined by Short, Williams, and Christie (1976) as the degree to which a person is aware of the presence of another person in a technology-mediated communication or learning setting. The definition was updated more recently to the degree to which someone is perceived as a "real person" in computer-mediated communication or learning (Gunawardena, 1995). It is a key concept in understanding and improving the degree of participation and success in online learning environments (Borup, West, & Graham, 2013; Sung & Mayer, 2013). Placed on a continuum of social presence,

[1] Note that the concept of social presence is similar, but not identical to King et al.'s (1998) concept of transactive peer tutoring. Although both concepts focus on the effects of taking the 'fellow learner(s)' into account, King's theory limits itself to highly interactive situations that allow for continuous interaction.

engaging in explaining in interactive situations (e.g., tutoring, small group discussions) is on the high end and self-explaining is on the low end because the explanations are directed at oneself (i.e., the student's own understanding). Explaining to present but merely listening (i.e., non-interacting) others and explaining to non-present others fall in between these two on the social presence continuum. Explaining to non-present others may seem odd, but has become quite common in online learning environments nowadays. For instance, people provide explanations to others who may not be online at the same time (and whom they often do not even know) in asynchronous text-based discussion forums (Andresen, 2009) or in demonstration ('how-to') videos (e.g., Spires, Hervey, Morris, & Stelpflug, 2012). Such video demonstrations or lectures are often recorded behind a webcam, or using a digital camera on a tripod, without an audience present.

Being aware of a recipient/listener and perceiving them as real (even if they are not present) may result in "productive agency", that is, the belief that one's actions can affect others (Okita & Schwarz, 2013; Schwartz, 1999; Schwartz & Okita, 2004). Okita and Schwartz argued and showed that collaborative learning and teaching are in part so effective because they foster learners' awareness that their actions can affect the learning of others, which stimulates them to contribute (more) and to keep on doing so in the face of difficulties.

The findings by Fiorella and Mayer (2013, 2014) and Hoogerheide et al. (2014b) have shown that explaining to non-present, fictitious other students on video can be effective for learning. It is unclear, however, whether the same would apply to explaining to non-present, fictitious other students in writing. Explaining in writing would be much easier to implement in the classroom as a learning activity. Moreover, as addressed below, the process of explaining in writing is very different from explaining on video, and therefore may be more or less advantageous relative to reading. This question was addressed in the present study.

The Present Study

If it is the act of explaining itself that produces beneficial effects on learning, then explaining to fictitious other students in writing would be expected to be effective compared to restudy, just like explaining on video was found to be (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014b). Indeed, with regard to engaging in *recall*, research conducted in the context of the testing effect (i.e., the finding that engaging in recall after an initial study phase is more effective than restudying; Roediger et al., 2011), has demonstrated that engaging in both oral and written recall of paired associates is effective for learning (Putnam & Roediger, 2013). Although recalling information and explaining it are different processes, prior research has proven written explanations during learning (e.g., Hilbert, Schworm, & Renkl, 2004; Schworm & Renkl, 2006) and during problem-solving (e.g., Aleven & Koedinger, 2002) to be effective compared to not explaining. The effectiveness of explaining in writing is perhaps to be expected, as, compared to reading, writing activities "can support more sophisticated elaboration and organizational strategies by linking new understandings with familiar ones, synthesizing knowledge, exploring relations and implications, and building outlines and conceptual frameworks' (Bangert-Drowns, Hurley, & Wilkinson, 2004, p. 32)." Moreover, writing can stimulate various metacognitive strategies, such as deliberate planning and monitoring the quality of the writing (Paris & Paris, 2001; Schraw, 1998).

On the other hand, because the act of writing is very different from the act of speaking in a camera, the form in which explanations are given may matter. Speaking into a camera may be higher in perceived social presence than writing. That is, the presence of a camera may give students a stronger feeling that they are communicating information to an actual other person (even though that person is not present at the moment) than writing does. Moreover, speaking allows for a high number of idea units to be expressed in a short amount of time, which is not the case for writing (Grabowski, 2007; Kellogg, 2007). But writing, in contrast to speaking, involves more deliberate planning and may therefore entice learners to think more about what is most important (i.e., key ideas/concepts/procedures) to explain to others. Indeed, it seems that writing results in less irrelevant or distorted idea units being (re)produced than speaking does (Horowitz & Newman, 1964; Kellogg, 2007). Moreover, explaining in writing may better enable learners to monitor whether the information is accurately presented, whereas speaking may impede output monitoring (Grabowski, 2007).

In sum, it is unclear whether explaining in writing would be more, less, or equally effective as explaining on video compared to a restudy control condition. We hypothesize that explaining on video would be more effective than restudying, while it is an open question whether explaining in writing would be more beneficial than restudying and whether there would be differences between explaining in writing and explaining on video. The present study addressed these questions in two experiments. Experiment 1 investigated the effects of explaining in writing compared to restudying and Experiment 2 compared explaining in writing to explaining on video and restudying.

Experiment 1

Experiment 1 aimed to replicate and extend the findings by Hoogerheide et al. (2014b), using the same materials and conditions (plus an additional control condition), but having students explain in writing rather than on video. In a 2 x 2 design participants studied a text on syllogistic reasoning (of which the content was new to them) with either a test or explanation study intention. Subsequently they either restudied the materials or explained them in writing to other students. Note that students were not told beforehand whether they would restudy or produce written or video explanations, and time on task was kept equal across conditions.

Given the mixed findings on the effects of studying with an explanation study intention (Hoogerheide et al., 2014b) or teaching expectancy (e.g., Bargh & Schul, 1980; Fiorella & Mayer, 2014; Nestojko et al., 2014; Renkl, 1995), we cannot formulate a directional hypothesis regarding the effects of study intention. Explaining is hypothesized to have beneficial effects on learning and transfer compared to restudying, at least when it was preceded by an explanation study intention (cf. the study by Hoogerheide et al. (2014b), in which explaining was always preceded by an explanation study intention).

We also analyzed perceived mental effort invested in the learning phase and in answering questions on the test. Such data, in combination with test performance measures, provide more insight in the learning process and the quality of learning outcomes, respectively (Van Gog & Paas, 2008). Hoogerheide et al. (2014b) investigated only effort invested during the test; they did not explore whether providing explanations is more effortful than restudying. Fiorella and Mayer (2013, 2014) only investigated effort invested in the learning phase, but found a mixed pattern of results, likely because the effort investment measurement was not presented directly after learning, but instead at the end of the experiment. We hypothesize that providing explanations might be more effortful than restudying in the learning phase (cf. germane cognitive load, Paas, Renkl, & Sweller, 2003; Paas & Van Gog, 2006; or desirable difficulties, Bjork & Bjork, 2011), but that this will also lead to higher learning outcomes, evidenced by higher test performance attained with equal or less effort investment on the test.

Method

Participants. Participants were 123 higher education students (81 female; M = 20.05, SD = 1.90), enrolled in the first year of a communication and media design (n = 37) or primary school teacher training (n = 86) program of a Dutch university of applied sciences.

Design. The experiment consisted of five phases: 1) pretest, 2) learning phase I, 3) learning phase II, 4) immediate posttest, and 5) delayed posttest. The experiment had a 2 X 2 design, with Study Intention (Test vs. Explanation; manipulated in learning phase I) and Explaining (No: Restudy vs. Yes: Writing Explanations; manipulated in learning phase II) as between-subject factors. Students were randomly assigned to one of the four conditions:

test study intention – restudy (n = 29), test study intention – explain in writing (n = 33), explanation study intention – restudy (n = 30), or explanation study intention – explain in writing (n = 31).

Materials

All the study and test materials were paper-based.

Pretest. The pretest presented eight syllogistic reasoning items and two Wason-selection task items to assess prior knowledge. The syllogistic reasoning items asked participants to assess whether the conclusion that followed from the two premises was logical (i.e., choose one of two answer options: valid or invalid). Two test items were used for each of the four forms of syllogistic reasoning, namely: affirming the antecedent (if *P* then *Q, P* therefore *Q)*, denying the antecedent (if *P* then *Q,* not *P* therefore not *Q)*, affirming the consequent (if *P* then *Q, Q* therefore *P)*, and denying the consequent (if *P* then *Q,* not *Q* therefore not *P)*. One of those items was prone to belief bias, whereas the other was not. Belief bias makes it more difficult to assess whether a conclusion is logically valid because the conclusion is in line with real world knowledge (George, 1995; Newstead, Pollard, Evans, & Allen, 1992).

Wason-selection tasks (Wason, 1966) require combining the two valid forms of syllogistic reasoning (i.e., affirming the antecedent and denying the consequent) to correctly test the validity of a rule. For example, when asked “If there is an A on one side, and a 2 on the other side” by turning two cards out of the four possibilities A, E, 1, and 2, then the correct answer would be to turn A (affirming the antecedent) and 1 (denying the consequent). People are, however, inclined to turn A and 2 instead. Thus, the pretest required participants to select the two correct forms of syllogistic reasoning out of four answer options.

Learning Phase I. In the first learning phase, participants received a 1930 words text (the same as used in Hoogerheide et al., 2014b). Participants studied the text for 12 minutes, which was equal to Hoogerheide et al. (2014b) in which this was based on a pilot study. This text addressed when a conclusion logically follows from two premises. After a general introduction, all four forms of syllogistic reasoning were explained using the same recurrent example: “If John sees a clown, then he is afraid. John sees a clown. Conclusion: John is afraid”. The last page presented a summary table of all four forms of syllogistic reasoning using the example “If this is an apple, then it is a fruit”. It was also indicated whether each form led to a valid or invalid conclusion.

Two versions of the study text were used in the present experiment, one for those who studied with a test study intention, and one for those who studied with an explanation study intention. These only differed in the study intention prompt placed on the first page and in the footer of each page: “Can you apply the information from this page to complete a test?” or “Can you explain the information on this page to a fellow student?”

Learning Phase II. The second learning phase had a duration of 8 minutes, which is 3 minutes longer than Hoogerheide et al. (2014b) to provide the explanation conditions with sufficient time. Half of all the students restudied the same text as in learning phase I for 8 minutes. The first page of this booklet, however, differed as it instructed participants to engage in a cued recall activity prior to restudying the text to ensure that all conditions engaged in retrieval practice (i.e., to rule out the possibility that beneficial effects of explaining are simply due to retrieving information from long-term memory, which is inherent in explaining). Using the table and the example that was on the last page of the study text (i.e., “If this is an apple, then it is a fruit”), but without the indications of which forms were valid, participants were asked to fill in the gaps in the table from memory. For example, for affirming the antecedent, the correct answer was: “it is an apple, therefore it is a fruit” and for denying the antecedent, the correct answer was: “it is *not* an apple, therefore it is *not* a fruit”, etcetera. The other half of the students engaged in the explanation activity by explaining what they had learned in writing as if explaining to a complete novice on the subject, with the help of the same table and example on the last page of the study text. In addition, participants in the explanation group were instructed to explain the error commonly made when judging whether a conclusion is valid (i.e., the belief-bias, although belief-bias was not explicitly mentioned).

Posttests. To assess learning, the immediate and delayed posttests presented eight conditional syllogistic reasoning items (one with and one without the belief-bias for each form) and two Wason selection tasks to assess transfer. An example of a conditional syllogistic reasoning test item (affirming the antecedent with belief-bias) is: If you are a Pokémon, then you belong in a pokeball. Pikachu belongs in a pokeball. Conclusion: Pikachu is a Pokémon. An example of a Wason selection task is: Which two cards would you have to turn to test the rule ‘If there is a Y on one side, then there is a 2 on the other side?’, with answer options X, Y, 2, and 7. Two parallel versions of the posttest (A and B) were created. These versions were structurally equivalent compared to each other and to the pretest, but different on surface features. On both posttests, participants were not only asked to select the correct answer, but also to explain their answer, making the posttests substantially more difficult.

Mental Effort. Mental effort was measured after each test item on the pretest and posttests and after the second learning phase, using a subjective 9-point rating scale (Paas, 1992), asking students to rate how much effort they invested in the preceding task, with answer options ranging from (1) very, very low effort to (9) very, very high effort.[2]

Procedure

The study was run in small groups with approximately 15 students per session, at a university of applied sciences. Within each session students were randomly assigned to one of the four conditions. The first session lasted 60 minutes. Every student received an envelope with four booklets and then received a general introduction. After the introduction, students were instructed to take out the first booklet containing the pretest, and to complete it. After each pretest item participants rated how much mental effort they invested in that item. Participants had 10 minutes to complete the pretest. When time was up, they were instructed to place the first booklet upside down at the corner of their table and to take out the second booklet containing the study text, for which they received 12 minutes. All participants were encouraged to fully use the available time and to learn as much as possible. After the experimenter indicated that the 12 minutes were up, participants placed the second booklet on the corner of their table. Then, they were instructed to take the third booklet out of their envelope and to follow the written instructions. Participants in the restudy conditions were instructed to fill in the gaps in the table from memory, after which they would turn the page and restudy the same text as in the second booklet. Again, participants were encouraged to use all available time and to learn as much as possible. The writing conditions were instructed to explain what they had just learned (instructions were provided in the booklet, see above). Participants had 8 minutes in total for the third booklet, in all conditions. When time was up, participants first indicated perceived effort investment in restudying or giving explanations, then returned booklet 3 to the corner of their desk and worked on the fourth booklet, which contained the immediate posttest. Half of the participants in each condition received version A as the immediate posttest while the other half received version B. Participants again rated perceived effort investment after every test item. Maximally 25 minutes were available for the immediate posttest. The delayed posttest was to take place one week later, at which participants who received version A as immediate test would now receive

[2] Perceived confidence was also measured (after the effort measures) because the second author, who conducted this study as part of the qualifications for her MSc degree, was interested in exploring that variable. However, because we had no hypothesis about it and did not measure it in the second experiment, those (null) results are not reported here. Details can be obtained from the first author.

version B and vice versa. Unfortunately, however, the delayed test session attendance was very low (n = 52) due to a scheduling error that was not under our control. Consequently, the analysis of the delayed posttest data would not be very useful and is not reported.

Data Analysis

Scoring was done using the same coding scheme as was used in the Hoogerheide et al. (2014b) study. The pretest was scored by assigning one point per correctly answered question, resulting in a maximum score of 10 points. The maximum score on the syllogistic reasoning items (i.e., the items that measured learning) on the immediate posttest was 56 points. Each belief bias item (four in total) could result in a maximum of eight points. One point could be earned for the correct choice on the multiple-choice question and seven points for the explanation. These seven points were comprised of: correctly recalling the form of syllogistic reasoning (one point), explaining correctly in abstract terms of p and q (one point), explaining correctly in concrete terms (two points), correctly concluding in the explanation whether a conclusion was valid or invalid (one point), and correctly explaining the belief-bias (two points). Each no belief-bias item (four in total) could result in a maximum of 6 points (scoring as on the other items without the two points for explaining the belief-bias). A total of 18 points could be earned on the Wason selection tasks, maximally 9 points per correctly answered item. These 9 points were comprised of one point for selecting the correct answer and two points per correct explanation for each of the four forms of syllogistic reasoning as applied to the rule in the Wason selection task. Two raters scored 10% of the tests. Because the inter-rater reliability was high (intra-class correlation coefficient of .90), the remainder of the tests was scored by one rater.

Average perceived mental effort investment was computed separately for the syllogistic reasoning items (i.e., learning performance) and the Wason selection tasks (i.e., transfer performance). One participant in the 'explanation intention – explaining' condition was removed from all analyses because of non-compliance with instructions on the immediate posttest. One other participant had a missing mental effort rating on the pretest, which was replaced with the series mean.

Results and Discussion

Performance data are presented in Table 1, and perceived mental effort data are presented in Table 2. At pretest, there were no differences among conditions, as one would expect after random assignment. An ANOVA showed no significant differences among conditions in pretest performance, F < 1, or perceived mental effort investment, $F(3, 118) = 1.51$, $p = .215$, $\eta_p^2 = .037$. The posttest data were analysed by 2 x 2 ANOVAs with the exception of the learning and transfer results, which were analysed by 2 x 2

ANCOVAs with students' pretest scores as a covariate. The nature of significant interactions was determined with follow-up Bonferroni-corrected t-tests.

As for the items that measured learning, students' pretest scores were a significant predictor, $F(1, 117) = 9.33$, $p = .003$, $\eta_p^2 = .074$, but there were no main effects of Study Intention or Explaining, nor an interaction effect (all *F*s < 1). In a similar vein, students' pretest scores were a significant predictor of performance on the items that measured transfer, $F(1, 117) = 10.73$, $p = .001$, $\eta_p^2 = .084$. There were no main or interaction effects (Study Intention and Explaining: both *F*s < 1; interaction: $F(1, 117) = 3.69$, $p = .057$, $\eta_p^2 = .031$).

The analysis of perceived mental effort investment in the second learning phase (booklet 3, restudying or writing explanations), showed a main effect of Explaining, $F(1, 118) = 6.29$, $p = .013$, $\eta_p^2 = .051$. This indicates that participants who gave explanations reported to have invested significantly more mental effort in the learning phase than those who restudied. There was no main effect of Study Intention, $F < 1$, nor an interaction effect, $F < 1$.

Table 1.
Mean (SD) of Learning and Transfer Test Scores per Condition in Experiment 1

	Test Intention – Restudy	Test Intention – Explain in writing	Explanation Intention – Restudy	Explanation Intention – Explain in writing
Pretest – Learning *(0-8)*	4.90 (1.45)	5.06 (1.14)	5.33 (2.02)	5.40 (1.48)
Pretest – Transfer *(0-2)*	0.66 (0.61)	0.76 (0.61)	0.80 (0.66)	0.87 (0.57)
Pretest – Total *(0-10)*	5.55 (1.43)	5.82 (1.36)	6.13 (1.53)	6.27 (1.41)
Immediate Posttest – Learning *(0-56)*	16.76 (9.18)	15.74 (6.69)	15.98 (7.58)	16.28 (7.18)
Immediate Posttest – Transfer (*0-18)*	2.38 (2.81)	1.44 (2.30)	2.10 (3.32)	2.97 (2.77)

Note. Whereas the pretest consisted of multiple choice items only, the posttest asked students not only to select the correct answer, but also to explain their answer, making the posttest substantially more difficult than the pretest.

On perceived mental effort investment in the posttest items that measured learning, there was no main effect of Study Intention, $F(1, 118) = 1.93$, $p = .167$, $\eta_p^2 = .016$, or Explaining, $F < 1$, nor a significant interaction effect, $F(1, 118) = 3.75$, $p = .055$, $\eta_p^2 = .031$. On perceived mental effort investment in the Wason selection tasks that measured transfer performance, no main effects of Study Intention or Explaining were found (both *F*s < 1). There was a significant interaction effect, $F(1, 118) = 5.74$, $p = .018$, $\eta_p^2 = .046$. To explore this interaction effect, two independent samples *t*-tests with Bonferroni-adjusted alpha levels of .025 were conducted that investigated effects of the test intention and explanation intention conditions separately. However, the test intention – restudy condition did not differ significantly (given the alpha-adjustment) from the test intention – explanation condition, $t(60) = 2.17$, $p = .034$, nor did the explanation intention – restudy differ significantly from the explanation intention – explain condition, $t(58) = 1.17$, $p = .247$.

Table 2.
Mean (SD) of Mental Effort (range 1-9) per Condition in Experiment 1

	Test Intention – Restudy	Test Intention – Explain in writing	Explanation Intention – Restudy	Explanation Intention – Explain in writing
Learning Phase 2	5.21 (2.27)	6.09 (2.08)	4.70 (2.53)	5.93 (2.42)
Pretest – Learning	3.34 (1.32)	2.91 (1.18)	2.55 (0.91)	3.28 (1.38)
Pretest – Transfer	4.21 (1.99)	4.02 (1.63)	4.28 (1.73)	3.57 (1.42)
Pretest – Total	3.51 (1.24)	3.13 (1.14)	2.90 (0.93)	3.34 (1.33)
Immediate Posttest – Learning	3.90 (2.06)	3.28 (1.44)	2.93 (1.27)	3.43 (1.55)
Immediate Posttest – Transfer	4.31 (2.19)	3.23 (1.73)	3.32 (1.46)	3.83 (1.93)

In sum, the results of Experiment 1 showed no benefit of studying a text with an explanation intention compared to a test-taking intention on learning outcomes. Interestingly, we found no evidence that actually providing explanations would be more effective than restudying. Explaining in the learning phase was perceived to be more effortful than restudying, but this additional effort investment did not seem to pay off, as it did not result in higher learning outcomes. Note however that it is possible that this additional effort investment would have been beneficial for learning or transfer measured on a delayed test. Students who explained after studying with an explanation intention did reach the highest transfer test score numerically (see Table 1), but the interaction effect was not statistically significant ($p = .057$). This may suggest that the effort invested in explaining positively affected students' deep comprehension of the material, and it is very well possible that the effects of deep comprehension would only show after a delay (cf. Fiorella & Mayer, 2013, 2014). This makes it even more unfortunate that we were unable to obtain delayed test data in Experiment 1 from a sufficiently large number of students. Alternatively, the beneficial effects of explaining in writing might just be small (when preceded by studying with an explanation intention), which would be similar to the finding that writing-to-learn assignments such as writing summaries or essays typically only yield small benefits (Bangert-Drowns et al., 2004).

Another potential factor contributing to the lack of effect, might have been the classroom setting in Experiment 1, which may have made it more difficult for students to concentrate than the individual study and test conditions in the study by Hoogerheide et al. (2014b). Therefore, a second Experiment was conducted in which we (a) made a direct comparison of explaining in writing and explaining on video to a restudy control condition, (b) did include a delayed test, and (c) tested students individually.

Experiment 2

In Experiment 2, students either studied a text with a test study intention and then engaged in a short recall activity (i.e., filling in the gaps in the Table) prior to restudying (Test Condition), or studied with an explanation study intention followed by explaining in writing (Explanation-Writing Condition) or followed by explaining on video (Explanation-Video Condition). Based on findings that explaining in front of a camera is more effective than restudy (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014b), we hypothesize that explaining on video is more effective than restudy on both an immediate and delayed test. Based on the findings from Experiment 1, we expect no differences between the explaining in writing condition and the Restudy Condition on the immediate test. However, it is possible that the beneficial effects of explaining in writing compared to restudy would show on a delayed test. Whether explaining orally on video would be more effective than explaining in writing, is an open question. We again measured (perceived) mental effort investment in Experiment 2 to investigate the efficiency of engaging in explaining. Additionally, because Hoogerheide et al. (2014b) found some tentative indications that explaining on video, although effective for learning and transfer, seemed to *reduce* students' perceived competence compared to restudy, this variable was also explored in Experiment 2. Perceived competence is an important variable to take into account because students' perceptions of their own competence are positively related to factors such as academic motivation and learning outcomes (Bong & Skaalvik, 2003; Law, Elliot & Murayama, 2012; Harter, 1990; Ma & Kishor, 1997). Whereas studying learning materials may foster the development of students' perceived competence (Hoogerheide, Loyens, & Van Gog, 2014a, 2016), explaining may not be as beneficial because it can confront learners with knowledge gaps, that is, what they do not know (Roscoe & Chi, 2007).[3]

[3] Note that the construct of perceived competence is similar, but not identical, to the construct of self-efficacy (Hughes, Galbraith, & White, 2011; Rodgers, Markland, Selzler, Murray, & Wilson, 2014). Although both reflect perceptions of one's own abilities, perceived competence focuses on the need to master personally challenging tasks and self-efficacy reflects more situation specific self-confidence.

Participants

Participants were 129 Dutch undergraduate students (M^{age} = 20.20, *SD* = 3.04; 99 female) who studied Psychology in a Problem-Based Learning curriculum. Participants received a monetary reward or course credits for their participation.

Design

Like Experiment 1, Experiment 2 also consisted of five phases: 1) pretest, 2) learning phase I, 3) learning phase II, 4) immediate posttest, and 5) delayed posttest. Participants were randomly allocated to one of three conditions, namely the Test Condition (i.e., test study intention – restudy; *n* = 42), the Explanation-Writing Condition (i.e., explanation study intention – explain in writing; *n* = 43), or the Explanation-Video Condition (i.e., explain study intention – explain on video; *n* = 44).

Materials and Procedure

The materials and procedure in Experiment 2 were almost identical to Experiment 1 with a few exceptions. First, some additional measures were added. Measures of perceived competence were added at the end of the pretest and start of the posttests (cf. Hoogerheide et al., 2014b), using an adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996). After the pretest, students were asked to indicate on a scale of 1 (not at all true) to 7 (very true): “I feel confident in my ability to learn an in-depth explanation of the eight items”, “I am capable of learning an in-depth explanation of the eight items”, and “I feel able to meet the challenge of performing well in learning an in-depth explanation of the eight items”. Prior to the posttests they were asked to indicate: “I feel confident in my ability to answer questions on a test”, “I am capable of answering questions on a test”, and “I feel able to meet the challenge of performing well answering questions on a test”. Moreover, we also asked participants to indicate perceived invested mental effort invested at the end of learning phase I (i.e., booklet 2) to explore whether study intention would already affect effort investment in the learning phase, which could not be inferred from the data from Experiment 1. Finally, because of the differences in writing and video creation, we asked participants to indicate to which degree they felt that they had enough time to explain the four forms of syllogistic reasoning (Explanation-Writing and Explanation-Video Condition) or to fill in the table on the first page and read the text (Test Condition) on a scale from 1 (to a very small degree) to 9 (to a very large degree).

A second difference with Experiment 1 was that participants in Experiment 2 were all seated in individual cubicles (as in the Hoogerheide et al., 2014b study). Third, in Experiment 1, where they had to write by hand, students had 8 minutes for explaining, whereas in the study by Hoogerheide et al. (2014b) students had only 5 minutes to create a video. In Experiment 2, we allowed students to type their explanations on the computer, which is faster than handwriting not only when copying information but also when writing from memory (even for “two-finger typists”; Brown, 1988). Nevertheless, we gave them some extra time compared to the Hoogerheide et al. (2014b) study: all three conditions received six minutes to either restudy or generate written or video explanations during learning phase II. Fourth, following Hoogerheide et al. (2014b, Experiment 2), the Explanation Conditions were no longer explicitly instructed to explain the common errors that people tend to make when judging whether a conclusion is valid or invalid (i.e., the belief-bias). This ensures that an increased focus on this bias would not be the cause of the expected benefits of providing explanations.

Data Analysis

Data were scored in the same manner as in Experiment 1. Four participants had to be removed from all analyses: One participant indicated high familiarity with the learning material from partaking in another experiment (Explanation-Video Condition), another failed to make a video (Explanation-Video Condition), the third did not follow instructions provided by the experimenter (Explanation-Writing Condition), and the last one received instructional materials from two conditions due to an experimenter error (Test Condition).

A further eight participants (two from the Test and Explanation-Video Condition and four from the Explanation-Writing Condition) who did not return for the Delayed Posttest were excluded from the analyses of learning, transfer, mental effort, and perceived competence on the posttests. One participant who did not fill in the mental effort rating after the first learning phase was removed from this analysis. In case of maximally two missing mental effort ratings on the tests, these were replaced with the series mean (two instances on the pretest; nine on the immediate posttest).

Results and Discussion

The learning and transfer scores can be found in Table 3, and the perceived mental effort investment and perceived competence scores are shown in Table 4. An ANOVA showed no significant differences among conditions in pretest performance or mental effort ratings, both *F*s < 1, nor in perceived competence, $F(2, 122) = 1.05$, $p = .353$, $\eta_p^2 = .017$.

With regard to learning, a repeated measures ANCOVA with Test Moment (Immediate vs. Delayed) as within-subjects factor, condition as between-subjects factor, and pretest scores as covariate, showed that students’ pretest scores were not a significant predictor of learning, $F < 1$. There was no main effect of Test Moment, $F < 1$, but there was a main effect of Instruction Condition, $F(2, 113) = 3.71$, $p = .027$, $\eta_p^2 = .062$. Bonferroni-corrected post-hoc tests showed that the Explanation-Writing Condition (M = 23.50; SD = 5.37) did not outperform the Test Condition (M = 22.03; SD = 5.39), $p = .709$, d = 0.193, but the Explanation-Video Condition did (M = 25.33; SD = 5.37), $p = .023$, d = 0.434. No significant difference was found between the Explanation-Writing and Explanation-Video Condition, $p = .405$, d = 0.241. There were no interaction effects (Test Moment * Pretest scores: $F < 1$; Test Moment * Instruction Condition: $F(2, 113) = 1.68$, p = .191, η_p^2 = .029). With regard to transfer, students' pretest scores were a significant predictor, $F(1, 113) = 9.55$, $p = .003$, $\eta_p^2 = .078$. There was no main effect of Test Moment or Instruction Condition, nor interaction effects, Fs < 1.

As for perceived mental effort invested in the learning phase, an ANOVA showed no significant effect of Instruction Condition, $F < 1$, on mental effort ratings in the first learning phase (booklet 2, test study intention or explanation study intention). There was a significant effect of Instruction Condition on the perceived mental effort investment in the second learning phase (booklet 3; restudying or writing explanations), $F(2, 122)$ = 21.48, $p < .001$, $\eta_p^2 = .260$. Students in the Test Condition (M = 2.90; SD = 1.96) reported having invested significantly less effort in this phase than both the Explanation-Writing Condition (M = 4.64; SD = 1.76), $p < .001$, d = 0.933, and the Explanation-Video Condition (M = 5.71; SD = 2.17), $p < .001$, d = 1.960. Furthermore, the Explanation-Video Condition reported having invested more effort in this phase than the Explanation-Writing Condition, $p = .042$, d = 0.542.

Perceived mental effort invested in the test was analysed with a repeated measures ANOVA with Test Moment (Immediate vs. Delayed) as within-subjects factor and condition as between-subjects factor. On perceived mental effort invested in solving the items measuring learning, a main effect of Test Moment was found, $F(1, 114) = 13.89$, $p < .001$, η_p^2 = .109. This indicated that participants, on average, reported to have invested less mental effort on the Delayed Posttest (M = 2.65; SD = 1.27) than on the Immediate Posttest (M = 3.02; SD = 1.42). There was no main effect of Instruction Condition, $F < 1$, nor a significant interaction, $F < 1$. On the items measuring transfer performance, invested mental effort ratings showed a similar pattern. That is, there was a main effect of Test Moment, $F(1, 114) = 55.37$, $p < .001$, $\eta_p^2 = .327$, with participants reporting less mental effort investment on the Delayed Posttest (M = 3.63; SD = 2.13) than on the Immediate Posttest (M = 4.82; SD = 2.20), but no main effect of Instruction Condition, F <1, nor a significant interaction effect, $F(2, 114) = 2.20$, $p = .116$, $\eta_p^2 = .037$.

Table 3.
Mean (SD) of Learning and Transfer Test Scores per Condition in Experiment 2

	Test Intention – Restudy *(Test Condition)*	Explanation Intention – Explain in writing *(Explanation-Writing Condition)*	Explanation Intention – Explain on video *(Explanation-Video Condition)*
Pretest – Learning *(0-8)*	5.68 (1.37)	6.00 (1.36)	5.95 (1.29)
Immediate Posttest – Learning *(0-56)*	22.49 (5.90)	24.86 (6.27)	26.01 (5.32)
Delayed Posttest – Learning *(0-56)*	21.53 (4.89)	22.17 (5.95)	24.66 (6.12)
Immediate Posttest – Transfer *(0-18)*	5.51 (3.45)	6.53 (3.08)	5.99 (3.61)
Delayed Posttest – Transfer *(0-18)*	6.22 (3.17)	6.84 (3.35)	6.15 (3.57)

Note. Whereas the pretest consisted of multiple choice items only, the posttests not only asked students to select the correct answer, but also to explain their answer, making the posttests substantially more difficult than the pretest.

As for perceived competence, an ANOVA on students' confidence in being able to learn the content of the materials before example study showed no significant effect of Instruction Condition, $F < 1$. As for students confidence in answering questions on a test, a repeated measures ANOVA showed a main effect of Test Moment, $F(1, 114) = 8.46$, $p = .004$, $\eta_p^2 = .069$, indicating higher perceived competence on the Immediate Posttest ($M = 5.88$; $SD = 0.86$) than on the Delayed Posttest ($M = 5.66$; $SD = 0.88$). There was no main effect of Instruction Condition, nor a significant interaction effect (both Fs < 1). So although the study by Hoogerheide et al. (2014b) seemed to indicate that explaining on video might reduce students' confidence in their own capabilities, no such indications were found here.

We also measured to what degree participants felt that they had enough time for the second learning phase (booklet 3; restudying or writing explanations). An ANOVA showed a main effect of Instruction Condition, $F(2, 122) = 69.07$, $p < .001$, $\eta_p^2 = .531$. There was no difference between the Test Condition ($M = 6.66$, $SD = 2.03$) and Explanation-Video Condition ($M = 7.10$, $SD = 1.74$), $p = .826$, $d = 0.232$, and the means suggest these students felt they had sufficient time. However, the Explanation-Writing Condition ($M = 2.86$, $SD = 1.66$) reported much lower scores than the Test Condition, $p < .001$, $d = 2.049$, and the Explanation-Video Condition, $p < .001$, $d = 2.493$. This indicates that students in this condition would have preferred to have more time for explaining. Although this might potentially explain why there was no benefit of the writing condition over the restudy condition, it seems that a more likely explanation lies in the differences between writing and video creation, as we will discuss below.

Table 4.
Mean (SD) of Mental Effort (range 1-9) and Perceived Competence (range 1-7) per Condition in Experiment 2

	Test Intention – Restudy *(Test Condition)*	Explanation Intention – Explain in writing *(Explanation-Writing Condition)*	Explanation Intention – Explain on video *(Explanation-Video Condition)*
Mental Effort			
Learning Phase 1	3.78 (1.33)	4.07 (1.76)	4.24 (1.80)
Learning Phase 2	2.90 (1.96)	4.64 (1.76)	5.71 (2.17)
Pretest	2.96 (1.17)	3.26 (1.19)	3.12 (1.04)
Learning Immediate Posttest	3.10 (1.50)	2.89 (1.23)	3.05 (1.54)
Learning Delayed Posttest	2.74 (1.31)	2.68 (1.23)	2.53 (1.29)
Transfer Immediate Posttest	4.81 (1.98)	4.66 (2.27)	4.96 (2.36)
Transfer Delayed Posttest	3.65 (2.06)	3.88 (2.29)	3.38 (2.05)
Perceived Competence			
Pretest	6.00 (0.95)	5.90 (1.03)	5.87 (0.99)
Immediate Posttest	5.86 (0.80)	5.78 (1.02)	6.00 (0.74)
Delayed Posttest	5.60 (0.96)	5.68 (0.85)	5.70 (0.84)

In sum, Experiment 2 replicated the findings from Experiment 1, showing that providing written explanations was a more effortful activity that did not contribute to learning outcomes compared to restudy. Surprisingly, Experiment 2 failed to replicate prior findings that explaining by making a video with a webcam would have a significant beneficial effect on transfer performance compared to restudy (Hoogerheide et al., 2014b). We did replicate prior findings that explaining on video was more beneficial for learning than restudy (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014b)–though it was not significantly better than explaining in writing. A potential explanation for the fact that explaining on video is more effective for learning than restudying while explaining in writing is not, is that explaining in front of a webcam may enhance feelings of social presence. That is, it may give students a stronger feeling that they are communicating information to an actual other person (even though that person is not present at the moment) than writing does. Increased feelings of social presence could be beneficial for learning. Students may, for instance, monitor whether the (imagined) audience will be able to understand the explanation, which would provide a good indicator of how well she understands it.

If this explanation holds true, then we should find more indications of audience-directed utterances (e.g., 'you') in the video explanations than in the written explanations. We explored this by counting the number of times participants used the self-other referential words 'me', 'you', 'us', 'we', 'your', and 'yourself' in their explanation, dividing this by the total number of words they used in their explanation and multiplying the result by 100 to get a percentage score. Counting such pronouns is a common method for assessing social presence in asynchronous computer-based communication and teacher-student interaction as they connote feelings of closeness and association (Rourke, Anderson, Garrison, & Archer, 1999; Sanders & Wiseman, 1990). Data from three participants in the Explanation-Video Condition were unavailable for this analysis as a result of a malfunction in the audio recording software. An independent samples *t*-test showed that the video explanations indeed contained a significantly higher percentage of those self-other referential words (M = 5.42%, SD = 1.73%) than written explanations (M = 2.16%, SD = 2.44%), $t(80) = 6.98$, $p < .001$, $d = 1.537$.

General Discussion

This study investigated whether studying a text with the intention to explain learned material to someone else would be more effective than studying to complete a test, and whether explaining to fictitious others in writing would yield the same benefits as explaining to fictitious others on video (Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014b). Regarding study intention, we did not find any indications in Experiment 1 that studying a text with an explanation intention would be more effective or efficient than studying with a test intention on an immediate test. Note that prior research has also found mixed results regarding the effectiveness of an explanation study intention that is not followed by actually providing explanations (Hoogerheide et al., 2014b; Bargh & Schul, 1980; Fiorella & Mayer, 2014; Nestojko et al., 2014; Renkl, 1995). All in all, there seems to be little evidence that studying with the intention of explaining the material to others helps learning unless it is actually followed by explaining –but not just any kind of explaining, as our results show.

Our experiments provided no evidence that explaining to a non-present fictitious other student in writing would be more beneficial for learning outcomes than restudy. Explaining in writing was actually less efficient for learning, in the sense that it required more effort than restudy while this additional effort investment did not pay off in terms of improved learning. One could argue in Experiment 1 that this additional effort was probably invested in more elaboration which would lead to deeper learning, the benefits of which might show only after a delay (Fiorella & Mayer, 2014). Experiment 2 did include a delayed test, yet still found explaining in writing to be less efficient than restudy.

Explaining to a non-present fictitious other student on video was also more effortful than restudy and even than explaining in writing, but this additional effort relative to restudy did pay off. That is, it resulted in better learning with a medium effect size (i.e., this effort was invested in processes that were germane to, or effective, for learning, e.g., Paas et al., 2003; Paas & Van Gog, 2006; see also research on desirable difficulties, e.g., Bjork & Bjork, 2011). The question is, then, why is explaining on video more effective than restudy while explaining in writing is not? We hypothesized that this might be due to increased feelings of social presence when explaining in front of a camera compared to producing written explanations. In terms of the social presence definition provided by Gunawardena (1995), producing video explanations might make the potential recipients feel “more real” (this has also been referred to as ‘immediacy’: Andersen, 1979; Wiener & Mehrabian, 1968). Consequently, students may be more inclined to take the perspective of their (imagined) audience into account while generating explanations, which may evoke several processes that could aid their own learning. For instance, imagining an audience may evoke students to believe that their actions (i.e., the explanations) can affect others (cf. productive agency; Okita & Schwarz, 2013; Schwarz, 1999). Consequently, they may monitor whether their explanations are comprehensible for their (imagined) audience, which provides a good indicator of how well they understand and explain it.

Moreover, if explaining on video stimulates learners to be aware of their potential audience, their level of arousal may increase (e.g., the Trier Social Stress Test also encompasses speaking for five minutes in front of a camera with the aim of inducing arousal; Kirschbaum, Pirke, & Hellhammer, 1993), which could affect their learning. It is well-established that the presence of an actual audience can affect how well people perform on a task (Aiello & Douthitt, 2001; Bond & Titus, 1983; Zajonc, 1965), and that arousal contributes to this audience effect (Aiello & Douthitt, 2001; Uziel, 2007). Importantly, situations that are located higher on the perceived social presence continuum seem to evoke stronger arousal responses. For example, being led to believe that another person in the room cannot see you decreases arousal levels compared to believing that they can (Myllyneva & Hietanen, 2015), and another person’s direct gaze leads to higher arousal levels than a person’s averted gaze (Helminen, Kaasinen, & Hietanen, 2011). Interestingly, no arousing effect of direct gaze occurs when pictures of people are presented on a screen (cf. Hietanen, Leppänen, Peltola, Linna-aho, & Ruuhiala, 2008; Pönkänen, Peltola, & Hietanen, 2011). With more credible manipulations, however, an imagined audience can also lead to more arousal. For example, Somerville et al. (2013) found that people who lay in a neuroimaging scanner experience higher arousal levels if they were led to believe that they were being watched by a peer via a camera embedded in the scanner than when they believed that the camera was off.

With regard to the relationship between arousal and learning, it has long been believed that there is an inverted U-shape function for the relationship between arousal and task performance (Salehi, Cordero, & Sandi, 2010; Yerkes & Dodson, 1908). Research indeed seems to indicate that relative to conditions of low or high arousal, moderate arousal levels can foster cognitive processes that are important for learning, such as memory, attention, and alertness (Arnsten, 2009; Diamond, Campbell, Park, Halonen, & Zoladz, 2007; Roozendaal, 2002; Sauro, Jorgensen, & Pedlow, 2003). Interestingly, Okita, Bailenson, and Schwarz (2007) demonstrated the link between social presence and arousal and learning for students learning how the human body deals with a fever. In their study, students who asked questions to a computer-based agent and then received scripted answers showed higher arousal levels when they were led to believe that the agent was controlled by an actual person than when they were led to believe that the agent was computer-controlled. Importantly, those who believed that the agent was an actual person performed better on a posttest, and students' posttest scores and arousal during learning were positively correlated.

In line with our hypothesis that explaining in front of a camera leads to increased feelings of social presence compared to producing written explanations, an explorative analysis of students' utterances showed that video explanations contained 2.5 times more self-other referential expressions (such as 'you' or 'we') than written explanations (and our measure corrected for explanation length). This finding resonates well with findings from research on asynchronous communication. Although an asynchronous communication situation is slightly different from explaining to a fictitious other, because learners may know each other (i.e., their audience) and may receive delayed replies, it is similar in the sense that messages are generated with the intention of being shared with others who are not present and cannot respond at that moment. Research on asynchronous communication has shown that social presence can be established using text only (e.g., asynchronous discussion forums; Andresen, 2009), although the lack of visual and vocal cues can make it difficult to do so (Garrison, Anderson, & Archer, 2000; Tu & McIsaac, 2002). The lack of vocal and visual cues associated with written communication has even been proposed as an explanation for the high attrition rate found in online education (Carr, 2000; Patterson & McFadden, 2009), possibly because learners feel isolated when such cues are not present in the learning environment (Palloff & Pratt, 2007). Consequently, asynchronous video communication has been proposed and is more frequently used as a means to increase the richness of communication (Borup et al., 2013). In the field of multimedia learning, the presence of social cues, such as a human voice compared to a machine-generated voice and a conversational speaking style opposed to a more formal one, has indeed been shown to positively affect the quality of learning outcomes. These benefits presumably arise because social cues induce a social response in the learner which leads to an increase in active processing (Mayer, 2014).

A potential limitation of the current study is that we cannot exclude the possibility that writing would also have been more beneficial than restudy if students would have had more time available for explaining. Compared to Hoogerheide et al. (2014b) and Fiorella and Mayer (2013, 2014), however, the time available for explaining had already been increased in the present experiments. Moreover, because the restudy and video condition indicated that they had sufficient time available, giving a writing group more time would result in unequal time spent on the task, and it would become unclear whether any potential benefits of writing would then be due to the explanation activity itself or the increased time on task. A second potential limitation is that we only used one type of learning task. Further research is needed to test whether beneficial effects of explaining on video can be generalized to other domains and types of tasks, although findings by Fiorella and Mayer (2013, 2014) suggest that it would. Their study showed that video explanations fostered comprehension when learning a short text about the Doppler Effect, which is very different from the syllogisms studied by our participants.

Despite these limitations, our findings add to the explanation literature by showing that in addition to providing self-explanations (e.g., Chi et al., 1989; Chi et al., 1994; Renkl, 1997, 2002), explanations to others in interactive situations (e.g., Cohen, 1994; Cohen et al., 1982; Johnson et al., 2007) or explanations to present others in non-interactive situations (Coleman et al., 1997), explaining to non-present, fictitious others is also effective for learning. This beneficial effect seems to be qualified by the form in which such explanations are provided: Explaining on video was effective, in writing it was not. These findings are also of interest for educational practice. Having students explain in writing is arguably much easier to implement, but it seems to yield little benefit. With cameras becoming ubiquitous (e.g., webcams, cameras in phones, tablets, laptops) and opportunities for storing and sharing video (online) becoming more affordable and accessible, video-based instruction is increasingly being used in educational practice. Moreover, a study procedure similar to the one used in this study can easily be implemented. The effectiveness of producing video explanations may even increase when students get the opportunity to edit and re-do their products (cf. learning by designing hypermedia: Lehrer & Romberg, 1996; Penner, Lehrer, & Schauble, 1998; or by designing “slow animations”: Hoban, Loughran, & Nielsen, 2011) or when they collaboratively create the videos (cf. Zahn et al., 2014; Zahn, Krauskopf, Hesse, & Pea, 2012).

To conclude, this study showed that explaining to fictitious others on video can be an effective learning activity compared to restudy, whereas explaining in writing is not. Considering that we found no direct differences between explaining in writing and on video and that we did not replicate the beneficial effect of explaining on video on transfer (cf. Hoogerheide et al., 2014b), it is important that these findings are replicated in future research. Such a replication could go hand in hand with a focus on mechanisms that make explaining on video more effective. We hypothesized and provided some tentative evidence that these might lie in feelings of social presence, and qualitative analysis of the explanation process data in future research could perhaps shed more light on this and other mechanisms that make explaining on video effective than restudy, but not explaining in writing (note that this would require log data of the writing process, instead of just the end product). Moreover, applied future research should investigate the effectiveness of providing video explanations when used in real classroom situations.

Chapter 8

Summary and Discussion

In recent years, learning from video modeling examples has become ubiquitous, both in formal and informal learning settings. Moreover, teachers are also increasingly creating video modeling examples for their students (e.g. http://www.wiskundeacademie.nl) or ask their students to create examples for fellow students (e.g. http://lvoorl.wikiwijs.nl). Yet, despite the widespread use and creation of video modeling examples nowadays, specific guidelines to help teachers and instructional designers to create videos that foster students' motivation and learning outcomes do not yet exist (Van Gog, 2013; of course there are well-established multimedia design guidelines that provide more general guidance, see Mayer, 2014). Therefore, the aim of the studies in the first part of this dissertation was to further our empirical knowledge of *whether and how the design of video modeling examples would affect students' motivation and learning.*

The second question addressed in this dissertation, relevant regarding the creation of video examples as a learning activity for students, was whether the act of explaining a learning task to others on video would have benefits for learning. In that case, the creation of video examples might have a double benefit, not only for the observer, but also for the model. Therefore, the studies in the second part of this dissertation were designed to investigate *whether providing explanations on video would yield benefits for motivation and learning*.

In this Chapter, the main findings from the studies presented in this dissertation are summarized and their theoretical relevance is discussed, followed by a discussion of recommendations for educational practice and a description of directions for future research. This is done separately for each part.

Part I: Effects of Example Design on Motivation and Learning

Summary of the Main Findings

Given the paucity of comparisons of different example types, it was as yet unclear whether the degree of model presence, and the design consequences this has, would affect self-efficacy and learning outcomes. Therefore, the question addressed in **Chapter 2** was whether and how the presence of the model in examples would affect cognitive and motivational aspects of learning. This was investigated by comparing written worked examples in which the model was absent, narrated examples in which the model could only be heard, and modeling examples in which the model was heard and seen. Students learned how to solve probability calculation problems by observing either two examples (Experiment 1) or one example (Experiment 2), with the form depending on their assigned condition. Effects on their self-efficacy and perceived competence were assessed, as well as effects on their learning outcomes in relation to invested mental effort. Effects on learning enjoyment and the willingness to receive similar instruction in the future were explored.

Results of both experiments showed that all three forms of example-based instruction were equally effective: test performance in all conditions improved substantially from the pretest to the immediate posttest, and performance remained high on a delayed posttest one week later. All three conditions were also equally effective at reducing effort investment in solving the problems from the pretest to the immediate posttest. Moreover, invested mental effort remained stable (at a lower level than the pretest) from immediate to delayed posttest after one week in Experiment 1, but increased somewhat in Experiment 2 while remaining at a lower level than the pretest. Effort investment in example study was low, and did not differ across conditions either. Regarding the motivational variables, it was found that all conditions were equally effective for enhancing students' self-efficacy and perceived competence from the pretest to the immediate posttest. Although students' confidence in their own abilities declined somewhat on the delayed posttest, it remained significantly above the perceived confidence levels prior to example study. The only significant difference among conditions in Experiment 1 was that those who had studied worked examples were more positive about receiving similar instruction in the future compared to students who studied video modeling examples with a visible model. In Experiment 2, those who had studied a video modeling example with a visible model seemed (numerically) more positive about receiving similar instruction in the future, but this was not statistically significant (p = .065). Possibly, the difference between the experiments is a consequence of removing the second example, which may have been redundant. In sum, worked examples and modeling examples with and without a visible model seem equally beneficial for cognitive and motivational aspects of learning –at least for the learning task (probability calculation) we used.

Chapters 3, **4**, and **5** all investigated the model-observer similarity (MOS) hypothesis. This hypothesis states that cognitive and affective aspects of learning are enhanced when students learn from a model that is more similar to them relative to a more dissimilar model. In **Chapter 3**, it was examined whether MOS in terms of gender would affect learning from video modeling examples that were otherwise identical in content. Male and female students learned how to solve probability calculation problems by observing a video modeling example with either a male or female model. The example content was kept identical across conditions. Effects on learning outcomes, effort investment, self-efficacy, and perceived competence were assessed. Based on the MOS hypothesis, it was expected that effects on motivation and learning would be more enhanced when studying a same-gender model than an opposite-gender model. Effects on students' learning enjoyment and the willingness to receive similar instruction in the future were explored.

Results showed that learning outcomes improved and effort investment in the test reduced from pretest to immediate posttest, but no MOS effects were found: There were no differences among conditions concerning increased test performance or reduced effort investment in the test. Effort invested in example study did differ between conditions, however: for male students it was less effortful to study a male model than a female model, and a male model was more effortful to study for female than male students. In a similar vein, a male model was also more enjoyable to learn from for male students than female students, and male students were more positive about receiving similar instruction in the future after studying a male model than female students. Students' self-efficacy and perceived competence improved from pretest to posttest, but again, no MOS effects were found. Males did show higher self-efficacy than females, and male models enhanced perceived competence more than female models for both male and female students.

The aim of the experiment presented in **Chapter 4** was to investigate whether similarity with the model in terms of age and putative expertise would affect students' learning and motivation when the example content was kept otherwise identical across conditions. Students first observed a short video in which an adult or a peer model introduced herself as having low or high expertise. Students then learned how to solve electrical circuit problems by observing two video modeling examples. Effects on learning and effort investment were examined, as well as on self-efficacy and perceived competence. Based on the MOS hypothesis, it was expected that students would benefit most from a low expertise peer model. Effects on learning enjoyment and students' assessment of the quality of explanations in the video modeling examples were explored.

Results showed that, in contrast to the MOS hypothesis, those who had studied *adult* models invested less effort in example study and learned more than those who studied peer models. Students who had observed adults also found the model's explanations of higher quality than those who observed peers, even though the adult and peer models provided the exact same explanations. Although students evaluated the explanations provided by the low expertise models as being of lower quality compared to the high expertise models, the expertise manipulation did not result in differences between the low or high expertise condition in mental effort or learning. Learning enjoyment or the degree to which self-efficacy and perceived competence were enhanced from pretest to posttest did not differ as a function of model age and expertise.

In **Chapter 5**, two experiments were described that addressed the open question of whether MOS in terms of gender, or in terms of age and expertise, would affect learning from text-based worked examples. Male and female students were led to believe via a short story and pictures that the worked examples were created by a male or female peer student (Experiment 1) or a peer student or teacher (Experiment 2). Notably, the content of the examples was kept identical across conditions. In both experiments, students learned how to solve electrical circuit problems with four worked examples. Effects on learning, mental effort, self-efficacy, and perceived competence were assessed. Effects on students' impression of the quality of the examples were explored in both experiments, and effects on learning enjoyment were explored in Experiment 2.

Experiment 1 showed no effects of the model's gender on learning outcomes, self-efficacy, perceived competence, or effort investment. Male students did show higher confidence in their own abilities than female students as well as higher test performance, attained with less effort investment. How students perceived the quality of the examples also did not differ across conditions. Like in Experiment 1, results of Experiment 2 showed no MOS effects. Males did show somewhat higher confidence in their own abilities as well as higher learning enjoyment than females, but this was not accompanied by a higher performance on the test. In sum, MOS in terms of gender or age and expertise does not seem to affect learning from text-based worked examples.

Discussion of the Main Findings

In general, the findings from Part I are in line with cognitive research on example-based learning (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Sweller, Ayres, & Kalyuga, 2011; Sweller, Van Merriënboer, & Paas, 1998; Van Gog & Rummel, 2010), by showing that studying video modeling examples (Chapters 2, 3, and 4) or text-based worked examples (Chapters 2 and 5) was very effective for learning. Students' problem-solving performance substantially increased from before (pretest) to after (posttest) example study. Moreover, the findings from Part I are also in line with social-cognitive research (Bandura, 1977, 1986; Renkl, 2014, Van Gog & Rummel, 2010), showing that learners' confidence in their own capabilities to perform the modeled task (self-efficacy and perceived competence) improved from studying video modeling examples (Chapters

2, 3, and 4) and worked examples (Chapter 2).[1] The effects of worked examples on students' motivation have received little attention (Renkl, 2014; Van Gog & Rummel, 2010), and our findings showed that worked examples, like modeling examples, enhance the confidence learners have in their own abilities to perform the modeled task. The more interesting questions, however, were whether motivational and learning outcomes would be affected differently depending on the degree of model presence and on the similarity between the model and the learner.

Effects of Model Presence. With regard to the effects of model presence, the study in Chapter 2 did not show any differences in terms of cognitive and motivational aspects of learning among written worked examples (in which the model is absent), narrated modeling examples in which the model was only heard, and modeling examples in which the model was both seen and heard. The fact that modeling examples did not lead to greater self-efficacy and perceived competence gains than worked examples was surprising, given that modeling examples contain more social cues due to the model's presence. There are several possible explanations for the lack of differences among conditions. First, perhaps the example content was insufficiently complex to find any differences among conditions that could be ascribed to differences in model presence and design consequences (e.g., more split attention in worked examples that consisted of written text and pictures). When task complexity is relatively low, ineffective design might be less of an issue, because sufficient working memory resources are then available to deal with the learning task as well as with the negative consequences of ineffective design (Paas, Renkl, & Sweller, 2003). Indeed, the high scores on the posttest after having studied just two examples (Experiment 1) or one example (Experiment 2) suggest that the learning task was not very complex.

Second, the nature of the modeled task may have played a role. The model's actions in this probability calculation problem mainly served to illustrate the problem solving steps. Different results might have been obtained when the modeled task inherently required some degree of human movement (e.g., origami), because in this case it might be more beneficial and more natural to observe a demonstration given by someone else than to read a written account of the actions accompanied by pictures (cf. Ayres, Marcus, Chan, & Qian, 2009; Wong, Marcus, Ayres, Cooper, Paas, & Sweller, 2009). For tasks that do not require learners to manipulate objects (such as solving probability calculation problems),

[1] Note that the study in Chapter 5 provided no information as to whether studying worked examples enhanced learners' self-efficacy and perceived competence because the phrasing of these measures was different before and after example study.

observing a demonstration by a model might not have added value relative to pictures of problem states. In addition, when the task is social in nature and/or requires models to interact with each other (e.g., negotiation skills: Gentner, Loewenstein, & Thompson, 2003; collaboration: Rummel & Spada, 2005), being able to see the model might be more beneficial for motivation and learning than only hearing the model or reading a written account of the interaction. In sum, although we did not find any differences among conditions, it cannot be definitively concluded from the present study that worked examples and modeling examples with different degrees of model presence are entirely equal in terms of motivational and learning outcomes for all types of tasks.

Effects of Model-observer Similarity. The question investigated in Chapters 3 to 5 was whether model-observer similarity (MOS) would affect motivation and learning. The MOS hypothesis states that the effectiveness of modeling examples is moderated by the degree to which learners perceive themselves as similar to the model in terms of age, gender, and expertise (Bandura, 1984; Schunk, 1987; Schunk & Zimmerman, 2007). Prior research on this question led to mixed findings, possibly because manipulations of the similarity between learners and the model across conditions often resulted in different content of the examples (e.g., Boekhout, Van Gog, Van de Wiel, Gerards-Last, & Geraets, 2010; Schunk & Hanson, 1985; Sonnenschein & Whitehurst, 1980), making it difficult to determine whether findings were due to MOS or differences in content. Therefore, the question whether MOS would affect motivation and learning was studied by varying *only* the model characteristics, while keeping the content of the examples equal across conditions.

None of the experiments showed any MOS effects on learning outcomes, self-efficacy and perceived competence, or effort investment. The experiment in Chapter 4 showed an opposite MOS effect with regard to age: adults were more effective models than peers, and more efficient in the sense that better test performance was attained with less effort investment during example study (for an elaboration on the concept of instructional efficiency, see Van Gog & Paas, 2008). The finding that, when example content is kept equal, adults are more effective models than peers is very interesting and potentially highly relevant for educational practice. A possible explanation is that students might have viewed the skill of troubleshooting electrical circuits as more appropriate for adults than peers, because adolescents tend to struggle with physics skills, even after continued direct instruction (Duit & Von Rhöneck, 1998; Fredette & Lockhead, 1980; McDermott & Shaffer, 1992).

It has been suggested that the degree to which students perceive a skill to be appropriate for a model can affect learning, and this 'task appropriateness view' may account for the mixed findings with regard to MOS (Bandura, 1986; Schunk, 1987; Zmyj & Seehagen, 2013). It indeed seems that students might have attributed less expertise to peer models than adult models, as those who had observed peers evaluated their model's explanations to be of lower quality than students who had observed adults, which is remarkable because they provided the exact same explanations. The idea that task-appropriateness rather than MOS determines the effects of model characteristics on motivation and learning resonates well with findings of other studies that manipulated gender and age of human models (e.g., Rodicio, 2012; Zmyj, Aschersleben, Prinz, & Daum, 2012) as well as with animated pedagogical agents (e.g., Arroyo, Woolf, Royer, & Tai, 2009; Moreno, Person, Adcock, Jackson, & Marineau, 2002; Rosenberg-Kima, Baylor, Plant, & Doerr, 2008).

Perceived task-appropriateness also seems to fit the finding that male models were found to enhance students' perceived competence more than female models in Chapter 3. That is, in mathematical tasks such as probability calculation, adolescents tend to ascribe more expertise to males than females (Ceci, Ginther, Kahn, & Williams, 2014; Steffens, Jelenec, & Noack, 2010), even if there are few (if any) actual differences in performance between the genders (Hyde, Fennema, & Lamon, 1990; Hyde, Lindberg, Linn, Ellis, & Williams, 2008).

Note that Chapter 5 showed that these effects of model age and gender did not apply to worked examples; here, adult models were not more effective than peer models and model gender did not affect perceived competence (or self-efficacy). A likely explanation is that video modeling examples contain a stronger social component than worked examples, which makes it easier for learners to identify with the model and engage in a social comparison (Mayer, 2014). Even when the worked examples used in Chapter 5 contained some degree of model presence in the form of an elaborate description of the model before example study, as well as the model's photo, name, and age during example study, the number of available social cues was still limited and could be ignored, which was much harder in the video modelling examples in Chapters 3 and 4. As such, MOS effects may not apply to learning from worked examples (see also Renkl, in press).

In sum, characteristics of the model in video modeling examples can affect motivation and learning, even when the content of the example is kept otherwise identical. However, these effects are not in line with the MOS hypothesis; instead, they can best be explained by perceptions of task-appropriateness. Theoretical views on the expected effects of MOS have also highlighted the importance of task-appropriateness, but did so by stating that similarity can help learners to determine whether behavior is appropriate (Bandura, 1994; Schunk, 1987).

This begs the question, is task appropriateness the most important factor, or does MOS have effects, and if so, under which conditions? One possibility is that MOS effects may occur for students learning 'neutral tasks', that is, tasks that are not (yet) perceived as more appropriate for one of the model types. This resonates well with findings from recent studies, showing that preschool children show more imitative behaviour after observing peer models relative to adults, unless adults are perceived as more knowledgeable (Zmyj & Seehagen, 2013).

Another question that remains is why the three manipulated model characteristics in Chapters 3 and 4 produced such a different pattern of results. That is, model age affected learning outcomes, model gender affected motivational outcomes, whereas (putative) model expertise did not have such effects. Although certain model characteristics may just be more crucial than others (Renkl, in press) and different characteristics may produce different effects, this mixed pattern of results may also, at least in part, be a result of differences across manipulations and/or limitations of the studies. For instance, in Chapter 4, each MOS factor (age and expertise) was tested using multiple models per condition, whereas in Chapter 3, only one male and one female model were used to test gender effects. As such, it cannot be ruled out that individual model characteristics may have affected the results. The finding that model gender had a motivational effect, but did not affect learning outcomes, may have had to do with the content of the examples. The probability calculation problems were not very difficult (as suggested by the performance data in Chapters 2 and 3), which makes it hard to find effects on learning outcomes. Had we used the physics problems from Chapter 4 to investigate effects of model gender, the outcomes might have been different. As for the lack of effect of the expertise manipulation, it should be noted that the age- and gender-related cues were continuously available and (likely) automatically processed during example study, but the expertise-related cues were not. As such, it is unclear whether expertise is less crucial (when content is kept equal) or whether our manipulation was simply not strong enough to establish effects on motivation and learning. Finally, a limitation of the studies presented here, is that perceptions of task-appropriateness or model similarity were not measured in any of the experiments. Therefore, it is impossible to tell whether the differential effects of model age, expertise, and gender, could result from the degree to which students' perceived the models as similar to themselves or the task as being appropriate for the models.

Practical Implications and Directions for Future Research

The studies presented in the first part of this dissertation showed that both video modeling examples and worked examples were very effective at fostering secondary education students' learning. This is in line with three decades of research (Atkinson et al., 2000; Renkl, 2014; Sweller et al., 1998; Sweller et al., 2011; Van Gog & Rummel, 2010), and as such, teachers and instructional designers can safely implement example-based instruction in educational practice. When designing video modeling examples, the degree to which the model is present in the example seems to be of limited importance, according to the findings from Chapter 2 (see also Kizilcec, Bailenson, & Gomez, 2015). However, when the model is present in a video example, findings from Chapters 3 and 4 suggest that characteristics of the model may affect motivation and learning. Most importantly, although future research would need to replicate this finding and establish whether it generalizes to other task types, it seems that video modeling examples provided by an adult model may be more effective for learning than examples provided by a peer model.

A first step for future research would be to systematically replicate and extend the studies using different types of learning tasks and learning materials. With regard to model presence (Chapter 2), for instance, it would be interesting to investigate whether, as has been shown with comparisons of dynamic vs. static animations (Ayres et al., 2009; Wong et al., 2009), observing a model's demonstration is more effective than reading a written account of the actions accompanied by pictures for tasks that inherently lend themselves to observational learning in the sense that they require human movement to complete.

With regard to Chapters 3 to 5, future research could investigate effects of model characteristics on motivation and learning using tasks that vary in their perceived appropriateness, while keeping the content of the examples otherwise identical. If the effectiveness of modeling examples indeed depends on learners' views of task-appropriateness, then, for example, students can be expected to benefit more from a peer model when learning 'a peer-appropriate task' and more from an adult model when learning 'an adult-appropriate task'. A neutral task might either show MOS effects, or no difference.

Lastly, considering that the effects of model age (Chapter 4) seem most promising for educational practice, another avenue for future research (assuming these findings would be replicated) would be to uncover what underlying process can explain why adults were more effective and efficient models than peers. One possible explanation could be that students just pay more attention to adults and keep their attention better sustained on the isomorphic example. Another possibility is that students focus more on task-irrelevant aspects of the video such as the model's appearance when studying a peer because they feel more interested in and attracted to peers than adults (Berscheid & Walster, 1969). These questions could be addressed using eye tracking to investigate students' attention allocation during example study.

Part II: Effects of Explaining To Fictitious Others on Video

In the second part, the question was addressed whether engaging in the two processes that are involved in acting as a peer model would foster students' learning and transfer. These two processes are: 1) studying learning materials with an explanation intention and 2) explaining learned materials to non-present others on video.

Summary of the Main Findings

In **Chapter 6**, two experiments were described in which it was examined whether the processes that are involved in acting as a peer model (i.e., studying with an explanation intention and providing explanations to non-present others on video) would foster learning outcomes for secondary education students (Experiment 1) and university students (Experiment 2) learning about syllogistic reasoning. Students read a text on syllogistic reasoning with the intention of completing a test (one group) or with the intention of explaining the content to others (two groups). One of the student groups who had studied with an explanation intention, subsequently explained the learning materials to a (non-present) other student via a webcam. The other explanation intention group and the test intention group restudied the text. Effects on learning, transfer, and effort investment were investigated, and effects on self-efficacy and perceived competence were explored.

Results of Experiment 1 (with secondary education students) showed that studying with an explanation intention was not more beneficial for learning and transfer than studying with a test intention. Studying with an explanation intention and subsequently explaining learning materials to others resulted in higher transfer performance than studying with a test intention. Although numerically the same trend was shown for learning, this difference was not statistically significant (p = .057). Conditions did not differ significantly in the degree to which self-efficacy, perceived competence, and effort investment in the test decreased from immediate to delayed posttest.

Results of Experiment 2 showed that, in contrast to Experiment 1, studying with an explanation intention enhanced learning more than studying with a test intention, possibly because the majority of students in Experiment 2 were university students enrolled in a problem-based learning curriculum and therefore used to providing explanations to other students. Again, only explaining to non-present others on video significantly fostered transfer performance relative to studying for a test. Explaining to others on video resulted in higher mental effort investment on the posttest than studying with an explanation intention without actually explaining. No differences were found in the degree to which students' self-efficacy and perceived competence decreased from immediate to delayed posttest.

The two experiments reported in **Chapter 7** extended the findings from the study reported in Chapter 6, by examining whether providing explanations to fictitious others in *writing* would foster learning outcomes for students' learning about syllogistic reasoning. In Experiment 1, students first read a text on syllogistic reasoning with a test intention or an explanation intention, and then restudied the text or provided explanations to non-present others in writing. In Experiment 2, the effects of explaining in writing, explaining on video, and studying with a test intention on learning, transfer, and effort investment were compared and effects on perceived competence were explored.

Results of Experiment 1 showed no beneficial effects of studying with an explanation intention compared to studying with a test intention. Moreover, explaining to fictitious others in writing did not enhance learning outcomes more than restudying (Experiment 1 and 2). Those who explained in writing even invested more effort in the learning phase than those who restudied. Explaining on video (Experiment 2) was also more effortful than restudying, but did translate into better learning. Surprisingly, given the findings in Chapter 6, it did not result in better transfer performance. The finding that explaining on video was more effective than restudy whereas explaining in writing was not, led to the hypothesis that explaining on video enhanced feelings of social presence (i.e., that students feel that they are conveying information to an actual person when they create a video, but not when they write). This hypothesis was explored by calculating the proportion of self-other referential words (e.g., 'you', 'me', and 'we'). Results showed that the video explanations indeed contained a higher percentage of such utterances, suggesting that explaining in front of a webcam provides students with enhanced feelings of social presence (i.e., addressing an audience) relative to explaining in writing.

Discussion of the Main Findings

The main question investigated in the second part of this dissertation (Chapters 6 and 7) was whether acting as a peer model for a video modeling example would foster learning outcomes for novices learning about syllogistic reasoning. In contrast to the large body of research on the effects of observational learning from video modeling examples (Van Gog & Rummel, 2010), it was still largely an open question whether the processes involved in acting as a model for a video modeling example, namely studying learning materials with the intention of explaining them to someone else and providing explanations to fictitious others on video, would also foster learning outcomes.

Studying with an Explanation Intention. Concerning the first process of acting as a model –studying learning materials with an explanation intention– prior research on the effects of studying with a teaching expectancy has shown mixed results (e.g., Ehly, Keith, & Bratton, 1987; Fiorella & Mayer, 2013, 2014; Fukaya, 2013; Renkl, 1995) and so did the studies in Chapter 6 and 7. Whereas studying with an explanation intention was equally effective as studying with a test intention in fostering learning and transfer for secondary education students (Chapter 6, Experiment 1) and higher professional education students (Chapter 7, Experiment 1), university students who studied with an explanation intention did show better learning (though not higher transfer performance) than students who studied with a test intention (Chapter 6, Experiment 2). The two experiments in Chapter 6 showed the same pattern of results across the immediate and delayed posttest. However, Fiorella and Mayer (2013, 2014) only found short-term benefits of studying with an explanation expectancy (i.e., at the immediate test) that decayed over time (i.e., had disappeared at the delayed test).

What might explain these mixed findings? It is possible that the experiment that did show a positive effect of studying with an explanation intention on learning comprised a 'false positive' finding, and that studying learning materials with an explanation intention simply does not foster learning outcomes relative to studying with a test intention. However, as mentioned above, research on the similar construct of teaching expectancy has also produced mixed results, and as such, it is more likely that the mixed findings can be attributed to other factors.

One factor that could affect whether students benefit from studying with an explanation expectancy is the degree of experience with explaining learned materials to others (cf. Nestojko, Bui, Kornell, & Bjork, 2014). Whereas secondary education and higher professional education students' experience might have been too low, the majority of university students that did benefit from an explanation intention were enrolled in a problem-based learning curriculum and therefore used to explaining learned materials to others and preparing to do so (cf. Loyens, Kirschner, & Paas, 2012). Although it is likely that experience plays a role, this 'experience hypothesis' cannot solely account for all findings, however. For example, Muis, Psaradellis, Chevrier, Leo, and Lajoie (2015) found that even primary school students, who were likely quite inexperienced with explaining learned materials to others, could benefit from studying with a teaching expectancy.

Next to experience with explaining, the strength of the prompts and instructions may also have contributed to the mixed effects. In both Chapters 6 and 7, the verbal and written prompts were arguably quite subtle (e.g., 'Can you explain the information on this page to a fellow student?'), which might have decreased the likelihood of students adopting a different study approach. Studies showing strong effects of teaching expectancy typically use more explicit and detailed instructions. In the study of Muis et al. (2015), for instance, students were explicitly instructed that they would have to create a teaching video in which they had to 'explain all steps involved, explain how they solved each step, and were told they could use all materials they created to solve the problem' (p. 10). A similar issue has been raised in the self-explanation literature, where the quality of the prompts might be crucial for determining whether students actually generate self-explanations (Schworm & Renkl, 2006). To compensate for the lack of self-explanation, various studies have prompted students to go 'beyond the text' (cf. Ploetzner, Dillenbourg, Preier, & Traum, 1999).

Explaining To Non-present Others on Video. As for the second process involved in acting as a model, which is actually explaining the materials, it can be concluded that providing explanations on video leads to lasting benefits. This was not only shown in the studies in Chapters 6 and 7, but also in studies by Fiorella and Mayer (2013, 2014). Although explaining to non-present others on video was more effortful than restudy (Chapter 7), this additional effort resulted in beneficial effects on learning and/or transfer for both secondary education students (Chapter 6, Experiment 1) and university students (Chapter 6, Experiment 2; Chapter 7, Experiment 2). These benefits were not simply a result of retrieval practice, which is inherent to explaining and known to enhance learning outcomes (Roediger, Putnam, & Smith, 2011), as those who restudied were provided with a recall activity before restudying in two of the experiments (Chapter 6, Experiment 2; Chapter 7, Experiment 2).

Note that in the first and second experiment in Chapter 7 it was investigated whether explaining to non-present others in writing, which would be easier to implement in educational settings than explaining on video, would also enhance learning outcomes compared to restudy. Results showed that generating written explanations did not enhance learning outcomes compared to restudying for higher professional education students (Experiment 1) or university students (Experiment 2). Explaining in writing was even less efficient in the sense that equal test performance was attained with more effort investment during example study (Van Gog & Paas, 2008).

The finding that explaining to non-present others in front of a camera is effective relative to restudy but explaining in writing is not, suggests that the benefits of explaining on video cannot simply be attributed to the act of generating explanations. Based on the data in Chapter 7, however, this alternative explanation cannot be ruled out, as explaining in writing might be more effective if more time is available for the explanation phase. In Experiment 2, those in the writing condition indeed indicated that they would have preferred to have more time for explaining. Note, though, that relative to Chapter 6 and the studies of Fiorella and Mayer (2013, 2014), both experiments in Chapter 7 already provided students with more time.

A more likely explanation for why only explaining on video was more beneficial than restudy is that explaining in front of a camera enhances feelings of social presence, that is, evokes learners to feel as if they are conveying their message to an actual audience (Gunawardena, 1995). Social presence is a key component of online learning (Sung & Mayer, 2013), and is positively associated with important factors such as learners' achievement (Mayer, 2005; Russo & Benson, 2005) and satisfaction (Gunawardena, 1995; Gunawardena & Zittle, 1997; Richardson & Swan, 2003). It indeed seems that explaining on video might have increased learners' social presence, as those who explained in front of a camera used significantly more self-other referential words such as 'you' and 'me' than those who explained in writing. There are several possible ways in which these enhanced feelings of social presence for those who explain in front of a camera could foster learning outcomes.

Firstly, higher levels of social presence could stimulate learners to consider their imagined audience while explaining, for instance by shifting their perspective to that of the recipient or monitoring whether the audience would be able to comprehend the message. This could provide learners with a good marker for the quality of their own knowledge, and may expose knowledge gaps and reasoning errors that could then be remedied during explaining.

Secondly, the social presence that learners might experience when generating explanations to fictitious others in front of a camera could enhance learners' arousal levels (explaining in front of a camera is a commonly used method to induce arousal;

Kirschbaum, Pirke, & Hellhammer, 1993) and thereby their learning. It is well-established in social psychology research that the presence of others can affect one's task performance, and that arousal is in part responsible for this effect (Aiello & Douthitt, 2001; Bond & Titus, 1983; Zajonc, 1965). Next to actual audiences, even imaginary audiences can induce arousal (except when the manipulations are not credible; cf. Hietanen, Leppänen, Peltola, Linna-aho, & Ruuhiala, 2008; Pönkänen, Peltola, & Hietanen, 2011). For example, it has been shown that people who lay in a FMRI scanner show enhanced arousal levels from merely being told that a peer would be monitoring them using a camera in the scanner (Sommerville et al., 2013).

As for the relationship between arousal and learning, research has shown that at least with moderate arousal levels, several processes that are determinants of and/or related to learning can be enhanced, such as memory capacity, attentional focus, and alertness (Arnsten, 2009; Diamond, Campbell, Park, Halonen, & Zoladz, 2007; Roozendaal, 2002; Sauro, Jorgensen, & Pedlow, 2003). Okita, Bailsenson, and Schwarz (2007) also showed that merely believing that an online computer-based agent was controlled by an actual person enhanced arousal and learning, and that those two measures were positively correlated. One of the possible mechanisms evoking arousal when explaining to someone else on video, even if s/he is not present, might be that it raises students' awareness that their actions can affect the imagined audience (cf. productive agency; Schwartz, 1999; Schwartz & Okita, 2004), stimulating them to focus and aim for high quality explanations rather than merely restating the problem (i.e., reflective-knowledge building rather than mere knowledge-telling; Roscoe & Chi, 2007, 2008).

Practical Implications and Directions for Future Research

The studies presented in part II of this dissertation showed that providing explanations to non-present others, at least on video, can be a powerful method for learning, just as providing explanations to one-self or in interactive situations can (Dunlosky et al., 2013; Fiorella & Mayer, 2015a, 2015b; Leinhardt, 2001; Lombrozo, 2012; Ploetzner et al., 1999; Richey & Nokes-Malach, 2015; Wilie & Chi, 2014). Considering that these days, cameras are everywhere, explaining on video is a promising instructional strategy that is easy to implement in educational practice. It is important to note that it is as of yet an open question to which degree these effects can be generalized to other types of learning tasks and learning materials, although findings Fiorella and Mayer (2013, 2014) indicate that the effects would generalize, as they showed very similar effects with very different materials (i.e., a short text about the Doppler effect).

Next to investigating the effects of explaining on video with different materials and tasks, another direction for future research would be to test whether the benefits of explaining on video can really be attributed to the video aspect. They might, alternatively, also 'simply' be a consequence of speaking words aloud (cf. the production effect; Bodner, Taikh, & Fawcett, 2014; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) or of generating explanations (Dunlosky et al., 2013; Fiorella & Mayer, 2015a, 2015b; Leinhardt, 2001; Lombrozo, 2012; Ploetzner et al., 1999; Richey & Nokes-Malach, 2015; Wylie & Chi, 2014). Note, though, that the former may be unlikely because so far production effects have only been shown with low complexity learning materials (e.g., word lists; Mama & Icht, 2016) and that the latter is unlikely because explaining in writing is not more beneficial than restudy (Chapter 7).

Assuming that the beneficial effects can indeed be attributed to the fact that explanations are directed to a camera, then the challenge is to explain exactly what it is that makes explaining on video effective for learning. As mentioned above, one hypothesis is that the benefits of explaining on video are in part due to feelings of social presence. Whether this is indeed the case, could be investigated by manipulating feelings of social presence while leaving the explaining activity untouched, for instance by manipulating the actual or virtual presence of an audience. Considering that explaining on video likely heightens learners' arousal levels (cf. Trier Social Stress Test; Kirschbaum et al., 1993), it would also be interesting to measure and actively manipulate students' arousal levels during explaining to see if certain levels of arousal are indeed associated with deeper learning (cf. Yerkes & Dodson, 1908). In addition to a self-report scale, online physiological measures such as skin conductance (Bach, Friston, & Dolan, 2010) and heart rate variability (Appelhans & Leucken, 2006) could be used because they provide a less biased, but still reliable, measurement of one's arousal.

Another interesting avenue for future research that could have important implications for educational practice is to examine whether the video modeling examples created by students can constitute effective educational materials. Prior research has shown that video modeling examples created by peer students can indeed be effective to learn from (e.g., Groenendijk et al., 2013a, 2013b), but these videos focused on a different topic and were created under different circumstances than those in the present dissertation. Before the video modeling examples may be beneficial for learning, those acting as a peer model might need additional time to plan, re-do, and edit the videos.

References

Aiello, J. R., & Douthitt, E. A. (2001). Social facilitation from Triplett to electronic performance monitoring. *Group Dynamics: Theory, Research, and Practice, 5*, 163–180. doi:10.1037//1089-2699.5.3.163
Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147-179. doi:10.1207/s15516709cog2602_1
Andersen, J. F. (1979). Teacher immediacy as a predictor of teaching effectiveness. In D. Nimmo (Ed.), *Communication yearbook 3* (pp. 543-559). New Brunswick, NJ: Transaction Books
Anderson, J. R. (1993). *Rules of the mind.* Hillsdale: Erlbaum.
Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 20*, 1322–1340.
Andresen, M. A. (2009). Asynchronous discussion forums: Success factors, outcomes, assessments, and limitations. *Technology & Society, 12*, 249-257.
Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience, 10*, 410–422.
Appelhans, B. M., & Leucken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology, 10*, 229-240. doi:10.1037/1089-2680.10.3.229
Arroyo, I., Woolf, B. P., Royer, J. M., & Tai, M. (2009). Affective gendered learning companion. In International conference on artificial intelligence and education. Brighton, England: IOS Press.
Atienza, F. L., Balaguer, I., & Garcia-Merita, M. L. (1998). Video modeling and imaging training on performance of tennis service of 9- to 12-year-old children. *Perceptual and Motor Skills, 87*, 519-529. doi:10.2466/pms.1998.87.2.519
Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181-214. doi:10.1037/0022-0663.94.2.405
Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*, 774–783. Doi:10.1037/0022-0663.95.4.774
Ayres, P., Marcus, N., Chan, C., & Qian, N. (2009). Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Computers in Human Behavior, 25*, 348-353. doi:10.1016/j.chb.2008.12.013
Ayres, P., & Sweller, J. (2005). The split-attention principle. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 135-146). Cambridge University Press, New York.
Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In: R. E. Mayer (Ed.), The Cambridge Handbook of Multimedia Learning (2nd edition, pp. 206-226). Cambridge University Press, New York.
Baars, M., Visser, S., Van Gog, T., De Bruin, A., & Paas, F. (2013). Completion of partially worked-out examples as a generation strategy for improving monitoring accuracy. *Contemporary Educational Psychology, 38*, 395-406. doi:10.1016/j.cedpsych.2013.09.001
Bach, D. R., Friston, K. J., & Dolan, R. J. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *International Journal of Psychophysiology, 76*, 52-55. doi:10.1016/j.ijpsycho.2010.01.011
Baddeley, A. (1986). *Working memory*. New York, NY: Oxford University Press.
Bandura, A. (1977). *Social learning theory*. Englewood Cliffs: Prentice Hall.
Bandura, A. (1981). Self-referent thought: A developmental analysis of self-efficacy. In: J. H. Flavell, & L. D. Ross, (Eds.), *Cognitive social development: Frontiers and possible futures* (pp. 200-239). Cambridge University Press, New York.
Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice Hall.
Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (pp. 71-81). New York: Academic Press. (Reprinted in H. Friedman [Ed.], Encyclopedia of mental health. San Diego: Academic Press, 1998).
Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52, 1-26*. doi:10.1146/annurev.psych.52.1.1
Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Greenwich, CT: Information Age Publishing.
Bandura, A., Ross, D., & Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology, 67*, 601-607. doi:10.1037/h0045550

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72*, 593-604. doi:10.1037/0022-0663.72.5.593

Baylor, A. L., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In J. C. Lester, R. M. Vicari, & F. Paraguacu (Eds.), *Intelligent tutoring systems* (pp. 592-603). Berlin: Springer.

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*, 95–115.

Becker, S., & Glidden, L. (1979). Imitation in EMR boys: Model competency and age. *American Journal of Mental Deficiency, 83*, 360-366.

Behrend, T. S., & Thompson, L. F. (2012). Using animated agents in learner-controlled training: The effects of design control. *International Journal of Training and Development, 16*, 263-283. doi:10.1111/j.1468-2419.2012.00413.x

Benware, C. A. & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal, 21*, 755-765. doi:10.2307/1162999

Berger, S. M. (1977). Social comparison, modeling, and perseverance. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 209-234). Washington, DC: Hemisphere.

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day.* Eugene, OR: International Society for Technology in Education.

Berscheid, E., & Walster, E. H. (1969). *Interpersonal attraction*. Reading, MA: Addison-Wesley.

Bjerrum, A. S., Hilberg, O., Van Gog, T., Charles, P., & Eika, B. (2013). Effects of modelling examples in complex procedural skills training: A randomised study. *Medical Education, 47*, 888-898. doi:10.1111/medu.12199

Bjork, E. L., & Bjork, R. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernbacher, R.W. Pew, L. M. How, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York: Worth.

Blessing, S., & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 576-598. doi:10.1037/0278-7393.22.3.576

Boekhout, P., Van Gog, T., Van de Wiel, M. W. J., Gerards-Last, D., & Geraets, J. (2010). Example-based learning: Effects of model expertise in relation to student expertise. *The British Journal of Educational Psychology, 80*, 557-566. doi:10.1348/000709910X497130

Bond, C. F., Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin, 94*, 264–292.

Bong, M. & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*, 1-40. doi:10.1023/A:1021302408382

Borup, J., West, R. E., & Graham, C. R. (2013). The influence of asynchronous video communication on learner social presence: A narrative analysis of four cases. *Distance Education, 34*, 48-63. doi:10.1080/01587919.2013.770427

Bottger, P. C. (1984). Expertise and air time as basis of actual and perceived influence in problem solving groups. *Journal of Applied Psychology, 69*, 214–221. doi:10.1037/0021-9010.69.2.214

Braaksma, M. A. H., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94*, 405-415. doi:10.1037/0022-0663.94.2.405

Brown, C. M. (1988). *Computer systems comparison of typing and handwriting in 'two-fingered typists'*. Human Factors and Ergonomics Society Annual Proceedings, 32, 381–385

Buunk, B. P., Zurriaga, R., Gonzalez-Roma, V., & Subirats, M. (2003). Engaging in upward and downward comparisons as a determinant of relative deprivation at work: A longitudinal study. *Journal of Vocational Behavior, 62*, 370-388. doi:10.1016/S0001-8791(02)00015-5

Carr, S. (2000). As distance education comes of age, the challenge is keeping the students. *Chronicle of Higher Education, 46*, 39-41.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*, 355–376.

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest, 15*, 75-141. doi:10.1177/1529100614541236

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460-473. doi:10.1037/0022-3514.66.3.460

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233-246. doi:10.1111/j.2044-8279.1992.tb01017.x

Chen, C., & Wu, C., (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108-121.

Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology, 104*, 103–125. doi:10.1037/a0030398

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182. doi:10.1207/s15516709cog1302_1

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477. doi:10.1207/s15516709cog1803_3

Chi, M. T. H., & VanLehn, K. A. (1991). The content of physics self-explanations*. Journal of the Learning Sciences, 1*, 69–105. doi:10.1207/s15327809jls0101_4

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (3rd ed.). San Francisco: Pfeiffer.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1-35.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N J: Erlbaum.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237-248. doi:10.2307/1162567

Coleman, E., Brown, A., & Rivkin, I. (1997). The effect of instructional explanations on formal learning from scientific texts*. Journal of the Learning Sciences, 6*, 347-365. doi:10.1207/s15327809jls0604_1

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453-494). Hillsdale: Erlbaum.

Collins, J.L. (1982). *Self-efficacy and ability in achievement behavior.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel patterns in fusiform face area differentiate faces by sex and race. *Plos One, 8*, e69684. doi:10.1371/journal.pone.0069684

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer*. Journal of Educational Psychology, 79*, 347-362. doi:10.1037//0022-0663.79.4.347

Craig, S. D., Gholson, B., Brittingham, J. K., Williams, J. L., & Shubeck, K. T. (2012). Promoting vicarious learning of physics using deep questions and explanations. *Computers & Education, 58*, 1042-1048. doi:10.1016/j.compedu.2011.11.018

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level reasoning effect: the role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction, 24*, 565-591. doi:10.1207/s1532690xci2404_4

Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers & Education, 49*, 809-821. doi:10.1016/j.compedu.2005.11.018

Crowley, K., & Siegler, R. S. (1999). Explanation and generalization in young children’s strategy learning. *Child Development, 70*, 304–316. doi:10.1111/1467-8624.00023

Davidson, E. S., & Smith, W. P. (1982). Imitation, social comparison, and self-reward. *Child Development, 53*, 928-932. doi:10.2307/1129130

Day, J. (2008). *Investigating learning with web lectures* (Doctoral dissertation). Available from Georgia Institute of Technology.

Day, J., & Foley, J. (2006). *Evaluating web lectures: A case study from HCI*. Paper presented at the conference on human factors in computing systems, Montreal, Canada. Retrieved June 6, 2014 from http://portal.acm.org/citation.cfm?id=1125493

Debono, K. G., & Harnish, R. J. (1988). Source expertise, source attractiveness, and the processing of persuasive information - a functional-approach. *Journal of Personality and Social Psychology, 55*, 541-546. doi:10.1037/0022-3514.55.4.541

Diamond, D. M., Campbell, A. M., Park, C. R., Halonen, J., & Zoladz, P. R. (2007). The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson Law. *Neural Plasticity.* doi:10.1155/2007/60803

Duit, R., & Von Rhöneck, C. (1998) Learning and understanding key concepts of electricity. In *Connecting Research in Physics Education with Teacher Education* (eds. A. Tiberghien, E. L. Jossem & J. Barajos). International Commission on Physics Education.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produced underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*, 271-280. doi:10.1016/j.learninstruc.2011.08.003

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions for cognitive and education psychology. *Psychological Science in the Public Interest, 14*, 4-58. doi:10.1177/1529100612453266

Ehly, S., Keith, T. Z., & Bratton, B. (1987). The benefits of tutoring: An exploration of expectancy and outcomes. *Contemporary Educational Psychology, 12*, 131-134. doi:10.1016/s0361-476x(87)80046-2

Ericsson, K. A., & Staszewski, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp.235-267). Hillsdale, NJ: Erlbaum.

Field, A. (2009). *Discovering statistics using SPSS*. London, England: Sage.

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*, 281-288. doi:10.1016/j.cedpsych.2013.06.001

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology, 39*, 75-85. doi:10.1016/j.cedpsych.2014.01.001

Fiorella, L., & Mayer, R. E. (2015a). Eight ways to promote generative learning. *Educational Psychology Review, 1*, 1-25.

Fiorella, L., & Mayer, R. E. (2015b). *Learning as a generative activity: Eight learning strategies that promote understanding*. New York: Cambridge University Press. doi:10.1017/CBO9781107707085

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*, 56-66. doi:10.1080/00461520.2012.748005

Forgasz, G. B., Leder, L. E., & Klosterman, P. (2004). New perspectives on the gender stereotyping of mathematics. *Mathematical Thinking and Learning, 6*, 389-420.

Frechette, C., & Moreno, R. (2010). The roles of animated pedagogical agents' presence and nonverbal communication in multimedia learning environments. *Journal of Media Psychology: Theories, Methods, and Applications, 22*, 61–72.

Fredette, N. H., & Lockhead, J. (1980). Students' conceptions of simple circuits. *The Physics Teacher, 18*, 194-198. doi:10.1119/1.2340470.

Fukaya, T. (2013). Does metacognitive knowledge about explanation moderate the effect of explanation expectancy? *Psychologia, 56*, 246–258. doi:10.2117/psysoc.2013.246

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2* ,87-105. doi:10.1016/S1096-7516(00)00016-6

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role of analogical encoding. *Journal of Educational Psychology, 95*, 393-408. doi:10.1037/0022-0663.95.2.393

George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology, 86*, 93-113. doi:10.1111/j.2044-8295.1995.tb02548.x

Gick, M. L. & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38. doi:10.1016/0010-0285(83)90002-6

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313-331. doi:10.1016/j.learninstruc.2005.07.001

Grabowski, J. (2007). The writing superiority effect in the verbal recall of knowledge: Sources and determinants. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and application* (pp. 165-179). Amsterdam: Elsevier.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York: Academic Press.

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10*, 17-32. doi:10.1002/(SICI)1099-0720(199611)10:7<17::AID-ACP435>3.3.CO;2-Z

Graesser, A., Person, N., & Magliano, J. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*, 495-522. doi:10.1002/acp.2350090604

Groenendijk, T., Janssen, T. M., Rijlaarsdam, G., & Van den Bergh, H. (2013a). Learning to be creative. The effects of observational learning on students' design products and processes. *Learning and Instruction, 28*, 35-47. doi:10.1016/j.learninstruc.2013.05.001

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & van den Bergh, H. (2013b). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *The British Journal of Educational Psychology, 83*, 3-28. doi:10.1111/j.2044-8279.2011.02052.x

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition, 14*, 53–82. doi:10.1075/pc.14.1.05gul
Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications, 1*, 147-166.
Gunawardena, C., & Zittle, F. (1997). Social presence as a predictor of satisfaction within a computer mediated conferencing environment. *American Journal of Distance Education, 11*, 8-26. doi:10.1080/08923649709526970
Harter, S. (1990). Causes, correlates, and the functional role of global self-worth: A life-span perspective. In: R. J. Sternberg & J. Kolligian (Eds.), *Competence Considered* (pp. 67-97)., Yale University Press, New Haven, CT.
Helminen, T. M., Kaasinen, S. M., & Hietanen, J. K. (2011). Eye contact and arousal: The effects of stimulus duration. *Biological Psychology, 88,* 124–130. doi:10.1016/j.biopsycho.2011.07.002
Hicks, D. J. (1965). Imitation and retention of film-mediated aggressive peer and adult models. *Journal of Personality and Social Psychology, 2*, 97-100. doi:10.1037/h0022075
Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction, 18*, 54-65. doi:10.1016/j.learninstruc.2006.10.008
Hilbert, T. S., Schworm, S., & Renkl, A. (2004). Learning from worked-out examples: The transition from instructional explanations to self-explanation prompts. In P. Gerjets, J. Elen, R. Joiner, & P. Kirschner (Eds.), *Instructional design for effective and enjoyable computer supported learning* (pp. 184-192). Tübingen, Germany: Knowledge Media Research Center.
Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., & Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach–avoidance motivational brain systems. *Neuropsychologia, 46*, 2423–2430. doi:10.1016/j.neuropsychologia. 2008.02.029
Hoban, G., Loughran, J. & Nielsen, W. (2011). Slowmation: Preservice elementary teachers representing science knowledge through creating multimodal digital animations. *Journal of Research in Science Teaching, 48*, 985-1009. doi:10.1002/tea.20436
Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*, 722-738. doi:10.1016/j.learninstruc.2007.09.013
Hoogerheide, V., Loyens, S. M. M., Jadi, F., Vrins, A., & Van Gog, T. (in press). Testing the model-observer similarity hypothesis with text-based worked examples. *Educational Psychology*. Advance online publication (2015), doi:10.1080/01443410.2015.1109609
Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014a). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior, 41*, 80-91. doi:10.1016/j.chb.2014.09.013
Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014b). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction, 33*, 108-119. doi:10.1016/j.learninstruc.2014.04.005
Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Does gender matter? *Instructional Science, 44*, 69-86. doi:10.1007/s11251-015-9360-y
Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review, 17*, 63-84. doi:10.1016/j.edurev.2015.11.002
Horowitz, M. W., & Newman, J. B. (1964). Spoken and written expression: An experimental analysis. *Journal of Abnormal and Social Psychology, 68*, 640-647. doi:10.1037/h0048589
Huang, X. (In press). Example-based learning: Effects of different types of examples on student performance, cognitive load and self-efficacy in a statistical learning task. *Interactive Learning Environments*. Advance online publication (2015), doi:10.1080/10494820.2015.1121154
Hughes, A., Galbraith, D., & White, D. (2011). Perceived competence: A common core for self-efficacy and self-concept? *Journal of Personality Assessment, 93*, 278-289. doi:10.1080/00223891.2011.559390
Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155. doi:10.1037//0033-2909.107.2.139
Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A., & Williams, C. (2008). Gender similarities characterize math performance. *Science, 321*, 494-495. doi:10.1126/science.1160364
Hyllegard, R., & Bories, T. L. (2009). Deliberate practice theory: Perceived relevance, effort, and inherent enjoyment of music practice: Study II. *Perceptual and Motor Skills, 109*, 431-40. doi:10.2466/PMS.109.2.431-440
Jakubczak, L. F., & Walters, R. H. (1959). Suggestibility as dependency behavior. *Journal of Abnormal and Social Psychology, 59*, 102-107.
Johnson, C. S., & Lammers, J. (2012). The powerful disregard social comparison information. *Journal of Experimental Social Psychology, 48*, 329-334. doi:10.1016/j.jesp.2011.10.010

Johnson, D. W., Johnson, T. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review, 19*, 15–29. doi:10.1007/s10648-006-9038-8

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments*. International Journal of Artificial Intelligence in Education, 11*, 47–78.

Juel, C. (1996). Learning to learn from effective tutors. In L. Schauble & R. Glaser (Eds*.), Innovations in learning: New environments for education* (pp. 49-74). Mahwah, NJ: Erlbaum.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-32. doi:10.1207/s15326985ep3801_4

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001).When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579-588. doi:10.1037//0022-0663.93.3.579

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science, 38*, 209-215. doi:10.1007/s11251-009-9102-0

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Education & Psychology, 96*, 558-568. doi:10.1037/0022-0663.96.3.558

Karpicke, J. D., & Blunt J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772-775. doi:10.1126/science.1199327

Kellogg, R. (2007). Are written and spoken recall of text equivalent? *American Journal of Psychology, 120*, 415-428. doi:10.2307/20445412

Kim, Y., & Baylor, A. L. (2015). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education, 15*, 95–115. doi:10.1007/s40593-015-0055-y

Kim, Y., Baylor, A. L., & Reed, G. (2003). The *impact of image and voice with pedagogical agents*. Paper presented at the E-Learn (World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education), Phoenix, Arizona.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*, 338–368. doi:10.3102/00028312031002338

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90*, 134–152.

Kirschbaum, C., Pirke, K. M., Hellhammer, D. H. (1993). The 'Trier Social Stress Test'-a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76-81. doi:10.1159/000119004

Kitsantas, A., Zimmerman, B. J., & Cleary, T. (2000). The role of observation and emulation in the development of athletic self-regulation. Journal of Educational Psychology, 91, 241-250.

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*, 724-739. doi:10.1037/edu0000013

Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In T. C. Urdan, & S. A. Karabenick (Eds.), *The Decade ahead: Theoretical perspectives on motivation and achievement* (pp. 1-33). Bingley, UK: Emerald Group Publishing Limited.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19,* 239–264. doi:10.1007/s10648-007-9049-0

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*, 708-721. doi:10.1016/j.learninstruc.2007.09.021

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32*, 22-36. doi:10.1016/j.learninstruc.2014.01.003

Koran, M. L., Snow, R. E., & McDonald, F. J. (1971). Teacher aptitude and observational learning of a teaching skill. *Journal of Educational Psychology, 62*, 219–228. doi:10.1037/h0031142

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, 131, 147-162. doi:10.1037/0096-3445.131.2.147.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 32*, 609-622. doi:10.1037/0278-7393.32.3.609.

Korving, H., Hernández, M., & De Groot, E. (2016). Look at me and pay attention! A study on the relation between visibility and attention in weblectures. *Computers & Education, 94*, 151-161. doi:10.1016/j.compedu.2015.11.011

Krauskopf, K., Zahn, C., & Hesse, F. W. (2012). Leveraging the affordances of YouTube: The role of pedagogical knowledge and mental models of technology functions for lesson planning with technology. *Computers & Education, 58*, 1194-1206. doi:10.1016/j.compedu.2011.12.010

Krautter, M., Weyrich, P., Schultz, J, H., Buss S, J., Maatouk, I., Junger, J., & Nickendei, C. (2011). Effects of peyton's four-step approach on objective performance measures in technical skills training: A controlled trial. *Teaching and Learning in Medicine: An International Journal, 23*, 244-50. doi:10.1080/10401334.2011.586917

Kühl, T., Scheiter, K., Gerjets, P., & Edelmann, J. (2011). The influence of text modality on learning with static and dynamic visualizations. *Computers in Human Behavior, 27*, 29-35. doi:10.1016/j.chb.2010.05.008

Kurby, C. A., Magliano, J. P., Dandotkar, S.,Woehrle, J., Gilliam, S., & McNamara, D. S. (2012). Changing how students process and comprehend texts with computer-based self-explanation training. *Journal of Educational Computing Research, 4*, 429–459. doi:10.2190/EC.47.4.e

Lachner, A., & Nückles, M. (2015). Bothered by abstractness or engaged by cohesion? Experts' explanations enhance novices' deep-learning. *Journal of Experimental Psychology: Applied*. doi:10.1037/xap0000038.

Law, W., Elliot, A. J., & Murayama, K. (2012). Perceived competence moderates the relation between performance-approach and performance-avoidance goals. *Journal of Educational Psychology, 104*, 806–819. doi:10.1037/a0027179

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*, 943-951. doi:10.1002/acp.1787

Lee, K. M., Liao, K., & Ryu, S. (2007). Children's responses to computer-synthesized speech in educational media: Gender consistency and gender similarity effects. *Human Communication Research, 33*, 310-329. doi:10.1111/j.1468-2958.2007.00301.x

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14*, 69-108. doi:10.1207/s1532690xci1401_3

Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook for research on teaching* (4th ed.) (pp. 333-357). Washington, DC: American Educational Research Association.

Lenhart, A. (2012). *Teens and video: Shooting, sharing, streaming and chatting*. Retrieved December 11, 2012, from pewinternet.org/Reports/2012/Teens-and-online-video/Findings.aspx.

Liew, T., Tan, S., & Jayothisa, C. (2013). The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions, task-related attitudes, and learning achievement. *Educational Technology & Society*, 16, 275-286.

Linek, S. B., Gerjets, P., & Scheiter, K. (2010). The speaker/gender effect: Does the speaker's gender matter when presenting auditory text in multimedia messages? *Instructional Science, 38*, 503-521. doi:10.1007/s11251-009-9115-8

Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input–process–output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology, 69*, 877–889. doi:10.1037/0022-3514.69.5.877

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*, 464-470. doi:10.1016/j.tics.2006.08.004

Lombrozo, T. (2012). *Explanation and abductive inference*. In K.J. Holyoak & R.G. Morrison (Eds.), Oxford Handbook of Thinking and Reasoning (pp. 260-276), Oxford, UK: Oxford University Press.

Loyens, S. M. M., Kirschner, P. A., & Paas, F. (2012). Problem-based learning. In K. Harris, S. Graham, & T. Urdan (Eds.), APA Educational Psychology Handbook: Vol. 3. *Application to learning and teaching* (pp. 403-425). Washington: American Psychological Association.

Ma, X., & Kishor, N. (1997). Attitude toward self, social factors, and achievement in mathematics: A meta-analytic view. *Educational Psychology Review, 9*, 89-120. doi:10.1023/A:1024785812050

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 671-685. doi:10.1037/a0018785

Mama, Y., & Icht, M. (2016). Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals the superiority of writing over vocalising. *Memory, 24*, 1–16. doi:10.1080/09658211.2014.986135

Mayer, R. E. (2005). Principles of multimedia learning based on social cues: Personalization, voice, and image principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 201-212). New York: Cambridge University Press.

Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment Principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 345-368). New York: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390-397. doi:10.1037/0022-0663.93.2.390

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320. doi:10.1037/0022-0663.90.2.312

McDermott, L., & Shaffer, P. (1992). Research as a guide for curriculum development: An example from introductory electricity. *American Journal of Physics, 60*, 994-1013. doi:10.1119/1.17003

McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176-2181). Austin: Cognitive Science Society.

McLaren, B. M., Van Gog, T., Ganoe, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in classroom experiments. *Computers in Human Behavior, 55*, 87-99. doi:10.1016/j.chb.2015.08.038

McMaster, K., Fuchs, D., & Fuchs, L. (2006). Research on peer-assisted learning strategies: The promise and limitation of peer-mediated instruction. *Reading and Research Quarterly, 22*, 5-25. doi:10.1080/10573560500203491

Montoya, R. M., & Horton, R. S. (2013). A meta-analytic investigation of the processes underlying the similarity-attraction effect. *Journal of Personal and Social Relationships, 30*, 64-94. doi:10.1177/0265407512452989

Moreno K. N., Person N. K., Adcock A. B., Eck, R. N. V., Jackson, G. T., & Marineau, J. C. (2002). *Etiquette and efficacy in animated pedagogical agents: the role of stereotypes.* Paper presented at the AAAI Symposium on Personalized Agents, Cape Cod, MA.

Moreno, R., & Flowerday, T. (2006). Student's choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology, 31*, 186-207. doi:10.1016/j.cedpsych.2005.05.002

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319-334. doi:10.1037/0022-0663.87.2.319

Muis, K. R., Psaradellis, C., Chevrier, M., Leo, I. Di, & Lajoie, S. P. (in press). Learning by preparing to teach: Fostering self-regulatory processes and achievement during complex mathematics problem solving. *Journal of Educational Psychology.* Advance online publication.

Myllyneva, A., & Hietanen, J. K. (2015). There is more to eye contact than meets the eye. *Cognition, 134*, 100–109. doi:10.1016/j. cognition.2014.09.011

Needham, D. R. & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition, 19*, 543-557. doi:10.3758/BF03197150

Nestojko, J. F., Bui, D. C., Kornell, N., & Bjork, E. L. (2014). Expecting to teach enhances learning and organization of knowledge in free recall of text passages. *Memory & Cognition, 42*, 1038-1048. doi:10.3758/s13421-014-0416-z

Newstead, S. E., Pollard, P., Evans, J. S., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition, 45*, 257-284. doi:10.1016/0010-0277(92)90019-E

Nievelstein, F., Van Gog, T., Van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology, 38,* 118–125. doi:10.1016/j.cedpsych.212.12.004

Nokes, T. J., Hausmann, R. G. M., VanLehn, K. A., & Gershman, S. (2011). Testing the instructional fit hypothesis: The case of self-explanation prompts. *Instructional Science, 39*, 645–666. doi:10.1007/ s11251-010-9151-4

Nosek, B. A, Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 10593-10597. doi:10.1073/pnas.0809921106

Okita, S. Y., Bailenson, J., & Schwartz, D. L. (2007). The mere belief of social interaction improves learning. In D. S. McNamara, & J. G. Trafton (Eds.), Th*e proceedings of the 29th meeting of the cognitive science society* (pp. 1355–1360). Nashville, USA: August.

Okita, S. Y., & Schwartz, D. L. (2013). Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *Journal of the Learning Sciences, 22*, 375–412. doi:10.1080/10508406.2013.807263

Orús, C., Barlés, M. J., Belanche, D., Casaló, L., Fraj, E., & Gurrea, R. (2016). The use of youtube as a tool for learner-generated content: Effects on students' learning outcomes and satisfaction. *Computers & Education, 95*, 254–269. http://doi.org/10.1016/j.compedu.2016.01.007

Ouwehand, K., Van Gog, T., & Paas, F. (2015). Effects of pointing compared with naming and observing during encoding on item and source memory in young and older adults. *Memory, 8211*, 1–13. doi:10.1080/09658211.2015.1094492

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education, 53*, 445-453. doi:10.1016/j.compedu.2009.03.002

Ozogul, G., Johnson, A. M., Atkinson, R. K., & Reisslein, M. (2013). Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education, 67*, 36-50. doi:10.1016/j.compedu.2013.02.006

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., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4. doi:10.1207/S15326985EP3801_1

Paas, F. & Van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*, 87-91. doi:10.1016/j.learninstruc.2006.02.004

Paas, F., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*, 737-743.

Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach*. Journal of Educational Psychology, 86*, 122–133.

Pajares, F. (2006). Self-efficacy during childhood and adolescence. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 339–367). Greenwich, CT: Information Age Publishing.

Palloff, R. M., & Pratt, K. (2007). Building online learning communities: Effective strategies for the virtual classroom (2nd ed.). San Francisco, CA: Jossey-Bass.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning classroom applications of research on self-regulated learning. *Educational Psychologist, 36*, 89–101. doi:10.1207/S15326985EP3602_4

Patterson, B., & McFadden, C. (2009). Attrition in online and campus degree programs. *Online Journal of Distance Learning Administration, 12.*

Penner, D. E., Lehrer, R., & Schauble, L. (1998). From physical models to biomechanics: A design-based modeling approach. *The Journal of the Learning Sciences, 7*, 429-449. doi:10.1207/s15327809jls0703&4_6

Ploetzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). Learning by explaining to oneself and to others. In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 103-121). Oxford: Elsevier.

Pönkänen, L. M., Peltola, M. J., & Hietanen, J. K. (2011). The observer observed: Frontal EEG asymmetry and autonomic responses differentiate between another person's direct and averted gaze when the face is seen live. *International Journal of Psychophysiology, 82*, 180–187. doi:10.1016/j.ijpsycho.2011.08.006

Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences, 23*, 125-133.

Putnam, A. L., & Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition, 41*, 36-48. doi:10.3758/s13421-012-0245-x

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, Cambridge, MA.

Renkl, A. (1995). Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. *Learning and Instruction, 5*, 21-36. doi:10.1016/0959-4752(94)00015-H

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*, 1-29. doi:10.1207/s15516709cog2101_1

Renkl, A. (2002). Learning from worked-out examples: Instructional explanations support self-explanations. *Learning and Instruction, 12*, 529-556. doi:10.1016/S0959-4752(01)00030-5

Renkl, A. (2011). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds*.), Handbook of research on learning and instruction* (pp. 272-295). New York: Routledge.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning*. Cognitive Science, 38*, 1-37. doi:10.1111/cogs.12086

Renkl, A. (in press). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed.). New York, NY: Routledge.

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin, 137*, 131-148. doi:10.1037/a0021705

Richey, J. E., & Nokes-Malach, T. J. (2015). Comparing four instructional techniques for promoting robust knowledge. *Educational Psychology Review, 27*, 181-218. doi:10.1007/s10648-014-9268-0

Richardson, J. C., & Swan, K. (2003). Examining social presence in online courses in relation to students' perceived learning and satisfaction. *Journal of Asynchronous Learning Networks, 7*, 68–88.

Ridgeway, C. L. (1987). Nonverbal behavior, dominance, and the basis of status in task groups. *American Sociological Review, 52*, 683–694. doi:10.2307/2095603

Robert, M. (1983). Observational learning of conservation: Its independence from social influence. *British Journal of Psychology, 74*, 1-10. doi:10.1111/j.2044-8295.1983.tb01838.x

Rodgers, W. M., Markland, D., Selzler, A. M., Murray, T. C., & Wilson, P. M. (2014). Distinguishing perceived competence and self-efficacy: An example from exercise. *Research Quarterly for Exercise and Sport, 85*, 527-539. doi:10.1080/02701367.2014.961050

Rodicio, H. G. (2012). Learning from multimedia presentations: The effects of graphical realism and voice gender. *Electronic Journal of Research in Educational Psychology, 10*, 885-906.

Rodriguez Buritica, J. M., Eppinger, B., Schuck, N. W., Heekeren, H. R., & Li, S. C. (in press). Electrophysiological correlates of observational learning in children. *Developmental Science.* Advance online publication. doi:10.1111/desc.12317

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1-36). Oxford: Elsevier.

Rohrbeck, C., Ginsburg-Block, M., Fantuzzo, J., & Miller, T. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review*. Journal of Educational Psychology, 95*, 240-257. doi:10.1037/0022-0663.95.2.240

Roozendaal, B. (2002) Stress and memory: Opposing effects of glucocorticoids on memory consolidation and memory retrieval. *Neurobiology of Learning and Memory, 78*, 578–595.

Roscoe, R. D. (2014). Self-monitoring and knowledge-building in learning by teaching. Instructional science, 42, 327-351. doi:10.1007/s11251-013-9283-4

Roscoe, R. D., & Chi, M. (2007). Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*, 534-574. doi:10.3102/0034654307309920

Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: the role of explaining and responding to questions. *Instructional Science, 36*, 321–350. doi:10.1007/s11251-007-9034-5

Rosekrans, M. A. (1967). Imitation in children as a function of perceived similarity to a social model and vicarious reinforcement. *Journal of Personality and Social Psychology, 7*, 307–315.

Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., & Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior, 24*, 2741-2756. doi:10.1016/j.chb.2008.03.017

Ross, S. M., & DiVesta, F. J. (1976). Oral summary as a review strategy for enhancing recall of textual material. *Journal of Educational Psychology, 68*, 689-695. doi:10.1037/0022-0663.68.6.689

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education, 14*, 51–70.

Rummel, N. & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences, 14*, 201–241. doi:

Ryalls, B., Gul, R., & Ryalls, K. (2000). Infant imitation of peer and adult models: Evidence for a peer model advantage. *Merrill-Palmer Quarterly, 46*, 188-202.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68-78. doi:10.1037//0003-066x.55.1.68

Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review, 22*, 379-392. doi:10.1007/s10648-010-9143-6

Salehi, B., Cordero, M. I., & Sandi, C. (2010). Learning under stress: The inverted-u-shape function revisited. *Learning & Memory, 17*, 522-530. doi:10.1101/lm.1914110

Sanders, J. A., & Wiseman, R. L. (1990). The effects of verbal and nonverbal teacher immediacy on perceived cognitive, affective, and behavioral learning in the multicultural classroom. *Communication Education, 39*, 341-353.

Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist, 18*, 42-50. doi:10.1080/00461528309529260

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort as a function of perceptions and attributions. *Journal of Educational Psychology, 76*, 647-658. doi:10.1037//0022-0663.76.4.647
Sauro, M. D., Jorgensen, R. S., Pedlow, C.T. (2003). Stress, glucocorticoids, and memory: A meta-analytic review. *Stress, 6*, 235–245.
Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113–125.
Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology, 73*, 93-105. doi:10.1037/0022-0663.73.1.93
Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist, 19*, 48-58. doi:10.1080/00461528409529281.
Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research, 57*, 149-174. doi:10.2307/1170234
Schunk, D. H. (1991a). *Learning theories: An educational perspective*. New York: Merrill.
Schunk, D. H. (1991b). Self-efficacy and academic motivation. *Educational Psychologist, 26*, 207-231.
Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 125-151). Mahwah, NJ: Erlbaum.
Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational psychology, 77*, 313-322. doi:10.1037//0022-0663.77.3.313
Schunk, D. H., & Hanson, A. R. (1989). Self-modeling and children's cognitive skill learning. *Journal of Educational Psychology, 81*, 155-163. doi:10.1037//0022-0663.81.2.155
Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Education & Psychology, 79*, 54-61. doi:10.1037/0022-0663.79.1.54
Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly, 23*, 7–25.
Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction, 14*, 293-305. doi:10.1016/j.learninstruc.2004.06.005
Schwartz, D. L. (1999). The productive agency that drives collaborative learning. In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 197-218). Oxford, UK: Pergamon.
Schwartz, D. L., & Okita, S. (2004). *The productive agency in learning by teaching.* Unpublished manuscript. Retrieved from http://aaalab.stanford.edu/papers/Productive_Agency_in_Learning_by_Teaching.pdf.
Schwarzer, R. (1992). *Self-efficacy: Thought control of action*. Washington, DC: Hemisphere.
Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. J. C. M. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior, 25*, 258–266. doi:10.1016/j.chb.2008.12.011
Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education, 46*, 426-445. doi:10.1016/j.compedu.2004.08.011
Sharpley, A., Irvine, J., & Sharpley, C. (1983). An examination of the effectiveness of a cross-age tutoring program in mathematics for elementary school children. *American Educational Research Journal, 20*, 103-111. doi:10.2307/1162677
Shipstone, D. M. (1984). A study of children's understanding of electricity in simple DC circuits. *European journal of science education, 6,* 185-198. doi:10.1080/0140528840060208
Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.
Simon, S., Ditrichs, R., & Speckhart, L. (1975). Studies in observational paired-associate learning: Informational, social, and individual difference variables. *Journal of Experimental Child Psychology, 20*, 81-104
Simon, S. J., & Werner, J. M. (1996). Computer training through behavior modeling, self-paced, and instructional approaches: A field experiment. *The Journal of Applied Psychology, 81*, 648-659. doi:10.1037/0021-9010.81.6.648
Singh, A., Marcus, N., & Ayres, P. (2012). The transient information effect: Investigating the impact of segmentation on spoken and written text. *Applied Cognitive Psychology, 26*, 848-853. doi:10.1002/acp.2885
Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review, 14*, 230–236. doi:10.3758/BF03194057
Smallwood, J., & Schooler, J. W. (2015). The Science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology, 66*, 487–518. doi:10.1146/annurev-psych-010814-015331

Somerville, L. H., Jones, R. M., Ruberry, E. J., Dyke, J. P., Glover, G., & Casey, B. J. (2013). The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. *Psychological Science, 24*, 1554-1562. doi:10.1177/0956797613475633

Sonnenschein, S., & Whitehurst, G. J. (1980). The development of communication: When a bad model makes a good teacher. *Journal of Experimental Child Psychology, 3*, 371-390. doi:10.1016/0022-0965(80)90101-0

Spanjers, I. A. E., Wouters, P., Van Gog, T., & Van Merriënboer, J. J. G. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior, 27*, 46-52. doi:10.1016/j.chb.2010.05.011

Spires, H. A., Hervey, L. G., Morris, G., & Stelpflug, C. (2012). Energizing project-based inquiry: middle grade students read, write, and create videos. *Journal of Adolescent & Adult Literacy, 55*, 483-493. http://dx.doi.org/10.1002/JAAL.00058

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology, 102*, 947-963. doi:10.1037/a0019920

Sternberg, R. J. (1987). Questioning and intelligence. Questing Exchange, 1, 11-13.

Stewart-Williams, S. (2002). Gender, the perception of aggression, and the overestimation of gender bias. *Sex Roles, 46*, 177-189. doi:10.1023/A:1019665803317

Strauss, S. S. (1978). The influence of peer and adult models upon object oriented children's performance of a task. *Dissertation Abstracts International, 39*, 3484.

Sung, E., & Mayer, R. E. (2013). Online multimedia learning with mobile devices and desktop computers: An experimental test of Clark's methods-not-media hypothesis. *Computers in Human Behavior, 29*, 639–647.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means–ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 463–474. doi:10.1037/0278-7393.8.5.463

Sweller, J. (1988). Cognitive load during problem-solving: Effects on learning. *Cognitive Science, 12*, 257-285. doi:10.1016/0364-0213(88)90023-7

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59-89. doi:10.1207/s1532690xci0201_3

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-295. doi:10.1023/A:1022193728205

Szpunar, K. K., Moulton, S. T., & Schacter, D. L. (2013). Mind wandering and education: from the classroom to online learning. *Frontiers in Psychology, 4*, 1–7. doi:10.3389/fpsyg.2013.00495

Tarmizi, R., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*, 424-436. doi:10.1037//0022-0663.80.4.424

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66-73. doi:10.1037/0022-0663.95.1.66

Töpper, J., Glaser, M., & Schwan, S. (2014). Extending social cue based principles of multimedia learning beyond their immediate effects. *Learning and Instruction, 29*, 10-20. doi:10.1016/j.learninstruc.2013.07.002

Traphagan, T., Kucsera, J. V., & Kishi, K. (2010). Impact of class lecture webcasting on attendance and learning. *Educational Technology Research and Development, 58*, 19-37. doi:10.1007/s11423-009-9128-7

Tu, C. H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. *The American Journal of Distance Education, 16*, 131-150. doi:10.1207/s15389286ajde1603_2

Uziel, L. (2007). Individual differences in the social facilitation effect: a review and meta-analysis. *Journal of Research in Personality, 41*, 579–601. doi:10.1016/j.jrp.2006.06.008

Van Blankenstein, F. M., Dolmans, D. H. J. M., Van der Vleuten, C. P. M., & Schmidt, H. G. (2011). Which cognitive processes support learning during small-group discussion? The role of providing explanations and listening to others. *Instructional Science, 39*, 189-204. doi:10.1007/s11251-009-9124-7

Van der Meij, H., & Van der Meij, J. (2013). Eight guidelines for the design of instructional videos for software training. *Technical Communication, 60*, 205-228.

Van der Meij, H., & van der Meij, J. (2014). A comparison of paper-based and video tutorials for software learning. *Computers & Education, 78*, 150–159. doi:10.1016/j.compedu.2014.06.003

Van Gog, T. (2011). Effects of identical example-problem and problem-example pairs on learning. *Computers & Education, 57*, 1775-1779. http://dx.doi.org/10.1016/ j.compedu.2011.03.019

Van Gog, T. (2013). *Voorbeeldig leren.* Inaugural lecture, Erasmus University Rotterdam.

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

Van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science, 36*, 1532-1541. doi:10.1111/cogs.12002.

Van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., Verkoeijen, P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review, 27*, 265-289. doi:10.1007/s10648-015-9297-3

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*, 212-218. doi:10.1016/j.cedpsych.2010.10.004

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16-26. doi:10.1080/00461520701756248

Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009b). The mirror-neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review, 21*, 21-30. doi:10.1007/s10648-008-9094-3

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science, 32*, 83-98. doi:10.1023/B:TRUC.0000021810.70784.b0

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*, 155-174. doi:10.1007/s10648-010-9134-7

Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education, 72*, 323-327. doi:10.1016/j.compedu.2013.12.004.

Vermunt, J. D. (1994). *Inventory of learning styles in higher education*. Leiden University, The Netherlands: ICLON-Graduate School of Education.

Walker, J. H., Sproull, L., & Subramani, R. (1994). Using a human face in an interface. In B. Adelson, S. Dumais, & J. Olson (Eds.), *Human factors in computing systems: CHI'94 conference proceedings* (pp. 85-91). New York: ACM Press.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology* (pp. 135-151). Harmondsworth, UK: Penguin.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13*, 21-39. doi:10.1016/0883-0355(89)90014-1

Wiener, M., & Mehrabian, A. (1968). *Language within language: Immediacy, a channel in verbal communication*. New York: Appleton-Century-Crofts.

Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychological values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology, 70*, 767-779. doi:10.1037/0022-3514.70.4.76

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review, 22*, 393-409. doi:10.1007/s10648-010-9136-5

Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction, 22*, 449-457. doi:10.1016/j.learninstruc.2012.05.004

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2009). Observational learning from animated models: Effects of modality and reflection on transfer. *Contemporary Educational Psychology, 34,* 1–8. doi:10.1016/j.cedpsych.2008.03.001

Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (Second edition, pp. 413–432). New York: Cambridge University Press.

Xeroulis, G, J., Park J., Moulton, C, A., Reznick, R, K., Leblanc, V, & Dubrowski, A. (2007). Teaching suturing and knot-tying skills to medical students: A randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery, 141*, 442-449. doi:10.1016/j.surg.2006.09.012

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation*. Journal of Comparative Neurology and Psychology, 18*, 459–482

Yi, M. Y., & Hwang, Y. (2003). Predicting the use of web-based information systems: Self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies, 59*, 431-449. doi:10.1016/S1071-5819(03)00114-9

Zahn, C., Krauskopf, K., Hesse, F.W., & Pea, R. (2012). How to improve collaborative learning with video tools in the classroom? Social vs. cognitive guidance for student teams. *International Journal of Computer-Supported Collaborative Learning, 7*, 259–284.

Zahn, C., Schaeffeler, N., Giel, K. E., Wessel, D., Thiel, A., Zipfel, S. & Hesse, F.W. (2014). Video clips for YouTube: Collaborative video creation as an educational concept for knowledge acquisition and attitude change. *Education and Information Technologies, 19*, 603-621.

Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269–274. doi:10.1126/science.149.3681.269.

Zhang, D., Zhou, Z., Briggs, R. O., & Nunamaker, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management, 43*, 15-27. doi:10.1016/j.im.2005.01.004.

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*, 660-668. doi:10.1037/0022-0663.94.4.660

Zmyj, N., Aschersleben, G., Prinz, W., & Daum, M. (2012). The peer model advantage in infants' imitation of familiar gestures performed by differently aged models. *Frontiers in Psychology, 3*, 250-266.

Zmyj, N. & Seehagen, S. (2013). The role of a model's age for young children's imitation: A research review. *Infant and Child Development, 22*, 622-641.

Nederlandse samenvatting

Summary in Dutch

Het leren van videovoorbeelden waarin iemand, het zogeheten model (als in 'rolmodel'), uitlegt en demonstreert hoe je een probleem oplost, is gemeengoed geworden in zowel informeel leren (denk aan leren klussen in huis via YouTube) als formeel leren (op school). Bovendien maken docenten steeds vaker zelf videovoorbeelden voor hun leerlingen (bijv. http://www.wiskundeacademie.nl), of vragen hun leerlingen om voorbeelden voor hun medeleerlingen te maken (bijv. http://lvoorl.wikiwijs.nl). Ondanks het veelvuldig gebruik van videovoorbeelden, is er een gebrek aan richtlijnen die docenten en ontwerpers van instructiematerialen kunnen helpen om videovoorbeelden zo te ontwerpen dat de motivatie van leerlingen en hun leeruitkomsten optimaal worden bevorderd (Van Gog, 2013; er zijn uiteraard wel richtlijnen die meer algemene handvaten bieden voor het ontwerpen van multimedia, zie Mayer, 2014). Het doel van de studies in het eerste deel van dit proefschrift was daarom om te onderzoeken *of en hoe de vormgeving van videovoorbeelden bepaalde motivationele aspecten (vertrouwen in eigen kunnen) en het leren kan bevorderen*.

De tweede vraag die in dit proefschrift werd onderzocht, was of het uitleggen van een leertaak aan anderen op video leeruitkomsten positief beïnvloedt. Dit is relevant voor docenten die het creëren van videovoorbeelden gebruiken als leeractiviteit voor hun leerlingen. Want als uitleggen leidt tot beter leren, dan is niet alleen het kijken naar, maar ook het creëren van videovoorbeelden een nuttige leeractiviteit. Daarom werden de studies in het tweede deel van dit proefschrift ontworpen om de vraag te onderzoeken of *het geven van uitleg op video bepaalde motivationele aspecten en leren bevordert*. In deze samenvatting van het proefschrift, wordt per hoofdstuk kort weergegeven wat er is onderzocht en wat de meest belangrijke bevindingen waren.

Deel 1: Effecten van de vormgeving van voorbeelden op motivatie en leren

Er is nog nauwelijks vergelijkend onderzoek gedaan naar effecten op motivationele aspecten en leeruitkomsten van verschillende typen voorbeelden. Daardoor is het nog onduidelijk of de mate waarin het model zichtbaar is in het voorbeeld, het vertrouwen van leerlingen in hun eigen kunnen, alsmede hun leeruitkomsten kan beïnvloeden. De vraag die in **Hoofdstuk 2** werd onderzocht was dan ook of en hoe de aanwezigheid van het model in voorbeelden cognitieve en motivationale aspecten van het leren kan beïnvloeden. Dit werd onderzocht door een vergelijking van geschreven tekstuele voorbeelden met plaatjes (model geheel afwezig), voorbeelden waarin de plaatjes zichtbaar waren en de uitleg van het model te horen was (model hoorbaar), en voorbeelden waarin het model zowel hoor- als zichtbaar was. Leerlingen leerden hoe ze kansrekeningproblemen konden oplossen met behulp van twee voorbeelden (Experiment 1) of één voorbeeld (Experiment 2), en de vormgeving van deze voorbeelden was afhankelijk van de conditie waarin de leerlingen zaten. Effecten op 'het vertrouwen in eigen kunnen' (Engels: *self-efficacy*) en 'waargenomen competentie' (Engels: *perceived competence*) werden onderzocht, evenals effecten op leeruitkomsten in relatie tot geïnvesteerde moeite. Effecten op plezier tijdens het leren en de bereidheid om soortgelijke instructie in de toekomst te ontvangen werden exploratief onderzocht.

De resultaten van beide experimenten lieten zien dat alle drie de vormen van voorbeelden even effectief waren: De testprestatie verbeterde aanzienlijk van pretest naar posttest, en de prestatie bleef hoog op de tweede posttest een week later. Alle drie de condities zorgden in gelijke mate voor een daling van de moeite die het oplossen van problemen kostte tussen de pretest en posttest. De moeite die het oplossen van problemen kostte bleef stabiel op het lagere niveau tussen de eerst posttest en de tweede posttest een week later. De moeite die leerlingen investeerden in het bestuderen van de voorbeelden was laag en verschilde niet over de condities. Inzake de motivationele variabelen, werd gevonden dat alle condities even effectief waren in het bevorderen van het vertrouwen in eigen kunnen van leerlingen evenals in het bevorderen van hun waargenomen competentie tussen de pretest en de posttest. Hoewel beide weer iets gedaald waren op de tweede posttest een week later, bleef het vertrouwen van studenten boven het niveau gemeten op de pretest. Het enige significante verschil tussen de condities in Experiment 1, was dat leerlingen die tekstuele voorbeelden hadden bestudeerd, positiever waren over het ontvangen van soortgelijke instructie in de toekomst dan leerlingen die voorbeelden hadden bestudeerd met een zichtbaar model. In Experiment 2 echter, leken de leerlingen die een voorbeeld hadden bestudeerd met een zichtbaar model juist positiever over het ontvangen van soortgelijke instructie in de toekomst dan leerlingen die een modelvoorbeeld zonder zichtbaar model of tekstueel voorbeeld hadden bestudeerd. Dit verschil was echter niet statistisch significant (p = .065). Het verschil tussen de twee experimenten is mogelijk een gevolg van het verwijderen van het tweede voorbeeld, dat mogelijk overbodig was in Experiment 1 (en dan is geschreven tekst makkelijker te negeren dan een video van iemand die de taak voordoet). Samenvattend lijken tekstuele voorbeelden en modelvoorbeelden met en zonder een zichtbaar model dus even effectief voor het bevorderen van cognitieve en motivationale aspecten van het leren—tenminste voor de leertaak (kansrekening) die we hebben gebruikt.

In de studies beschreven in Hoofdstuk 3, 4 en 5 werd de *model-observer similarity* (MOS) hypothese onderzocht. Deze hypothese stelt dat cognitieve en motivationele aspecten van leren meer bevorderd worden door een op de leerling gelijkend model dan door een minder op de leerling gelijkend model. In **Hoofdstuk 3** werd een experiment beschreven waarin onderzocht werd of MOS in termen van geslacht het leren van videovoorbeelden zou beïnvloeden. De voorbeelden waren verder identiek qua inhoud, alleen het geslacht van het model verschilde. Jongens en meisjes leerden hoe ze kansrekeningproblemen konden oplossen met behulp van een videovoorbeeld gepresenteerd door ofwel een mannelijk ofwel een vrouwelijk model. Effecten op leeruitkomsten, cognitieve belasting, vertrouwen in eigen kunnen, en waargenomen competentie werden onderzocht. Gebaseerd op de MOS hypothese, verwachtten we positieve effecten op motivatie en leren wanneer het geslacht van het model in overeenstemming is met het geslacht van de leerling. Effecten op het plezier tijdens het leren en de bereidheid om soortgelijke instructie in de toekomst te ontvangen werden exploratief onderzocht.

De resultaten lieten zien dat de testprestatie toenam en de moeite die het kostte om de testopgaven te maken afnam van de pretest naar de posttest, maar er waren geen MOS effecten: De condities verschilden niet in de mate waarin de testprestatie verbeterde en de geïnvesteerde moeite verminderde. Er was wel een effect op de hoeveelheid moeite die het bestuderen van het videovoorbeeld de leerlingen kostte: Voor jongens kostte het minder moeite om een voorbeeld door een mannelijk model te bestuderen dan een voorbeeld door een vrouwelijk model, en het bestuderen van een voorbeeld door een mannelijk model kostte meer moeite voor meisjes dan jongens. Soortgelijke resultaten werden gevonden voor het leerplezier: Jongens vonden het prettiger om van een mannelijk model te leren dan meisjes, en jongens waren positiever over het ontvangen van soortgelijke instructie in de toekomst na het leren van een mannelijk model dan van een vrouwelijk model. Zowel het vertrouwen in eigen kunnen als de waargenomen competentie namen toe van de pretest naar de posttest, maar wederom werd er geen MOS effect gevonden. Jongens gaven wel blijk van meer vertrouwen in het eigen kunnen dan meisjes, en een mannelijk model bevorderde de waargenomen competentie meer dan een vrouwelijk model, voor zowel jongens als meisjes.

Het doel van het experiment in **Hoofdstuk 4** was om het effect van MOS in termen van expertise en leeftijd te onderzoeken. Opnieuw werd de inhoud van de voorbeelden gelijk gehouden. Leerlingen bestudeerden eerst een korte video waarin een volwassen vrouw of een leeftijdsgenote zichzelf introduceerde als iemand met weinig of veel expertise in natuurkunde. De leerlingen bestudeerden vervolgens twee videovoorbeelden waarin door dat zelfde model werd voorgedaan hoe ze fouten in elektrische schakelingen konden opsporen. Effecten op leeruitkomsten, cognitieve belasting, vertrouwen in eigen kunnen, en waargenomen competentie werden onderzocht. Gebaseerd op de MOS hypothese, zou verwacht worden dat studenten het meest profijt hebben van een model van dezelfde leeftijd met weinig expertise in natuurkunde. Effecten op leerplezier en de perceptie van de kwaliteit van de voorbeelden werden exploratief onderzocht.

Resultaten lieten zien dat, in tegenstelling tot de MOS hypothese, de leerlingen die voorbeelden van een volwassen model bestudeerden minder moeite investeerden in het bestuderen van de voorbeelden, maar meer leerden dan leerlingen die voorbeelden van een leeftijdsgenote bestudeerden. Leerlingen die voorbeelden van een volwassen model hadden bestudeerd, vonden de kwaliteit van de uitleg van hun model ook beter dan degenen die een voorbeeld door een leeftijdsgenote hadden bestudeerd, ook al gaven alle modellen precies dezelfde uitleg. Ondanks dat leerlingen die voorbeelden zagen van modellen met zogenaamd hoge expertise, de uitleg van een hogere kwaliteit vonden dan de leerlingen die voorbeelden van modellen met zogneaamd weinig expertise bestudeerden, leidde de expertise-manipulatie niet tot verschillen in de hoeveelheid moeite die het bestuderen van voorbeelden en oplossen van problemen hen kostte, noch in de leeropbrengsten. Het plezier tijdens het leren en de mate waarin het vertrouwen in eigen kunnen en waargenomen competentie toenamen, was niet afhankelijk van de leeftijd of zogenaamde expertise van het model.

In **Hoofdstuk 5** werden twee experimenten beschreven waarin de open vraag werd onderzocht of MOS in termen van geslacht, (Experiment 1) of in termen van leeftijd en expertise (Experiment 2), het leren van tekstuele voorbeelden zou beïnvloeden. De manipulatie deed jongens en meisjes middels een kort verhaal en foto’s geloven dat de voorbeelden gemaakt waren door een mannelijke of vrouwelijke leeftijdsgenoot (Experiment 1) of door een leeftijdsgenoot danwel docent (Experiment 2). Wederom was de inhoud van de voorbeelden verder identiek. In beide experimenten bestudeerden leerlingen vier tekstuele voorbeelden waarin werd uitgelegd hoe ze fouten in elektrische schakelingen konden opsporen. Effecten op leeruitkomsten, cognitieve belasting, vertrouwen in eigen kunnen, en waargenomen competentie werden onderzocht. Effecten op de perceptie van de kwaliteit van de voorbeelden werden exploratief onderzocht, evenals effecten op plezier tijdens het leren in Experiment 2.

De resultaten van Experiment 1 gaven geen blijk van effecten van het geslacht van het model in relatie tot het geslacht van de leerling, op leeruitkomsten, vertrouwen eigen kunnen, waargenomen competentie, of cognitieve belasting. Jongens gaven wel blijk van meer vertrouwen in eigen kunnen dan meisjes, en lieten ook een betere testprestatie zien, die zij bereikten met minder moeite. Er was geen verschil tussen de condities in de waargenomen kwaliteit van de voorbeelden. In Experiment 2 werden eveneens geen MOS effecten gevonden. Ook in dit experiment, gaven jongens blijk van wat meer vertrouwen in hun eigen kunnen dan meisjes, maar dit ging niet gepaard met betere leeruitkomsten. Kortom: MOS in termen van geslacht of de combinatie van leeftijd en expertise, lijkt het leren van tekstuele voorbeelden niet te beïnvloeden.

Samengevat lijkt de mate waarin voorbeelden motivatie en leren bevorderen niet afhankelijk van de aanwezigheid/zichtbaarheid van het model in het voorbeeld (Hoofdstuk 2). Toekomstig onderzoek zal echter moeten uitwijzen of dit ook voor andersoortige taken geldt. Als het model zichtbaar is in een videovoorbeeld, dan kunnen cognitive en affective aspecten van het leren beïnvloed worden door de leeftijd en geslacht van het model, zelfs als de inhoud van het voorbeeld verder hetzelfde is (Hoofdstuk 3 en 4; dit geldt niet voor tekstuele voorbeelden, zie Hoofdstuk 5). De gevonden effecten zijn echter –met uitzondering van het effect op moeite in Hoofdstuk 3- niet in lijn met de MOS hypothese. Mogelijk komen deze effecten voort uit de mate waarin lerenden de taak als gepast zien voor het model. Een mannelijk model had vermoedelijk meer effect op de toename in waargenomen competentie dan een vrouwelijk model in Hoofdstuk 3, omdat adolescenten vaak meer expertise toeschrijven aan mannen dan aan vrouwen voor wiskundetaken. Wat betreft de natuurkunde taken in Hoofdstuk 4, waren volwassen modellen waarschijnlijk effectiever en efficiënter om van te leren dan leeftijdgenoten, omdat adolescenten meer expertise toeschrijven aan volwassenen als het gaat om complexe taken. Toekomstig onderzoek zal moeten onderzoeken of deze effecten gerepliceerd worden met andere soorten taken en of de effecten inderdaad voortkomen uit percepties van de gepastheid van de taak (bijvoorbeeld: Laten taken die als gepaster voor vrouwen of jongeren gezien worden, tegenovergestelde effecten zien?).

Deel II: Effecten van het uitleggen aan denkbeeldige anderen op video

In het tweede deel van dit proefschrift werd de vraag onderzocht of de twee processen die onderdeel uitmaken van het fungeren als model leren en transfer konden bevorderen. Die twee processen zijn: 1) het bestuderen van leerstof met de intentie om het een medeleerling uit te kunnen leggen en 2) het geleerde daadwerkelijk uitleggen op video aan een (niet aanwezige) medeleerling/medestudent.

In **Hoofdstuk 6** werden twee experimenten beschreven waarin onderzocht werd of de processen die inherent zijn aan het fungeren als model (d.w.z. studeren met een uitlegintentie en het daadwerkelijk uitleggen) leiden tot betere leeruitkomsten voor middelbare scholieren (Experiment 1) en universitaire studenten (Experiment 2). De leerstof bestond uit een tekst over syllogistisch redeneren. De lerenden bestudeerden deze tekst met de intentie om toetsvragen over de inhoud te kunnen beantwoorden (één groep) of met de intentie om de inhoud aan een ander uit te kunnen leggen (twee groepen). Van de twee groepen deelnemers die studeerden met een uitlegintentie, legde één groep vervolgens de geleerde materialen daadwerkelijk uit aan een medeleerling/medestudent (die niet aanwezig was) en die uitleg werd opgenomen met een webcam. De andere uitlegintentie groep en de testintentie groep bestudeerden nogmaals de tekst. Effecten op retentie en transfer en cognitieve belasting werden onderzocht en effecten op vertrouwen in eigen kunnen en waargenomen perceptie exploratief onderzocht.

De resultaten van Experiment 1 (met middelbare scholieren) lieten zien dat studeren met een uitlegintentie niet tot een betere testprestatie leidde dan studeren met een testintentie. Studeren met een uitlegintentie en vervolgens daadwerkelijk de geleerde materialen uitleggen aan anderen leidde wel tot een betere prestatie op transfer items dan studeren met een testintentie. Hoewel numeriek dezelfde trend zichtbaar was voor prestatie op items die retentie maten, was dit verschil tussen condities niet statistisch significant (p = .057). De condities verschilden ook niet in de mate waarin het vertrouwen in eigen kunnen, de waargenomen competentie en de moeite die de test kostte veranderde van de eerste posttest direct na het leren tot de tweede posttest een week later.

De resultaten van Experiment 2 lieten zien dat, in tegenstelling tot Experiment 1, studeren met een uitlegintentie tot een betere prestatie op retentie items leidde dan studeren met een testintentie, mogelijk doordat de meerderheid van de universitaire studenten in Experiment 2 een opleiding met probleemgestuurd onderwijs volgde en daardoor gewend was aan het geven van uitleg aan medestudenten. Wederom leidde het daadwerkelijk uitleggen aan anderen tot significant beter transfer prestatie dan studeren voor een test. Studenten die uitlegden op video investeerden meer moeite in de test dan degenen die studeerden met een uitlegintentie zonder daadwerkelijk uitleggen. Er werden geen verschillen gevonden in de mate waarin het vertrouwen in eigen kunnen en de waargenomen competentie veranderde van de posttest meteen na het leren tot de tweede posttest een week later.

De twee experimenten die in **Hoofdstuk 7** werden gerapporteerd, bouwden voort op de bevindingen van de studie uit Hoofdstuk 6, door te onderzoeken of het ‘op schrift’ uitleggen aan een medestudent ook een positief effect op retentie en transfer zou hebben. In Experiment 1 lazen studenten eerst een tekst over syllogistisch redeneren met een test- of uitlegintentie en bestudeerden vervolgens de tekst nogmaals of schreven een uitleg aan een (denkbeeldige) medestudent. In Experiment 2 werden de effecten onderzocht van uitleggen op schrift of uitleggen op video, vergeleken met studeren met een testintentie.

De resultaten van Experiment 1 lieten geen voordeel zien van studeren met een uitlegintentie ten opzichte van studeren met een testintentie. Ook een uitleg schrijven aan denkbeeldige anderen leidde niet tot betere testprestatie dan herstuderen (Experimenten 1 en 2). Degenen die uitlegden op schrift investeerden zelfs meer moeite in de leerfase dan degenen die herstudeerden (Experimenten 1 en 2). Uitleggen op video (Experiment 2) kostte ook meer moeite dan herstuderen, maar dit resulteerde wel in betere prestatie op retentie items. Verrassend genoeg gegeven de bevindingen in Hoofdstuk 6, resulteerde uitleggen op video echter niet in betere transfer prestatie. De effecten op retentie leidden tot de hypothese dat de verklaring voor het feit dat uitleggen op video wel effectiever was dan herbestuderen maar uitleggen op schrift niet, zou kunnen liggen in gevoelens van sociale aanwezigheid (d.w.z. dat lerenden het gevoel hebben dat ze informatie overbrengen aan een echt persoon wanneer ze een video maken, maar niet wanneer ze schrijven). Deze hypothese werd exploratief onderzocht door de proportie van persoonlijke voornaamwoorden in de uitleg te berekenen (bijv. ‘jij’, ‘ik’ en ‘wij’). De resultaten lieten zien dat de videouitleg inderdaad een hogere proportie aan persoonlijk voornaamwoorden bevatte, wat suggereert dat uitleggen voor een camera het gevoel van sociale aanwezigheid (d.w.z. spreken voor een publiek) bevordert, ten opzichte van uitleggen in geschreven vorm.

In conclusie, de studies beschreven in deel 2 lieten zien dat studeren met een uitlegintentie niet consistent tot betere leeruitkomsten leidt dan studeren met een testintentie. Alleen wanneer de geleerde informatie vervolgens daadwerkelijk wordt uitgelegd op video neemt de testprestatie toe vergeleken met herbestuderen van de informatie (Hoofdstukken 6 en 7). Hoofdstuk 7 liet verrassend zien dat uitleggen in geschreven vorm niet tot betere leeruitkomsten leidt dan herstuderen, wat suggereert dat de voordelen van uitleggen op video niet simpelweg een gevolg zijn van het geven van uitleg. Een mogelijke verklaring voor waarom gesproken uitleg voor een camera wel tot meer leren leidde en geschreven uitleg niet, is dat uitleggen op video lerenden het gevoel geeft dat de uitleg aan een 'echt publiek' wordt gegeven. Dit zou lerenden kunnen stimuleren om het publiek in acht te nemen, bijvoorbeeld door te monitoren of de uitleg coherent genoeg is en dat kan bevorderlijk zijn voor hun eigen begrup van de stof. Tevens kan de aanwezigheid van een videocamera spanning opleveren, en een bepaalde mate van spanning kan bevorderend zijn voor leren (Arnsten, 2009; Diamond, Campbell, Park, Halonen, & Zoladz, 2007; Roozendaal, 2002; Sauro, Jorgensen, & Pedlow, 2003). Aangezien digitale camera's vandaag de dag alomtegenwoordig zijn, is het uitleggen van geleerde informatie op video een veelbelovende en gemakkelijk te implementeren leeractiviteit voor de onderwijspraktijk.

Dankwoord

Acknowledgements in Dutch

Vijftien jaar geleden geloofde ik dat ik slechts één 'diploma' zou behalen, een rijbewijs. Ik herinner mij het nog als de dag van gisteren. Als je me toentertijd had verteld dat ik een universitaire opleiding zou afronden, had ik je niet geloofd. Laat staan als je me had verteld dat ik zou gaan promoveren. Ik ben van ver gekomen en dit promotietraject en proefschrift waren dan ook niet mogelijk geweest zonder de hulp van velen. Ik maak graag van deze gelegenheid gebruik om jullie te bedanken.

Allereerst natuurlijk mijn promotoren, Sofie en Tamara. Ik wil jullie beiden ontzettend bedanken voor de persoonlijke en prettige supervisie. Met jullie fijne persoonlijkheden en humor maakten jullie onze wekelijkse meetings steevast het hoogtepunt van mijn werkweek. Met als hoogtepunt misschien wel de verrukkelijke pannenkoeken bij Sofie thuis! Sofie, ik heb in het bijzonder gewaardeerd dat je altijd mijn wel en wee in acht hebt gehouden, en dat ik als groentje met allerlei praktische en theoretische vragen langs mocht lopen (of als ik laat op de avond simpelweg iets te eten wenste). Jouw perfectionisme en oog voor detail hebben de presentaties en manuscripten altijd in kwaliteit doen toenemen. Tamara, ik kan niet in woorden uitdrukken hoe dankbaar ik ben voor alle kansen die je mij hebt gegeven. Ik heb zóveel van je geleerd. Met name ben ik dankbaar voor je enorme drive en betrokkenheid (ondanks je drukke agenda, sorry en bedankt, Bas!) waardoor je mij stimuleerde om het optimale eruit proberen te halen. Het is mooi om te zien hoe een onderzoeker met de focus op het leren van goede voorbeelden, zelf zo een goed voorbeeld is.

Fred, ook jou wil ik in het bijzonder bedanken voor de kansen die je mij hebt gegeven op zowel onderwijs- als onderzoeksgebied en de prettige samenwerking en betrokkenheid over de jaren heen. Ik besef me dat zonder jou dit alles niet mogelijk was geweest.

All the defense committee members —prof.dr. Fred Paas, prof.dr Alexander Renkl, prof.dr. Liesbeth Kester, prof.dr. Gert Rijlaarsdam, dr. Peter Verkoeijen, and Alfons ten Brummelhuis— thank you so much for the time and effort you have invested in reading the dissertation, formulating questions, and the defense itself. I feel very honored and grateful to be able to defend my dissertation to such a wonderful group of experts.

Een van fijne aspecten van mijn promotietraject was de vrijheid die ik kreeg om samenwerkingsverbanden aan te gaan en onderzoeksideeën te ontwikkelen met andere onderzoekers. Dit hield het werk divers en gaf mij de kans om ook van anderen te kunnen leren. Naast Fred wil ik dus ook graag An, Anita, Lydia, en Margot bedanken voor de prettige samenwerking. Logan, thank you as well for the inspiring collaboration. Jullie hebben mij een betere onderzoeker gemaakt.

Het instituut voor psychologie aan de Erasmus Universiteit Rotterdam is jaren lang een prettig thuis geweest, voornamelijk door alle fijne collega's. Paranimfen en kamergenoten Marien en Jesper, bedankt voor de ontzettend fijne jaren. Ik beschouw jullie als broers en als het belangrijkste dat ik aan de promotieperiode over heb gehouden. Jullie aanwezigheid motiveerde mij om vroeg naar werk te reizen (want dan was Marien er vaak al) en pas laat weer naar huis te gaan (want dan was Jesper er meestal nog). We hebben samen niet alleen hard gewerkt, we hebben ook legendarische herinneringen gecreëerd. Zo vernietigden we elkaar regelmatig met Fifa, maakten we bowlingbanen onveilig, en lieten we AIO-etentjes ontsporen. Tim, dit alles geldt eveneens voor jou. Wie had gedacht dat er nog een gamende onderzoeker zou zijn met een liefde voor katten? Het is fijn om mensen om mij heen te hebben waar ik goede gesprekken mee kan voeren, maar ook een hoop plezier mee kan beleven. Ik kijk dan ook uit naar alle nieuwe herinneringen die we nog zullen vormen. Chantal, Hannah, en Kayin, allen bedankt dat jullie partner (nog steeds) af en toe buiten mag komen spelen.

Verder wil ik graag de gehele HLP-vakgroep bedanken voor alle inhoudelijke feedback tijdens de onderzoeksbijeenkomsten en pubgroepen, en de gezelligheid onder werktijd en daarbuiten, namelijk Bonnie, Charly, Daniel, Fred, Gerdien, Gert-Jan, Huib, Iris, Jacqueline, Jan, Kim, Lisette, Lysanne, Mario, Marit, Marloes, Margina, Margot, Martine, Monique, Nicole, Noortje, Peter, Lydia, Remy, Silvia, Sofie, Stefan, Stijn, Tamara, Tim, en Wim. Ik ben blij dat ik onderdeel uit heb mogen maken van deze bijzondere groep en de vele gezellige momenten staan in mijn geheugen gegrift. Fijn dat een aantal van jullie onderdeel uitmaakt van mijn nieuwe thuis in Utrecht. Kim en Martine, ik zal nooit vergeten hoe jullie als senior-aio's er voor mij zijn geweest tijdens de mijn eerste twee conferenties in Toulouse en München. Deze hulp reikte zo ver, dat toen ik per ongeluk een cheeseburger had besteld in het buitenland (zoals iedereen –behalve Marien– weet, heb ik een hekel aan kaas), Kim de helft van haar cheeseburger zonder kaas ruilde voor de helft van mijn cheeseburger mét kaas. Dat is nog eens collegialiteit. Gerdien, Marit, en Marloes, bedankt voor de fijne koffie- en theemomenten die mij de broodnodige afleiding gaven van het werk. Jullie zijn bijzondere personen. Silvia, I appreciate our talks about both research and life in general, and hope that we will be able to bring our research ideas to life. Guus en Marise, bedankt voor de hulp en betrokkenheid.

Ook buiten de eigen vakgroep heb ik het getroffen met leuke collega's. Killer Whales en Fifa Crew (Ali†, Daniel, Gert-Jan, Jan, Jesper, Marien, Mario, Stefan, Tim, en Wim), bedankt voor de fijne voetbaltijd. Ondanks dat we niet zo goed waren, of misschien zelfs erg slecht, heerste er een mooi kameraadschap onder deze mannengroep. (sarcasme) In het bijzonder heb ik gewaardeerd dat jullie anderen nooit hebben verteld dat ik een opponente een beenblessure heb bezorgd (/sarcasme). Debby and Ivo, thank you for being particularly helpful and kind during my rookie days. Marijntje, je was de meest knetterende korte-termijn kamergenoot die ik kon wensen! Jasmien, Michelle, Samantha, en Shalini, jullie uiteraard ook bedankt voor de gezellige koffie- en theemomenten. Michelle, ik ben blij dat we weer collega's zijn in Utrecht. Andrea, Danielle, Gloria, Huib, Keri, Generaal Luuc, Rolf, and Stephan, thank you for the wonderful conversations.

Over de jaren heen heb ik mogen genieten van een hoop ondersteuning, zoals statistische hulp van Marieke, Peter, en Samantha, lab- en videohulp van Christiaan, Marcel, en Gerrit Jan, en praktische ondersteuning bij onderwijs- en onderzoekswerkzaamheden van Angelique, Ilona, Iris, Janine, Jolien, Marit, Marja, Mirella, Ricardo, Shalini, en Wies. Allen ontzettend bedankt! Het is een luxe als er ongeacht de hulpvraag een expert op de afdeling rondloopt, waar je ook nog een goed gesprek mee kan voeren.

Mijn dank gaat uiteraard ook uit naar Kennisnet en in het bijzonder Alfons ten Brummelhuis, daarnaast wil ik Rola Hulsbergen en Mees van Krimpen van het 'Leerlingen voor Leerlingen' project bedanken voor hun betrokkenheid. Bedankt dat jullie dit project mogelijk hebben gemaakt. Verder wil ik graag alle docenten bedanken die het mogelijk hebben gemaakt om mijn onderzoek uit te kunnen voeren, in het bijzonder: Rinus van der Hoek (Krimpenerwaard College), Arjan Bijleveld (Develstein College), en Vincent van Dam (Calvijn Groene Hart). Zonder jullie enthousiasme en behulpzaamheid had dit proefschrift er niet gelegen. Jullie zijn geweldig. Dit geldt eveneens voor de studentassistenten die mij ondersteund hebben met de dataverzameling en dataverwerking (Eveline, Kirsten, Myrthe, Pim, Susan, en Tim) en de getalenteerde studenten wiens mooie onderzoek tot publicaties heeft geleid of waarschijnlijk nog zal leiden (Anna, Anniek, Fedora, Justine, Lian, Marleen, en Myrto).

I would like to thank some of the people that I have been fortunate to meet outside of the Erasmus University. Firstly, I would like to express my gratitude to Anne, Steffi, and Tamara for giving me the opportunity to serve as junior coordinator of the EARLI special interest group 7. I have learned a lot from working alongside you on the SIG activities, as well as from our collaborations with the past and present coordinators of special interest group 6 (Bjorn, Jean-Michel, Katharina, and Nikol). Alexander, thank you for your kindness and willingness to discuss research findings (and football) with me. Your comments and thoughts are always of great value. Anne-Marie and Mona, I appreciate our friendship and that you introduced Hsinya and me to the delicacy that is Tim-Tams. Little Bear still speaks fondly of his excursion in Bochum! Maria, I really value our friendship as well. I am happy and excited that you will be joining us in Utrecht. Now we can finally go to Feyenoord again! Siuman and Roxette, thank you both for the wonderful conversations about research and life. Christian, Ferdi, Ioulia, Jimmie, Julia, Michael, Myrto, Olaf, Paul, Quint and all the others, I look forward to sharing drinks with you again at the next conference.

Door de jaren heen heb ik veel steun mogen ontvangen van mensen 'buiten de wetenschap'. Firstly, my secondary education friends group comprised of Alex, Calvin, Chas, Chris, Carlos, Fung, Jay Jay, Jordy, Kelly, Kristofer, Mike, and Wessel. For many years now, we have been gaming, eating, and traveling together. Thank you for all the epic moments of awesomeness. Chas, Chris, Jay Jay, en Kelly, ik ben jullie in het bijzonder zoveel dank verschuldigd. Jullie zijn al 20 jaar als familie voor mij en ik heb geen idee wat er zonder onze vriendschap van mij terecht zou zijn gekomen. Samen met familie Lai (vader en moeder en oma Lai, Carroll, Chieroll, en natuurlijk Janice, Olivia, PoYee, en SeungKue), hebben jullie mij de wereld laten zien, leren genieten van lekker eten, en een extra thuis gegeven. Bedankt dat ik altijd welkom ben. Eveline, je bent niet alleen de afgelopen jaren mijn studentassistent en videomodel geweest, je bent ook een ontzettend belangrijke vriendin voor Hsinya en mij. Dank je voor al je hulp in vele vormen; je bent zó bijzonder voor ons! Cookie, thank you for your kindness and support (and Japanese cookies and presents for Maximus). You are a great friend. Ricardo en Monica, jullie ook bedankt voor de langstaande vriendschap en fijne momenten. Ik besef mij dat mijn (liefde voor mijn) werk ervoor heeft gezorgd dat ik jullie allen niet zo vaak zie als ik zou willen, maar dat maakt jullie niet minder belangrijk. Ik zal mijn leven proberen te beteren.

Moeder Elles, dank je voor je onvoorwaardelijke liefde en steun. Je bent een bijzonder persoon en ik hecht ontzettend veel waarde aan onze bijzondere band. En ik vergeef je voor de koude kip en het kaas-toetje. Vader Hans, jij ook bedankt voor je liefde en steun en in het bijzonder dat je een fijn thuis hebt gecreëerd vanuit waar ik mij heb kunnen ontwikkelen. Het is fijn dat jullie er altijd voor ons zijn en ik hoop dat wij nog vele jaren van jullie mogen genieten. Ik ben trots op jullie. Zusjes Leonie, Denise, en Natasja, hartelijk bedankt voor jullie hulp en steun. Als ik tegen problemen aanliep, zoals een gebrek aan modellen voor videovoorbeelden of proefpersonen voor onderzoeken, dan hoefde ik maar te gillen, en losten jullie het probleem voor mij op. Leonie, ik wil van deze kans gebruik maken om jou in het bijzonder te bedanken. Je bent er altijd voor mij geweest, en hebt mij geholpen om de persoon te worden die ik nu ben. Ik ben trots op hoe ver we allemaal gekomen zijn. Pieter en Gertjan, bedankt voor julie hulp hierbij. Jullie zijn lieve partners. Ome Jan en Tante Marian, bedankt dat jullie er altijd voor mij zijn geweest. Ook mijn familie zie ik niet zo vaak als ik zou willen. Maar weet dat jullie, en dit geldt ook voor alle lieve neefjes en nichtjes, ontzettend belangrijk voor mij zijn.

I would like to extend my gratitude to my family in law. Papa and Mama, thank you for how sweet you are to us and for your kind support. I look forward to seeing you again, hopefully soon in our new house. And I will try to find the time to learn Chinese. But for now: 鳳梨 我愛你! Nadia, thanks for your kindness, for all the lovely cat-talks and series-evenings, and for helping me with the front cover. You are an incredibly talented artist. Please do not forget to believe in yourself.

Lastly, my little family. General Maximus, thank you for often not allowing me to sleep, biting me, and making sure that I do not work too hard (by frequently stealing my chair and reminding me that it is playtime) and do not play too many games (by blocking my view of the computer screen and occassionally just sitting on the power button so my computer turns off). Hsinya, by the time that this dissertation is printed, we will be together for more than 5 girlfriends. I absolutely love our time together; we really have a great balance of serious conversations and serious silliness. Words cannot express just how important you are to me, and how grateful and blessed I feel that I can always count on your love, understanding, and support. You are the partner I have always dreamt of. <3

Nu het proefschrift helemaal af is, kijk ik met trots terug op de afgelopen jaren. Voor mij staat dit proefschrift symbool voor veel meer dan enkel het promoveren. Het is de bekroning op een lange tijd waarin ik hard heb gewerkt om mijzelf te ontwikkelen tot de persoon die ik nu ben. Deze tijd begon ooit toen ik ondanks leerplichtigheid op zestien-jarige leeftijd van de middelbare school afging. Moeder Elles en dhr. Bos, ontzettend bedankt dat dit toen mogelijk was. Het was vast een moeilijke beslissing, maar weet dat het de juiste was. Ik realiseer me dat er nog veel te leren valt, en gelukkig heb ik in Utrecht een fijne plek vanuit waar ik verder kan groeien. Lieve Utrecht collega's, bedankt voor de fijne ontvangst. In particular, Shakila, thank you for being a sweet roomie, and Anke, Barbara, and Milou, thank you for the great conversations.

Het is lichtelijk ironisch dat ik nu regelmatig veel langer bezig ben met de reis naar en van mijn werk, omdat ik het enige 'diploma' mis die ik vroeger dacht dat ik zou behalen, een rijbewijs. En alhoewel ik regelmatig de luxe heb van een privé chauffeuse (Tamara), zal ik na het promoveren gaan beginnen met lessen. Dan is de cirkel rond.

Publications

In press

Hoogerheide, V., Loyens, S. M. M., Jadi, F., Vrins, A., & Van Gog, T. (in press). Testing the model-observer similarity hypothesis with text-based worked examples*. Educational Psychology.* Advance online publication (2015), doi:10.1080/01443410.2015.1109609

2016

Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction, 44*, 22-30. doi:10.1016/j.learninstruc.2016.02.004

Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heiltjes, A., & Van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology, 44*, 95-106. doi:10.1016/j.cedpsych.2016.02.005

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Does gender matter? *Instructional Science, 44*, 69-86. doi:10.1007/s11251-015-9360-y

2015

Van Gog, T., Kester, L., Dirkx, K., & **Hoogerheide, V.**, Boerboom, J., & Verkoeijen, P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review, 27*, 265-289. doi:10.1007/s10648-015-9297-3

2014

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior, 41*, 80-91. doi:10.1016/j.chb.2014.09.013

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction, 33*, 108-119. doi:10.1016/j.learninstruc.2014.04.005

Mavilidi, M., **Hoogerheide, V.**, & Paas, F. (2014). A quick and easy strategy to reduce test anxiety and enhance test performance. *Applied Cognitive Psychology, 5*, 720-726. doi:10.1002/acp.3058

2012

Hoogerheide, V., & Paas, F. (2012). Remembered utility of unpleasant and pleasant learning experiences: Is all well that ends well? *Applied Cognitive Psychology, 26*, 887-894. doi:10.1002/acp.2890

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2012). Observationeel leren van videovoorbeelden. *Weten wat werkt en waarom* (p. 17-22). Zoetermeer, The Netherlands: Kennisnet.

Submitted

Hoogerheide, V., Vink, M., Finn, B., M., Raes, A. K., & Paas, F. (2016). *How to bring the news… Peak-end effects in children's affective responses to peer assessments of their social behavior*. Manuscript submitted for publication.

Fiorella, L., Van Gog, T., **Hoogerheide, V.**, & Mayer, R. E. (2016). *It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task.* Manuscript submitted for publication.

Van Gog, T., Kusuma, L. A., Loyens, S. M. M., **Hoogerheide, V.**, Baars, M., Heijltjes, A., & Mamede, S. (2016). *Effects of reflection, examples, and reflection examples on learning and transfer of reasoning and judgment tasks.* Manuscript submitted for publication.

Presentations

2016

Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M. M., & Van Gog, T. (2016, August). *Learning from video modeling examples: Content kept equal, adults are more effective models than peers*. Paper presented at the EARLI Special Interest Group 6 and 7, Dijon, France.

Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M. M., & Van Gog, T. (2016, June). *Example-based learning: A test of the model-observer similarity (MOS) hypothesis.* Poster presented at the 9th International Cognitive Load Theory Conference, Bochum, Germany.

2015

Hoogerheide, V., Loyens, S. M. M., Jadi, F., Vrins, A., & Van Gog, T. (2015, August). *Testing the model-observer similarity hypothesis with worked examples.* Paper presented at the EARLI conference, Limassol, Cyprus.

Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heiltjes, A., & Van Gog, T. (2015, August). *Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them.* Paper presented at the EARLI conference, Limassol, Cyprus.

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2015, August). *Learning from video modeling examples: The effect of gender.* Paper presented at the JURE 2015, the biannual pre-conference of the Junior Researchers of EARLI, the European Association for Research on Learning and Instruction, Limassol, Cyprus.

Hoogerheide, V. (2015, April). *Learning from creating and watching video modeling examples.* Masterclass given at TinQwise, Hilversum, The Netherlands.

Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heijltjes, A., & Van Gog, T. (2015, March). *Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them.* Poster presented at the International Convention for Psychological Science, Amsterdam, The Netherlands.

<u>2014</u>

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014, August). *Learning from video modeling examples: Does gender matter?* Poster presented at the EARLI Joint Special Interest Group 6 and 7, Rotterdam, The Netherlands.

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014, August). *Comparing the effects of worked examples and modeling examples on learning.* Poster presented at the EARLI Special Interest Group 2, Rotterdam, The Netherlands.

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014, June). *Effects of worked examples and modeling examples on learning outcomes.* Paper presented at the 7th International Cognitive Load Theory Conference, Taipei, Taiwan.

<u>2013</u>

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2013, August). *Effects of creating video-based modeling examples on learning and transfer.* Poster presented at the JURE 2013, the biannual pre-conference of the Junior Researchers of EARLI, the European Association for Research on Learning and Instruction, Munich, Germany.

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2013, June). *Effects of study intention and creating video-based modeling examples on learning and transfer.* Paper presented at the 6th International Cognitive Load Theory Conference, Toulouse, France.

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2013, April). *Beter leren door het maken van een videovoorbeeld?* Paper presented at the Kennisnet Conference ‘When ICT works in education’, The Hague, The Netherlands.

Curriculum Vitae

Vincent Hoogerheide was born in Rotterdam, The Netherlands, on July 11th, 1984. After high school, he obtained his Occupational Therapy degree at the Zadkine College (2006) followed by a propaedeutic Primary School Teacher diploma at the Pabo Thomas Moore (2007). During these years, he worked with intellectually disabled adults, people with dementia, and primary school children, respectively. Afterwards, he started studying Psychology at Erasmus University Rotterdam, during which he obtained his Bachelor's degree in Developmental and Educational Psychology with a specialization in Criminology (2010) and his Master's degree in Educational Psychology (2011; cum laude). After a brief period of working as a researcher at Erasmus University Rotterdam, Vincent started to work as a PhD candidate in March 2012 at the Institute of Psychology at Erasmus University Rotterdam. His PhD project was funded by Kennisnet and focused on the effects of learning from observing and creating video modeling examples. During his PhD project, he presented his work at various international conferences and helped organizing the EARLI Joint Special Interest Group 6 and 7 meeting in Rotterdam (2014) and Dijon (2016) as part of his junior coordinatorship for the EARLI SIG7. He was involved in various teaching activities, including coordinating a master's course and a practical. After three and a half years as a PhD candidate, Vincent started working as a Postdoctoral Researcher at the Department of Education at Utrecht University.