Educational Psychology
An International Journal of Experimental Educational Psychology

ISSN: 0144-3410 (Print) 1469-5820 (Online) Journal homepage: http://www.tandfonline.com/loi/cedp20

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

Vincent Hoogerheide, Sofie M.M. Loyens, Fedora Jadi, Anna Vrins & Tamara van Gog

EDUCATIONAL PSYCHOLOGY, 2017
VOL. 37, NO. 2, 112–127
http://dx.doi.org/10.1080/01443410.2015.1109609

Routledge
Taylor & Francis Group

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

Vincent Hoogerheide[a], Sofie M.M. Loyens[a,b], Fedora Jadi[a], Anna Vrins[a] and Tamara van Gog[a,c]

[a]Institute of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands; [b]Roosevelt Center for Excellence in Education, University College Roosevelt, Middelburg, The Netherlands; [c]Department of Education, Utrecht University, Utrecht, The Netherlands

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) modelling examples in which an instructor (the 'model') provides a demonstration. The model-observer similarity (MOS) hypothesis predicts that the effectiveness of modelling 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.

ARTICLE HISTORY
Received 14 July 2015
Accepted 14 October 2015

KEYWORDS
Example-based learning; observational learning; model-observer similarity; gender; expertise; worked examples

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 *modelling 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 & 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 modelling examples, design choices concern, for instance, which

CONTACT Vincent Hoogerheide hoogerheide@fsw.eur.nl
The experiments were conducted in several secondary education schools in The Netherlands.

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 modelling depends in part on the degree to which observers perceive a model to be similar to them (Schunk, 1987). Modelling 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 modelling examples have emerged, such as video modelling 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 visualisations (i.e. video, animation) accompanied by a male or female voice-over. Surprisingly, in view of the idea that the gender appropriateness of the behaviour 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 Tai (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 modelled task (e.g. Schunk & Hanson, 1985). When weaker students have no prior experience with the modelled 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 favourable 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 optimise 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 post-test 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 et al., 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 × 2 design, with Gender Observer (Male vs. Female) and Gender Model (Male vs. Female) as between-subject factors. 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 was 128. Participants were 147 secondary education students ($M^{age} = 14.43$, $SD = .58$; 74 male) in their third year of pre-university education. 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. 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. $U = R \times I$; $R = U/I$; $I = U/R$) 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 1 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.

Post-test

The post-test 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 nine-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, Loyens, and van Gog (2014) 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.

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

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 1–7 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 post-test 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 post-test, 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 post-test, 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.

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 analysed using 2 × 2 ANOVAs, with Gender Observer (Female, Male) and Gender Model (Female, Male) as between-subject factors, unless otherwise specified. Interactions were further analysed 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 post-test 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 neither a 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 post-test 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.

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 (range 0–10)	3.28 (1.43)	2.69 (1.29)	2.47 (1.47)	2.39 (1.35)
Performance post-test (range 0–8)	4.77 (2.37)	4.53 (2.54)	3.45 (2.31)	3.86 (2.44)
Mental effort study phase (range: 1–9)	3.16 (1.71)	2.85 (1.47)	3.55 (1.75)	3.72 (1.69)
Mental effort post-test (range 1–9)	3.88 (1.89)	4.01 (1.92)	4.77 (2.15)	5.22 (1.62)
Self-efficacy pretest (range 1–9)	5.41 (1.79)	5.70 (1.66)	5.17 (1.49)	4.71 (1.76)
Self-efficacy post-test (range 1–9)	6.30 (1.79)	6.51 (1.64)	5.62 (1.42)	5.39 (1.68)
Perceived competence pretest (range 1–7)	5.38 (1.28)	5.37 (1.22)	4.97 (1.15)	4.47 (1.23)
Perceived competence post-test (range 1–7)	5.63 (1.42)	5.61 (1.18)	5.05 (1.23)	4.65 (1.52)

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 (range 1–9)	6.73 (1.35)	6.78 (1.18)	6.28 (1.50)	6.75 (1.11)
Perceived attractiveness (range 1–7)	3.89 (1.58)	3.41 (1.91)	3.70 (1.75)	4.56 (1.50)
Perceived friendliness (range 1–7)	5.32 (1.13)	5.42 (1.18)	5.00 (1.53)	5.56 (1.30)
Perceived intelligence (range 1–7)	5.22 (1.42)	5.28 (1.30)	4.95 (1.49)	5.69 (1.04)

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 neither a main effect of Gender Model nor an interaction effect, both Fs < 1.

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: $F(1,142) = 2.45$, $p = .119$, $\eta_p^2 = .017$).

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 × 2 design, with Gender Observer (Male vs. Female) and Model Expertise (Peer vs. Teacher) as between-subject factors. Participants were 130 secondary education students ($M^{age} = 14.63$, $SD = .68$; 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.

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,

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

Tan, & Jayothisa, 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 × 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, or an interaction, both Fs < 1. There were no significant effects on post-test performance (Gender Observer: $F(1,126) = 2.88$, $p = .092$, $\eta_p^2 = .022$; Model Expertise and interaction: Fs < 1).

Mental effort

There were no significant effects on mental effort during example study (Gender Observer and Model Expertise: Fs < 1; interaction, $F(1,126) = 3.32$, $p = .071$, $\eta_p^2 = .026$) or mental effort invested in completing

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 (range 0–10)	2.28 (1.24)	2.31 (1.27)	1.76 (1.07)	1.45 (1.15)
Performance post-test (range 0–8)	3.62 (2.54)	4.24 (2.39)	4.55 (2.42)	4.80 (2.57)
Mental effort study phase (range: 1–9)	4.00 (1.94)	3.42 (1.66)	3.50 (1.73)	4.02 (1.43)
Mental effort post-test (range 1–9)	4.46 (2.40)	4.29 (2.05)	5.02 (2.09)	4.77 (1.66)
Self-efficacy pretest (range 1–9)	5.08 (1.91)	5.81 (1.74)	5.09 (1.48)	4.91 (1.04)
Self-efficacy post-test (range 1–9)	5.59 (2.08)	6.43 (1.46)	5.59 (1.84)	5.47 (1.17)
Perceived competence pretest (range 1–7)	4.92 (1.52)	5.63 (.97)	4.57 (1.44)	4.42 (1.16)
Perceived competence post-test (range 1–7)	5.03 (1.33)	5.47 (1.11)	4.87 (1.39)	4.83 (1.21)

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 (range 0–10)	4.89 (2.86)	5.66 (2.33)	4.59 (2.24)	3.87 (2.18)
Example quality (range 1–9)	6.33 (2.06)	6.60 (1.31)	6.72 (1.22)	6.57 (1.25)
Perceived attractiveness (range 1–7)	3.67 (1.90)	4.29 (1.95)	3.72 (1.25)	3.03 (1.38)
Perceived friendliness (range 1–7)	5.53 (1.13)	5.17 (1.69)	5.31 (.89)	4.87 (1.43)
Perceived intelligence (range 1–7)	5.26 (1.80)	5.46 (1.58)	5.41 (1.09)	5.60 (1.13)

the post-test tasks (Gender Observer, $F(1,126) = 2.00$, $p = .159$, $\eta^2_p = .016$; Model Expertise and interaction, Fs < 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: $F(1,91) = 1.86$, $p = .178$, $\eta^2_p = .020$; Model Expertise: $F < 1$; interaction: $F(1,91) = 1.91$, $p = .170$, $\eta^2_p = .021$), or after the acquisition phase (Gender Observer: $F(1,124) = 2.62$, $p = .108$, $\eta^2_p = .021$; Model Expertise: $F(1,124) = 1.47$, $p = .228$, $\eta^2_p = .012$; interaction: $F(1,124) = 2.60$, $p = .109$, $\eta^2_p = .021$).

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, $F(1,126) = 11.74$, $p < .001$, $\eta^2_p = .085$, 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, $F(1,126) = 1.46$, $p = .228$, $\eta^2_p = .011$, nor an interaction effect, $F(1,126) = 3.54$, $p = .062$, $\eta^2_p = .027$. On perceived competence after the acquisition phase, there was no longer a main effect of Gender Observer, $F(1,125) = 3.14$, $p = .079$, $\eta^2_p = .025$, and there was no effect of Gender Model, $F < 1$, nor an interaction, $F(1,125) = 1.16$, $p = .284$, $\eta^2_p = .009$.

Lesson enjoyment, example quality and model evaluation

The analysis of lesson enjoyment scores showed a main effect of Gender Observer, $F(1,125) = 5.91$, $p = .017$, $\eta^2_p = .045$, 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, $F(1,125) = 3.01$, $p = .085$, $\eta^2_p = .023$. There were no effects on students' perceptions of the quality of the examples (all Fs < 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).

Discussion

As in Experiment 1, there were no effects of MOS on post-test 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 post-test 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 modelling 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 et al., 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 post-test performance when assessed straight after the learning phase, whereas post-test 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 optimise 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.

Acknowledgements

The authors would like to thank Roel Jansen, Norbert van Veen, Mylene Vossen and Wim van der Aalst for facilitating this study, and Dr Peter Verkoeijen and Dr Samantha Bouwmeester for their advice regarding statistical analyses.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Arroyo, I., Woolf, B. P., Royer, J. M., & Tai, M. (2009). Affective gendered learning companion. In *Proceedings of the international conference on artificial intelligence and education* (pp. 41–48). Brighton: IOS Press.

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). New York, NY: Cambridge University Press.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

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. *The Journal of Abnormal and Social Psychology, 67*, 601–607. doi:10.1037/h0045550

Baylor, A. L. & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. *Lecture Notes in Computer Science: Intelligent Tutoring Systems, 3220*, 592–603. doi:10.1007/978-3-540-30139-4_56

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.

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

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. *British Journal of Educational Psychology, 80*, 557–566. doi:10.1348/000709910X497130

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

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

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

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

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

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). New Haven, CT: Yale University Press.

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

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

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.

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.

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: Emerald Group Publishing Limited.

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

Lachner, A., & Nückles, M. (2015). Bothered by abstractness or engaged by cohesion? Experts' explanations enhance novices' deep-learning. *Journal of Experimental Psychology: Applied, 21*, 101–115. doi:10.1037/xap0000038

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

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.

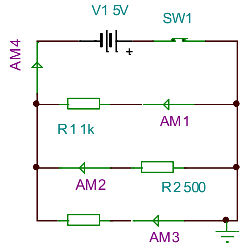
Linek, S. B., Gerjets, P., & Scheiter, K. (2010). The speaker/gender effect: Does the speaker's gender matter when presenting auditory text in multimedia messages? *Instructional Science, 38*, 503–521. doi:10.1007/s11251-009-9115-8

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, NY: Cambridge University Press.

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.
Moreno, R., & Flowerday, T. (2006). Students' 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
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
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
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*, 1–37. doi:10.1111/cogs.12086
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.
Rosenberg-Kima, R., Baylor, A., Plant, A., & Doerr, C. (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
Schunk, D. (1987). Peer models and children's behavioral change. *Review of Educational Research, 57*, 149–174. doi:10.3102/00346543057002149
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
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
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.
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
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
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., & 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
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
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

Appendix 1

Example 1 (Maarten, 15, 4VWO)



1. Determine how this circuit <u>should</u> 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: $I_t = I_1 + I_2 + I_3$

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

This means you should measure:

AM1 = 5 *mA* AM2 = 10 *mA* AM3 = 50 *mA* AM4 = 65 *mA*

2. Suppose the ammeters indicate the following measurements:

AM1 = 5 *mA* AM2 = 7.14 *mA* AM3 = 50 *mA* AM4 = 62.14 *mA*

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 = 7.14 mA instead of 10 mA. Thus, R_2 has a higher resistance than the indicated 500Ω. The actual resistance of R_2 can be calculated using the <u>measured</u> current:

$$R_2 = \frac{U}{I_2} = \frac{5\text{ V}}{7.14\text{ mA}} = 0.7\text{ k}\Omega = 700\ \Omega$$