Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction

Nicole C. Krämer a,*, Bilge Karacora a, Gale Lucas b, Morteza Dehghani c, Gina Rüther a, Jonathan Gratch b

a University of Duisburg-Essen, Forsthausweg 2, 47057, Duisburg, Germany
b USC Institute for Creative Technologies, 12015 Waterfront Drive, Playa Vista, CA 90094-2536, USA
c University of Southern California, 3620 South McClintock Ave., Los Angeles, CA 90089-1061, USA

ABSTRACT

While numerous research endeavors address the effects of pedagogical agents, the role of the agent’s gender and its rapport behavior has been neglected. We hypothesize that a minimal amount of behavioral realism induced by display of rapport is necessary for any social effects to occur in human-computer interaction. Further, in line with results from STEM research on female role models, we assume that especially for female learners a same sex agent will be beneficial. In a 2 (student gender) × 2 (agent gender) × 2 (rapport behavior yes/no) between subjects design, we investigate whether virtual agents can help enhance participants’ performance, effort and motivation in mathematics. Female and male participants (N = 128) interacted with a male or female virtual agent that either displayed rapport or no rapport. Our results confirm the expected main effect of rapport. However, against expectations, our results do not support the assumption that a same sex agent is beneficial for female learners. Participants’ performance and effort were significantly enhanced when interacting with an agent of opposite gender that displayed rapport. Our results have implications on designing agents for education and training purposes.

© 2016 Published by Elsevier Ltd.

1. Introduction

A common goal in educational endeavors is to increase interest and performance in science and math related subjects. Compared to the need for future experts in the realm, there is a lack of students in STEM fields (Science, Technology, Engineering and Mathematics) and especially women are underrepresented (Hill, Corbett, & Rose, 2010). It is therefore of high societal relevance to encourage women to enter STEM related fields. However, previous findings suggest that due to sociocultural factors like implicit and explicit gender stereotypes, a gender difference exists in STEM fields especially with regard to motivation and willingness to enter the corresponding disciplines (Correll, 2004; Nosek, Banaji, & Greenwald, 2002a; Nosek, Banaji, & Greenwald, 2002b; Steele & Aronson, 1995). These stereotypes are mainly perceived and learned through
interaction with other social beings (Eccles, Wigfield, & Schiefele, 1998; Jacobs, 1991). Women seem to underestimate their competence in mathematics (and other STEM fields) because of the socially common belief that women have lower abilities in those fields than men (Beyer & Bowden, 1997; Bong, 1999). Apparently, educating women about the inappropriateness of those gender stereotypes is not sufficient since those beliefs can continue to stay implicitly present in their minds (Nosek et al., 2002a, 2002b). According to the expectation-value-model (Eccles et al., 1983), women underestimate their own competence and have lower expectation for success and therefore tend to have a lower preference for a career in STEM fields.

Recently, there has been an upsurge of interest in educational technology which exploits social and motivational factors that enhance math performance in general, and reduces gender inequality in particular (Baylor & Ryu, 2003; Kim, 2004). Specifically, pedagogical agents have been suggested as a means to individually motivate and instruct students. These autonomous agents have been widely expected to transfer the benefits of human–human tutoring and instructional communication to the area of computer-supported learning. Future systems are anticipated to enhance e-learning programs for individual learning with a human-like motivator or to support teachers in the classroom by attending to small groups of students (Lester, Towns, Callaway, Voerman, & Fitzgerald, 2000; Moreno, 2004). Advocates of embodied automatic tutors claim that one major advantage of these pedagogical agents is that they can, or will be able in the future, communicate via verbal and nonverbal means, thus facilitating and personalizing the interaction with an e-learning program (e.g., Lester et al., 2000). Furthermore, increased motivation is expected: Baylor and Ryu (2003) suggest that the key advantage is that human-likeness creates more positive learning experiences and provides a strong motivating effect (Krämer & Bente, 2010). This assumption is also in line with the phenomenon that people often treat computers as social actors (Reeves & Nass, 1996). Here, it has been demonstrated that people automatically display social behavior towards computers and virtual agents even if they are convinced that the artificial interlocutor does not warrant human-like treatment (Krämer, 2005; Nass & Moon, 2000). Therefore, it can be assumed that psychological factors which improve people’s performance in traditional face-to-face instructional settings (Andersen, 1979) can be successfully simulated by technologies in form of virtual learning companions or virtual instructors — at least in the sense that people will tend to react to them in a similar way as they would towards human instructors.

One factor that has been shown to be beneficial in terms of facilitating interaction, relationship building and instructional communication is rapport. In social psychology, rapport is described as the establishment of a positive relationship among interaction partners by rapidly detecting and responding to each other’s nonverbal behavior (Gratch, Wang, Gerten, Fast, & Duffy, 2007a). This includes displaying behaviors that indicate positive emotions (such as head nods and smiles), showing mutual attentiveness (such as mutual gaze) and certain coordination behaviors (such as postural mimicry and synchronized movement, Tickle-Degnen & Rosenthal, 1990). Not only is rapport beneficial in human–human-interaction but also within human–agent-interaction: Niewiadomski, Mancini, Hyniwska, and Pelachaud (2010) have shown that when an agent displays appropriate and socially adapted emotional expressions it is perceived as more human-like than an agent that shows human expressions which are inappropriate or not socially adapted. This indicates that rapport is an important feature in order for the agent to be perceived as human-like and for any social effects to occur.

In this study, our goal is to focus on the agent’s social role and social effects within instructional settings. Therefore, we do not employ a virtual agent as a tutor as in most other research on pedagogical agents (see Graesser, Wiemer-Hastings, Wiemer-Hastings, Kreuz, & Tutoring Research Group, 1999; Lester et al., 2000). Instead, the agent’s role is to motivate by observing the learners’ success and — by asking about the experiences — giving the learner the opportunity to self-reflect on one’s own performance. We therefore do not primarily focus on the role of the agent in terms of guiding and instructing cognitive processes but rather address the agent’s social role and the corresponding processes (Krämer & Bente, 2010). Although traditionally, research on pedagogical agents has been dominated by a cognitive perspective, recent approaches explicitly state the importance of complementing this by a social psychological framework. For instance, Kim and Baylor (2006) argue that learning environments should provide situated social interaction since it is well documented that the cognitive functioning of learners is framed by social contexts. Kim and Baylor (2006) sum up that teaching and learning are highly social activities. They conclude that the reason for the weak impact of virtual agents (compared to human tutors) concerning learning outcomes might lie in the lack of empathetic social encouragement and caring aspects.

In this paper, we seek to address two goals. First, we aim to show that employing socio-emotional cues can enhance math performance in a human-computer setting. Specifically, we aim to show that the presence of an agent which provides rapport and fosters self-reflection can improve performance on standardized math tests. By this, we seek to provide further evidence that people do treat computers as social actors and help elucidate the design principles that foster this effect. Second, we strive to contribute to the question whether the gender of the agent matters and influences the effects in interaction with the student’s gender. Here, we do not strictly follow up on the research on competent role models and their potential to motivate young women to engage in STEM fields but — in order to provide a new foundation to the discussion — rather focus on the effects of the mere presence of a pedagogical agent which is either of same or different sex as the learner.

2. Theoretical background

The importance of motivating factors in instructional processes has long been demonstrated. Especially the expectation to excel in terms of self-efficacy plays a crucial role: according to the expectancy-value-model of academic motivation, an individual’s engagement and performance with regard to an academic task is best predicted by the two factors expectation and value of success (Eccles, Wigfield, Harold, & Blumenfeld, 1998; Pintrich & Schrauben, 1992; Pintrich & Schunk, 1996). Marks
(2000) defines engagement as a psychological process which is mainly defined by the factors attention, interest, investment and effort put to a certain task or subject. Engagement has been validated as a reliable predictor for academic performance with high engagement being associated with good grades and test results (Arhar & Kromrey, 1993; Connell & Wellborn, 1991; Mounts & Steinberg, 1995; Voelkl, 1995). The expectation of success with regard to a certain task is strongly determined by an individual’s belief of having the required capabilities to conclude that task successfully. This expectation is a critical factor when it comes to an individual’s motivation to engage and thrive to perform well in a task; the more an individual believes to have the necessary capabilities to do well, the more they are motivated to engage and actually be successful — which seems to be true for STEM related subjects in particular (Greene, DeBacker, Ravindran, & Krows, 2002; Singh, Granville, & Dika, 2002).

This means, that, according to the expectation-value-model, the subjective perception of personal task-specific competence and self-efficacy strongly influence an individual’s behavior with regard to a certain task (Eccles et al., 1998). Previous studies showed that children’s self-evaluation of their own competence was a stronger predictor for their performance in different subjects than their actual previous performance (Eccles, 1994; Eccles et al., 1983; Parsons, Adler, & Kaczala, 1984; Wigfield & Eccles, 1992). These results have implications for the motivation and success in STEM related subjects since it has been shown that girls and women tend to have a lower esteem of their own STEM related competences and a lower expectation of success (Meece, Wigfield, & Eccles, 1990). One of the most important goals for teaching in STEM related fields therefore is to strengthen students’ and especially female students’ self-evaluation of their own competence. Two aspects which might be influential in this context are teacher gender and teacher (rapport) behavior which will be discussed in the following sections.

2.1. Influence of social aspects in instructional settings

With regard to the expectancy-value-model, Eccles et al. (1983) stress the impact of the social environment on an individual’s academic motivation. Previous research repeatedly showed that social actors play a significant role when it comes to self-evaluation of personal competence and academic value beliefs. Different authors (Eccles & Jacobs, 1986; Grolnick, Gurland, Jacob, & Decourcy, 2002; Grolnick & Ryan, 1989; Wigfield & Eccles, 1992) concluded in their research that individual’s academic motivation was strongly influenced by their perceived support and encouragement from parents and teachers. These perceptions seemed to have an even stronger impact on performance-related decisions and academic engagement than their actual performances. Additionally, other social processes play a role: a number of studies showed that competent female role models in STEM, especially mathematics, contributed successfully to minimizing stereotype threats (Aronson, Fried, & Good, 2002; Marx & Roman, 2002; Mallam, 1993; McIntyre, Paulson, & Lord, 2003; McIntyre et al., 2005). Until now, there have been very few studies specifically focusing on the effect of role models. However, numerous studies show that social comparison with a competent person who is considered similar to oneself has a positive influence on one’s academic performance by inspiring (Lockwood & Kunda, 1997; Tesser, 1986), improving self-evaluation, increasing motivation (Blanton, Crocker, & Miller, 2000; Collins, 1996; Major, Sciacchitano, & Crocker, 1993; Taylor & Lobel, 1989) and even impacting student’s academic goals (Hackett, 1985; Lockwood & Kunda, 1997). This is in line with previous classical social psychology research which shows that similarity of personal characteristics (like gender, ethnicity, age) is a key factor with regard to a role models’ influence on motivation and perceived self-efficacy in a learning environment (Bandura, 1997). Marx and Roman (2002) tried to use those positive effects of role models to counteract the influence of stereotypical beliefs on female’s academic performance in mathematics. Their research showed that girls’ performance in a mathematical test showed no significant difference compared to the boys’ performance when they were instructed by a competent female tutor. The authors argued that female role models can help overcome mathematically talented girls’ (explicit and implicit) stereotypical beliefs of being less mathematically talented as boys. They further explained that the female role model is perceived as a kind of counter evidence that girls have lower competence and therefore a lower chance of success in mathematics than boys do. Also, girls’ fear of confirming that stereotype in front of others is minimized. Therefore, it has been concluded that especially for girls, a same sex tutor which might serve as a role model is important and beneficial. There are — albeit a few — studies which directly test this assumption: a series of experiments revealed that exposure to female STEM experts promoted positive implicit attitudes and stronger implicit identification with STEM, greater self-efficacy in STEM, and more effort on STEM tests (Stout, Dasgupta, Hunsinger, & McManus, 2011). The results were driven by greater subjective identification which predicted enhanced self-efficacy.

However, other studies show that academic motivation and engagement does not significantly vary as a function of the teacher’s gender (Martin & Marsh 2005). Moreover, Betz and Sekaquaptewa (2012) demonstrated that the counter-stereotypic-yet-feminine success of female role models may actually be demotivating, particularly to young girls. Results showed that feminine STEM role models reduced middle school girls’ current math interest, self-rated ability, and success expectations relative to gender-neutral STEM role models.

The studies so far did not address the question of which mechanisms exactly contribute to the effects of role models: it is not sufficiently clear whether the driving factor is primarily the role models’ content-related competence, the fact that there is less fear of confirming the stereotype in front of others or merely the presence of a same-sex person which could lead to a more comfortable situation. In order to take a first step towards disentangling these explanations, the present study will focus on the mere presence of a same-sex person.
2.2. Influence of social signals within instructional communication

It has long been shown that the quality of a relationship between students and their teachers strongly influences student’s motivation to engage in those academic subjects and is very important to students’ self-concept (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Learner & Kruger, 1997). With regard to mathematics, Chouinard and Karsenti (2005) even showed that the feeling of being supported by teachers is more influential for students than feeling supported by their own parents. In social psychology, rapport has been repeatedly shown as a correlate of a positive relationship between an individual and their instructor, which leads to an increase of motivation and engagement with regard to a certain task (Granitz, Koernig, & Harich, 2009; Thomas, Ribich, & Freie, 1982). Rapport is said to have positive influences in “negotiations, management, psychotherapy, teaching and caregiving” (Gratch, Wang, Gerten, Fast, & Duffy, 2007a, p.1). Rapport has been defined as verbal and nonverbal codependence between interaction partners (Stumm & Pritz, 2000). Tickle-Degnen and Rosenthal (1990) described rapport as consisting of three essential components: mutual attention, positivity and coordination (e.g., in terms of turn taking). Also, rapport occurs on three levels. On the behavioral level both interaction partners align their body movements (e.g. posture shifts, nods), on the emotional level both conversation partners feel comfortable and perceive the interaction as rewarding. Finally, on the cognitive level there is a shared understanding (Gratch et al., 2007a; Tickle-Degnen & Rosenthal, 1990). Rapport is a sign of quality within an interaction and does not arise in every interpersonal conversation. When rapport occurs people, for instance, are more responsive to what the counterpart is saying, broaden their variety of topics, keep more eye-contact, smile more frequently and increase proximity (Cassell, Gill, & Tepper, 2007; Grahe & Bernieri, 1999). Rapport happens both on a verbal and on a nonverbal level. However, several researchers (Grahe & Bernieri, 1999; Tickle-Degnen & Rosenthal, 1990) argued that the nonverbal cues are more essential indicators (and drivers) for rapport. In line with these results, Andersen (1979) demonstrated that in instructional communication immediate social cues such as smiling and nodding displayed by the instructor had a positive effect on the participant’s interest and motivation resulting in a better performance.

2.3. Virtual agents in instructional settings

Eccles and Midgley (1989) summarize that students are highly motivated when the learning environment is adapted to their individual need, interests and capabilities. Using virtual figures as tutors or instructors can help to achieve this kind of individual learning environment. Their variable features can be adapted to the individual needs and goals of certain users and user groups. In addition, virtual figures can be used independently from time and location without any human oscillations like tiredness, time pressure, negative mood etc. This is one reason why the development and evaluation of so called pedagogical agents has been flourishing in the last 20 years. Johnson, Rickel, and Lester (2000) found evidence that the use of virtual agents influences students’ performance in a positive way. A number of studies showed that the mere presence of an autonomous virtual figure can increase interest and attention with regard to a specific learning situation thereby enhancing the learning performance (Baylor & Ryu, 2003; Lester et al., 2000). However, numerous studies demonstrate that it is not first and foremost the agent’s presence which is influential but that the agent’s behavior is decisive. Garau, Slater, Pertaub, and Razzaque (2005) conducted a study showing that only participants who interacted with an agent with behavior that was responsive to their movements experienced a sense of personal contact with the agent which influenced them to behave more socially considerate as opposed to interacting with a static or moving but unresponsive agent. A certain amount of behavioral realism is indeed necessary for a virtual agent to be perceived as human-like and therefore trigger social reactions in their human interaction partner; this behavioral realism is created by a virtual agent’s appropriate emotional and behavioral reactions towards their human interaction partner (Blascovich, 2002; Von der Pütten, Krämer, Gratch, & Kang, 2010). A virtual agent needs to be capable of showing appropriate emotional reactions towards a human interaction partner in order to cause a feeling of support and encouragement which can have a motivational effect (Elliott, Rickel, & Lester, 1997; Kim, 2004). In line with these assumptions, Andersen’s (1979) findings that an instructor’s immediate social cues (e.g. smiling and nodding) increase motivation and learning, were replicated for human–computer-interaction (Kim, 2004). She demonstrated that positive emotional facial expressions displayed by virtual agents had a positive effect on participants’ motivation and engagement to work on a given task. When it comes to human–computer-interaction the display of appropriate emotional reactions and immediate social cues by a virtual agent creates a feeling of rapport in the user and contributes to a positive relationship (Gratch et al., 2006, 2007a; Krämer & Bente, 2010; Wang & Gratch, 2009).

2.4. The influence of virtual agents dependent on their gender

For the area of STEM, research has shown that virtual agents can enhance females’ preference for STEM fields in general (Eisenhart, 2008; Turner & Lapan, 2005). Of interest is whether differential characteristics of female or male virtual agents can be expected to affect this preference. In line with the results from face-to-face instructional communication, several studies demonstrated that individuals were influenced more strongly by a virtual agent with the same ethnicity (Baylor & Kim, 2004; Pratt, Hauser, Urgay, & Patterson, 2007) or the same gender (Baylor, 2005; Baylor & Kim, 2004; Guadagno, Blascovich, Bailenson, & McCall, 2007; Kim, Baylor, & Shen, 2007) compared to a virtual agent with different characteristics. Plant, Baylor, Doerr, and Rosenberg-Kima (2009) showed that (male and female) undergraduate students reported more positive attitudes and higher interest in engineering after the interaction with a female virtual character (compared to prior to the
interaction, a control group and those interacting with a male agent). Further studies demonstrate that female agents have a stronger influence on students’ self-efficacy with regard to engineering and were more efficient in counteracting negative gender stereotypes (Baylor & Plant, 2005). Especially agent models that were similar to the young women (i.e., young, female and “cool”) tended to be the most effective for positively influencing the women’s self-efficacy (Rosenberg-Kima, Baylor, Plant, & Doerr, 2008).

However, the findings are not consistent. In a study with female participants and female agents, Zanbaka, Ulinski, Goolkasian, and Hodges (2007) demonstrate that the presence of a virtual agent inhibits the performance of participants on a mathematical task. Hayes, Ulinski, and Hodges (2010) found a similar decrease in performance with regard to male participants interacting with an agent of the same gender. However, the study demonstrated that when male participants interacted with an agent of the opposite gender their performance improved. The authors argued that participants experienced a stronger feeling of “being in the room with the agent” when interacting with an agent of opposite gender.

Although in sum there is evidence that female virtual agents have advantages especially for female students, more research is needed to fully understand the mechanisms — especially with regard to rapport and immediacy behaviors — and to clarify inconsistent findings.

2.5. The current study

In our study, we examine whether the presence of a virtual agent can enhance motivation and performance in STEM subjects and whether this influence depends on the virtual agent’s gender and its rapport behavior. As mentioned above, the virtual agent is not a tutor in the sense that s/he instructs the student how to learn but is supposed to motivate students by giving the chance to reflect on one’s own performance and by giving feedback. Based on the literature summarized above, we hypothesize that women will especially benefit from a female agent and that rapport will lead to positive effects:

**H1.** When the agent displays rapport behavior, participants perceive a stronger feeling of rapport compared to the agent displaying no rapport behavior.

**H2.** When the agent displays rapport behavior, motivation, effort and performance is higher than when the agent displays no rapport behavior.

**H3.** For women, the interaction with a female virtual agent causes higher increase in motivation, engagement and performance than when interacting with a male virtual agent.

3. Method

The goal of the study was to investigate whether a virtual agent can motivate participants and help to improve their performance in a mathematical task. For this purpose, we examined whether beneficial effects occur when participants interact with virtual agents and how gender of the agent as well as of the learner and the rapport displayed by the agent influence these effects. In order to test the assumptions that a female agent will be especially beneficial for women and that rapport will yield increased motivation and performance, we designed an experiment in which we manipulated virtual agents’ gender and rapport behavior. We recruited men and women and had them perform two mathematical tasks, one before interacting with an agent and one during the interaction. Each participant was either paired with an agent of matching gender or of opposing gender. We also investigated participants’ motivation and increase of effort to solve the math problems.

3.1. Participants

We recruited 128 participants (53.1% females) from the greater Los Angeles area. Their age ranged from 18 to 34 years with an average age of 23.85 (SD = 3.07). 16.0% of participants had a high school education, 5.3% completed some college, and 78.7% had a college education. Data was not collected on SAT completion or score, and college education included either two or four year degrees. Participants were recruited by responding to recruitment posters posted on craigslist.com and were paid $30. Data were collected between November 2011 and June 2012.

3.2. Design

We used a $2 \times 2$ full factorial between subjects design, with the first variable being the gender (male/female) of the agent and the second variable being whether or not the agent displayed rapport (rapport/no rapport). Participants were randomly assigned to one of the four conditions (it was, however, considered that participant gender was distributed equally across conditions). In all four conditions the agent showed a minimum of human-like nonverbal behavior by blinking and performing slight occasional posture shifts. In the conditions with no rapport, there was no additional nonverbal behavior added. In the rapport-condition, the agent additionally smiled and nodded corresponding to the participants’ behavior (verbal and nonverbal input). Additionally, gender of the participants was considered as a variable (see Table 1).
3.3. Procedure

After an explanation of the study and obtaining consent, participants were led to a private room where they completed the experiment individually. Participants were seated at a desk with two monitors that were positioned next to each other. They were then instructed to work on a mathematical task for 10 min, which was presented as a computer-based survey on one of the monitors. The second monitor was turned off at this point. To minimize self-presentation concerns, the anonymity of task performance results and the non-competitiveness of the task were emphasized. The experimenter left the room for the duration of the working period. Participants’ answers were not reviewed by the experimenter and they were not given any performance feedback. Next, the experimenter provided detailed (verbal and written) instructions on the following interaction with the virtual agent. It was emphasized that the agent is a computer program and that participants will be alone during their interaction period. The agent was then launched and displayed on the second monitor. The experimenter would leave the room before the interaction started and then would control the virtual agent’s speech over a separate computer in a different location, without the participants’ knowledge. The agent’s nonverbal behaviors were automated by the system according to the condition (rapport/no rapport) as described above. First, the agent asked how the participants estimated their previous performance on a 5-point Likert-scale (very poor to very good). Next, they were asked to rate how difficult they thought the previous math problems were on another 5-point Likert-scale (easy to hard). This was followed by the second task period that was part of the interaction. The agent would instruct the participants to load the task on the other monitor and work on it for a time period of 10 min. During that time the agent reminded the participants twice of the time remaining (5 min and 1 min left) and also let them know when time was up. Afterwards, the agent asked the participants to estimate their performance with regard to the second task on the same scale. During the last part of the interaction, the agent interviewed the participants concerning their experiences while working on the tasks. Then, the agent would announce the end of the interaction and would disappear from the monitor. Finally, situational motivation and demographic variables were measured in a post survey without the virtual agent visible or the experimenter present. Subsequently, participants were debriefed. During debriefing it was made sure that participants had not been aware during the experiment that the experimenter or any other human was involved and/or had any part in the interaction. The experiment lasted about 60 min.

3.4. Mathematical tasks

Two mathematical tasks were presented to the participants. Each task consisted of a set of 24 math problems comparable to original GRE and SAT math items (examples: The child care center charges $11 an hour plus a daily $3 drop-off fee. How many hours of childcare did Robert pay for if he dropped his son off 3 days last week and paid $130 at the end of the week? B) What is the value of the expression $3 + 4 \times 6 - 2(5 \times 8)$?). The math problems were selected out of a larger set of questions pretested with regard to their difficulty. By pretesting the problems we made sure that both tasks have approximately the same levels of difficulty and require the same sets of skills. Also, tasks had enough number of questions to prevent ceiling effects due to participants finishing all the math problems in less than 10 min. To avoid an improvement in performance in the second task due to simple learning/practice, the math problems were modified with regard to their wording and surface features. This way the first and the second task appeared distinct from each other while they still each required the same set of skills.

3.5. Rapport agent

Participants interacted with a female or male virtual agent with a human-like appearance. Four different characters were used: 2 male and 2 female (see Fig. 1) to control for possible effects of particular agents. We used the Rapport Agent developed by Gratch et al. (2006). To create rapport with the participant, the agent displayed positive listening behaviors (such as nodding and smiling) that correspond to the verbal and nonverbal behavior of a human speaker. Previous studies of the rapport agent have shown that it is highly capable of creating the experience of rapport comparable with a face-to-face condition (Gratch et al., 2006, 2007a, 2007b). To produce listening behaviors, the Rapport Agent first collects and analyzes audiovisual features from the speaker’s voice (silence, speech) and upper-body movements (head nod, smile, eye gaze) in real time. This happens via a microphone and a Videre Design Small Vision System stereo camera, which was placed in front of the participants to capture their movements. Watson, an image-based tracking library, uses images captured by the stereo camera to track the participants’ head position and orientation. Acoustic features are derived from properties of the pitch and intensity of the speech signal using a signal processing package, LAUN (Gratch et al., 2006). The Rapport Agent displays...
behaviors that show that the animated character is “alive” (eye blinking, breathing), and listening behaviors such as posture shifts and head nods automatically triggered by the system corresponding to participants’ verbal and nonverbal behavior. This allows the agent to provide contingent feedback while the speaker is speaking by following a response model (Huang, Morency, & Gratch, 2011) to decide which behavioral response would be most appropriate (such as head nod or smile). The different animations are converted into Behavior Markup Language (BML) (Kopp et al., 2006), sent to an action scheduler (to determine the duration of each animation) and passed on to Smartbody, an animation system that blends the different animations naturally into each other (Thiebaux & Marsella, 2007). The commercial game engine Gamebryo then renders the animations and displays them to the user.

3.6. Measures

Improvement in performance, increase in effort, motivation and rapport were measured as dependent variables by using a computer-based survey and audio recordings of participants’ verbal behavior during the interaction.

3.6.1. Improvement of performance

We calculated participants’ performance improvement by the difference in their performance before and during the interaction with the agent. For this purpose, we subtracted the number of math problems they solved correctly in the first task from the number they solved correctly in the second task. We used this relative performance improvement, assuming that participants have indifferent pre-knowledge, education level, practice and talent affecting their absolute performance but not their relative score during the manipulation.

3.6.2. Increase in effort

By subtracting the number of solved math problems in the first task from the number solved in the second task, we calculated the increase in effort. Here, we did not take into account whether the solutions were correct or not. This variable was interpreted as an additional indicator for their situational motivation.

3.6.3. Motivation

To measure participants’ motivational state with regard to the mathematical tasks we used the Situational Motivation Scale (SIMS) by Guay, Vallerand, and Blanchard (2000). The scale includes 16 items asking for the motivation to work on the tasks and comprises four subscales of four items each: Intrinsic Motivation (“Because I think the activity is interesting”), Identified Regulation (“Because I am doing it for my own good”), External Regulation (“Because I am supposed to do it”) and Amotivation (“There may be good reasons to do this activity, but personally I don’t see any”). All items are rated on a seven-point Likert-scale (1 = corresponds not at all to 7 = corresponds exactly) and were summed for further analysis (M = 60.20, SD = 11.94).

3.6.4. Rapport

The Rapport Scale (Kang & Gratch, 2012) with 39 items was used in order to assess how participants perceived the interaction and the Rapport Agent per se (e.g., “I felt I had a connection with the listener”; “The interaction was frustrating”; “The listener’s body language encouraged me to continue talking”). People were asked to evaluate the interaction on an 8-point Likert scale (strongly disagree — strongly agree). In order to conduct further analyses the 39 items were all combined into one variable (M = 3.72, SD = 0.66).
4. Results

We conducted all analyses using Statistical Package for the Social Sciences (SPSS, Version 18). Before starting the analyses of hypotheses, we verified that agent appearance did not affect the results. As anticipated, there were no significant differences between agents with the same gender but different appearances. Therefore, the data was collapsed for further analysis. Also, variances were not significantly different for any test (ps > 0.51), therefore the equal variances assumption holds for all outcome variables. Visual inspection of residuals through a kernel density plot demonstrated normality of residuals. As residuals are normally distributed, and thus the data do not violate this assumption, it is appropriate to use ANOVA.

First, we analyzed the effect of rapport on improvement in performance using a between-subjects t-test. Participants who interacted with an agent that displayed rapport behaviors improved their performance more (M = 1.65, SD = 1.88) than those who interacted with an agent that did not engage in rapport behaviors (M = 0.57, SD = 2.17, t(126) = 3.00, p = 0.003, d = 0.59). Participants who interacted with the rapport-agent also increased their effort (in the sense that they attempted to solve more problems in the second working period compared to the first; M = 1.75, SD = 2.42) more than those who interacted with an agent that did not engage in rapport behaviors (M = 0.79, SD = 2.54, t(126) = 2.19, p = 0.03, d = 0.42). Although results confirmed this expected effect of rapport on improvement in performance and effort, there were no effects of rapport on self-reported rapport (t(125) = -0.40, p = 0.69, d = 0.07) or motivation (t(126) = 0.06, p = 0.95, d = 0.01). Therefore, there is no support for H1 suggesting an improvement on self-reported rapport while H2 is supported for performance and effort (but not for motivation).

To test H3 stating that the effect of female role models among women in the STEM domain could be moderated by presence of a same-sex agent who displayed rapport behaviors, we conducted a 2 (participant gender: male vs. female) × 2 (agent gender: male vs. female) × 2 (rapport: rapport vs no rapport) ANOVA on performance improvement. This analysis yielded a significant main effect of rapport (F(1,120) = 10.23, p = 0.002, d = 0.59), as described above. The three-way interaction was also significant on the 10% level (F(1,120) = 3.35, p = 0.07, d = 0.33). As depicted in Fig. 2a, among men, there is only an effect of rapport among those exposed to the female agent, however, no effect of rapport among those exposed to the male agent. Likewise, for women, rapport only has an effect among those exposed to the male agent (Fig. 2b).

Although unexpected, the three-way interaction suggests an effect of rapport for participants exposed to agents of the opposite gender. To test this, we ran an exploratory 2 (rapport: rapport vs no rapport) × 2 (gender matched agent: opposite vs same gender) ANOVA on improvement in performance. This analysis revealed an interaction on the 10% level of significance (F(1,124) = 3.64, p = 0.059, d = 0.35), reflecting a significant effect of rapport on performance improvement among participants exposed to agents of the opposite gender.

![Fig. 2. a. Effects of rapport and agent gender on performance for men. b. Effects of rapport and agent gender on performance for women.](image-url)
participants who interacted with an opposite gender agent (M = 1.97, SD = 1.75 vs. 0.12, SD = 2.34; t(124) = 3.43, p = 0.001, d = 0.60), but no significant effect of rapport among those who interacted with an agent of the same gender (M = 1.37, SD = 1.97 vs. 0.89, SD = 2.01; t(124) = 1.10, p = 0.32, d = 0.18; see Fig. 3). The confidence intervals around a) the difference between rapport and no rapport of male agent for men (Fig. 2a), b) the difference between rapport and no rapport of female agent for women (Fig. 2b), and c) the difference between rapport and no rapport of same gender agent (Fig. 3) all contain zero (−0.91 to 1.91, −0.76 to 1.74, and −0.44 to 1.40, respectively), confirming that differences are not significant.

We confirm that this rapport X gender-matching interaction remains for both men and women with a 2 (participant gender: male vs. female) × 2 (gender matched agent: opposite vs same gender) × 2 (rapport: rapport vs no rapport) ANOVA. Indeed, this effect holds for both males and females as no significant three-way interaction emerges from this analysis (F(1,120) = 0.40, p = 0.53, d = 0.11), while the two-way rapport X gender-matching interaction remains (F(1,120) = 3.66, p = 0.058, d = 0.35), as does the main effect of rapport (F(1,120) = 10.46, p = 0.002, d = 0.59).

We conducted 2 (participant gender: male vs. female) × 2 (agent gender: male vs. female) × 2 (rapport: rapport vs no rapport) ANOVAs on effort improvement, as well as on self-reported rapport and motivation. While the main effect of rapport on effort improvement (F(1,120) = 5.23, p = 0.02, d = 0.42) was again significant, there were no other significant effects for effort improvement (Fs < 1.56, p > 0.21, d < .23), and no significant effects at all for self-reported rapport (Fs < 1.84, p > 0.17, d < 0.25) or motivation (Fs < 2.80, p > 0.10, d < 0.31).

Therefore, there is no support for H3, but interestingly, performance yielded significant results, while there were no significant effects of agent gender for self-reported rapport, motivation and effort improvement. The significant results, however, are opposite of what was assumed in the corresponding hypothesis.

5. Discussion

Our results yield several interesting insights. The aim of the study was to demonstrate the effectiveness of rapport during instructional communication in fostering performance and effort in STEM learning as well as to find further evidence for the advantage of same sex instructors for women. While the former was clearly supported, the latter was not only unsupported but the opposite of our assumptions was substantiated. In the following, we will discuss these results in greater depth.

As hypothesized, rapport was able to foster performance and effort. Those participants who interacted with an agent who displayed rapport behavior showed a better performance and a larger increase in effort compared to those participants who interacted with an agent who did not show any rapport. This yields further support to the notion that by providing reciprocity and positivity (Gratch et al., 2007a; Tickle-Degnen & Rosenthal, 1990), rapport is able to motivate students in a similar way as positive nonverbal behavior in teaching contexts does (Andersen, 1979). Likewise, research in social psychology has shown that establishing rapport between people and their instructors in face-to-face interactions increases desirable outcomes such as motivation and improvement in task success (Granitz et al., 2009; Thomas et al., 1982). Our results show that rapport has a similar positive effect on performance in human-computer-interaction.

However, although the availability of rapport behavior influenced the participants’ behavior, there was no influence on subjective experience, neither on self-reported motivation nor on perceived rapport. Here, it has to be considered that other studies also tended to show more pronounced effects on participants’ behavior instead of participants’ perception and subjective experiences (Gratch et al., 2006; Huang et al., 2011; Von der Pütten et al., 2010). Therefore, rapport seems to have the potential to be influential even without people consciously noticing it. The same might be true here: while rapport behavior of the agent impacts participants’ behavior, they do not notice rapport to be present. It is possible that more pronounced differences in agent rapport behavior between conditions would allow the rapport to be perceived, but this remains to be tested in future research.

Additionally, it is important to consider the effect sizes observed. We find a medium (d = 0.59) effect of rapport behavior on participants’ performance and a small (d = 0.42) effect of rapport behavior on participants’ effort. These effect sizes are...
consistent with the range from medium to small effects found in other work on computer human-interaction (e.g., Gratch et al., 2007a; Lucas, Gratch, King, & Morency, 2014). While the effect on performance is larger than the effect on effort, given that effect sizes are merely estimates of effect size based on samples, it is possible that -in the population- the effect sizes are even more similar than they appear to be here. So, readers should hesitate to draw the conclusion that the effect on performance is larger than the effect on effort and further research should be considered first to replicate this difference. While these are not large effects by standard conventions (Cohen, 1988), the manipulation used is also fairly subtle. Participants in the rapport and non-rapport conditions experienced exactly the same experimental setup; the only difference was the non-verbal and verbal behavior of the rapport agent included rapport-building. The small to medium effects are more impressive in light of the subtlety of the manipulation.

While the main effect of rapport was in line with the hypotheses, the results with regard to the gender of the agent was opposite to our hypothesis. However, our hypothesis regarding the potential influence of the virtual instructor’s gender was based on inconsistent prior research. Although several studies demonstrate the positive effect of female role models in STEM fields (Cheryan, Siy, Vichayapai, Drury, & Kim, 2011; Stout et al., 2011), others indicate that academic motivation and engagement does not significantly vary as a function of the teacher’s gender or even show that female teachers may hinder girls’ STEM success (Betz & Sekaquaptewa, 2012; Martin & Marsh 2005). The latter is now supported by our data in the sense that the dyads which are NOT gender-matched lead to most performance and effort. It has to be acknowledged, though, that the role of the virtual agent here, was not the one of a competent role model as mentioned in the literature but the role of a motivating interviewer. Therefore, we can at least conclude that the mere presence of a woman in a learning situation is not sufficient to lead to increased motivation and learning. This does not preclude, however, that results might be different when presenting a competent role model.

More specifically, the results indicate that when rapport was displayed, participants’ improvement was higher when they interacted with an agent of opposite gender than with an agent of matching gender. This indicates that a facilitation of performance only occurs under a certain gender condition, i.e. opposing gender, and when rapport is displayed by the agent. Participants’ performance improved most when they interacted with a virtual agent of opposite gender that displayed rapport and it improved least with an agent of the opposite gender who did not display rapport. This indicates that rapport has an effect on participants’ improvement in performance only when the agents’ gender does not match their own. Rapport is therefore especially beneficial when it is shown by an agent of opposite sex (e.g. a male agent talking to a woman), while an agent of opposite sex who does not display rapport, is detrimental. While this seems to be plausible for women (who might be positively surprised and especially encouraged by a rapport-displaying male agent asking her about her experiences and success, while they are appalled by a male character who does not show rapport), the pattern is more difficult to explain for men. Here, men might not have been pleasantly surprised because reciprocity and positivity is attributed to women qua stereotype. Instead, it is more likely that they would simply react especially positive towards a rapport displaying woman compared to a rapport displaying man — and additionally do not want to disgrace themselves in the presence of a responsive woman.

However, the specific mechanisms have to be targeted in greater detail in future studies. As yet, it is not sufficiently clear whether rapport fosters self-reflection and by this self-efficacy, or whether the rapport directly influences motivation which increased effort to perform well by attempting to solve a higher number of math problems.

5.1. Limitations

The study certainly has some limitations: the most important limitation probably is the laboratory situation which questions the generalizability to every day instructional communication. Especially as solving the math tasks did not have actual consequences for the participants’ future, their behavior and experiences might have been very different from real life situations.

Also, our results are confined to the specific learning situation that was tested: the virtual agent merely acted as an observer who triggered reflections on self-efficacy. The results therefore cannot be transferred to a classic tutor system. With regard to clarifying the mechanisms of female competent role models it can, however, be concluded that the mere presence of a woman is not sufficient for women to feel and learn better.

Additionally, the sample was mixed and heterogeneous in age and education. Although this usually is a strength, it could have been detrimental here because some participants might have been too different from the group of people this research is meant to be most relevant for, which is high school students or first-year college students.

With regard to sample size, power analyses reveal power of 85% for performance, but only 59% for effort. Therefore, although we had a sufficient sample size to detect a medium effect, we were underpowered (by conventional standards of 80% power) to detect a small effect such as the one we found for effort. Thus, if future research efforts were undertaken to replicate the current results (especially for effort), we suggest an increase in sample size. Specifically, traditional estimates of power would lead one to recommend a sample size of 212 (106 per condition); however, other perspectives on power suggest that an even greater number of participants may be needed for replication (e.g., Maxwell, Lau, & Howard, 2015).

Also, we have to mention and consider some violations to normal distribution in the data: Shapiro-Wilk tests reveal that, while rapport and motivation are normally distributed (Shapiro-Wilks > 0.98, ps > 0.19), improvement in performance and effort were not (Shapiro-Wilk = 0.96, p = 0.001, and Shapiro-Wilk = 0.97, p = 0.02). Despite this violation of normality assumption for ANOVA for performance and effort, we were able to find significant effects for these variables. On the variables
where we found no significant effects (rapport and motivation), normality held; thus, finding non-significance with rapport and motivation cannot be attributed to violating this assumption. With regard to the appearance of the agent, several aspects have to be reflected critically. First, although we tried to alleviate the effects of the specific appearance by using two different male and two different female characters, the characters still represent a specific type of virtual agent that may not resemble all participants. This may be important since it has been demonstrated that virtual characters that resemble the user are assessed more positive (Iacobelli & Cassell, 2007) and therefore are probably more influential. Specifically an “in-group” member agent is perceived as more intelligent and competent – even when it has the same narratives as a virtual character that does not resemble the user. In our study, several participants, for example with different ethnicity from the agents, might not have felt similar to the agent.

Finally, our results are derived from an interaction which lasted less than an hour. Especially in the learning and instruction realm, usually instructor-learner-relation exist for longer periods of time. Therefore, long-term studies are needed.

6. Conclusion

In summary, contributions of this work are three-fold. First, the study adds to literature on human-computer-interaction by showing that virtual agent rapport is an important factor for achieving desirable outcomes such as effort and performance with regard to mathematical tasks. Virtual Agents with rapport can contribute to improve people’s performance, specifically with standardized math tests. This observation may support the development of useful and effective applications in mathematical education and training, such as virtual instructors, tutors or learning companions. Second, the study shows that an agent of opposing gender, not a same sex agent, is most successful in boosting participants’ performance. In this context, the results of this study may motivate future research on gender differences with regard to effects of specific gender conditions. Finally, this work makes a methodological contribution to the fields of experimental psychology and human-computer-interaction.

Acknowledgements

This research was supported by the National Scientific Foundation under grant # IIS-0916858. Two of the authors received support by the PROMOS Program of the German Academic Exchange Service (DAAD).

References
