A Gender Matching Effect in Learning with Pedagogical Agents

in an Immersive Virtual Reality Science Simulation

*Guido Makransky[1], *Philip Wismer[2], Richard E. Mayer[3]

[1] Department of Psychology, University of Copenhagen, Copenhagen, Denmark

[2] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark,

Lyngby, Denmark

[3]Department of Psychological and Brain Sciences, University of California Santa Barbara, CA

* The first and the second author contributed equally to this study.

Abstract

The main objective of this study is to determine whether boys and girls learn better when the characteristics of the pedagogical agent are matched to the gender of the learner while learning in immersive virtual reality. Sixty-six middle school students (33 females) were randomly assigned to learn about laboratory safety with one of two pedagogical agents: Marie or a drone, who we predicted serve as a role models for females and males respectively. The results indicated that there were significant interactions for the dependent variables of performance during learning, retention, and transfer, with girls performing better with Marie ($d = 0.98$, $d = 0.67$, and $d = 1.03$; for performance, retention, and transfer respectively), and boys performing better with the drone ($d = -0.41$, $d = -0.45$, $d = -0.23$, respectively). The results suggest that gender specific design of pedagogical agents may play an important role in VR learning environments.

Key words: pedagogical agents, immersive virtual reaility, multimedia learning, social agency theory, virtual learning

1. Introduction

1.1 Objective and Rationale

The goal of this study is to determine how to create online pedagogical agents that are effective for learning in immersive virtual reality (VR). Specifically, we are interested in whether boys and girls learn better in immersive VR when the characteristics of the onscreen pedagogical agent is matched to the gender of the learner.

An onscreen pedagogical agent is a character rendered on a screen who is intended to facilitate learning of the presented material (Kim & Baylor, 2006; Moreno, Mayer, Spires, & Lester, 2001; Veletsianos & Russell, 2014). The character can be presented as an animation of a cartoon creature or a video of a human. The representation can be displayed on a desktop computer, laptop computer, tablet, smartphone, or head-mounted display (HMD) in virtual reality (VR) or augmented reality. The instructional material can cover any topic, with the goal that the onscreen pedagogical agent is designed to help students learn. In this study, we focus on an animation displayed on a HMD in VR on the topic of laboratory safety.

Virtual reality (VR) is an artificial environment that projects the user into a 3D generated space (Blascovich & Bailenson, 2011). Immersion can be regarded as an objective measure of the extent to which the VR system presents a vivid virtual environment while shutting out the physical world (Cummings and Bailenson, 2016). Therefore, the term immersive VR is typically used in the literature to refer to VR administered through a HMD. Immersive VR is increasingly being used in education due to heavy investment by large technology companies which has made the technology increasingly affordable, and a recent report predicts that VR and related technologies could reach 15 million learners by 2025 (Goldman Sachs, 2018). The main affordance of using immersive VR for learning is that the high level of immersion leads to a

higher sense of presence (Makransky & Lilleholt, 2018; Makransky et al., 2017; Parong & Mayer, 2018), which is the subjective sensation of "being there" in the virtual environment (Lee, 2004). Although higher presence does not in itself lead to better learning (Makransky, Mayer, & Terkildsen, 2017; Moreno & Mayer, 2002; Parong & Mayer, 2018), previous research suggests that certain instructional design principles may be particularly relevant for immersive VR (e.g., Makransky, Mayer, & Terkildsen, 2017).

The practical rationale for focusing on learning with onscreen pedagogical agents is that instructional content is increasingly being delivered in the form of computer-based lessons with onscreen agents who explain or model for the learner, including animated pedagogical agents in online lessons, animated pedagogical agents for learning in immersive virtual reality, and human agents in instructional video. The theoretical rationale is to better understand the conditions under which people will accept computers as social partners (Reeves & Nass, 1996). We focus on the role of onscreen pedagogical agents in immersive VR because this venue is not as well studied as some others; thus, we can increase the domain of study.

1.2 Literature Review

Although the instructional effectiveness of onscreen pedagogical agents has been a topic of interest for the past 20 years (Cassell, Sullivan, Prevost, & Churchill, 2000; Dehn & Mulken, 2000; Heidig & Clarebout, 2010; Johnson & Rickel, 2000; Johnson & Lester, 2016; Mayer & DaPra, 2012; Moreno, Mayer, Spires, & Lester, 2001; Schroeder & Adesope, 2013; Schroeder, Adesope, & Gilbert, 2013; Veletsianos & Russell, 2014; Wang, Li, Mayer, & Liu, 2018), an important remaining issue concerns how best to render the basic characteristics of the agent, such as gender, ethnicity, and age (Hoogerheide, van Wermeskerken, van Nassau, & van Gog, in press; Hoogerheide, Loyens, & van Gog, 2016; Kim & Baylor, 2006; Moreno & Flowerday,

2006; Ozogul, Johnson, Atkinson, & Reisslein, 2013; Rosenberg-Kima, Plant, Doerr, & Baylor, 2010, Baylor & Kim, 2004). Kim & Lim (2013) found that learner gender was a significant factor in the learner's evaluations of a pedagogical agent with females holding more positive attitudes toward agents. However, with regard to the gender of the agent, there is some evidence in the literature that students learn better in STEM subjects with male agents than female agents regardless of the students' demographics. For instance, Johnson (2013b) found that a female agent only improved learning outcomes for low performing students in an engineering simulation, while multiple studies reported that male agents improve learning benefits for all students or outperform female agents in similar contexts (e.g., Kim, Baylor, & Shen 2007; Moreno et al., 2010; Ozogul et al., 2011, Exp. 1). This suggests that students might hold stereotypical views about the agent's gender, i.e. that male agents are more competent STEM teachers. According to this hypothesis students learn best when the characteristics of the agent matches the stereotype (Johnson, 2013b).

An alternative view is that students learn best when the characteristics of the agent are instead matched to the gender of the student, which we refer to as the *gender matching hypothesis*. Although some studies have found that students reported that they preferred agents that are similar to themselves, such as preferring an agent with the same gender (Johnson 2013a), studies generally have failed to find support for the gender matching hypothesis both with instructional video (Hoogerheide, Loyens, and van Gog, 2016; Hoogerheide, van Mermeskerken, van Nassau, & van Gog, in press) and animated agents (Johnson et al., 2013a; Moreno & Flowerday, 2006; Ozogul, Johnson, Atkinson, & Reisslein, 2013). That is, when faced with a female agent, no differences in learning outcomes were found between male and female

participants, even though female participants generally have more positive attitudes towards

pedagogical agents (Johnson et al., 2013b; Kim & Lim, 2013).

Should we give up on the gender matching hypothesis? Part of the problem with the

foregoing studies is that even though gender was varied, the opposite gender agent still may have

displayed appealing characteristics that all students would perceive as similar to themselves,

such as their age, ethnicity, or the way they dressed. Furthermore, previous studies have been

primarily conducted with college students, who have been found to have a lower preference for

agents that match their own gender, as opposed to younger participants (Johnson et al., 2013b).

In the present study, we seek to further test the gender matching hypothesis by designing an

agent that was intended to appeal to girls rather than boys--a young woman in a white lab coat

named Marie who could serve as a role model--and an agent that we predict to appeal more to

boys rather than girls--a hovering robot we called the drone. Such non-traditional, mentor-like

role models have previously been shown to enhance students' concentration and focus, as well as

their transfer and self-efficacy scores (Baylor & Kim, 2004; Johnson 2013b; Moreno et al., 2002;

Thisgaard & Makransky, 2017). We also extended the domain of inquiry beyond instructional

video with human agents and desktop animated agents to instruction in immersive VR, which is

intended to be a more engaging context of learning (e.g., Makransky & Lilleholt, 2018).

In summary, a major gap in the existing literature on the role of pedagogical agents'

gender is that students typically learned with pedagogical agents who displayed appealing

features that both boys and girls could identify with, rather than with pedagogical agents

specifically designed for mainly one gender to identify with. The present study fills that gap by

comparing learning outcomes by boys and girls who learn with pedagogical agents that exhibit

characteristics designed to seem appealing specifically to boys more than to girls (e.g., a robot-like drone) or girls more than to boys (e.g., a young female scientist).

1.3 Theory and Predictions

The matching hypothesis is that students learn better with onscreen pedagogical agents that they can identify with. In the present experiment, we examine a specialized version of the matching hypothesis, the gender matching hypothesis, which posits that girls will learn better with an onscreen pedagogical agent that has characteristics they can identify with, and boys will learn better with a pedagogical agent that has characteristics they can identify with. In this study we intend to address a number of gaps in the literature related to the choice of pedagogical agents when learning about STEM. In particular, we created an onscreen pedagogical agent that was intended to appeal to girls--Marie, who is rendered as a young, female scientist in a white lab coat (as shown in the right panel of Figure 1)--and an onscreen agent that we predict to appeal more to boys—a drone, which is rendered as a futuristic, hovering robot (as shown in the left panel of Figure 1). The reason we expect the drone to appeal to boys is that it resembles agents from modern computer games (e.g., Higs from *Robinson: The Journey* or Wheatley from *Portal 2*) and exhibits superhero characteristics that boys tend to identify with, such the ability to fly around. We inserted the onscreen agents in an instructional VR simulation aimed at teaching middle school students about lab safety because most previous research in this field has been done with university students. Furthermore, we use immersive VR in order to increase the psychological fidelity and social presence of interacting with a pedagogical agent when learning about STEM. According to the gender matching hypothesis, we predict that girls will learn better with Marie than the drone as the pedagogical agent, whereas boys will learn better with the drone than with Marie as a pedagogical agent.

Overall, the gender matching hypothesis yields three specific predictions concerning each of three primary dependent variables in the experiment. First, we measured problem-solving performance during the learning phase of the experiment. Hypothesis 1 is that girls will score higher on learning performance with Marie than with the drone whereas boys will score higher on learning performance with the drone than with Marie. Second, we measured learning outcomes with a knowledge test--covering the basic information in the lesson--and with a transfer test--which required students to apply what they had learned to new situations. Hypothesis 2 is that girls will score higher on the knowledge test if they had learned with Marie rather than the drone, whereas boys will score higher on the knowledge test if they had learned with the drone rather than Marie. Hypothesis 3 is that girls will score higher on the transfer test if they had learned with Marie rather than the drone, whereas boys will score higher on the transfer test if they had learned with the drone rather than Marie.

The matching hypothesis is consistent with social agency theory (Mayer, 2014), which focuses on the impact of social cues in instructional messages. According to social agency theory, the first link is that social cues in an instructional message (such as an onscreen pedagogical agent you can identify with) can prime a social response in the learner (such as feeling that the instructor is a social partner). The next link is that when students see the instructor as a social partner and feel as if they are in a conversation with the instructor, this motivates cognitive activity aimed at trying harder to make sense out of what the instructor is saying. The final link is that when students engage in deeper cognitive processing during learning, such as mentally arranging the material into a coherent structure and integrating it with relevant prior knowledge activated from long-term memory, this results in desirable learning outcomes such as measured by posttests. Overall, having a lesson with an onscreen agent that

you can relate to causes a social response in learners that makes them exert more effort to

understand the material and therefore construct better learning outcomes. Schroeder, Adesope,

and Gilbert (2013) report that the embodiment of the agent does not need to be anthropomorphic

in order to create social agency. Thus, it is possible that a non-humanoid agent can trigger the

same social responses, as a human agent. In their meta-analysis review Schroeder, et al., (2013)

found that the pedagogical agent's form (e.g., humanoid, non-humanoid, actual human, and

mixed agent form) did not result in significant differences for learning.

The matching hypothesis is inspired by, and consistent with the model-observer similarity

hypothesis in social-cognitive learning theory (Bandura, 2001; Schunk, 1987). This theory

posits that when students view an instructional video modeling how to perform a task, "the more

similar learners perceive themselves to be to the model in terms of characteristics such as age,

expertise, and gender, the greater the self-efficacy and thereby learning gains" (Hoogerheide, van

Wermeskerken, van Nassau, & van Gog, in press). Kim and Baylor (2006, p. 569) have shown

how Bandura's social-cognitive theory can apply to learning with onscreen pedagogical agents in

which "pedagogical agents as learning companions (PALs)...might provide an opportunity to

simulate...social interaction in computer-based learning." For example, girls may see a female

onscreen pedagogical agent as a role model who influences their motivation to exert effort to

learn (Rosenberg-Kima, Plant, Doerr, & Baylor, 2010).

The matching hypothesis is also inspired by and consistent with similarity-attraction

theory (Bryne, 1971) and social identity theory (Tajfel, 1982), which posit that people are

attracted to, identity with, and seek to affiliate with others who appear to be similar to them. In

short, "people are more attracted to others who match their personality and other human

characteristics than those who mismatch" (Moreno & Flowerday, 2006, p. 190). A core idea is

that "when two individuals share certain attributes, such as demographic characteristics...communication between them is more likely to be effective" (Qiu & Benbasat, 2009, p. 673). In short, Ozogul, Johnson, Atkinson, and Reisslein (2013, p. 38) have shown how the similarity attraction hypothesis applies to learning with pedagogical agents: "The similarity-attraction hypothesis in the context of learning with animated pedagogical agents would predict increased learning and more positive perceptions the greater the similarity between the learner and the agent."

2. Method

2.1 Participants

The sample consisted of 66 7th (33) and 8th (33) grade students (33 males, and 33 females), between the ages of 13 and 16 who were in a classroom setting at a science camp. Students had been selected by their teachers to take part in the science camp based on their interest in natural sciences. A total of 23 students reported that they had never used immersive VR before, 36 said that they had used it but for less than 2 hours, and seven said that they had used immersive VR for more than 2 hours. There were no significant differences on any of the dependent variables based on previous VR use.

2.2 Procedure

The study took place during two one week-long science camps, where students participated in different mandatory workshops, one of which was the VR-workshop. The sessions in the two different camps, followed the same setup: ID numbers were randomized and distributed to students as they arrived. Prior to playing the simulations, all students were given the pretest in a lecture hall, followed by a five-minute oral introduction on how to use the VR headsets, and how to navigate in the VR lab. A total of 21 Samsung Gear VR headsets with

matching phones had been set up, so each participant only had one version of the simulation.

Students were randomly assigned to one of the two conditions: Marie (17 boys, 16 girls) or the

drone (16 boys, 17 girls). They played the VR-simulations individually, and 9 to 12 students

participated at the time. Immediately after finishing the simulation, students' score from the

simulation was recorded, and they were given the posttest.

2.3 Materials

2.3.1 VR Simulations

The simulations were built in collaboration with the EdTech company Labster. The

experience was optimized for immersive VR, so players could use the full potential of the virtual

space which featured circular workbenches. Also, the likelihood of motion sickness was reduced

by high frame-rates and by excluding tracking camera movements. The simulations' content was

kept identical across conditions, with the main learning objective being laboratory safety. Four

types of knowledge based on Mayer's (2008) knowledge taxonomy were the target by the

simulations including: facts (e.g., the definition of important hazard symbols in the lab), concepts

(e.g., it may be dangerous to wear contact lenses because chemicals may get trapped behind

them), procedures (e.g., the step by step process of what you should do in case of a simple spill

of a corrosive chemical), and beliefs (e.g., building self-efficacy by providing positive feedback

after successfully completing a task).

There were three types of activities in the VR simulations. The first activity was

receiving information relevant for completing the tasks from the pedagogical agent. Written

information and illustrations were also provided on a LabPad (a type of virtual tablet) that

students used to read about theory, and see visualizations, pictures, and assignments. The second

type of activity was to perform tasks in the virtual environment. Tasks included removing

inappropriate items from the lab, dealing with acid spills, and identifying hazardous situations (as exemplified in the top of Figure 2). The third type of activity was answering multiple-choice questions regarding lab safety (as exemplified in the bottom of Figure 2). The simulations used multiple choice questions with explanatory feedback as a way of priming appropriate metacognitive processing during learning (Makransky et al., 2006; Mayer, 2016) based on literature that has shown the benefits of retrieval practice (e.g., Adescope et al., 2017).

2.3.2 Pedagogical Agents

The two pedagogical agents had the role of guiding the students through the lab by giving instructions, explaining theoretical concepts, and asking questions. This is in line with the modality and guided activity principles of instructional design (Mayer, 2014). Both embodied agents follow the player through the experience and display similar types of behaviour, body movements, and facial expressions. The agents use lip-sync approximation to match the voiced text. In the case of the drone, the "lip" consists of a blue ring around its eye (as shown in Figure 1). Additionally, both agents exhibit micro-expressions when resting, such as eye and head movements to add an additional layer of realism, which was found to positively affect learning outcomes (Baylor, 2004). All pedagogical agents used the same modern text-to-speech voice (NeoSpeech).

2.3.3 Measures

2.3.3.1 Pretest. The pretest questionnaire contained demographic questions concerning gender and grade, an item asking about prior immersive VR experience, and a knowledge test. The latter assessed students' knowledge about lab-safety with 12 multiple-choice questions with four response options (e.g. "What should you always do before getting rid of a strong acid?" A) Neutralize it, B) Get your supervisor, C) Dilute it with water, D) Open a window").

2.3.3.2 Performance Score in the Simulation. There were 11 multiple-choice question within

the VR simulations. Each question had four possible answers, one of which was correct. The

pedagogical agent provided short explanatory feedback when a question was answered correctly.

If a student answered a question incorrectly, they were asked to try again. The four response

options were shuffled randomly to ensure that students would not just select a new option

without engaging in the content when answering incorrectly. Students were only able to proceed

to the next step in the simulation once they got the answer correct. Students received 5 points for

a correct answer on the first try and got a reduction of 1 point for each additional attempt until

the value was 0. Student's performance in the VR simulation was calculated by adding up their

score on the 11 items.

2.3.3.3 Posttest. The posttest included a scale assessing social presence, a knowledge test, and a

transfer test. An adaptation of the social presence sub-dimension from the Multidimensional

Presence Scale (Makransky, Lilleholt, & Aaby, 2017) was used to assess social presence. The

scale consists of five items (e.g., "I had a sense that I was interacting with others in the virtual

environment, rather than a computer simulation.") on a five-point Likert scale. The knowledge

test consisted of the same items as in the pretest. The transfer test consisted of three open-ended

questions: "1) What should you do, if you spill chemicals on the floor?" "2) What should you

do, if you spill chemicals on yourself or in your eyes?" "3) What should you do if a person

catches fire?" The transfer questions were blind-marked by a lab-safety expert based on a pre-

defined scoring key. Students could get up to 3 points for question 1, and 4 points for questions

2 and 3 due to their higher level of complexity.

2.4 Apparatus

The VR simulations were administered on Samsung Galaxy S7 or S8 phones, and stereoscopically displayed through a Samsung Gear VR head-mounted display (HMD). The touch pad on the right side of the HMD is used to emulate the left-click function of a computer mouse. In order to interact with elements in the virtual learning environment, they are selected with the centered dot-cursor inside the simulation. The HMD features rotational tracking, but no positional tracking of the player. Hence, head movement is used to change the participant's field of view and dynamically render the 360-degree virtual space. Clicking on workbenches and holograms with the cursor then moves the player's physical position. The pre- and posttests were taken on students' own computers or smart pads through a SurveyMonkey internet link.

3. Results

3.1 Do the Groups Differ on Basic Characteristics?

A preliminary question concerns whether the groups differ on their pre-existing knowledge of lab safety. A 2 x 2 analysis of variance with type of agent and gender as factors showed that the mean pretest knowledge score of the drone group ($M = 6.88$, $SD = 2.53$) did not differ significantly from the mean pretest knowledge score of the Marie group ($M = 7.88$, $SD = 1.60$), $F(1,62) = 3.536$, $p = 0.311$; the mean pretest knowledge score of girls ($M = 7.56$, $SD = 1.87$) did not differ significantly from the mean pretest knowledge score of boys ($M = 7.18$, $SD = 2.43$), $F(1, 62) = 0.622$, $p = 0.575$; and there was no significant interaction between type of agent and gender, $F(1, 62) = 1.062$, $p = 0.307$. We conclude that the groups did not differ on their knowledge of lab safety before the start of the experiment.

3.2 Do Girls and Boys Benefit from Different Kinds of Onscreen Agents?

The primary goal of the study is to determine whether girls and boys learn better in an immersive virtual reality science simulation with different kinds of onscreen agents. First,

concerning performance during learning, we explore whether girls perform better during learning with Marie and boys perform better during learning with a drone (hypothesis 1) as predicted by the matching hypothesis. The first line of Table 1 shows the mean learning performance score (and standard deviation) of the four groups. A 2 x 2 ANOVA was conducted with type of agent and gender as the factors and learning performance score as the dependent variable. Overall, the mean learning performance score of students who learned with Marie ($M = 47.59$, $SD = 4.16$) did not differ significantly from the mean learning performance score of students who learned with the drone ($M = 46.11$, $SD = 5.34$), $F_{(1,62)} = 0.208$ $p = 0.728$; and the mean learning performance score for girls ($M = 45.98$, $SD = 5.27$) did not differ significantly from the mean learning performance score for boys (M = 47.71, SD = 4.21), $F_{(1,62)} = 0.287$, $p = 0.687$. However, consistent with the predictions of the matching hypothesis, there was a significant interaction between type of pedagogical agent and gender, $F_{(1,62)} = 8.045$, $p = .006$, in which girls performed better during instruction with Marie than with the drone ($d = 0.98$) whereas boys performed better during instruction with the drone than with Marie ($d = -0.41$). When we included pretest knowledge score as a covariate, an ANCOVA yielded the same pattern of results, including a significant interaction, $F_{(1,61)} = 7.280$, $p = .009$. This pattern of results confirms hypothesis 1.

Next, concerning learning of the presented information, we explore whether girls learn better with Marie and boys learn better with a drone (hypothesis 2) as predicted by the matching hypothesis. The second line of Table 1 shows the mean posttest knowledge score (and standard deviation) of the four groups. A 2 x 2 ANOVA was conducted with type of agent and gender as the factors and posttest knowledge score as the dependent variable. Overall, the mean posttest knowledge score of students who learned with Marie ($M = 9.36$, $SD = 1.22$) did not differ significantly from the mean posttest knowledge score of students who learned with the drone (*M*

= 9.24, SD = 1.32), $F(1,62)$ = 0.30, p = 0.890; and the mean posttest knowledge score for girls

(M = 9.27, SD = 1.26) did not differ significantly from the mean posttest knowledge score for

boys (M = 9.33, SD = 1.29), $F_{(1,62)}$ = 0.007, p = 0.947. However, consistent with the predictions

of the matching hypothesis, there was a significant interaction between type of agent and gender,

$F_{(1,62)}$ = 5.011, p = .029, in which girls learned better with Marie than with the drone (d = 0.67)

whereas boys learned better with the drone than with Marie (d = -0.45). When we included

pretest knowledge score as a covariate, an ANCOVA yielded the same pattern of results,

including a significant interaction, $F_{(1,61)}$ = 4.312, p = .042. This pattern of results confirms

hypothesis 2.

Third, concerning ability to apply the presented information to solve new problems in a

transfer test, we explore whether girls perform better on the transfer test if they learned with

Marie and boys show better transfer if they learned with a drone (hypothesis 3) as predicted by

the matching hypothesis. The third line of Table 1 shows the mean transfer score (and standard

deviation) of the four groups. A 2 x 2 ANOVA was conducted with type of agent and gender as

the factors, and transfer score as the dependent variable. Overall, the mean transfer score of

students who learned with Marie (M = 4.53, SD = 1.42) did not differ significantly from the

mean transfer score of students who learned with the drone (M = 4.18, SD = 1.37), $F(1,62)$ =

0.333, p = 0.667; and the transfer score for girls (M = 5.01, SD = 1.25) did not differ significantly

from the mean transfer score for boys (M = 3.70, SD = 1.22), $F(1,62)$ = 3.90, p = 0.299.

However, consistent with the predictions of the matching hypothesis, there was a significant

interaction between type of agent and gender, $F(1,62)$ = 5.287, p = .025, in which girls

transferred better with Marie than with the drone (d = 1.03) whereas boys transferred better with

the drone than with Marie (d = -0.23). When we included pretest knowledge score as a

covariate, an ANCOVA yielded the same pattern of results, including a significant interaction, $F_{(1,61)} = 4.126$, $p = .047$. This pattern of results confirms hypothesis 3.

Overall, girls performed better on tasks during learning and demonstrated better learning outcomes in terms of knowledge and transfer test scores if they learned with Marie rather than with the drone, and boys show the opposite pattern. These three interactions constitute the major findings of this experiment.

3.3 Do Girls and Boys Differ on Their Experience of Social Presence from Different Kinds of Onscreen Agents?

Finally, we examined whether girls and boys differed in the degree to which they reported a feeling of social presence for Marie versus the drone. The fourth line of Table 1 shows the mean social presence rating (and standard deviation) of the four groups. A 2 x 2 ANOVA was conducted with type of agent and gender as the factors, and social presence rating as the dependent variable. Overall, the mean social presence rating of students who learned with Marie ($M = 3.15$, $SD = 0.59$) did not differ significantly from the mean social presence rating of students who learned with the drone ($M = 2.94$, $SD = 0.75$), $F(1,62) = 0.462$, $p = 0.620$; and the mean social presence rating for girls ($M = 3.12$, $SD = 0.61$) did not differ significantly from the mean social presence rating for boys ($M = 2.98$, $SD = 0.75$), $F(1,62) = 0.21$, $p = 0.727$. However, there was a marginally significant interaction between type of agent and gender, $F(1,62) = 3.793$, $p = .056$, in which girls rated the two agents equivalently ($d = -0.17$) whereas boys gave higher social presence ratings for Marie than for the drone ($d = 0.76$).

4. Discussion

4.1 Empirical Contributions

In contrast to previous research on gender matching (Hoogerheide, Loyen, & van Gog, 2016; Hoogerheide, van Wermesken, van Nassau, & van Gog, in press; Moreno & Flowerday, 2006; Ozugul, Johnson, Atkinson, & Reisslein, 2013), in the present study girls and boys learned better with different kinds of onscreen agents across three different measures of learning. Specifically, on learning performance, on a knowledge posttest, and on a transfer posttest, there was an interaction in which girls obtained higher scores with Marie as the pedagogical agent and boys learned better with a drone as the pedagogical agent. Some key differences between prior studies and the present one are that the present study involved learning in immersive VR, which may have made the pedagogical agents more salient, and that we compared learning with Marie (intended for girls) versus the drone (intended for boys).

The results related to social presence suggest that boys reported higher social presence with Marie than the drone, whereas there was no difference for girls. The fact that boys had higher social presence with Marie but learned less with her suggests that they payed less attention to the learning material in her presence. There are several examples of studies that have investigated learning and presence in immersive VR that suggest that higher presence does not necessarily lead to more learning (e.g., Makransky, Terkildsen & Mayer, 2017; Moreno & Mayer, 2002; Parong & Mayer, 2018). It is possible that the role of presence in developing learning is not simple but depends on a number of other variables and instructional design features. Presence can facilitate learning through positive affective outcomes, such as enjoyment and motivation, to the extent that instructional design elicits and promotes appraisal and intrinsic value of the educational content (Pekrun, 2006; Plass and Kaplan, 2016).

On the other hand, presence may be a factor in lowering learning outcomes by inhibiting reflective thinking, as learners could become completely engrossed in the social interactions and

the virtual environment to the point where they do not reflectively make sense of the material

(i.e., engage in generative processing). In the given context it is possible that by focusing on the

embodied characteristics of Marie, boys experienced greater social presence, but focused less on

the learning material, which then lead to lower performance, learning, and transfer. Thus,

Marie's characteristics may have drawn the boys' attention away from the learning material.

This is consistent with the coherence principle (Mayer & Fiorella, 2014) of multimedia learning

which has been identified as being important for VR environments in previous literature

(Makransky, Terkildsen & Mayer, 2017; Moreno & Mayer, 2002).

Overall, these results tie back to and contribute to the existing literature by showing that

gender matching can be an effective instructional design strategy when the pedagogical agent for

boys is designed specifically to be appealing to boys and the pedagogical agent for girls is

designed specifically to be similar to and appealing to girls.

4.2 Theoretical Implications

The findings support the gender matching hypothesis, the similarity attraction theory

(Bryne, 1971; Tajfel, 1982) and the model-observer similarity hypothesis (Bandura, 2001;

Schunk, 1987) from which it is derived. Consistent with the similarity attraction theory, people

learn better when they perceive the instructor--in this case, the onscreen pedagogical agent--as

similar to them (such as sharing important characteristics related to gender for girls and boys).

In the realm of online multimedia learning, the results help extend social agency theory (Mayer,

2014) by demonstrating that having an onscreen pedagogical agent that the learner can identify

with, may serve as a social cue that motivates the learner to exert more effort to understand the

material. Social agency theory holds that aspects of the instructor--such as the instructor's

appearance, gesture, voice, and conversational style--can serve as social cues that trigger a social

response in the learner in which the learner sees the instructor as a social partner who is working

with the learner. This causes the learner to exert more effort to make sense of the incoming

information, and therefore build meaningful learning outcomes that better support transfer test

performance.

4.3 Practical Implications

The most important practical implication of this study is that instructional designers

should consider how to prime the learner's sense of social identification with the onscreen

pedagogical agent while learning in immersive virtual reality. In particular, female learners may

better identify with one kind of pedagogical agent whereas male learners may better identify with

a different kind of pedagogical agent. Although more research is needed to pinpoint which

features of pedagogical agents boost the social response of female and male learners

respectively, the present study suggests that girls identify better with a young woman in a white

lab coat in a science lab whereas boys identify better with a flying, robotic entity.

In short, this work suggests that designers should consider building some pedagogical

agents specifically for boys and others specifically for girls. Agents for boys should be easy to

identify with for boys, whereas agents for girls should be easy to identify with for girls.

The finding that higher presence did not necessarily lead to higher learning when seen

across gender and conditions suggests that the relationship between presence and learning

outcomes is complicated when learning in immersive VR. The practical implications of these

findings are that it is important to appropriately design instructional material for immersive VR,

where the affective learning outcome benefits from the high levels of presence afforded by this

platform are maintained, while the media's adverse effects on learning are minimized.

Generative learning strategies, such as learning by teaching or summarizing, have been

advocated and shown to facilitate learning by encouraging learners to select, organize and integrate the essential information by putting it into their own words, thereby fostering reflective knowledge building (i.e. generative processing; Fiorella & Mayer, 2016; Parong & Mayer, 2018). Using these strategies with pedagogical agents in immersive VR could lead to promising learning and motivational outcomes. An example could be to investigate if it is possible to design an immersive VR simulation that encourages social agency and generative learning strategies such as summarizing or learning by teaching through contextualized interaction with pedagogical agents as peers.

4.4 Limitations and Future Directions

This study involves a single experiment on one topic (i.e., lab safety) with a low sample size and an immediate posttest. Future research is needed to determine whether the interactions observed in this study can be replicated. In addition, studies that involve different context, larger sample size, and/or delayed posttests are required. It also would be useful to obtain measures of how well learners identified with (or perceived similarity with) the onscreen pedagogical agent to better test the mechanisms proposed to underlie the gender matching hypothesis. In short, work is needed to determine which features of a pedagogical agent have a positive impact on boys more than on girls and which aspects have a positive impact of girls more than on boys.

Furthermore, it would be interesting to investigate whether students hold stereotypical views towards the agents, e.g., if students perceive Marie as less competent because of her gender, and if this perception depends on the students' gender and age. This study used a young study population of 7th and 8th grade students. Johnson, et al. (2013a) found that students preferred an agent of the same gender and age, but this effect gets lower as students get older. Therefore, it possible that students at this age may have naïve concepts of science and science

teachers. Future research should investigate how stereotypes toward females representing an often male dominated domain of science labs and science safety may change with age.

Also, it would be useful to compare students who learned with onscreen agents as opposed to having no image of an agent on the screen in immersive VR, as was done in studies involving desktop presentations (Mayer & DaPra, 2012; Wang, Li, Mayer, & Lui, in press).

The sample in this study was participating in a science camp where the students had been selected because they were interested in STEM. Therefore, students were highly engaged in the content of the simulation, and took the experiment very seriously. Future research should investigate if the results generalize to more traditional classroom settings. Finally, it would be useful to determine whether the gender matching hypothesis works the same way in immersive VR (in which the learner wears a head mounted display and moves in virtual space) and in desktop VR (in which the learner sits in front of a computer screen).

References

Adescope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A
meta-analysis of practice testing. *Review of Educational Research*, 87 (3), 659-701.

Bandura, A. (2001). Social-cognitive theory: An agentic approach. *Annual Review of
Psychology, 52,* 1-26.

Baylor, A. L., & Kim, Y. (2004, August). Pedagogical agent design: The impact of agent
realism, gender, ethnicity, and instructional role. In *International Conference on Intelligent
Tutoring Systems* (pp. 592-603). Springer, Berlin, Heidelberg.

Blascovich, J., & Bailenson, J. (2011). *Infinite reality.* New York: HarperCollins.

Bryne, D. (1971). *The attraction paradigm.* New York: Academic Press.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational
agents.* Cambridge, MA: MIT Press.

Dehn, D. M, & van Mulken, S. (2000). The impact of animated interface agents: A review of
empirical research. *International Journal of Human-Computer Studies, 52,* 1-22.

Goldman Sachs (2018). Profiles in innovation report. Retrieved from
http://www.goldmansachs.com/our-thinking/pages/technology-driving-innovation-
folder/virtual-and-augmented-reality/report.pdf

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student
motivation and learning? *Educational Research Review, 6,* 27-54.

Johnson, W.L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty
years later. *International Journal of Artificial Intelligence in Education, 26,* 25-36.

Johnson, A. M., DiDonato, M. D., & Reisslein, M. (2013a). Animated agents in K-12

    engineering outreach: Preferred agent characteristics across age levels. *Computers in*

    Human Behavior, *29*(4), 1807-1815.

Johnson, A. M., Ozogul, G., Moreno, R., & Reisslein, M. (2013b). Pedagogical agent signaling

    of multiple visual engineering representations: The case of the young female agent.

    Journal of Engineering Education, *102*(2), 319-337.

Johnson, W. L., & Rickel, J. W. (2000). Animated pedagogical agents: Face-to-face interaction

    in interactive learning environments. *International Journal of Human-Computer Studies,*

    11, 47-78.

Kim, Y., Baylor, A. L., & Shen, E. (2007). Pedagogical agents as learning companions: the

    impact of agent emotion and gender. *Journal of Computer Assisted Learning*, *23*(3), 220-

    234.

Kim, Y., & Lim, J. H. (2013). Gendered socialization with an embodied agent: Creating a social

    and affable mathematics learning environment for middle-grade females. *Journal of*

    educational psychology, *105*(4), 1164.

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 Mermeskerken, M., van Nassau, H., & van Gog, T. (in press). 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? *Computers in Human Behavior, 00*, 000-000.

Kim, Y., & Baylor, A. (2006). A social-cognitive framework for pedagogical agents as learning

    companions. *Educational Technology Research & Development, 54*, 569-596.

Lee, K. M. (2004). Presence, explicated. *Communication Theory, 14 (1),* 27-50.

Makransky, G., & Lilleholt, L. (2018). A Structural Equation Modeling Investigation of the

> Emotional Value of Immersive Virtual Reality in Education. *Educational Technology*

> *Research and Development.* 1-24. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the Multimodal

> Presence Scale for virtual reality environments: A confirmatory factor analysis and item

> response theory approach. *Computers in Human Behavior*, *72*, 276–285.

> https://doi.org/10.1016/j.chb.2017.02.066

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2017). Adding immersive virtual reality to a

> science lab simulation causes more presence but less learning. *Learning and Instruction*.

> https://doi.org/10.1016/j.learninstruc.2017.12.007

Makransky, G., Thisgaard, M. W., & Gadegaard, H. (2016). Virtual simulations as preparation

> for lab exercises: Assessing learning of key laboratory skills in microbiology and

> improvement of essential non-cognitive skills. *PloS one*, *11*(6), e0155895.

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* (2nd ed; pp. 345-370). New York: Cambridge University Press.

> https://doi.org/10.1075/idj.16.1.13pel

Mayer, R. E. (2016). The role of metacognition in stem games and simulations. In H. F. O'Neil,

> E. L. Baker, & R. Perez (eds.), *Using games and simulations for teaching and assessment:*

> *Key issues* (pp. 183–205). New York: Routledge. https://doi.org/10.4324/9781315817767

Mayer, R. E., & DaPra, S. (2012). An embodiment effect in computer-based learning with

    animated pedagogical agents. *Journal of Experimental Psychology: Applied, 18, 239-*

    252.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia

    learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity

    principles. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (2nd

    ed; pp. 279–315). New York: Cambridge University Press.

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.

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments:

    Role of methods and media. *Journal of Educational Psychology*, *94*(3), 598–610.

    https://doi.org/10.1037//0022-0663.94.3.598

Moreno, R., Mayer, R. E., Spires, H., & Lester, J. (2001). The case for social agency in

    computer-based teaching: Do students learn more deeply when they interact with

    animated pedagogical agents? *Cognition and Instruction, 19*, 177-214.

Moreno K. N., Person N. K., Adcock A. B., Eck R. N. V., Jackson G. T. & Marineau J. C.

    (2002) Etiquette and efficacy in animated pedagogical agents: the role of stereotypes.

    Paper presented at the *AAAI Symposium on Personalized Agents*, Cape Cod, MA.

Moreno, R., Reisslein, M., & Ozogul, G. (2010). Using virtual peers to guide visual attention

    during learning: A test of the persona hypothesis. *Journal of Media Psychology*:

    Theories, methods, and applications, 22(2), 52–60.

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.

Ozogul, G., Reisslein, M., & Johnson, A. M. (2011). Effects of visual signaling on precollege students' engineering learning performance and attitudes: Peer versus adult pedagogical agents versus arrow signaling. *Proceedings of the 118th Annual Conference and Exposition of the American Society for Engineering Education* (AC 2011-344).

Qiu, L., & Benbasat, I. (2010). A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human-Computer Studies, 68,* 669-688.

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology.* Advance online publication. http://dx.doi.org/10.1037/edu0000241

Reeves, B., & Nass, C. (1996). *The media equation.* New York: Cambridge University Press.

Rosenberg-Kima, R. B., Plant, E. A., Doerr, C. E., & Baylor, A. L. (2010). The influence ofcompter-based models' race and gender on female students' attitudes and beliefs towards engineering. *Journal of Engineering Education, 99,* 35-44.

Schroeder, N. L., & Adesope, O. O. (2014). A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education, 46*, 229-251.

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49*, 1-39.

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational*

    Research, 57, 149-174.

Tajfel, H. (1982). *Social identity and intergroup relationships*. New York: Cambridge

    University Press.

Thisgaard, M., & Makransky, G. (2017). Virtual learning simulations in high school: Effects on

    cognitive and non-cognitive outcomes and implications on the development of STEM

    academic and career choice. *Frontiers in psychology*, *8*, 805.

Veletsianos, G., & Russell, G. S. (2014). Pedagogical agents. In J. M. Spector, M. D. Merrill, J.

    Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and*

    technology (4th ed; pp. 759-770). New York: Springer.

Wang, F., Li, W., Mayer, R. E., & Liu, H. Animated pedagogical agents as aids in multimedia

    learning: Effects on eye fixations during learning and learning outcomes. *Journal of*

    Educational Psychology, 110, 250-268.

Table 1: Table of the results from the two group (drone/Marie) by two (female/male) ANOVA.

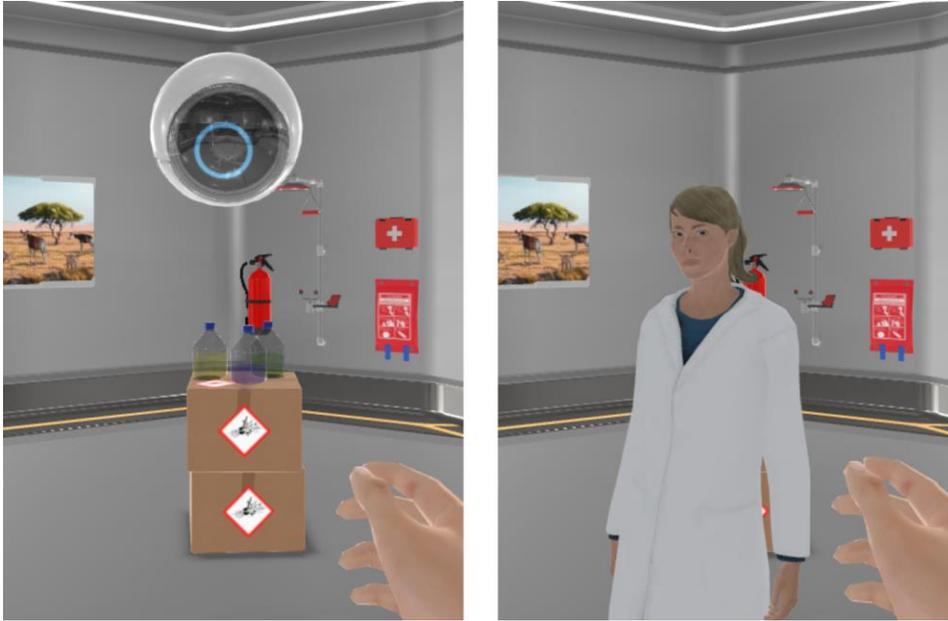| Scale	Females		Males		Significance		
	Drone M (SD)	Marie M (SD)	Drone M (SD)	Marie	conditionp	gender p	interaction p
Performance	43.76 (5.67)	48.34 (3.65)	48.56 (3.69)	46.88 (4.59)	0.728	0.687	.006
Knowledge	8.88 (1.17)	9.69 (1.25)	9.63 (1.41)	9.06 (1.14)	0.890	0.947	.029
Transfer	4.50 (1.26)	5.56 (1.01)	3.84 (1.43)	3.56 (1.00)	0.667	0.299	.025
Social presence	3.16 (0.73)	3.06 (0.48)	2.70 (0.72)	3.24 (0.69)	0.620	0.727	.056

Figure 1: Screen shots from the immersive VR lab safety simulation showing the two

pedagogical agents: Left: Hovering robot called the drone. Right: Female, humanoid assistant
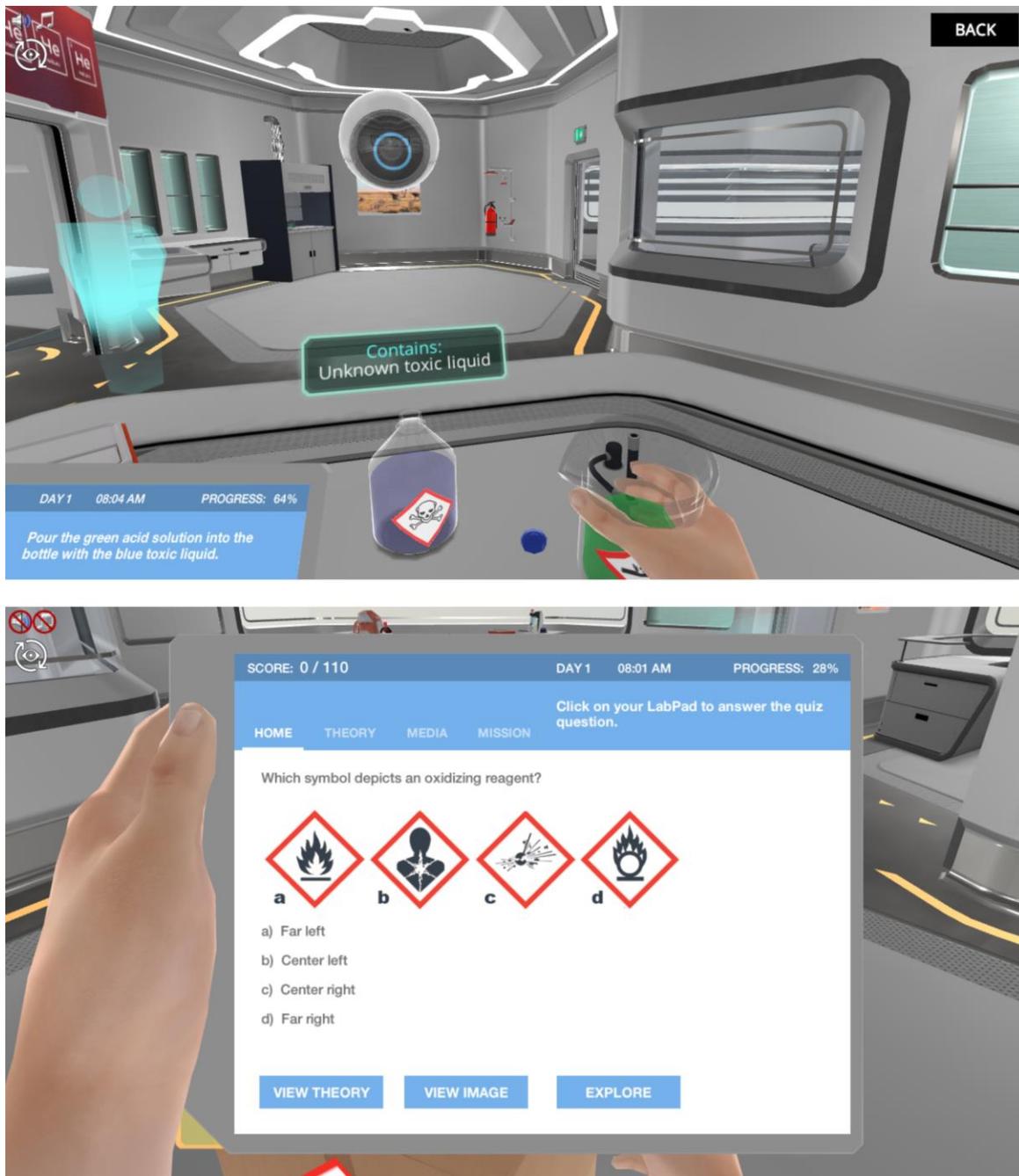
called Marie.

Figure 2: Screen shots from different types of activities in the virtual lab. Top: Oral instructions given by the agent, while the user is performing a task. Bottom: The LabPad is emphasized to answer quiz-questions, read the theory or look at images.