Computers in Human Behavior 89 (2018) 457–464

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence?

Vincent Hoogerheide*, Margot van Wermeskerken, Hilke van Nassau, Tamara van Gog

Department of Education, Utrecht University, The Netherlands

A R T I C L E I N F O

Article history:
Received 10 March 2017
Received in revised form
30 October 2017
Accepted 12 November 2017
Available online 13 November 2017

Keywords:
Model-observer similarity
Task-appropriateness
Modeling
Gender
Video
Multimedia learning

A B S T R A C T

In video modeling examples, a human model demonstrates and explains how to perform a task. The model-observer similarity (MOS) hypothesis predicts that learners who perceive themselves to be more similar to the model will learn more from the examples and show greater self-efficacy gains. Findings have been mixed, however; possibly because perceptions of task-appropriateness might affect learning and self-efficacy independently of similarity views. Therefore, we examined whether the effectiveness of modeling examples on troubleshooting electrical circuits, a task typically perceived as more appropriate for males than females, would differ as a function of the gender of the model and the observer. Secondary education students ($N = 159$) watched two video modeling examples, either by a male or a female model. The example content was kept equal. A manipulation check confirmed that students perceived same-gender models as more similar to them than opposite-gender models. They also perceived the task as more appropriate for males than females. Males performed somewhat better than females and showed higher confidence gains; however, model gender did not affect students' test performance, self-efficacy and perceived competence gains, effort investment, learning enjoyment, or perceived explanation quality. Our findings suggest that there is no need to take the model's gender into account when designing video modeling examples.

1. Introduction

Watching video modeling examples in which another person (the model) demonstrates and explains how to complete a task is an effective instructional strategy for novices to learn new skills and to enhance their self-efficacy (Renkl, 2014; Schunk, 1987; Van Gog & Rummel, 2010). There is a rapidly growing interest in the use of video modeling examples in both informal (e.g., YouTube, Vimeo) and formal (e.g., as part of massive open online courses or flipped classrooms) learning settings because video modeling examples have become relatively easy and cost efficient to create and to share with others in online platforms (Van der Meij & van der Meij, 2014). Despite the popularity of video modeling examples however, there is a paucity of evidence-based guidelines for how they can best be designed to stimulate learning and self-efficacy.

The popularity of video modeling examples and the scarceness of design guidelines have spurred studies in which it was investigated whether self-efficacy or learning gains are affected by various design aspects of the video, such as whether the video portrays an already drawn diagram from the start or a model drawing that diagram in 'lecture-style' video examples (Fiorella & Mayer, 2016), the perspective (first vs. third person) in which the video is presented in examples that involve demonstrations with objects (Fiorella, Van Gog, Hoogerheide, & Mayer, 2017), and the presence (vs. absence) of the model or the model's face in the video (e.g., Hoogerheide, Loyens, & Van Gog, 2014; Van Gog, Verveer, & Verveer, 2014; Wang & Antonenko, 2017). Other recent developments include examining whether, if the model is visible in the video, students' attention and learning are affected by gaze and gesture cues provided by the model (Beege, Schneider, Nebel, & Rey, 2017; Ouwehand, Van Gog, & Paas, 2015), and the presence (vs. absence) of the model's hands (Castro-Alonso, Ayres, & Paas, 2014).

Last but not least, a design question that garnered attention in early research on modeling (see Schunk, 1987) has regained

---

* Corresponding author. Department of Education, Utrecht University, P.O. Box 80140, 3508 TC, The Netherlands.
E-mail address: v.hoogerheide@uu.nl (V. Hoogerheide).

https://doi.org/10.1016/j.chb.2017.11.012

interest in the context of video modeling examples: whether and how the effectiveness of video modeling examples for self-efficacy and learning is affected by characteristics of the model in the video, such as the model's age, expertise, and gender. The hypothesis that motivated most of these studies was the model-observer similarity hypothesis.

1.1. Model-observer similarity

According to the *model-observer similarity (MOS) hypothesis* (Bandura, 1994; Schunk, 1987), the benefits of video modeling examples rely in part on model characteristics because modeling evokes learners to engage in a social comparison with the model (Berger, 1977; Johnson & Lammers, 2012). The hypothesis states that the more similar learners perceive themselves to be to the model in terms of characteristics such as age, expertise, and gender, the greater the influence on self-efficacy and thereby learning gains. Or in the words of Bandura (1994):

> “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 be more willing to pay attention to a more similar model (Bandura, 1971; Maccoby & Wilson, 1957) and the benefits of perceived similarity are held to be even more pronounced when learners consider the learning task to be (more) appropriate for the model (Schunk, 1987).

Empirical findings have been mixed however. For example, concerning model gender, a few early studies found evidence in favor of the MOS hypothesis. Bussey and Bandura (1984, Experiment 1) examined whether young children benefited more from video modeling examples portraying several male models playing a game or several female models playing a game, and found that both girls and boys showed greater imitative behavior after viewing same-gender than opposite-gender models. However, the majority of studies did not show greater benefits of a same-gender compared to an opposite-gender model on self-efficacy and learning outcomes. For example, Schunk and Hanson (1985) provided primary school students with video modeling examples in which either a male or female model demonstrated how to solve subtraction problems, and found no performance or self-efficacy differences between boys or girls who viewed a male or a female model. Schunk, Hanson, and Cox (1987, Experiment 2) replicated these null-findings with a sample of primary school students who learned how to solve fraction problems; again, a male and female model were equally beneficial for self-efficacy and learning.

1.2. Task-appropriateness

A possible explanation for these mixed effects might be that task-appropriateness perceptions are not only a catalyst for stronger MOS effects as has been argued in the MOS literature (e.g., Schunk, 1987), but can also affect self-efficacy and learning independently of similarity views. Various older and recent findings provide evidence in favor of this *‘task appropriateness hypothesis’* over the MOS hypothesis. For instance, Hoogerheide, Van Wermeskerken, Loyens and Van Gog (2016) recently found that secondary education students who studied adult models demonstrating how to troubleshoot faulty electrical circuits attained better learning outcomes (but not higher self-efficacy) and perceived the quality of the explanations to be higher than those who studied peer models, even though adults and peers gave the exact same explanations. It is unlikely that these findings can be attributed to individual characteristics of the models, because multiple models were used in each condition. Hoogerheide and colleagues hypothesized that this effect could possibly be explained by the age-appropriateness of the modeled task, because students tend to struggle with such difficult physics tasks (Duit & Von Rhöneck, 1998) and might pay more attention to adult models than peer models.

With regard to gender, several studies showed, in line with the task-appropriateness hypothesis, that a male model was more beneficial than a female model for both male and female learners when the task was likely to be viewed as more appropriate for males than females. This was the case for social behavior and cognitive skills. For instance, Bandura, Ross, and Ross (1963) and Hicks (1965) found that boys around the age of four who viewed a male model displaying aggression towards a doll showed greater imitative aggression than boys who viewed a female model, and that girls also displayed more imitative aggression after viewing a male than female model on the most ‘severe’ measures of aggression (e.g., aggressive gunplay). Similarly, with regard to cognitive skills, Garcia-Rodicio (2012, Experiment 2) recently investigated whether university students (mostly females) benefitted more from a multimedia presentation (without a visible model) about geology with a male voice-over or a female voice-over, and found that the male model led to better learning. In contrast, Linek, Gerjets, and Scheiter (2010, Experiment 1) found that male and female university students benefited most from and preferred a female voice-over compared to a male voice-over when learning how to solve probability calculation problems.

Finally, Hoogerheide, Loyens, and Van Gog (2016) provided secondary education students with a video modeling example in which a male or a female model demonstrated how to solve a probability calculation problem, and found no effects of the model's gender on self-efficacy and learning outcomes. However, a male model was more conducive to perceived competence than a female model for male and female students, which is a construct close to self-efficacy that focuses more on learners' need to be competent rather than the confidence in specific abilities (Rodgers, Markland, Selzler, Murray, & Wilson, 2014). Moreover, they found effects of model gender on learning enjoyment and effort invested during example study, showing that it was more enjoyable for male students to study a male model than a female model, that male students reported lower effort investment levels in studying a male than a female model, and that a male model was less effortful to study for male students than female students. However, because only one male and female model (or voice) were used per condition in these studies (e.g., Garcia-Rodicio, 2012; Hoogerheide, Van Wermeskerken et al., 2016), these effects could possibly be a result of individual model characteristics.

In sum, MOS studies have produced mixed findings, and one possible explanation might lie in learners' perceptions of task-appropriateness. The idea that task-appropriateness perceptions matter when learning from modeling is not entirely novel, as it has previously been argued that task-appropriateness views can strengthen MOS effects, both with regard to learning social behavior and cognitive skills (Schunk, 1987). Moreover, Bandura (1986) already proposed that how much attention learners pay to a model may vary depending on the perceived functional value of the demonstrated behavior, that is, whether learners believe that the demonstrated behavior leads to success or not. According to Bandura, similarity can act as a cue for the functionality of behavior, but there are other cues that may affect learners' attention, such as how appropriate the task is considered to be for the model.

Based on the currently available literature however, it is difficult to draw firm conclusions about whether and under what conditions model characteristics matter. That is, in many MOS studies, similarity manipulations also affected the models' actions, so what the models' did or said, making it difficult to assess whether any differences could be attributed to perceptions of similarity with the model or the different example content. Moreover, almost every study used only one model per condition, making it difficult to estimate whether any findings could be attributed to the manipulated model characteristic(s) or to individual characteristics of the models. Furthermore, perceptions of similarity to the model and of appropriateness of the task were only measured in a handful of MOS studies, which makes it difficult to tell whether findings resulted from MOS or task-appropriateness.

1.3. The present study

The purpose of the present study was to test the MOS hypothesis and task-appropriateness hypothesis with regard to model gender while 1) keeping the example content equal by scripting the model's explanations, movements, and gestures to rule out the possibility that any findings could be attributed to differences in the models' behavior, 2) using multiple models per condition to rule out the possibility that any findings could be attributed to individual characteristics of the models, and 3) using a learning task (i.e., a physics task, more specifically on troubleshooting electrical circuits) that is likely to be perceived as more appropriate for male than female models (Marchand & Taasoobshirazi, 2013).

The main question was whether the benefits of video modeling examples would depend on the gender of the observer and of the model. We assessed test performance and self-efficacy and perceived competence gains as the main outcome measures. If the MOS hypothesis is true, then male and female students would show greater test performance and higher self-efficacy and perceived competence gains when they observed a same-gender model (i.e., male students benefit more from a male model and female students from a female model). If the task-appropriateness hypothesis is true, then both male and female students would perform better on the test and show higher self-efficacy and perceived competence gains when they observed a model they considered to be more task appropriate, which is assumed to be the male model (i.e., male and female students benefit more from a male model). Because findings have been mixed, we could not formulate clear expectations as to which pattern of results would be more likely.

We additionally explored the effects of student and model gender on several variables. We measured perceived effort investment, because in conjunction with test performance, effort invested in example study and the posttest can provide valuable information about the cognitive efficiency of the conditions (Van Gog & Paas, 2008). We also asked students directly after example study to indicate their learning enjoyment, because learning enjoyment can provide an important indication as to whether examples would be used outside the experimental context (Yi & Hwang, 2003). Moreover, we asked students directly after example study to evaluate the quality of the model's explanations, because perceived explanation quality seems to be associated with how much expertise learners attribute to their model, and might therefore be affected by task-appropriateness perceptions as well (Hoogerheide, Loyens et al., 2016). Lastly, we checked whether our manipulation was successful by asking (at the end of the experiment) how similar learners perceived themselves to be compared to the model and whether they perceived the skill of explaining and demonstrating how to troubleshoot electrical circuits to be more appropriate for males or females.

2. Method

2.1. Participants and design

Participants were 159 Dutch secondary education students (84 females, 75 males; age: $M = 15.06$, $SD = 0.87$) in their third or fourth year of pre-university or general secondary education (i.e., the highest and second highest level of secondary education in the Netherlands, respectively). The experiment comprised a 2 × 2 design, with Model Gender (female vs. male) and Gender Observer (female vs. male) as between-subjects factors. Participants were quasi-randomly (i.e., matched for gender; for instance, within each class, for each male student allocated to the male model condition, another male student was allocated to the female model condition) allocated to the female model (43 female, 37 male students) or male model (41 female, 38 male students) condition. At the time of the experiment, participants were novices concerning the modeled task of troubleshooting electrical circuits, as troubleshooting electrical circuit problems had not yet been taught in the curriculum.

2.2. Materials

The materials were presented in Qualtrics (a survey software tool; www.qualtrics.com) and based on materials used in prior studies (e.g., Hoogerheide, Loyens et al., 2016; Van Gog, Kester, & Paas, 2011).

2.2.1. Pretest

The pretest was a conceptual prior knowledge test that consisted of seven open-ended questions about troubleshooting and parallel circuits principles, such as “What is probably going on if you do not measure any current in a parallel branch of the circuit?” and “What do you know about the total current in parallel circuits?”.

2.2.2. Introductory text

A brief introductory text provided a drawing of an electrical circuit, an explanation of the abbreviations and of the components in a circuit drawing and of Ohm's law using the three different iterations of the formula ($R = U/I$; $I = U/R$; $U = I^{*}R$).

2.2.3. Video modeling examples

Eight video modeling examples were created in total, two examples per model. Each video presented a lecture-style example with the model standing next to a screen on which PowerPoint slides were projected (see Fig. 1). The first example (240 s) started with a circuit drawing that was presented on the screen containing three parallel branches. The circuit indicated the resistance of each resistor and the voltage of the power source. The model explained, using this circuit drawing, that the current that should be measured in all three parallel branches and overall if the circuit were functioning correctly, could be calculated using Ohm's law. The model then demonstrated step-by-step how to calculate the current in each branch (using the information on voltage and resistance presented in the drawing) and the sum of the currents in the branches, that is, the total current. The presentation supported this explanation by displaying a smaller version of the same circuit drawing accompanied by Ohm's law and the worked-out problem-solving steps. Next, the presentation showed the measured currents as well as additional faulty ammeter measurements. The model explained that there was a discrepancy in one of the branches (lower current) and explained that this meant the resistance in that branch must be higher than it should be if the circuit were functioning correctly. The model then demonstrated how to calculate the actual resistance, using the slide displaying the

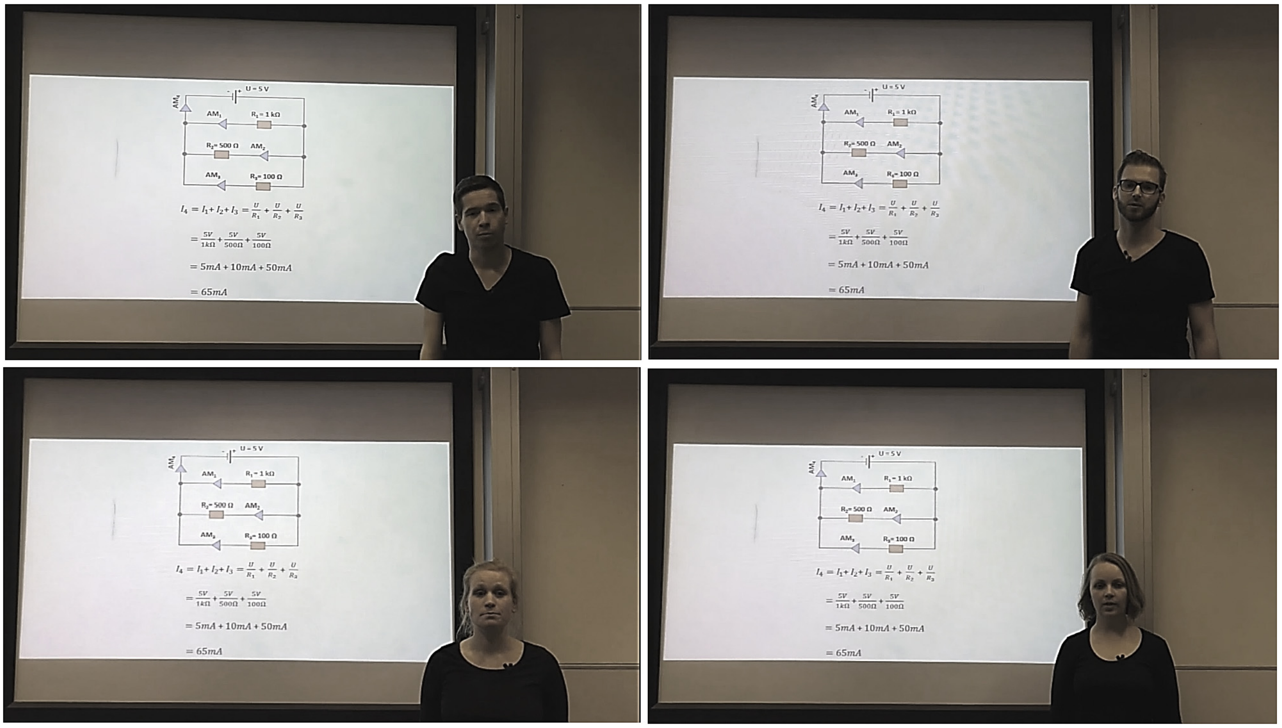

Fig. 1. Male models (top row) and female models (bottom row).

measured currents, Ohm's law, and the worked-out problem-solving steps. The second example (244 s) followed the exact same procedure, except with different surface features (e.g., different resistor values) and a different fault (i.e., *higher* current was measured in one parallel branch, which is indicative of *lower* resistance in that branch).

The models were all Caucasian, around 30 years old, and wore blue jeans and a black t-shirt. They were provided with a training before the final versions of the videos were recorded and it was ensured that each model provided the same explanation (by an autocue placed underneath the camera) and allocated the same amount of time to each part of the video (so that the videos had approximately the same length). At various points in time, the models gestured to elements in the presentation slides; which was scripted and trained to prevent differences among conditions.

2.2.4. Posttest

The posttest consisted of two troubleshooting problems. The first problem contained one fault (i.e., lower current which is indicative of higher resistance) and the second problem contained both faults that were encountered in the learning phase (i.e., lower current in one branch which is indicative of higher resistance and higher current in another branch which is indicative of lower resistance). Both problems reminded participants that ‘The current (U) is expressed in volt (V), resistance (R) is expressed in Ohm (U), and power (I) is expressed in amperes (A).’

2.2.5. Self-efficacy and perceived competence

Participants were asked to rate their self-efficacy before and after the learning phase by indicating how confident they were that they mastered the skill of detecting and solving electrical circuit problems on a scale of 1 (not at all confident) to 9 (very, very confident). Perceived competence was measured before and after the learning phase using an adapted version of the Perceived Competence Scale for Learning (Williams & Deci, 1996). That is, participants were asked before and after the learning phase to indicate on a scale of 1 (not at all true) to 7 (very true) to which extent three different items applied to them: ‘I feel confident in my ability to detect and solve electrical circuit problems’, ‘I feel able to meet the challenge of performing well in detecting and solving electrical circuit problems’, and ‘I am capable of detecting and solving electrical circuit problems’.

2.2.6. Invested mental effort

Participants were asked to rate how much mental effort they invested in studying each video modeling example or in solving each posttest problem on a 9-point rating scale (Paas, 1992) ranging from 1 (very, very low effort) to 9 (very, very high effort).

2.2.7. Learning enjoyment

Participants were asked to rate how enjoyable studying both video modeling examples was for them on a 9-point rating scale (Hoogerheide et al., 2014) ranging from 1 (lowest) to 9 (highest).

2.2.8. Perceived explanation quality

Participants were asked to rate the quality of the explanations after studying both video modeling examples, on a 9-point rating scale (Hoogerheide, Loyens et al., 2016) ranging from 1 (very, very bad quality) to 9 (very, very good quality).

2.2.9. Perceived similarity and perceived task-appropriateness

After the posttest, participants were asked to indicate who they believe to be better at demonstrating and explaining how to troubleshoot electrical circuits, ‘men’ or ‘women’, and to rate on a scale of 1 (not at all) to 9 (very much so) to what degree they perceived themselves to be similar to the person who demonstrated the electrical circuits troubleshooting tasks. Participants were provided a still from one of the video modeling examples to provide a reminder of what the model looked like.

2.3. Procedure

The experiment lasted circa 45 min and was conducted in a computer lab at participants' schools. Participants were first provided with a calculator, a headset, and an A4-page containing a link to the condition-specific Qualtrics environment. After logging into the learning environment, participants were first provided with a YouTube video that could be used to test the quality of their headset and broadband. After this test, the study progressed in 4 'blocks'. The first block consisted of a demographic questionnaire (e.g., educational level and age) and the pretest, which were followed by self-efficacy and perceived competence scales. Participants received 6 min for completing Block 1. Block 2 provided the brief introductory text, which participants studied for 2 min. The experimenter emphasized that participants needed this information to be able to understand the videos later on and that participants were not allowed to take notes. Block 3 presented the two video modeling examples. Each example was followed by the invested mental effort scale, and after the second mental effort scale, participants were provided with the learning enjoyment, explanation quality, self-efficacy, and perceived competence rating scales. Finally, Block 4 presented the posttest, for which participants received 12 min. Both posttest tasks were followed by the invested mental effort scale, and at the end of the posttest, participants were provided with the perceived similarity and perceived task-appropriateness rating scales.

2.4. Data analysis

We scored test performance using straightforward coding schemes that had been developed and used in prior research (e.g., Van Gog et al., 2011). Participants could earn ten points in total for the conceptual prior knowledge test and eight points for the posttest. The first task of the posttest was worth three points (one for calculating the correct value of the ammeters, one for indicating which resistor was faulty and what the fault was, and one for indicating the actual value of the faulty resistor) and the second task five points (extra points for correctly indicating the second faulty resistor and for calculating its resistance correctly). On both the pretest and posttest, incomplete or partially correct scores were given partial credit. Moreover, averages were calculated for invested mental effort in example study, invested mental effort in the posttest, and perceived competence before and after the learning phase.

3. Results

Two male students were removed from the sample for failing to complete the experiment within the given timeframe (one in the male model and one in the female model condition) leaving a final sample of 157 students (84 females; age: $M = 15.06$, $SD = 0.87$).[1]

We first checked whether the two male models had differential effects on the outcome measures, and found no significant differences; neither did the two female models have differential effects on the outcome measures. Therefore, we proceeded to analyze the data at condition level.

We then checked whether our manipulation had been successful, that is, whether students indeed perceived same-gender models as more similar to them than opposite-gender models, and the physics task as more appropriate for males than females. Concerning participants' perceived similarity to the models, a 2 × 2 ANOVA with Model Gender (female vs. male) and Gender Observer (female vs. male) as between-subjects factors showed, as one would expect, a significant interaction between Gender Observer and Model Gender, $F(1, 153) = 4.17$, $p = 0.043$, $\eta_p^2 = 0.027$, showing that male students perceived themselves to be more similar to male ($M = 3.00$, $SD = 1.87$) than female models ($M = 2.33$, $SD = 1.82$), whereas female students perceived themselves to be more similar to female ($M = 2.60$, $SD = 1.80$) than male models ($M = 2.10$, $SD = 1.70$). There were no main effects of Gender Observer, $F(1, 153) = 1.20$, $p = 0.274$, $\eta_p^2 = 0.008$, or of Model Gender, $F < 1$. As for students' perceptions of task-appropriateness, we conducted two binomial tests, one for male students and one for female students, to check if the answers to the question whether 'males' or 'females' would be better at demonstrating and explaining how to troubleshoot electrical circuits significantly deviated from what would be expected by chance (i.e., 50% males). Results showed that the answer possibility 'males' was indeed picked significantly more frequently than would be expected based on chance by both male students (88%; $t(72) = 9.72$, $p < .001$, $d = 2.292$) and female students (75%; $t(84) = 5.26$, $p < .001$, $d = 1.155$).

Thus, we proceeded with analyzing the main research and explorative questions. Unless otherwise indicated, the analyses consisted of 2 × 2 ANOVAs with Model Gender (female vs. male) and Gender Observer (female vs. male) as between-subjects factors. Table 1 shows the test performance, self-efficacy, perceived competence, invested mental effort, learning enjoyment, and perceived explanation quality scores per condition.

3.1. Test performance

We checked whether the four different conditions showed comparable performance on the pretest. There was no interaction between Model Gender and Gender Observer, $F < 1$. There was no main effect of Model Gender, $F < 1$, but there was a main effect of Gender Observer, $F(1, 153) = 10.28$, $p = 0.002$, $\eta_p^2 = 0.063$, showing that the male students ($M = 3.25$, $SD = 1.51$) performed significantly better than female students ($M = 2.49$, $SD = 1.40$).

Our main question was whether male and female students attained better test performance scores after example study when they had observed a same gender model (as the MOS hypothesis would predict) or a male model (as the task-appropriateness hypothesis would predict). This was tested for male and female students separately because of the pre-existing differences in prior knowledge between male and female students, with two independent samples *t*-tests. Concerning male students, there was no significant difference in posttest performance between those who had observed a male model ($M = 3.30$, $SD = 2.69$) or a female model ($M = 3.65$, $SD = 2.75$), $t < 1$. There was no significant difference in test performance between the female students who had observed a male ($M = 2.44$, $SD = 2.74$) or a female model ($M = 2.63$, $SD = 2.82$) either, $t < 1$.[2]

3.2. Self-efficacy and perceived competence

Self-efficacy and perceived competence were analyzed using a mixed factors ANOVA with Model Gender (female vs. male) and Gender Observer (female vs. male) as between-subjects factors and Test Moment (before vs. after learning) as within-subjects factor. Concerning self-efficacy, there were no interaction effects, $F\text{s} < 2.20$

---

[1] We checked, and removing those students did not affect the outcomes of the analyses on which data from these students was available.

[2] We reran these analyses with non-parametric Mann-Whitney *U* tests, because of the large standard deviations and violations of the normality assumption. We again found a significant main effect of Gender Observer ($U = 2405.50$, $z = -2.35$, $p = 0.019$) but no main effect of Gender Model ($U = 2934.00$, $z = -0.52$, $p = 0.602$).

Table 1
Means (SD) of pretest performance (range 0–10), posttest performance (range 0–8), self-efficacy (range 1–9), perceived competence (range 1–7), invested mental effort (range 1–9), learning enjoyment (range 1–9), and perceived explanation quality (range 1–9) per condition.

	Female Model		Male Model	
	Female Student	Male Student	Female Student	Male Student
Pretest Performance	2.38 (1.38)	3.24 (1.42)	2.61 (1.43)	3.26 (1.62)
Posttest Performance	2.63 (2.82)	3.65 (2.75)	2.44 (2.74)	3.30 (2.69)
Self-efficacy Before Learning	2.67 (1.44)	4.00 (2.06)	2.88 (1.86)	3.84 (2.30)
Self-efficacy After Learning	4.14 (1.98)	5.08 (1.99)	4.24 (2.31)	4.68 (2.55)
Perceived Competence Before Learning	2.57 (1.33)	3.43 (1.48)	2.84 (1.51)	3.13 (1.89)
Perceived Competence After Learning	3.39 (1.60)	3.98 (1.53)	3.46 (1.80)	3.79 (1.86)
Invested Mental Effort in Example Study	4.88 (2.19)	4.74 (2.09)	5.12 (2.20)	4.78 (2.27)
Invested Mental Effort in Posttest	6.00 (2.34)	4.96 (2.65)	6.23 (2.36)	5.24 (2.71)
Learning Enjoyment	2.74 (1.54)	2.78 (1.81)	2.80 (1.54)	2.89 (1.78)
Perceived Explanation Quality	5.40 (1.22)	5.17 (1.54)	5.46 (1.43)	5.38 (1.16)

(i.e., *ps* > .15). There was a main effect of Test Moment, *F* (1, 153) = 89.95, $p < .001$, $\eta_p^2 = 0.370$, showing that students' self-efficacy increased from before ($M = 3.27$, $SD = 2.05$) to after learning ($M = 4.51$, $SD = 2.23$). There was a main effect of Gender Observer, *F* (1, 153) = 7.90, $p = 0.006$, $\eta_p^2 = 0.049$, indicating that male students showed higher self-efficacy than female students, which is in line with their performance (see section 3.1). There was no main effect of Model Gender, $F < 1$.

Perceived competence showed the same pattern of results: there were no significant interaction effects and there was no main effect of Model Gender, *Fs* < 1, but students' perceived competence increased from before ($M = 2.97$, $SD = 1.58$) to after learning ($M = 3.64$, $SD = 1.71$), as indicated by a main effect of Test Moment, *F* (1, 153) = 56.20, $p < .001$, $\eta_p^2 = 0.269$. Moreover, in line with their higher test performance (section 3.1) male students showed higher perceived competencethan female students, *F* (1, 153) = 4.41, $p = 0.037$, $\eta_p^2 = 0.028$.

3.3. *Invested mental effort*

The analysis of mental effort invested in example study showed no significant interaction effect between Model Gender and Gender Observer, nor main effects of Model Gender or Gender Observer, *Fs* < 1. As for the reported effort investment in the posttest tasks, there was no significant interaction effect, $F < 1$, but there was a main effect of Gender Observer, *F* (1, 153) = 6.39, $p = 0.012$, $\eta_p^2 = 0.040$, with male students ($M = 5.10$, $SD = 2.67$) reporting to have invested less mental effort than female students ($M = 6.11$, $SD = 2.34$), which is again in line with their higher performance (section 3.1). There was no main effect of Model Gender, $F < 1$.

3.4. *Learning enjoyment*

The analysis on students' learning enjoyment showed no significant interaction effect and no significant main effect of Gender Observer or Model Gender, *Fs* < 1.

3.5. *Perceived explanation quality*

There were no interaction or main effects in students' ratings of the quality of the model's explanation in the videos (all *Fs* < 1). [3]

4. Discussion

The question addressed in this experiment was whether the benefits of video modeling examples would depend on the gender of the observer and the model. More specifically, we tested the model-observer similarity (MOS) hypothesis, which would predict greater benefits of watching a similar-gender model on learning and self-efficacy and perceived competence compared to a opposite-gender model for male and female students, and the task-appropriateness hypothesis, which would predict greater benefits of watching a male model than a female model for male and female students when the learning task is viewed as more appropriate for men than women.

Students perceptions indeed allowed us to test both the MOS and task-appropriateness hypothesis, as a manipulation check confirmed that male and female students perceived same-gender models as more similar to themselves than opposite-gender models, and perceived the task of troubleshooting electrical circuits as more appropriate for males than females. In contrast to both hypotheses, however, the model's gender did not significantly affect students' test performance, self-efficacy, or perceived competence. The model's gender also did not affect students' perceived effort investment in example study or the posttest tasks, learning enjoyment, or the perceived quality of the explanations provided in the videos. Thus, our findings do not provide support for either the MOS or the task-appropriateness hypothesis, and instead suggest that there would be no need to take the model's gender into account either when designing video modeling examples, even when the task is perceived as being more appropriate for males than females. These null-findings are in line with some prior studies that investigated whether the benefits of video modeling examples would depend on the model's gender (e.g., George, Feltz, & Chase, 1992; Schunk & Hanson, 1985; Schunk et al., 1987; Experiment 2), but not with other studies that did find that model gender mattered (e.g., for perceived competence, perceived effort investment, and learning enjoyment: Hoogerheide, Van Wermeskerken et al., 2016; for imitative behavior: Bandura et al., 1963; Hicks, 1965; for self-efficacy: Weeks et al., 2005).

One possible explanation for why we found no effects of model gender but several other studies did (e.g., Bandura et al., 1963; Hoogerheide, Van Wermeskerken et al., 2016; Maccoby & Wilson, 1957; Weeks et al., 2005; see also; Garcia-Rodicio, 2012) might lie in the number of models used. That is, in the present study we used two male and two female models to enhance the likelihood that the

---

[3] Because of the pre-existing differences in prior knowledge between male and female students, we used independent-samples *t*-tests to check whether the results would change if we test the effect of Model Gender on self-efficacy, perceived competence, invested mental effort, learning enjoyment, and explanation quality for male and female students separately. The conclusions remained the same, *ps* > .050. Moreover, we explored whether the results change if we would rerun all analyses only including those students who indicated that they perceived the task of troubleshooting electrical circuits as more appropriate for males than females ($n = 127$). Results again showed no significant interaction effects between Model Gender and Gender Observer or main effects of Model Gender.

findings could be attributed to the manipulated model characteristic (i.e., gender) rather than to individual model characteristics (e.g., attractiveness, speech). Yet, previous studies that found effects of model gender almost without exception used only one male and one female model, which makes it difficult to rule out effects of individual model characteristics.

Another possible explanation might be that the importance of model gender is moderated by learners' age and gender-related developmental processes. Children typically develop a clear conception of what it means to be a boy or a girl between the ages three to five (Martin & Halverson, 1981). Around that age, the model's gender may start to matter, as children start actively searching and positively valuing those behaviors that are congruent with the acquired concept of gender (Bussey & Bandura, 1992), and almost exclusively socializing with same-gender peers (Maccoby, 1998). The importance of model gender may drop again during the early teenage years however, as teenagers develop their own identity inspite of their own and others' gender (Steensma, Kreukels, de Vries, & Cohen-Kettenis, 2013) and start befriending opposite-gender peers (McDougall & Hymel, 2007). In line with this explanation, prior research consistently found model gender effects on learning for children ages four to six (e.g., Bandura et al., 1963; Bussey & Bandura, 1984; Experiment 1; Hicks, 1965), yet showed mixed effects when using a sample of early teenagers (e.g., Schunk & Hanson, 1985; Schunk et al., 1987), adults (e.g., George et al., 1992), or even elderly (Weeks et al., 2005). Additionally, whether or not model gender is important for other age groups may depend on situational interest. It is imaginable that model gender only affects learning for those whose interest is low, as those who are very interested in the topic might continue to pay attention to the video modeling example regardless of who their model is.

Although the MOS and task-appropriateness hypothesis do not seem to account for how much male and female students benefit from observing a male or female model in video modeling examples, at least in an adolescent sample, other model characteristics might matter. Actually, Hoogerheide, Loyens et al. (2016) used the same materials and measures as in the present study and found that secondary education students (around 15 years of age) who studied adult models reported lower effort investment in example study and attained greater learning outcomes than students who studied peer models (while using multiple models per condition and keeping the example content identical). Some model characteristics might be more relevant for learning processes and outcomes than others (Renkl, 2017). Therefore, it would be fruitful for future research to test the MOS and task-appropriateness hypothesis with other model characteristics, such as age and expertise.

Concerning observer gender, we found that male students showed higher self-efficacy, perceived competence, and test performance than female students, and reported lower effort investment in completing the posttest tasks. These findings are in line with prior example-based learning studies in which physics materials were used (Hoogerheide, Loyens, Jadi, Vrins, & Van Gog, 2017; Hoogerheide, Loyens et al., 2016) and likely reflect a gender gap in physics at the secondary education level, where females tend to perform worse than males (Hazari, Sonnert, Sadler, & Shanahan, 2010; Lorenzo, Hirschfield Crouch, & Mazur, 2006).

Strengths of our study were that we used multiple models per condition, kept the example content identical, and measured students' perceptions of task-appropriateness and similarity. A potential limitation of our study is that learners' perceived similarity to the model was rather low, which might have made it more difficult for students to identify with the model. This limitation might apply to many MOS studies, yet often go unnoticed because in only a handful studies, similarity perceptions were actually measured. Indeed, of the few studies in which similarity perceptions were measured, several found low levels of perceived similarity (e.g., George et al., 1992; Schunk et al., 1987) – and in some studies perceived similarity ratings did not differ significantly between a similar-gender and an opposite-gender model (e.g., George et al., 1992; Schunk & Hanson, 1985; Schunk et al., 1987). Students' perceived similarity to the model might have been greater, if we had used coping models whose demonstration contain errors that get corrected later on rather than mastery models providing an ideal account of how to solve the problem. Findings of Schunk and Hanson (1985) and Schunk et al. (1987) indeed suggest that novice learners perceive coping models as more similar to themselves than mastery models, but considering that neither study found effects of model gender on learning outcomes or self-efficacy, it seems unlikely that the use of coping models would have changed our findings.

An interesting avenue for future research would be to investigate whether there would be no need to take the model's gender into account either when the learning task is perceived as more appropriate for females rather than for males. Because of personality differences between boys and girls (e.g., girls score higher than boys on the personality traits agreeableness and neuroticism; Costa, Terracciano, & McCrae, 2001), it is possible that boys may be less willing to pay attention to and therefore learn less from opposite-gender models than girls when the learning task is perceived as more appropriate for the other gender. A related suggestion is that it would be interesting to measure how much attention students pay to the models and how well they follow the model's explanation by using eye tracking (cf. Van Wermeskerken, Ravensbergen, & Van Gog, this issue; Van Wermeskerken & Van Gog, 2017). Also, it would be interesting to investigate whether model gender would affect test performance, self-efficacy, and perceived competence, when students are able to choose their own model from a selection of models that differ in terms of gender and attractiveness. Findings in the animated agent literature suggest that students' attention to agents and therefore learning may vary depending on perceived attractiveness to the agent (Moreno & Flowerday, 2006) and whether or not students were provided a choice of agent (Ozogul, Johnson, Atkinson, & Reisslein, 2013).

Overall, our findings contribute toward an understanding of whether and when model characteristics affect how much students benefit from video modeling examples. This question garnered attention in early research on modeling (Schunk, 1987) and recently regained interest now that video modeling examples have become ubiquitous in formal and informal learning settings. The model's gender does not seem to be an important factor to take into account when designing video modeling examples, even when the task is perceived as being more appropriate for males than females, at least for adolescent learners learning cognitive skills. Future research will have to uncover whether different results would emerge with a younger student population.

References

Bandura, A. (1971). Analysis of modeling processes. In A. Bandura (Ed.), *Psychological modeling* (pp. 1–62). Chicago: Aldine-Atherton.

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.), *Vol. 4. Encyclopedia of human behavior* (pp. 71–81). New York: Academic Press.

Bandura, A., Ross, D., & Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology, 67*, 601–607. https://doi.org/10.1037/h0045550.

Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2017). Look into my eyes! Exploring the effect of addressing in educational videos. *Learning and Instruction, 49*, 113–120. https://doi.org/10.1016/j.learninstruc.2017.01.004.

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.

Bussey, K., & Bandura, A. (1984). Influence of gender constancy and social power on sex-linked modeling. *Journal of Personality and Social Psychology, 47*, 1292–1302.

Bussey, K., & Bandura, A. (1992). Self-regulatory mechanisms governing gender development. *Child Development, 63*, 1236–1250.

Castro-Alonso, K. C., Ayres, P., & Paas, F. (2014). Learning from observing hands in static and animated versions of non-manipulative tasks. *Learning and Instruction, 34*, 11–21. https://doi.org/10.1016/j.learninstruc.2014.07.005.

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322–331. https://doi.org/10.1037//0022-3514.81.2.322.

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

Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology, 108*, 528–546. https://doi.org/10.1037/edu0000065.

Fiorella, L., Van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology, 109*, 653–665.

Garcia-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.

George, T., Feltz, D., & Chase, M. (1992). Effects of model similarity on self-efficacy and muscular endurance: A second look. *Journal of Sport & Exercise Psychology, 14*, 237–248.

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching, 8*, 978–1003. https://doi.org/10.1002/tea.20363.

Hicks, D. J. (1965). Imitation and retention of film-mediated aggressive peer and adult models. *Journal of Personality and Social Psychology, 2*, 97–100. https://doi.org/10.1037/h0022075.

Hoogerheide, V., Loyens, S. M. M., Jadi, F., Vrins, A., & Van Gog, T. (2017). Testing the model-observer similarity hypothesis with text-based worked examples. *Educational Psychology, 27*, 121–127. https://doi.org/10.1080/01443410.2015.1109609.

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. https://doi.org/10.1016/j.chb.2014.09.013.

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Does gender matter? *Instructional Science, 44*, 69–86.

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. https://doi.org/10.1016/j.learninstruc.2016.02.004.

Johnson, C. S., & Lammers, J. (2012). The powerful disregard social comparison information. *Journal of Experimental Social Psychology, 48*, 329–334. https://doi.org/10.1016/j.jesp.2011.10.010.

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. https://doi.org/10.1007/s11251-009-9115-8.

Lorenzo, M., Hirschfield Crouch, C., & Mazur, E. (2006). Reducing the gender gap in the physics. *American Journal Of Physics, 74*, 118–122.

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.

Maccoby, E. E., & Wilson, W. C. (1957). Identification and observational learning from films. *Journal of Abnormal and Social Psychology, 55*, 76–87.

Marchand, G. C., & Taasoobshirazi, G. (2013). Stereotype threat and women's performance in physics. *International Journal of Science Education, 35*, 3050–3061. https://doi.org/10.1080/09500693.2012.683461.

Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development, 52*, 1119–1134.

McDougall, P., & Hymel, S. (2007). Same-gender versus cross-gender friendship conceptions: Similar or different? *Merrill-Palmer Quarterly, 53*, 347–380.

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. https://doi.org/10.1016/j.cedpsych.2005.05.002.

Ouwehand, K., Van Gog, T., & Paas, F. (2015). Designing effective video-based modeling examples using gaze and gesture cues. *Educational Technology & Society, 18*, 78–88.

Ozogul, G., Johnson, A. M., Atkinson, R. A., & Reisslein, M. (2013). Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education, 67*, 36–50. https://doi.org/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. https://doi.org/10.1037/0022-0663.84.4.429.

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

Renkl, A. (2017). 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 (in press).

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. https://doi.org/10.1080/02701367.2014.961050.

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research, 57*, 149–174. https://doi.org/10.2307/1170234.

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*, 313–322. https://doi.org/10.1037//0022-0663.77.3.313.

Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*, 54–61. https://doi.org/10.1037/0022-0663.79.1.54.

Steensma, T. D., Kreukels, B. P. C., de Vries, A. L. C., & Cohen-Kettenis, P. T. (2013). Gender identity development in adolescence. *Hormones and Behavior, 64*, 288–297. https://doi.org/10.1016/j.yhbeh.2013.02.020.

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. https://doi.org/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. https://doi.org/10.1080/00461520701756248.

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*, 155–174. https://doi.org/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. https://doi.org/10.1016/j.compedu.2013.12.004.

Van Wermeskerken, M., Ravensbergen, S. J., & Van Gog, T.. (under review). What are the odds? Effects of seeing an instructor in video-modeling examples on probability calculation on attention and learning. Computers in Human Behavior.

Van Wermeskerken, M., & Van Gog, T. (2017). Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education, 113*, 98–107. https://doi.org/10.1016/j.compedu.2017.05.013.

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. https://doi.org/10.1016/j.compedu.2014.06.003.

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71*, 79–89. https://doi.org/10.1016/j.chb.2017.01.049.

Weeks, D. L., Borrousch, S., Bowen, A., Hepler, L., Osterfoss, M., Sandau, A., et al. (2005). The influence of age and gender of an exercise model on self-efficacy and quality of therapeutic exercise performance in the elderly. *Physiotherapy Theory and Practice, 21*, 137–146. https://doi.org/10.1080/09593980500213902.

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. https://doi.org/10.1037/0022-3514.70.4.76.

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. https://doi.org/10.1016/S1071-5819(03)00114-9.