Computers & Education 63 (2013) 176–196

Contents lists available at SciVerse ScienceDirect

Computers & Education

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

Learning by teaching with virtual peers and the effects of technological design choices on learning

Sandra Y. Okita*, Selen Turkay, Mihwa Kim, Yumiko Murai

Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027, USA

A R T I C L E I N F O

Article history:
Received 22 June 2012
Received in revised form
3 December 2012
Accepted 3 December 2012

Keywords:
Virtual reality
Teaching/learning strategies
Cooperative/collaborative learning
Computer-mediated communication
Interactive learning environments

A B S T R A C T

Advancements in technology have brought about new forms of learning and online instruction that allow communication through virtual representations without physically meeting in person. This study builds on previous work involving recursive feedback that tests the hypothesis that an important facet of learning-by-teaching is the opportunity to watch one's pupil perform. Sixty graduate students examined the value of *recursive feedback* that occurred when tutors observed their pupil subsequently apply what they had been taught. The study took place in the virtual environment Second Life where adults tutored another adult about human biology through their virtual representations. The tutors who observed their pupil avatar interact with an examiner exhibited superior learning relative to several control conditions that included learning-by-teaching elements but not recursive feedback. The second study examined the effect of popular design choices on recursive feedback during learning-by-teaching (e.g., customization, look-alike features). The customization condition involved tutoring a pupil avatar that the participant customized prior to the study and observing the pupil avatar answer questions. The doppelgänger look-alike condition involved tutoring a pupil avatar that looked like the participant and observing the pupil avatar answer questions. Results showed that conscious awareness of look-alike features and the extent to which one customizes the pupil avatar influences learning.

1. Introduction

Advancements in communication technologies have developed opportunities for students to learn both online and offline. Students are no longer restricted to physical space, as they can remotely access online classes from afar. Learning environments can use virtual reality technology to share real-time audio, video, text, and graphical information among learners (Dede, Whitehouse, & Brown-L'Bahy, 2002). In the past, collaborative learning methodologies (e.g., peer learning) were mostly restricted to face-to-face interactions by sharing physical space. Now, successful non-commercial virtual reality environments (e.g., Second Life, Active Worlds) provide space to support online collaborative activities without meeting in person. Learners can remotely interact through their avatar representations, create virtual objects together, and engage within various thematic environments.

The features and design choices available in virtual environments can provide an ideal learning situation tailored to each student's interest. For example, providing the learner with an option to observe the class from different points of view (e.g., front of the class, first person, third person, bird's-eye view), or automatically shifting the seats in virtual classrooms based on student attention level (Bailenson, Yee, et al., 2008). Multi-user virtual environments have also provided features that permit novel investigations of what it means to be social, as people communicate through virtual avatars instead of meeting in person. For example, one way to address learner's needs may be to change the virtual representation of the teacher (or peers) so the learner can feel comfortable and communicate in the most optimal way (e.g., same gender, age, ethnicity, multicultural or specific cultural appearance). Although technology can provide environments difficult to achieve in the real world, successful use of technology often requires a compatible method of learning to achieve benefits. Careful considerations are required as learning can easily vary depending on learning situations and choices of technology.

* Corresponding author. Tel.: +1 212 678 4165, +1 650 799 6439 (mobile); fax: +1 212 678 8227.
E-mail addresses: so2269@columbia.edu, okita@tc.columbia.edu (S.Y. Okita), st2282@columbia.edu (S. Turkay), mk2734@columbia.edu (M. Kim), ym2336@columbia.edu (Y. Murai).

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2012.12.005

In this paper we focus on recursive feedback, which refers to the information that flows back to tutors when observing their pupils' subsequent performances. Recursive feedback came to our attention during research on Learning-by-Teaching (LBT). The desired side effect of LBT was that students would learn especially well when they took on the role of the teacher (e.g., Palincsar & Brown, 1984). Teaching was a good context for recursive feedback, because teachers could learn by observing how their students applied their teachings. Of special interest to our current work are the effects of teaching on the tutor's own learning, as research has found that peer tutoring influences tutors' academic development, self-concept, responsibility, and deep understanding of content knowledge (Cohen, 1986; Sherin, 2002). Our research follows previous work on recursive feedback during LBT where adults tutored another adult pupil in a face-to-face interaction. The study found that observing their pupil independently interact with others exhibited superior learning relative to several control conditions that included LBT elements but not recursive feedback (Okita & Schwartz, submitted for publication). A separate study also tested the effects of recursive feedback during LBT where participants tutored a computer-controlled Teachable Agent (Biswas, Schwartz, Bransford, & TAG-V, 2001; Blair, Schwartz, Biswas, & Leelawong, 2007). Here, adults tutored a pupil agent where teaching meant programming the agent's reasoning. Participants observed their pupil agent's performance as it competed in a logic game with another agent. Participants who taught and observed their agent showed higher performance on logical reasoning, than those who played the game themselves. Here, "agents" refers to virtual graphic representations controlled by a computer, and "avatars" as virtual computer graphic representations controlled remotely by humans.

The current study is a direct application of the previous findings to an online virtual learning environment (i.e., Second Life), to see whether recursive feedback can be observed even if tutors never physically meet their pupils. The tutors interact with their pupils through virtual graphic representations (i.e., avatars). This "never-in-person" interaction is quite common among online communities, as people prefer to stay anonymous in multi-user gaming environments. The findings will have important implications for similar situations, as more virtual reality learning environments and distance-learning programs become available. Although technology may have increased online learning and collaboration, many researchers are still concerned whether the limits in social interaction will impede learning, motivation, and course completion rates (Hara & Kling, 2001; Sikora & Carroll, 2002).

This paper will first give an analytic review of the LBT literature and recursive feedback. The first experiment will examine whether the value of recursive feedback extends to online virtual environments where the tutor and pupil never physically meet. We attempt to do this by experimentally isolating the effects of recursive feedback during LBT. Another important characteristic of virtual reality technology is how people can create an ideal social situation for learning by making specific design choices (e.g., customization, look-alike features). To see the implications for learning, the second experiment explores the influence of design choices on recursive feedback during LBT, specifically customization and look-alike (doppelgänger) features. To avoid confusion going forward, we will use the terms teacher or tutor when referring to the student engaged in tutoring, and use the term pupil to refer to the student receiving the tutoring. A general discussion section, future research, and summary will follow.

2. Three phases of learning by teaching

The idea of LBT is not new. LBT in the form of peer tutoring or students teaching another student has yielded strong results across a broad range of educational practices in the form of tutoring, cooperative learning, peer learning, and reciprocal teaching (Dufrene, Noell, Gilbertson, & Duhan, 2005; Graesser, Person, & Magliano, 1995; Palincsar & Brown, 1984; Rittschof & Griffin, 2001). There are many variations of peer-tutoring that involve mixed and matched ability groupings, ages, and genders, as well as the degree of support students receive to tutor well (Chi, Silver, Jeong, Yamauchi, & Hausmann, 2001). Research has found positive effects in tutor learning, such as academic development (Morgan & Toy, 1970), self-concept (Fantuzzo, Riggio, Connelly, & Dimeff, 1989), social listening skills (Greenwood, Carta, & Hall, 1989), responsibility (Cohen, 1986), and deeper understanding of content knowledge (Sherin, 2002).

As a way to understand the source of the benefits experienced by student tutors, we decompose LBT into three phases: Preparing to Teach, Teaching, and Recursive Feedback. We have highlighted the positive effects of the preparation and teaching phases from the LBT literature. Although we emphasize them separately, the three phases can easily be interwoven over various time frames, as there are different ways that students might teach each other. The analytic separation helps to pinpoint possible causes. Here, the literature is used to set up the hypotheses at test, and is linked to the description of the experimental conditions and the rationale.

2.1. Prepare to teach

The first phase is preparing to teach. Research showed that the preparation phase had demonstrable benefits for learning, as students who prepared to teach learned more than students who prepared to take the test themselves (Bargh & Schul, 1980). A similar comparison was conducted with a longer two-week intervention, where Benware and Deci (1984) found that the benefits of teaching expectancy appeared in higher-level items that required greater conceptual understanding, while no difference was found with low-level rote material. The effects of learning-by-teaching are often attributed to the teaching activity itself, but just the expectancy to teach can invite cognitive processes (e.g., generation and motivation) that contribute to deep learning.

For example, in advance of interacting with pupils, tutors need to think about what they will convey. This encourages tutors to actively retrieve what they already know, and prepare for any shortcomings. This effort at retrieval creates a generation effect (Slamecka & Graf, 1978), which is known to improve future retrieval. Karpicke and Blunt (2011) showed that the retrieval involved in taking a test improves students' subsequent ability to draw inferences, because students could more easily remember the facts needed for the inferences. Another is motivation, where preparing to teach seems more intrinsically motivating than preparing to study oneself. Deci and Ryan (1980) found that when students approach the material with an anticipation of using it, students are motivated to become more fully involved. When students are given materials to be tested, the learning becomes passive, where factual knowledge may be covered, but less effort is seen in interpretation and application. Strong motivational aspects of LBT also need to be used with care, as they can sometimes elevate into performance anxiety (Renkl, 1995). Garbarino (1975) also mentions that taking an already interesting task like tutoring, and mixing it with other extrinsic incentives can interfere with the student's intrinsic motivation.

2.2. Learning by teaching

The second phase involves actual teaching. Numerous studies have looked at whether different tutoring approaches, and monitoring abilities of tutors (Chi et al., 2001) can further extend benefits from tutoring (cf. Roscoe & Chi, 2007 for a more comprehensive review). The findings suggest that LBT provides opportunities to reformulate and extend the use of peer tutoring. LBT triggers similar cognitive processes that regulate the benefits of preparing to teach, except that they occur in real-time interactions (e.g., giving explanations and responding to questions).

When tutors interact with pupils, they need to give explanations and respond to pupil's questions. Ploetzner, Dillenbourg, Praier, and Traum (1999), describe this as an explanation effect referring to benefits of explaining ideas rather than receiving them. In a classic correlational study of cooperative interactions, Webb (1989) found that students who provided explanations learned more than those who received them. Providing explanations seemed most effective when it involved constructing causal connections between procedures and their effects (Siegler, 2002), or between structural, functional, and behavioral aspects of systems and sub-systems (Chi, 2000). Siegler (1995) conducted a study with children that explored whether active explaining is correlated to learning. Results show positive correlations between the amount of explanation and the amount of learning. Regardless of student ability, having students explain more about the answer resulted in greater performance than feedback alone.

In addition to explanations, tutors often answer questions. Borko et al. (1992) provide a classical example where questions have a significant effect on tutor learning. A pre-service teacher was asked by a student why fractions get inverted for multiplication, and the teacher realized that she did not know the conceptual reason. Questions asked by pupils can help teachers recognize and repair their own gaps in understanding or to consider new extensions to their understanding. Roscoe and Chi (2007) also found that "tutees [questions] were responsible for about two-thirds of tutors' reflective knowledge-building activity" (p. 23). The quality of the questions students ask can also have a measurable effect on tutor learning (Uretsi, 2000).

Explaining your own reasoning, and responding to questions, can easily exercise its effects on cognitive processes in generation, metacognition, and motivation. As tutors teach, they generate various explanations (e.g., re-phrasing, apply different instructional methods) while observing their student's response. Tutors are also trying to make sense of the real-time feedback their pupils are showing toward them (e.g., confused face, wrong answer, questions). Trying to explain one's reasoning, or explaining what is written in textbooks (Chi, deLeeuw, Chiu, & LaVancher, 1994), helps tutors generate a greater number of strategies compared to preparing to teach with no feedback. In relation to metacognitive processes, when tutors are responding to pupil's questions, more effort is made to search for in-depth explanations, and identifying gaps in their own understanding that lead to deeper cognitive processing in tutors. Through the interaction with pupils, tutors may identify approaches that are more effective than their initial choice, and decide to replace ineffective ones. Generating more responses and increasing the degree of engagement will naturally motivate the tutor, as learning becomes more enjoyable when what you are studying makes more sense (Siegler, 2002).

2.3. Observing recursive feedback

The third phase involves tutors learning from observing recursive feedback. Tutors teach and then receive feedback by observing their pupils' post-instruction performance. Recursive feedback is the information that flows back to tutors when observing their pupils' subsequent performances. One way to describe *recursive feedback* is to compare with *direct feedback*.

There are many characterizations and categorizations of feedback, which is often defined as a direct consequence of a person or agent's behaviors, thoughts, or actions (e.g., Hattie & Timperley, 2007; Kluger & DeNisi, 1998; Lawton et al., 2012; Mory, 2004; Narciss, 2008; Nicol & Macfarlane-Dick, 2006). In the top panel of Fig. 1, we refer to the *direct feedback* loop. The information or reinforcement can come from another person, a computer program, a book, or any aspect of the environment that can signal a discrepancy from a standard. The bottom panel of Fig. 1 shows *recursive feedback* that is not a direct consequence of tutor behavior. There are two key elements. Recursive feedback is

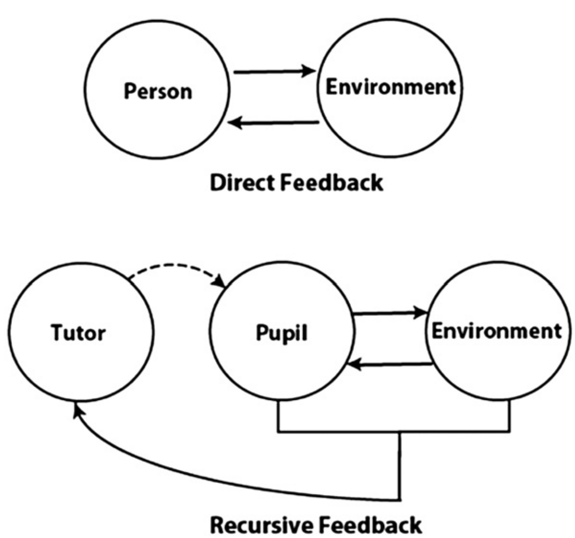

Fig. 1. Direct feedback (top panel) and recursive feedback (bottom panel).

generated by the independent interaction of the pupil and the environment. Recursive feedback occurs when tutors observe their pupils independently use what they have been taught in a relevant performance context. The second element is the recursive mapping. Teachers need to map the independent interactions of their pupils to their own understanding of the task and the desired standard of performance, otherwise informative feedback will not flow through to the teacher.

In the first phase, prepare to teach, the tutor's preparations are based on the anticipated consequences of their actions (i.e., tutoring). In the second phase, teaching, tutors see the consequences of their actions during the interaction with their pupil. When teaching, tutors need to give an immediate response to the student, as the interaction is in real time. Tutors do not have much time to think about what they observe in the student, or reflect on their understanding. In the third phase tutors teach and then receive recursive feedback by observing their pupils' post-instruction performance. When the tutor is receiving recursive feedback, the tutor is not engaged in a dual process of teaching and observing the pupil. The tutors may have more cognitive capacity to interpret the performance of their pupil, and the information being exchanged during the pupil's subsequent interaction. The tutor can search more deeply for explanations, and interpret the implications of feedback for their own performance and understanding. The uptake of recursive feedback differs from pure observational learning (e.g., Bandura & Jeffery, 1973), as the tutor does not simply observe the performance of the pupil; rather, the tutor also has a standard of performance that is based on his or her own understanding of the task. When the pupil falls short of this standard, this puts the tutor in a position where they must make sense of the observed phenomena, and explain to themselves what they have seen. The feedback reflects what the tutor taught the pupil, and in the process, the tutor's own understanding.

2.4. *Strengths of recursive feedback in addressing challenges in direct feedback*

It is important to recognize that feedback is not always effective. For example, the delivery of praise for having exceeded a standard can demotivate student learning. Generic praise often does not provide specific information for improvement and it may also lead students to lower their efforts because they may feel they have already exceeded the standard. In general, there are two major threats to feedback success.

The first threat is affective. Students often interpret feedback with respect to themselves rather than as task feedback (Kluger & DeNisi, 1996). Task feedback refers to information relevant to performing or understanding the task or topic. It is a natural inclination for feedback to trigger the involvement of a student's ego. Students like positive feedback precisely because it makes them feel good about themselves, even if it has little effect on learning (Hattie & Timperley, 2007). Hattie and Timperley (2007) explain that when a student takes feedback episodes as directed at the self, the episodes "have a negative effect on learning because they include or lead to self-handicapping, learned hopelessness, or social comparison. The related [task] feedback itself is usually discounted or dismissed, and goals of low challenge are adopted" (p. 97). Similar results were found as Mangels, Butterfield, Lamb, Good, and Dweck (2006) observed how pupils frequently attribute negative feedback to their fixed intelligence rather than as an informative signal for how they can improve.

The second threat is cognitive complexity and interpreting the implications of feedback for one's performance or understanding. This is related to prior knowledge dependence. When a task is simple, students can more readily map the feedback into the components and goals of a task, because they have the prior knowledge that makes the task seem "simple." In contrast, for a complex task, prior knowledge is typically lower, so it is more difficult for learners to determine the precise implications of the feedback with respect to improving their understanding or performance (Chi et al., 1994). Kluger and DeNisi (1998) state that their "analyses suggest that the effects of FIs [feedback interventions] grow more positive either as the task becomes more subjectively familiar or as it becomes objectively familiar" (p. 71).

The recursive aspect of feedback during LBT may help mitigate both threats. On the affective side, recursive feedback may circumvent ego involvement, because tutors focus on their pupils rather than solely themselves. In their studies of the protégé effect, Chase, Chin, Oppezzo, and Schwartz (2009) described how students were likely to acknowledge their Teachable Agents' errors and to express negative affect (e.g., "Oh no! I'm sorry I didn't teach you that"). In contrast students who answered the same questions themselves were inclined to show little affect upon errors. They suppressed uptake of the negative feedback, because, presumably, it had implications for themselves. The authors proposed that teaching creates an ego-protective buffer, because the pupil rather than the tutor, takes immediate responsibility for being wrong. As a result, tutors can pay attention to information that is relevant to improve learning because the pupil rather than the tutor takes the immediate responsibility for being wrong. In addition, teaching brings a sense of responsibility, so that tutors are motivated to pay attention to task-level feedback to improve their own understanding so they can teach better. On the cognitive side, LBT helps develop the prior knowledge needed to make recursive use of feedback. In the preparation for teaching and the teaching phases, the tutors have a chance to develop their knowledge. As a result of developing prior knowledge and knowing what they taught, tutors can make better sense of the implications of feedback for their own understanding. Also, because the goal is to improve the pupil, the tutor ideally thinks how to reorganize their teaching (and understanding) so they can tutor more effectively.

In previous work, one study examined whether recursive feedback supported effective tutor learning in face-to-face interaction with adults. College students who taught and observed their pupil exhibited superior learning of human biology relative to several control conditions that included LBT but not recursive feedback (Okita & Schwartz, submitted for publication). When given an additional set of questions that required deep understanding, participants in the recursive feedback condition were more successful in answering the questions. Participants who took on the role of tutor, experienced recursive mapping that improved their own learning. Another study examined whether the benefits of recursive feedback extend to teaching computer-controlled agents in regular classrooms. Here, LBT referred to students teaching computer-controlled agents. High school students played games where they induced logical rules to describe the relations between causes and outcomes. Students taught their agent the governing rule and received recursive feedback when they observed their agent play a prediction game against a second "evil" agent. On a posttest, students, who taught and observed their agent, exhibited greater abilities to use logic to solve novel problems compared to control conditions in which students received direct feedback by playing against the evil agent themselves (Okita & Schwartz, submitted for publication). Moving forward, the current studies attempt to explore whether the value of recursive feedback extends to online virtual environments where students tutor other students through their virtual representations (i.e., avatars), but not in person. Here, LBT refers to students teaching other students through their virtual representations.

2.5. Multi-user virtual environments as unique learning environments for recursive feedback

People learn from various sources. Traditional sources involve learning from humans or objects (e.g., books), while recent sources may involve computerized people (e.g., pedagogical agents and avatars) and/or computerized instructions (e.g., intelligent tutoring systems). Social interactions also occur in various settings. Traditional settings involve face-to-face interactions in both formal and informal environments (e.g., classroom, private tutor), while recent settings can involve online learning environments (e.g., video conferencing systems such as Adobe Connect, virtual reality environments such as Second Life). Virtual reality and communicative technology seem to have the potential to strengthen recursive feedback during LBT at two levels, the environment level, and the social interaction level. At the environment level, the use of technology-mediated tools can create unique learning situations. Virtual reality technology has created successful learning environments such as Second Life and Active Worlds which allow users to share graphical information in real-time. Virtual environments allow students to build, participate, and share their creations with others within the environment. Learners can build their own simulated world (e.g., ecosystem) rather than passively participate in a given environment. Students can visualize, reason, and directly manipulate their environment to identify causal chains linked to their actions. The graphic nature of virtual environments offer a powerful tool to create a learning environment tailored to fit learner needs.

The growing number of applications that use virtual environments extends to military training. In the past, virtual military training environments were mostly limited to flight simulations. Current applications include combat training, negotiation with foreign civilians in the Middle East, and creating personalized environments to treat post-traumatic stress disorder (PTSD) (Rizzo et al., 2009), fear of heights (Hodges, Anderson, Burdea, & Hoffmann, 2001), fear of spiders (Juan et al., 2005), and eating disorders (Riva, Bacchetta, Baruffi, Rinaldi, & Molinari, 1999). Treating PTSD and phobias involves a delicate balance of the virtual and the real, where familiarity is important but the circumstances cannot be too real or too fake. The learning environment should be just familiar enough to trigger a response and let the learner imagine and “fill in the blanks.” The virtual environment can also be used as an assessment tool to examine student behavior and performance. Rizzo et al. (2006) created a virtual classroom simulation environment and assessed student behavior for rehabilitation purposes. The virtual classroom simulation environment controlled different stimuli in (and around) the classroom (e.g., sound of footsteps, classmates giggling, passing notes, teacher writing on chalkboard, noise in the hallway, car passing outside the window) so that attention processes could be systematically assessed in children with attention-deficit/hyperactivity disorder (ADHD). The head-tracking sensor not only revealed the student’s susceptibility to certain stimuli, but also displayed the student’s first-person perspective, allowing the teacher to experience firsthand what ADHD students experience. Although virtual reality technology has the potential to provide the learner with an ideal environment, we mustn’t forget that learning inextricably depends on successful interaction with humans (MacKenzie & Wajcman, 1985; Stahl & Hesse, 2006).

At the social interaction level, the use of virtual reality technology can create unique social situations that may have interesting implications for learning. Computerized people often have a human-like feature that elicits a social response. Using virtual reality technology, peers can be represented by a computer graphic character that they remotely control in a virtual reality environment (e.g., an avatar). While human facilitation can indeed be effective, students may prefer specific learning situations/environments that are not readily available in a physical environment. For example, the teacher can be represented differently (e.g., with or without eye contact, same gender, age, or ethnicity of the student) to communicate with a student in the most optimal way (Beall, Bailenson, Loomis, Blascovich, & Rex, 2003). The student can also experience different points of view in the classroom (e.g., multicultural atmosphere, first person, third person, bird’s-eye view), or be seated at desks that move based on their attention level (Bailenson, Yee, et al., 2008). Teachers can set up quests for learners (Barab, 2006) in virtual environments, or have students engage in interactive stories with computerized story-telling agents (Ryokai, Vancelle, & Cassell, 2002). Studies have shown that believing that a real person is remotely controlling the avatar leads to more learning, than thinking that the virtual character is controlled by a computer (Okita, Bailenson, & Schwartz, 2007). An example that combined the unique characteristic of virtual reality at both the environment and social interaction level was the study by Bailenson, Patel, et al. (2008). In their study, students were trained tai chi movements in a virtual learning environment. The student and instructor were physically apart but remotely connected online through the Internet. The student and instructor engaged in tai chi movements through their virtual avatars. The student-controlled avatar followed the tai chi movements of the instructor-controlled avatar in real time. The student was also able to overlap his avatar onto the instructor’s avatar to adjust his movements in real time. Such unique teacher–pupil interactions may only be possible in an online virtual environment. The critical question is whether the value of recursive feedback extends to online virtual environments where students interact through virtual avatars, and not in person.

2.6. Overview of the two studies

The first study examined whether similar effects of recursive feedback extends to tutor learning in virtual environments. The study attempted to demonstrate that recursive feedback was important for maximizing the benefits of LBT. College students tutored other students through virtual avatars. The topic of learning was causal pathways by which humans maintained a fever. The primary treatment employed all three phases of LBT, while the control conditions consisted of LBT models that did not include recursive feedback. For example, in one control condition, the tutor avatar prepared and taught, but did not observe their pupil avatar. The experimental design isolated the potential value of recursive feedback through an extraction study that systematically removed various phases of the full LBT cycle.

The second study examined implications of design choices on recursive feedback during LBT, specifically customization and look-alike (doppelgänger) features. Technology has helped people develop an ideal learning situation through design choices and preferences of virtual avatars and agents, however little is known about its implications for learning.

3. Experiment 1: recursive feedback during LBT through virtual representations

In the first study, we isolated the potential value of recursive feedback through an ablation study that systematically removed various phases of the full LBT cycle. Fig. 2 represents the experimental design of the study. The conditions were treated as a 2×2 between-subjects factorial where Preparing to Teach served as the control condition for each of the factors. The study comprised four conditions created by

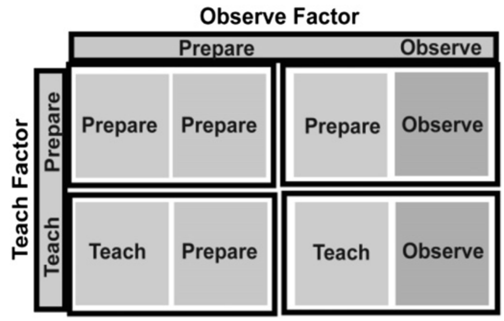

Fig. 2. Study design.

crossing the factors of Teach and Observe. The topic of learning was the mechanisms by which humans maintain a fever, which they initially learned from a short passage. The study measured how much the tutors learned from the passage using a posttest.

The panel in Fig. 3 shows a graphic representation of the research design. Here, we will give a brief description of each of the four conditions to better prepare the reader for the details presented later in the Methods section. The Teaching factor in session 2 is whether students prepared or taught (P v. T). The Observe factor in session 3 is whether students prepared or observed (P v. O). This created the four conditions of PTO (recursive feedback condition), PPP (baseline prepare condition), PTP (no recursive feedback condition), and PPO (control for observing a response from an unfamiliar pupil). Crossed with these between subjects factors are the within-subjects factor question sets on the posttest (See bottom of Fig. 3, Materials). Set A is associated with the comparison of preparing versus teaching (P v. T). Set B is associated with preparing versus observing (P v. O). Set C consists of inferential and applied questions that assess overall understanding rather than picking up specific facts that participants learned when teaching and observing.

The first branch at the top of the panel represents the recursive feedback condition, where the tutors completed all three phases of LBT, Prepare (P) + Teach (T) + Observe (O) labeled as the PTO condition. We describe this condition first. In session 1 (P), the tutors read the passage, and were told that they would have to teach a pupil on the topic of the passage. During this time, tutors prepared to teach. In session 2 (T), tutors taught and interacted with a pupil through their virtual avatars. Tutoring was done vocally, using headsets to communicate. The tutors received five questions (Set A) that guided them on what they tutored about. The pupil was a confederate whom we will refer to as Pupil X. All tutors interacted with the same Pupil X who provided the same quality of interaction and information for all of the tutors. In session 3 (O), Pupil X vocally responded to five new questions (Set B) asked by an outside interviewer (a confederate), while the tutor watched through a virtual window in a separate room. The interviewer simply queried Pupil X without any indication of whether her answers were correct or not. Pupil X followed a scripted response to the interviewer, and controlled the response across all conditions. Finally, in session 4, the tutors took a posttest that included all the questions in Sets A and B, plus five new questions (Set C).

There were four reasons for separating the question sets across the sessions. The first was to help keep the teaching interactions comparable across conditions by using the questions to focus the interactions. The second was to ensure that the quality of questions was constant across all four conditions. As mentioned above, the quality of questions raised by a pupil can have an effect on tutor learning

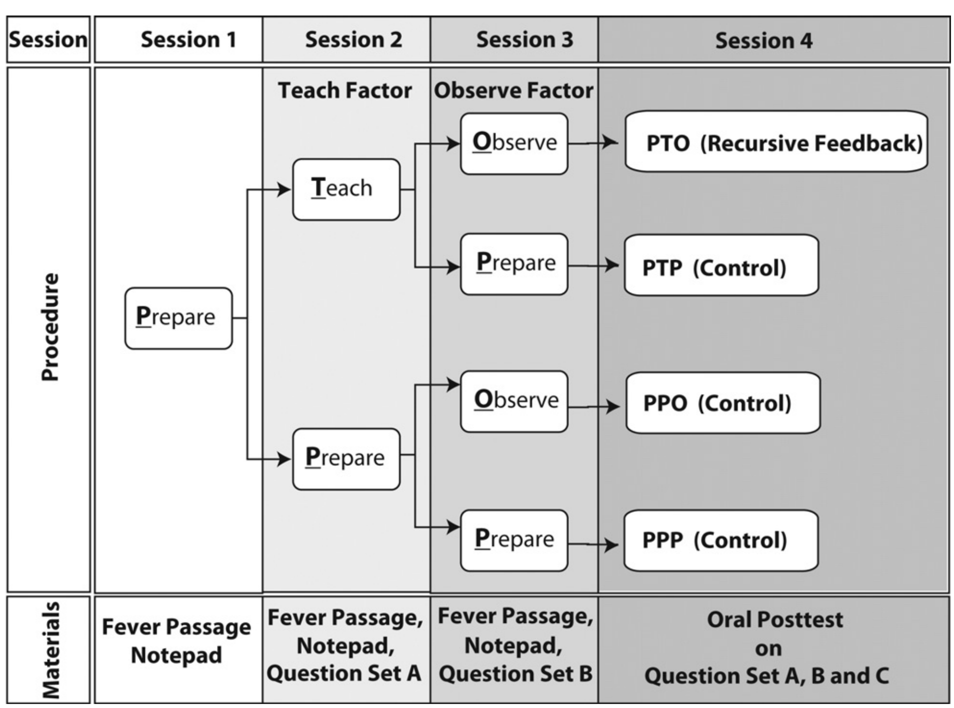

Fig. 3. Study procedure.

(Roscoe & Chi, 2008); consequently, we wanted to control this so it was not a factor in the results. The third reason we introduced questions at different points in the study was to gain some purchase on how the different phases influence learning. We predicted that the tutors in the full LBT condition would begin to show their learning advantage for Set B, because these questions drive the recursive feedback when the tutors observe their pupil. Finally, Set C provides a test of the tutors' general mental model of fever mechanisms, because these questions were never broached during the learning phases. We also predicted that participants in the PTO condition would do relatively well on Set C, because their overall mental model would have improved as a result of the recursive feedback earlier.

The fourth branch at the bottom represents the baseline control condition Prepare (P) + Prepare (P) + Prepare (P) labeled as the PPP condition. Rather than teach or observe, the tutors in this control condition continued preparation to teach throughout sessions 1–3. To keep the total time spent on the task constant, these participants received the same amount of time as the other conditions for each question set. As reviewed earlier, preparing to teach can have strong benefits for learning, so this condition made for a stronger control than simply asking the graduate students to learn for themselves.

The most important control condition involves the second branch Prepare (P) + Teach (T) + Prepare (P) labeled as the PTP condition. Tutors in this condition prepared and taught, but they did not observe their pupil, so they did not receive recursive feedback. Instead, for session 3, they prepared for another teaching session with the help of question Set B. One might reasonably predict that these students would exhibit the greatest learning. After interacting with Pupil X, they would have a chance to concentrate on improving their own understanding and prepare in light of the prior interactions. In contrast the tutors who received the recursive feedback in the PTO condition did not receive a subsequent chance to restudy, which conceivably could have put them at a relative disadvantage for learning. Our prediction, however, was that the benefits of recursive feedback would be stronger than the opportunity to restudy.

Finally, the third control condition is represented by the third branch Prepare (P) + Prepare (P) + Observe (O) labeled as the PPO condition. In this condition students prepared twice, and then observed Pupil X. They had not taught Pupil X, so observing (unfamiliar) Pupil X responding to questions did not serve as recursive feedback, although it did include an opportunity to learn through observation. If the tutors in this condition do not perform as well as the PTO condition, then we can conclude that the benefit of recursive feedback by watching your pupil perform was not simply the result of observing any pupil's response that may hold important information or models of thinking.

3.1. Methods

Thirty-nine graduate students from a private university in New York City voluntarily participated in the study. The graduate student participants were in the field of education, humanities, and social sciences, without any expertise in biology. Participants were randomly assigned to one of four conditions. During the study, the participants never met face-to-face physically with their pupil, as all interactions were remotely controlled through virtual avatars.

The study design was a 2 $\times$ 2 between-subjects factorial where Preparation to Teach (P) served as the control condition for each of the factors Teach and Observe. The study comprised four conditions created by crossing the factors of Teach and Observe. The Teach factor was implemented in session 2 with question Set A, where participants either taught Pupil X or worked on their own and continued to prepare for teaching (P v. T). Pupil X was the same female avatar controlled by a confederate throughout the study. The Observe factor was implemented in session 3 with question Set B, where participants observed Pupil X answer questions or they worked on the questions themselves in preparation to teach (P v. O). Participants never received feedback on any answers generated from the questions sets. Participants were never told they would have to answer the questions again at posttest. All participants completed a posttest by answering question Sets A, B, and the new Set C.

Participants completed the study one at a time. In session 1, all participants studied a short passage on fever for 10 min. The passage appeared on a virtual white board in Second Life. The participants were told that they would eventually tutor a student about fever based on the passage, and that they should prepare to teach. The passage was available for consultation until the posttest. Participants were allowed to type in notes on an electric notepad that appeared in the virtual environment. The notepad was available until the posttest. Only the participant could see the notepad. After preparing, participants began session 2. For conditions that teach in session 2 (PTO and PTP condition), tutors were introduced to Pupil X who was a confederate across conditions. Pupil X studiously avoided providing any new information or asking targeted questions in response to her tutors, the comments made were consistent across conditions. The tutors were told that Pupil X had "some questions that she needs to be tutored on, but also reflect on other issues that seem important from the passage. Please feel free to look at the passage while you tutor." The questions provided to tutors in session 2 were from question Set A.

For conditions that prepare but do not teach in session 2 (PPP and PPO condition), participants also received question Set A and were told to study them because, "These are examples of some questions that you will need to address when you tutor. Please feel free to look at the passage as you examine the questions." The participants prepared to teach or tutored for 5 min. Participants were not told that they would have to answer these questions again in the posttest. Careful consideration was made to minimize the information in Pupil X's responses. To control for this, a script was used so that Pupil X provided the same responses and questions during all the tutoring (session 2) and observation during the question answering with the interviewer (session 3). In other words, the exact same information was available to those participants who observed Pupil X (as their pupil) and those who observed Pupil X (a random student).

Participants then began session 3. For conditions where participants observed in session 3 were the PTO and PPO conditions. In both conditions, Pupil X moved to another room where the interviewer (confederate) asked questions from Set B. Pupil X answered question Set B, while participants observed the interaction in a separate room through a virtual window. Participants who had not taught Pupil X were told, "Now you will watch (unfamiliar student) Pupil X answer some questions about fever. Her answers can be right or wrong. Please feel free to consult the passage as you watch." Participants who had taught Pupil X were told that their pupil would be answering some questions about fever. They were now going to observe the question session through a virtual window, and they could consult the passage as they watched. Thus, participants who had taught Pupil X thought they were seeing their pupil answer the questions. Participants who had not taught thought they were seeing a random pupil answer the questions. As Pupil X answered the questions, she stated that she was not sure whether here answers were right or wrong. In fact, Pupil X did not provide any complete answers, and the interviewer provided no hints as to the quality of the answers. For conditions that prepare to teach but do not observe in session 3 (PPP and PTP condition), participants received question Set B, and they were asked to study them in the same way as before to prepare for subsequent teaching. After finishing the treatments, in session 4, all participants completed a posttest where they answered question Sets A, B, and C (15 questions). The posttest was administered vocally and transcribed for analysis.

3.2. Materials and experimental setting

The experimental setup was technologically demanding as the participant, confederate, interviewer (for testing Pupil X), and experimenters were simultaneously logged on to Second Life, controlling each of their virtual avatars. Below, we describe the virtual environment established for each of the four conditions. The participants had a default standard avatar as their virtual representation where the genders of the avatars were matched (e.g., male avatar for male participant, female avatar for female participant). The participants were wearing a microphone headset that allowed them to communicate vocally with other avatars (e.g., Pupil X avatar, experimenter avatar for any trouble shooting). Participants were also able to control the movement of their avatar. Most of the time participants would be interacting with the environment from a first-person point of view, and would not be able to see themselves from a third-person perspective. The experimenter controlled the participants' point of view. The Pupil X avatar was always a female avatar controlled by the same confederate across conditions. An experimenter avatar was present in the back of the room, out of the participant's sight, and only intervened when the participant had technical difficulty within the environment.

Below in Fig. 4, the screen shots for each of the sessions in the PTO condition are displayed. Starting from the left, in the first session Prepare (P) participants prepare to teach. The passage appears on a white board along with a notepad in case the participant wants to make any notes. The passage and notepad is available throughout the session until the posttest. In the second session Teach (T), the male tutor avatar (participant) is seen tutoring a female avatar (Pupil X). Question Set A appears on a white board along with the passage, and notepad (not shown in the image, but present). In the third session Observe (O), the screen shot shows the interviewer testing Pupil X on question Set B, while the tutor is observing his student Pupil X's real-time interaction from a window in another room. The last image on the far right is the layout of the PTO environment.

Below in Fig. 5, the screen shots for each of the sessions for the PPP condition are displayed. Starting from the left, in the first session Prepare (P) participants prepare to teach with the passage and notepad. In the image you also see the experimenter avatar that was on stand-by across all conditions in case the participant had any technical difficulty. In the second session Prepare (P), the avatar (participant) is presented with question Set A that appears on a white board. In the third session Prepare (P), question Set A disappears, and question Set B appears on the white board. The last image on the far right shows the layout of the PPP environment.

Below in Fig. 6, the screen shots for each of the sessions in the PTP condition are displayed. Starting from the left, in the first session Prepare (P) participants prepare to teach. The passage appears on a white board along with a notepad. In the second session Teach (T), the tutor avatar (participant) is seen tutoring a pupil avatar (Pupil X). Question Set A appears on a white board along with the passage, and notepad (not shown, but present). In the third session Prepare (P), question Set A disappears, and question Set B appears on the white board. The last image on the far right is the layout of the PTP environment.

Below in Fig. 7, the screen shots for each of the sessions for the PPO condition are displayed. Starting from the left, in the first session Prepare (P) participants prepare to teach with the passage and notepad. In the second session Prepare (P), the avatar (participant) is presented with question Set A that appears on a white board. In the third session Observe (O), the image shows the interviewer testing Pupil X on question Set B, while the tutor is observing his student Pupil X's interaction (video) from a window in another room. The last image on the far right is the layout of the PPO environment.

3.3. Measures

The fever passage explained how the human body gets and maintains a fever. The passage explained the mechanisms that trigger the fever response (e.g., macrophages), the mechanisms that introduce more heat into the body (e.g., shivering), and the mechanisms that prevent the body from releasing heat (e.g., blocking sweat).

The posttest consisted of question Sets A, B, and C. Each question set had five questions. Question Sets A and B were largely factual and similar in content (e.g., "What processes cause the body to increase temperature?"). Question Set C used inference and applied questions that were more difficult (e.g., "Why does a dry nose mean a dog might have a fever?").

3.3.1. Coding

In Table 1 we display a sample coding for one of the questions. We scored each question on a 0–2 point scale: 0 indicated an incorrect or no answer; 1 indicated a partially correct but incomplete answer; and 2 indicated a precise and detailed answer. To ensure reliability of the coding system, a secondary coder scored 50% of the transcripts with 95% agreement.

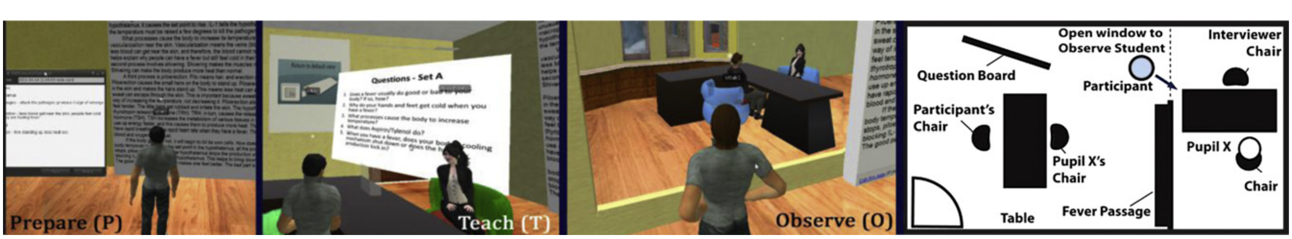

Fig. 4. Virtual environment setup in Second Life for PTO condition.

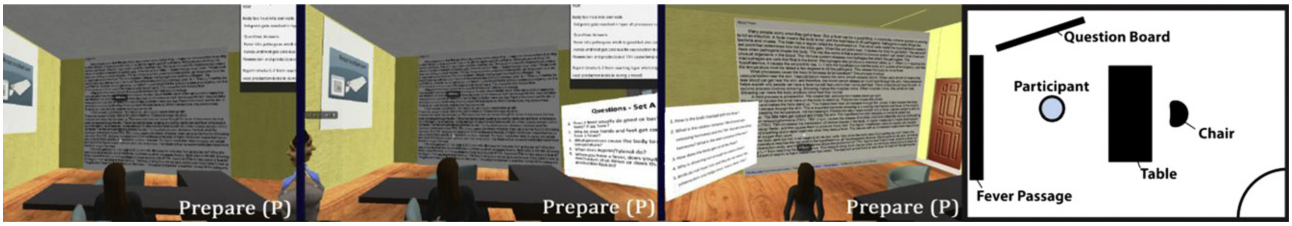

Fig. 5. Virtual environment setup in Second Life for PPP condition.

3.3.2. Organizing data analysis by posttest, teach, and observe session

We applied the coding scheme to three separate data analyses. The first analysis looked at the posttest performance comprising question Sets A, B, and C. Thus, for each set of five questions, the maximum score was 10 points with a possible total score of 30 points. The second analysis focused on the teaching session, where we coded the participant's response when tutoring the questions from Set A. Then, we coded if there were any additional relevant information that Pupil X introduced in response to the participant's teaching, meaning anything beyond what the tutor had told the pupil. The third analysis focused on the observation session, where we coded Pupil X's answers to question Set B. We coded the quality of the answers given by Pupil X to examine if any information provided by Pupil X may have influenced what the tutor learned.

3.4. Results

The hypothesis was that participants who received recursive feedback would do better than those in the other three control conditions on the posttest. More specifically the first hypothesis was that participants who taught would do better than participants who did not teach on question Set A. This hypothesis received support. The second hypothesis was that participants who received recursive feedback would do better than those in the other three control conditions on question Set B. This hypothesis received support. The third hypothesis was that the students who received recursive feedback would also do better on question Set C. Question Set C involved inferential questions that only appeared on the posttest and were designed to evaluate overall depth of understanding. This hypothesis also received support. Receiving recursive feedback during LBT was effective compared to control conditions that included LBT elements but not recursive feedback.

The first analysis looked at the posttest performance comprising of question Sets A, B, and C. Fig. 8 shows the average scores for participants in each condition on each question set. Participants in the PTO condition who received recursive feedback performed better on the posttest compared to the other three control conditions. To test these effects statistically, we conducted a multivariate analysis of variance. Teach (Teach/Prepare) and Observe (Observe/Prepare) were crossed between-subjects factors. The question set (A, B, C) was a within-subject factor for the dependent measure of accuracy on the posttest score. An omnibus test indicated a significant three-way interaction of Teach X Observe X Test; $F(2, 34) = 3.4$, Hotelling's T, 0.2, $p < .05$. Planned contrasts indicated that this interaction derived from the relative gain of the PTO condition on question Sets B and C compared to the other conditions; $F(1, 35) = 4.9$, $p < .05$. When aggregating all three question sets into a single measure, there was also a main effect of the opportunity to Teach; $F(1, 35) = 30.5$, $p < .01$. Looking at Fig. 8, one can see that both Teach conditions (PTO and PTP) showed relatively superior performance across the three problem sets. There was also an overall main effect of Observe compared to Prepare; $F(1, 35) = 9.98$, $p < .01$. However, this latter result was driven by the PTO condition and cannot be taken as a strong endorsement of the idea that observing works in all cases.

The second analysis looked at the teaching session where we coded the participant's response in tutoring question Set A, not the response from the posttest session. An important question was whether Pupil X introduced information in response to the tutor that might have inadvertently improved the scores of the tutor. We coded the answers that the participants taught to Pupil X during the teaching session, and we coded the information that Pupil X introduced during the session. We used the same coding system that we used for the posttest. Then, we coded if there were any additional relevant information that Pupil X introduced in response to the tutor, meaning anything beyond what the tutor had told her. Participants and Pupil X received a coding from 0 to 2 for a maximum possible score of 10 points across the five items of question Set A. Table 2 shows the mean score from the teaching sessions (not the posttest), broken out by those participants who subsequently observed Pupil X and those who did not. As one can see, Pupil X was artful at not introducing information. T-tests indicated that the amount of information introduced by the participants; $t(16) = 0.36$, $p = .73$, or by Pupil X, $t(16) = 0.06$, $p = .96$, did not significantly differ by condition. Thus, the advantage of the PTO condition over the PTP condition on the posttest was not due to any differences in the information provided by Pupil X in the teaching session.

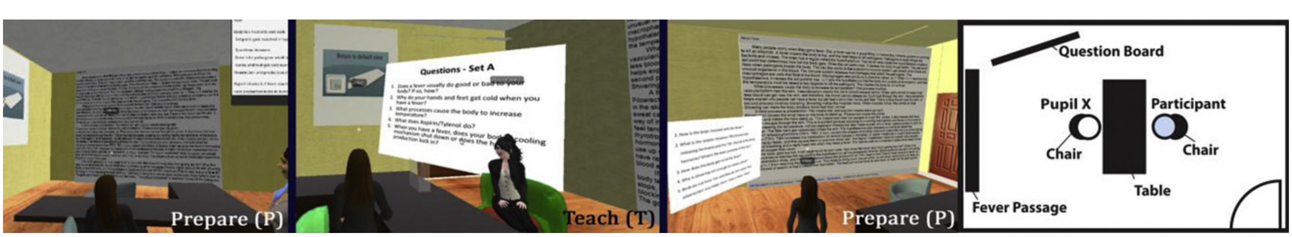

Fig. 6. Virtual environment setup in Second Life for PTP condition.

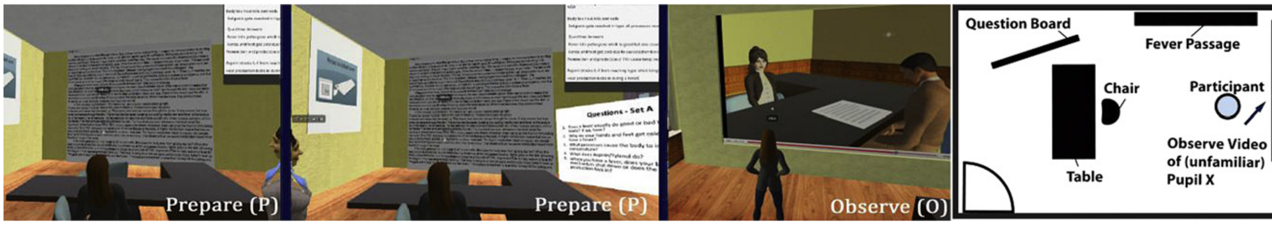

Fig. 7. Virtual environment setup in Second Life for PPO condition.

In the third part of the analysis we examined the observation session and coded Pupil X answers to question Set B, to see if any information provided by Pupil X may have influenced what the tutor learned. Pupil X received a score of 1 point for four questions, and for the remaining question 0 points. Pupil X did not introduce any incorrect information – she was simply incomplete or vague. We then compared this score to the participant's posttest scores on the four questions where Pupil X received a score of 1. On the posttest, the PTO participants gave answers that included Pupil X's information 91% of the time, whereas the PPO participants included Pupil X's information only 48% of the time, $t(17) = 3.51$, $p < .01$.

Out of a possible score of 10 points, the PTO participants were also more likely to go beyond the information they heard from Pupil X (4 points). For example, if a student scored 6 points on the posttest, that would mean the participant was able to score an additional 2 points (provide 33% more information on his or her own) beyond the 4 points given from Pupil X. Overall, the PTO participants were able to provide 42% more information on his or her own compared to the PPO participants who could only provide 20% more information on his or her own. Thus, the PTO participants appear to have been more prepared to learn from observing Pupil X, and they were more prepared to elaborate upon what Pupil X had to offer. It should be remembered, however, that we cannot be sure when the PTO participants developed their understanding of the question Set A or B questions, given that they also did much better on the question Set C, which did not directly involve Pupil X. Question Set C, which no participants saw prior to the posttest, is one indication that Pupil X was not giving question-specific answers that led to an advantage in working with her.

3.5. Discussion

The current experiment isolated one element of learning by teaching: the chance to see how one's pupil performs. We hypothesized that recursive feedback can be valuable for conceptual learning and that it should be an integral component of instructional models that include LBT. The participants who received recursive feedback by observing their pupil's subsequent performance (PTO) outperformed three control conditions with LBT that did not receive feedback. Even though interaction was limited through virtual avatar representations, and never in person, the results were in accord with a set of detailed predictions, as with the previous human and computerized agent study (Okita & Schwartz, 2006; Okita & Schwartz, submitted for publication).

First, we found that teaching and then observing one's pupil had additive value to LBT. For question Set A, the PTO recursive feedback condition showed an advantage, which makes sense because these are the questions from the teaching session. The PTO condition started to truly separate itself on question Set B. Question Set B was covered when the participants observed their student performing and, thus, it was exactly where we would have predicted the differences to start showing up. Interestingly, the benefits carried over to the questions in question Set C, which were new questions that had not been broached during the learning experiences. The results imply that the greatest positive advantage for the PTO condition appeared after participants had gone through the recursive feedback phase of LBT. This suggests that the PTO participants developed a fuller model of temperature regulation. Importantly, the students in the PPO condition observed the same video of Pupil X working on question Set B. However, they showed minimal benefits for observing compared to the otherwise comparable students who worked on question Set B themselves (the PPP condition). Evidently, the participants who taught their student had been prepared to learn by observing their student subsequently.

Second, we found that teaching other students around a set of questions led to gains compared to preparing to teach by working on the same questions. These results showed up even though the confederate Pupil X did not contribute much information in the teaching interaction besides statements like, "I think I understand," and "Oh, I didn't know that." The amount of information from Pupil X was minimal and equivalent across conditions. Notably, the benefits of teaching lasted into the later portions of the study, so that even the teaching participants who subsequently prepared to teach by studying on their own (instead of observing) showed advantages on questions from the later parts of the study. Thus, the results support the basic idea that interacting with a pupil can be helpful for learning.

The study was not designed to isolate specific psychological factors that might have contributed to the effect, but it is still possible to detect hints of why recursive feedback was successful. Our leading interpretation is that the recursive feedback loop enabled the participants to invest themselves in Pupil X's performance, while their memory of what they had previously taught enabled them to map the

Table 1
Sample coding.

Scoring method (0–2 point scale)		
0 = incorrect/no answer	**1** = partially correct but incomplete	**2** = precise and detailed
Why is shivering not enough to create a fever?		
0 points	"Because it's not enough, you need more."	
1 point	"Shivering alone creates heat, but the brain is not involved with the temperature set point."	
2 points	"You can create heat with shivering, but you also need a mechanism that doesn't let that heat escape, so you need the hypothalamus to raise the set point."	

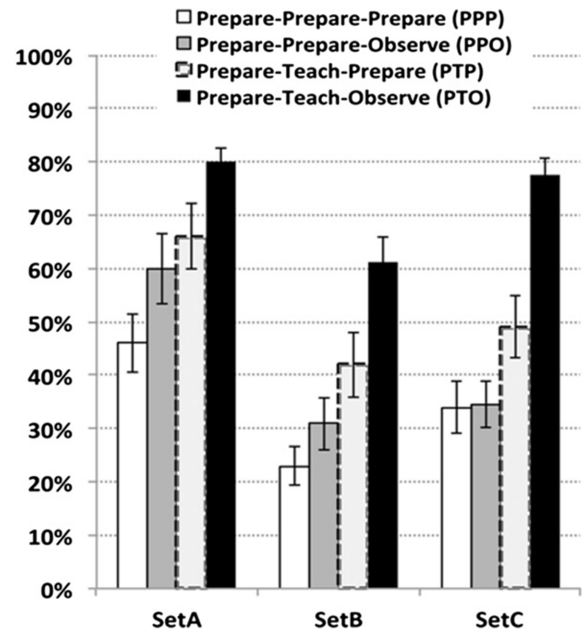

Fig. 8. Average percentage accuracy by condition.

feedback to their own understanding of fever mechanisms. One explanation may be that recursive feedback mitigated ego threats and difficulties with interpreting feedback (Chase et al., 2009). Recursive feedback appeared to engage task-related affect as well as abilities to interpret and build on feedback. The attribution was to their teaching (feedback directed to the task) rather than their own knowledge competence (feedback directed to their egos). In terms of cognitive effects, the recursive participants built on Pupil X's answers 42% of the time compared to 20% for those participants who observed Pupil X but had not taught her. This implies that the tutors were able to use the feedback to figure out what was missing, which they then integrated into their understanding of the mechanisms of fever. A second possible explanation is that people find it easier to monitor someone else who is doing their thinking for them, compared to having to fully generate, formulate, and monitor their own thoughts (Okita & Jamalian, 2012).

Another general concern may be whether the difference in the number of avatars present in each session influenced participant's performance on the posttest. More specifically whether performance was an effect of social facilitation (Zajonc, 1965) rather than LBT or recursive feedback. To make sure that this was not the case, question Sets A, B, C were introduced at different points of the study, but only tested at the posttest to gain insight on how the different interactions during the sessions influenced participant's learning. The posttest scores from Set A, helped keep the teaching conditions (three avatars present) comparable across the non-teaching conditions (two avatars present). Even with equal number of avatars in the teaching conditions, the PTO recursive feedback condition performed significantly better on Set A, than the other teach condition PTP. The posttest scores from Set B, helped keep the observing conditions (four avatars present) comparable across the non-observing conditions (two avatars present). Even with equal number of avatars in the observing conditions, the PTO recursive feedback condition, but still performed significantly better on Set B, than the other observe condition PPO. Most importantly, the results from Set C showed that this was not a social facilitation account. While question Set A, B were introduced during the sessions where the number of avatars present in the room varied, Set C was introduced for the first time at posttest, where all conditions had an equal number of two avatars in one room (two avatars present: participant's avatar and examiner avatar). Set C was never broached during the learning sessions for any participants (so social facilitation could not have influenced Set C prior to posttest), and still the PTO recursive feedback condition performed significantly higher than the other three control conditions (PTP, PPO, PPP).

Similarly, another concern may be whether the performance differences were due to a simple encoding and retrieval memory account, where multimodality learning provided more cues for later retrieval of information. The concern that the results were due to a simple encoding and retrieval memory account is unlikely. The results from question Set C confirmed that this was not the case. While question Set A, B were introduced during the sessions, question Set C was introduced for the first time at posttest. Meaning all conditions had no prior encoding of Set C. Still, the recursive feedback condition showed performance that was significantly higher than the other three control conditions. Questions Set C was never broached during the learning phases so encoding could not have influenced participant's answers for Set C at posttest.

During the study, many students showed interest in wanting to customize the avatar. An important characteristic in virtual reality technologies is that people can create their own ideal learning situation by making specific design choices (e.g., customization, look-alike

Table 2
Average information introduced during the teaching session.

Condition	Participant (SD)	Pupil X (SD)
PTO	6.38 (1.69)	0.88 (0.35)
PTP	6.10 (1.59)	0.90 (1.19)

features). In this first study, we did not allow our participants to customize avatars, as we wanted to control the features in the environment. However, customization and look-alike avatars are popular features within virtual environments, and research has shown that customizing avatars influences user behavior (Turkay & Adinolf, 2010). However, little is known about its implications for learning. Seeing that the PTO group had the strongest learning situation among the four conditions, we wanted to examine whether the implementation of various design choices could make the learning situation even stronger. To examine the effect of design choices on participant's learning, an additional study was conducted where students completed one of two conditions: (a) Customization: Teach (the pupil avatar that the participant customized prior to the study) and then observe the pupil avatar answer questions or (b) Doppelgänger: Teach (the pupil avatar that looks like the participant) and then observe the pupil avatar answer questions.

4. Effect of design choices on recursive feedback during LBT

When examining the social aspects of technology enhanced learning situations (e.g., Arnseth & Ludvigsen, 2006; Crook, 2002; Säljö, 2004), some researchers have found that students are sensitive to certain technical features (e.g., adaptive prompts, feedback). These features were implemented to assist learning, but students avoided responding to real-time prompts and feedback because they felt it disrupted the communication flow with their online peers (Kumar, Rosé, Wang, Joshi, & Robinson, 2007; Van den Bossche, Gijselaers, Segers, & Kirschner, 2006). This was an interesting case of how a specific design choice played out and how learners oriented themselves to, and made sense of, tools and artifacts presented in the environment (Stahl & Hesse, 2006).

Online virtual environments permit novel investigations of what it means to be social, and provides unique ways to examine avatars as virtual representations, through which people remotely communicate. As more biologically inspired artifacts begin to replicate human appearances, virtual avatars and agents present an array of interesting design choices that may influence learning. Common design choices include look-alike avatars/agents and customized avatars/agents. Design choices and features in technology can address learner needs and preferences by changing the virtual representation of the teacher or pupil so that the participant can feel comfortable and communicate with the teacher or pupil in the most optimal way (e.g., same gender, age, ethnicity, cultural appearance). This led to our initial research interest whether providing the learner with their preferred social environment, would have positive implications for learning.

4.1. Doppelgängers and look-alike avatars

Virtual reality technology allows for the easy creation of virtual look-alikes from photos mapped onto avatars using graphical software applications. A user can remotely control these virtual look-alikes in real time, or the look-alikes can act independently through a computer program. Virtual look-alikes are common in the commercial video game industries, as game technology has features that allow players to build look-alike avatars, but once built, these avatars are usually controlled by the game application. For example, the Microsoft Avatar Kinect has many features with which the players can create avatars that resemble themselves. Similarly, the Nintendo Wii system allows users to create cartoon-like 3-D avatars and agents that can be customized to look like the player. However, a simple cartoon character on a screen can be just as effective, according to McCloud (1993), who demonstrated that simple cartoon figures leave more room for interpretation and elicit greater empathy from the audience. De Bruine (2002) found that when using such look-alikes, participants developed more trust in agents that looked similar.

Our interest in look-alike features came from the extensively documented work by Bailenson (2012) and colleagues on how "doppelgänger" virtual humans influence human behavior. According to Bailenson, facial similarity and mimicking user gestures can persuade user's decision making (Ratan & Bailenson, 2007) and voting behavior in presidential elections (Bailenson, Iyengar, Yee, & Collins, 2008). Industry has found virtual look-alikes useful in marketing products, while others study virtual look-alikes in consumer behavior, health communication, and digital media studies (e.g., motivation, false memories). Similarity can range from physical appearances, gender, beliefs and behaviors, where some features can be easily recognizable through observations, and others need to be revealed through verbal interactions. The likelihood of learning increases as participants successfully identify with the look-alike avatars (Andsager, Bemker, Choi, & Torwel, 2006; Bandura, 2001; Bandura & Huston, 1961). Two constructs seem to influence participant behavior with look-alike avatars; visualizing and identifying with the look-alike avatar.

In the area of health communication, Fox and Bailenson (2009) conducted a study in a virtual environment that compared participants watching their virtual doppelgänger exercise, someone else's doppelgänger exercise, or their look-alike doppelgänger stand still. Participants who watched their look-alike doppelgänger exercise reported higher self-efficacy and intention to adopt the exercising behavior compared to the other conditions. As participants exercised in the physical world, their look-alike avatars also exercised. The next day, a follow up interview revealed that the look-alike doppelgänger group had exercised 1 h more than the other group participants in the study. Participants who interacted with and observed their look-alike avatar seemed to better connect the cause and effect of exercise and weight loss than those who interacted with the standard avatar. Results revealed that it was not just the direct mapping of their own behavior to avatar behavior that led to the behavioral difference, but the look-alike appearance in the avatar helped participants identify, visualize, and reason how specific causal chains were linked to their actions. Similar effects were also found on healthy eating behaviors (Fox, Bailenson, & Binney, 2009) where participants viewed their look-alike avatar eating candy and gaining weight or eating carrots and losing weight. The study found that the eating behavior of the look-alike avatar affected the eating behavior of participants.

The strength of virtual environments is that technology can assist visualization to emphasize similarities to help participants identify more with look-alike avatars. For example, Ahn and Bailenson (2011) placed participants' heads on people depicted on billboards using fictitious brands, holding up a soft drink with a brand label on it. When participants saw his or her avatar wearing a certain brand of clothing, they were more likely to recall and prefer that brand. In other words, if one observed his avatar as a product endorser, he was more likely to remember and embrace the product. After the study, participants expressed better memory of, as well as a preference for the brand, even though it was obvious that their faces had been placed in the advertisement. In a study with elementary school students, Segovia and Bailenson (2009) created a fictitious situation where children's look-alike agents were swimming with whales in a virtual aquarium (a virtual situation easily created with technology, but highly unlikely in the physical world). Segovia built look-alike avatars for children and had them watch their look-alike agent swimming with whales in a virtual aquarium, to see whether children would develop false memories and mistake their look-alike's actions for their own actions. The children were interviewed a week later in the physical world. Results showed that over half of the elementary school children who observed their look-alike agents swimming, believed that they had physically gone swimming with whales compared to children who observed standard avatars swimming.

As these studies have explored how facial similarity with a virtual agent or avatar influences human behavior, little research has been done to examine the implications for learning a specific content area. Moreno and Flowerday (2006) examined how animated pedagogical agents in a science education multimedia application influenced student learning and affect when the agent shared some characteristics similar to the student (e.g., gender, skin color). Giving the learner the opportunity to interact with a look-alike avatar may help the learner identify more with the avatar, which may possibly lead to more learning.

4.2. Avatar customization

Virtual reality technology often provides choices for users to customize their agent and avatar characters. Customizing and building an avatar often involves projecting ideas onto a standard avatar. Users have the option to customize the avatar by selecting different attributes (e.g., gender, body shape, hair, skin color, clothing). Yee and Bailenson (2007) found that avatar customization influences not only the users' mental representations, but also their behavior and preferences. Bailey, Wise, and Bolls (2009) conducted a study that examined whether customization of avatars influenced children's physiological responses and subjective feelings of presence while playing advergames (e.g., major snack food companies creating free advertising video games on their Web sites to target children). The study compared avatar assignment, avatar choice, and avatar customization, and found that avatar customization elicited the highest skin conductance level and feeling of presence during the advergame play, thereby influencing consumer preferences. Giving students the opportunity to customize and interact with their avatars may provide a sense of ownership and responsibility that leads to motivation.

Customization brings a sense of ownership and responsibility, as participants have contributed ideas into the creation of the avatar. Contributing triggers ownership, and motivates learners to pay attention to the task-level feedback and improve their own understanding (Cordova & Lepper, 1996). For example, Langer (1975) has found that giving individuals choices has led to better performance, more satisfaction, and more intrinsic motivation when performing tasks. Cordova and Lepper (1996) found similar results with children where simple choices, such as naming a task, naming a player, or naming the "math space mission," personalized the experience, created ownership over the "math space mission", and elicited a positive reaction on math learning, intrinsic motivation, and enjoyment. Iyengar and Lepper (1999) also found similar results in a cross-cultural study where choice was an important factor for intrinsic motivation and learning.

Much of the previous studies involve customizing your representation, or the representation of an agent controlled by a computer program. In this study, participants in the customization condition were specifically asked to "customize the avatar their pupil will be using." This explores an unfamiliar territory of customizing someone else's representation where ownership and responsibility are blurred. The customized pupil avatar is controlled by another independent human, which is different from an agent controlled by a computer program. The motivation was to see whether creating one's own ideal social environment had implications for student learning. Since virtual environments can allow two people to see the same or different views, we feel this situation is close to the current multi-user virtual environments. Also, such design choices can support more learner-centered environments where virtual teachers and classmates can be represented differently (e.g., same gender, age, level of attraction, ethnicity, wearing uniforms, competitive appearance) to help the learner communicate and learn in the most optimal way (e.g., comfortable, competitive, familiar).

We hypothesized that both design choices would further extend the benefits of recursive feedback during LBT, since both look-alike and customization help develop stronger connections with the pupil avatar (e.g., identifying with pupil, responsibility toward the pupil). To examine the effect of design choices (e.g., customization, look-alike avatars), an additional follow-up study was conducted with two conditions: (a) Customization: Teach the pupil avatar that the participant customized, and then observe the pupil avatar answer questions, and (b) Doppelgänger look-alike: Teach the pupil avatar that looks like the participant, and then observe the pupil avatar answer questions. Our prediction was that customization would lead to higher performance, as customizing involves contributing and projecting ideas onto the avatar, which adds another layer involvement to the tutor–pupil relation. In comparison, the look-alike condition is less direct, and depends on how much the participant can identify with the avatar's appearance.

5. Experiment 2: implications of design choices on learning

Design choices and preferences are important characteristic of virtual reality technologies where people can create their own ideal learning situation by making specific design choices (e.g., customization, look-alike features). However, little is known about whether such a learning environment has positive implications for learning. The second experiment explores the implications of design choices on recursive feedback during LBT, specifically customization and look-alike (doppelgänger) features.

5.1. Methods

Twenty graduate students from a private university in New York City voluntarily participated in the study. The graduate students were in areas of education and social sciences, with no expertise in biology. Participants never met face-to-face with their pupil, as all interactions were conducted through virtual avatars.

The students were randomly assigned to one of two recursive feedback conditions (see Fig. 9), PTO (look-alike) or PTO (customization). The results of the two conditions were later compared with the PTO condition from experiment 1 as the control condition. The procedures of the two conditions were identical to the PTO condition from the earlier study (see Fig. 3 and Section 3 for description). The difference in the PTO (look-alike) condition was that the Pupil X avatar looked like the participant. Only in the PTO (look-alike) condition, did the pupil and participant's gender match. For male participants, the pupil was a male look-alike avatar controlled by a male confederate, and for female participants, the pupil was a female look-alike avatar controlled by a female confederate. The look-alike avatars were created prior to the

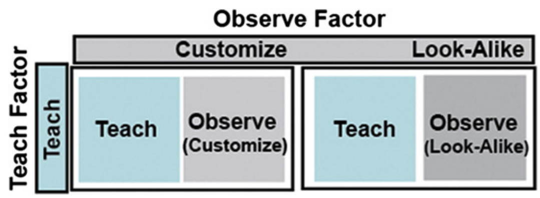

Fig. 9. Study design for customization and look-alike study.

study from the participant's photo, and the commercial software AvMaker 1.0.4 from CyberExtruder, Inc., that created 3D head models of the participants. The participants were consciously aware that their pupil avatar shared their facial similarities.

The only difference in the PTO (customization) was that the participants were asked to customize a standard default avatar. Participants were given the following instructions, "This is the avatar your pupil will be using. Please feel free to customize this avatar using the tools here in the drop down menu. You will have five minutes." The participants were presented with a standard default female avatar and asked to freely customize. Participants had the option of changing the hair, skin and eye color, body shape, clothes, and shape of facial features. All participants used the full 5 min. In the PTO (customization) the pupil avatar was always female to be consistent with the first experiment. No other special instructions were given. As in the first experiment, the participants never met face-to-face physically with their pupil. All interactions occurred through virtual avatars that were remotely controlled by the participant or research staff (e.g., experimenter, Pupil X confederate, and interviewer).

5.2. Materials and experimental setting

The experimental setup was identical to the PTO condition in the earlier study (see Section 3.2). The participant, confederate (Pupil X), interviewer (for testing Pupil X), and experimenter (technical support) were simultaneously logged on to Second Life, on stand-by, or controlling each of their virtual avatars. Below, we describe the virtual environment established for the PTO (customization) and PTO (look-alike) conditions.

Fig. 10 shows a screen shot of the virtual environment established for the PTO (customization) condition. The setting is identical to the PTO condition from the earlier study. Starting from the left, in the first session Prepare (P) participants prepare to teach. The passage appears on a white board along with a notepad, and is available throughout the session until posttest. In the second session Teach (T), the image shows the participant's customized avatar (Pupil X) seated in front of the male tutor avatar (participant). Question set A appears on a white board along with the passage, and the participant tutors Pupil X. In the third session Observe (O), the screen shot image shows the interviewer testing Pupil X on question set B, while the tutor is observing his Pupil X's real-time interaction from a window in another room. The last image on the far right is the layout of the PTO (Customization) environment.

The environment setting for the PTO (look-alike) condition is identical to the PTO (customization) condition. The only difference is that the participant will be tutoring and observing Pupil X which is a look-alike avatar created from the participant's photo. Fig. 11 shows a glimpse of the process when creating look-alike avatars. The look-alike avatars were created from a photo of the participant taken prior to the session. The commercial software AvMaker 1.0.4 from CyberExtruder, Inc., was used to create 3D head models of participants that were then used as avatar skins and uploaded for use in Second Life. The software located the anatomical points of the head image, and small adjustments were made to fit the face structure template in Second Life.

5.3. Measures

The same measures were used in the follow-up study. Each question set had five questions. Question Sets A and B were largely factual (e.g., "What processes cause the body to increase temperature?"). Question Set C used inference questions (e.g., "Why does a dry nose mean a dog might have a fever?"). The participants in the customization condition had an additional measure that calculated the number of times the students customized their avatars in order to examine the implications on learning performance.

5.4. Results

To test these effects statistically, we conducted a multivariate analysis of variance. The PTO type: PTO (control) vs. PTO (customization) vs. PTO (look-alike) was the between-subjects factor. The question set (A, B, C) was a within-subject factor dependent measure, which was the accuracy score on the posttest. There was no significant interaction effect between the question set and PTO type.

Fig. 10. Virtual environment setup for the PTO (customization) condition.

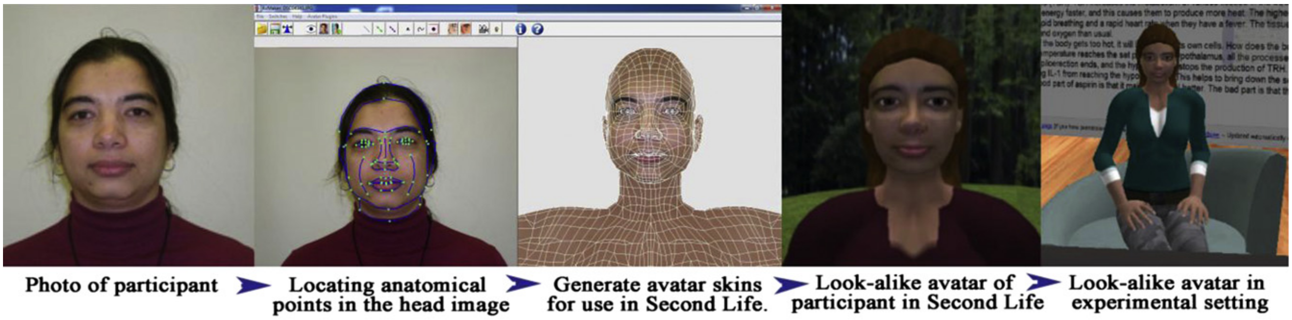

Fig. 11. Avatar of participant for PTO (look-alike) condition.

When aggregating the question sets (A, B, C) into a single measure, there was a main effect on the between-subject factor PTO type; $F(2, 25) = 9.25$, $p < .01$. Looking at Fig. 12, one can see that the PTO (control) showed relatively superior performance across the other two conditions, more so for the PTO (look-alike) condition. A post-hoc Tukey's HSD test showed that PTO (control) was significantly higher compared to the PTO (look-alike) condition ($p < .01$, Cohen's $d = 2.32$, effect-size $r = 0.76$) and somewhat higher than the PTO (customization) condition but not significant ($p = .182$, Cohen's $d = 0.96$, effect-size $r = 0.43$). The difference between PTO (look-alike) and PTO (customization) approached significance ($p = .058$, Cohen's $d = 0.96$, effect-size $r = 0.43$).

Again, we addressed an important question of whether Pupil X introduced correct answers that might have inadvertently improved the scores of those participants who either taught or observed Pupil X. The results for question Set C, which no participants saw prior to the posttest, was one indication that Pupil X was not giving question-specific answers that led to an advantage in working with Pupil X. We further looked at the information that Pupil X introduced in the sessions.

We began with question Set A, which was the focus of the teaching session. We coded the answers that the participants taught to Pupil X during the teaching session, and we coded the information that Pupil X introduced during the session. We used the same coding system used for the posttest. Participants and Pupil X received a coding from 0 to 2 for a maximum possible score of 10 points across the five items of question Set A. Table 3 shows the mean score from the teaching sessions (not the posttest), broken out by those participants who subsequently observed Pupil X and those who did not. As can be seen, Pupil X was artful at not introducing information. The ANOVA tests indicated that the amount of information introduced by the participants; $F(2, 25) = 0.64$, $p = .54$, or by Pupil X, $F(2, 25) = 0.20$, $p = .83$, did not significantly differ by condition. Thus, the advantage of the PTO (control) condition on the posttest was not due to any differences in the information provided by Pupil X in the teaching session.

For four questions, Pupil X received a score of 1 point and for the remaining question 0 points. Pupil X did not introduce any incorrect information – she was simply incomplete or vague. We examined posttest scores for the four questions where Pupil X received a score of 1. On the posttest the PTO (control) condition, participants gave answers that included Pupil X's information 91% of the time, whereas the PTO (look-alike) condition participants included Pupil X's information 64% of the time and PTO (customization) participants included Pupil X's information 89% of the time.

However, the PTO (control) participants were able to provide 42% more information beyond what Pupil X showed in the observation session. Compared to the control condition, the PTO (look-alike) participants could only provide additional 17% more information, and the PTO (customization) participants provided 35% more information. Thus, the PTO (control) participants appear to have been more prepared

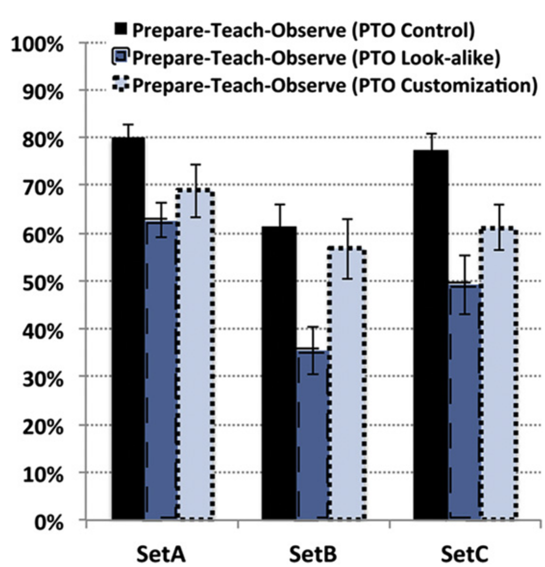

Fig. 12. Average percentage accuracy by PTO (previous study-control), PTO (customization), and PTO (look-alike conditions).

Table 3
Average information introduced during the PTO sessions.

Condition	Participant (SD)	Pupil X (SD)
PTO (control)	6.38 (1.69)	0.88 (0.35)
PTO (look-alike)	5.82 (1.78)	1.09 (1.14)
PTO (customization)	6.67 (1.66)	0.89 (0.78)

to learn from observing Pupil X, and were more prepared to elaborate on what Pupil X had to offer. It should be remembered, however, that we cannot be sure when the PTO participants developed their understanding of the question Set A or B questions, given that they also did much better on the question Set C questions, which did not directly involve Pupil X.

The study also examined the implications of design choices on student performance by seeing whether there were any learning differences attributable to the number of times the participant customized his or her avatar. Although the sample size was too small to generalize any findings, there were some promising trends that might be interesting for future studies. The level of customization was categorized by the number of times participants made changes to the standard default avatar presented to them. In session 2, the customized avatar was introduced to the participants as Pupil X. The number of participants and the average, minimum, and maximum number of customizations determined the categorization: low level (0–15 times), medium level (16–30 times), and high level (30+ times) of customization. To test these effects statistically, we conducted a multivariate analysis of variance. The customization levels (low vs. medium vs. high) were between-subjects factors. The question set (A, B, C) was a within-subject factor dependent measure, which was the accuracy score on the posttest. There was no significant interaction effect between the question set and customization level.

When aggregating the question sets (A, B, C) into a single measure, there was a main effect on the between-subject factor customization level; $F(2, 6) = 14.08$, $p < .01$. Looking at Fig. 13, one can see that the medium level of customization showed relatively superior performance across question Sets A, B, and C, compared to the other two groups, more so for the low level group. A post-hoc Tukey's HSD test showed that the medium level was significantly higher when compared to the low level group ($p < .01$, Cohen's $d = 3.89$, effect-size $r = 0.89$) and somewhat higher than the high level group approaching significance ($p = .058$, Cohen's $d = 2.40$, effect-size $r = 0.77$). There was no significant difference between the low level and high level groups ($p = .123$, Cohen's $d = 2.19$, effect-size $r = 0.74$).

To examine how similar the look-alike avatars were to the participants, we conducted a manipulation check. Thirty-five random graduate students filled out a survey that consisted of a matching task on perceived similarity between participants and their look-alike avatars. Students were given a matching task consisting of face photos of participants from the look-alike condition, and a snap shot of their look-alike avatars listed in random order. Additional photos of people with similar facial traits were added to the list as decoys to avoid process of elimination. The manipulation check, showed 92% accuracy in similarity rating where students correctly matched (both male and female) look-alike avatars to their respective participants.

A manipulation check was also conducted for the customized avatars to check whether customized avatars looked different or similar to the participants. This was important to check so that the learning effect was due to customization, and not facial similarity. The same 35 randomly selected graduate students were given a second matching task consisting of face photos of participants, and a snap shot of the customized avatars listed in random order. Students were instructed to match the customized avatar and the participant who they think customized the appearance. Again, additional snapshots of random avatars were added to the list as decoys to avoid process of elimination. The customized avatar manipulation check, showed an average of 8% accuracy rate in identifying the correct match for both male and female participants and their customized avatars. The customized avatars did not look similar to the participants, and students were not able to identify which participant customized which avatar.

5.5. Discussion

Results showed that the performance of the PTO (control) group had the highest performance compared to the two experimental groups. The experimental groups did assist learning, however the PTO (customization) group performed better compared to the PTO (look-alike) group. The literature shows that participants trust avatars that share similar features (Baumeister, 1998; De Bruine, 2002, for a review), and

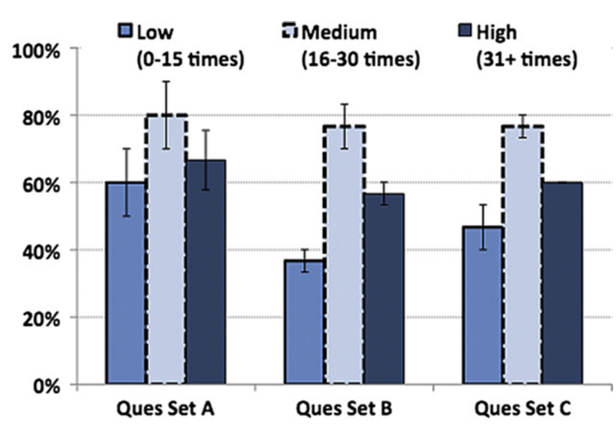

Fig. 13. Average percentage accuracy by number of times participant customized a feature on pupil X avatar.

observing the behavior of such avatars, influences participant behavior (Fox & Bailenson, 2009; Fox et al., 2009). These findings led to our initial hypothesis that facial similarity would further extend the value of recursive feedback, as participants would identify more with the pupil avatar. However, our findings were not as strong as expected. Several features in our study may have influenced how the participant identified with the look-alike avatar. We entertain three possibilities for the limited outcome, 1) the look-alike feature restrained the ego-protective buffer from taking effect, 2) the verbal response triggered a mismatch between the look-alike appearance and voice, and 3) the human tendency to evaluate one's appearance and performance favorably rather than objectively warranted had impeded learning.

Initially recursive feedback was thought to circumvent ego involvement, because tutors focused on their pupils rather than solely on themselves. As a result, tutors could pay attention to information that was relevant to improve learning because the pupil, rather than the tutor, took immediate responsibility for being wrong. One possibility for the limited learning outcome was that teaching and receiving recursive feedback from a look-alike avatar did not create an ego-protective buffer, as participants were consciously aware of their facial similarity with the pupil avatar. Tutors therefore took immediate responsibility for being wrong and could not pay attention to the task-level feedback to improve their own understanding (Kluger & DeNisi, 1996). We had entertained the possibility of a worst-case scenario where the results from the PTO (look-alike) condition (See Fig. 12) may look similar to the PPP condition from experiment 1, as participants may perceive the interaction with the look-alike avatar as the same as working on their own. Even if this was the case, the results reflected the effects of teaching and observing (just not recursive feedback). This would explain the PTO (look-alike) condition's higher performance on the posttest compared to PPP and PPO conditions from experiment 1, and scoring close to the PTP condition (See Fig. 8). This may be the reason why the other two PTO conditions (control condition and customization condition) showed higher performance on the posttest. Sharing no facial resemblance, the ego-protective buffer had taken effect, and participants focused on task-level feedback and improved their own understanding.

The second possibility was the difference in interaction and feedback. In many of Bailenson et al.'s doppelgänger studies, the interaction and feedback between the participants and their look-alike avatars were mostly non-verbal (e.g., pressing a button and the avatar makes a choice) or physical behavioral interactions (e.g., as participants exercised in the physical world, their look-alike avatars exercised in the virtual environment). In comparison, our study was based solely on verbal exchanges and observations (e.g., tutoring a pupil on biology). Although the avatar may have looked like the participant, the avatar was an independent pupil, and controlled by a confederate who had a different voice. The pupil's verbal exchanges were done in the voice of the confederate. Hearing the voice of the confederate coming from a look-alike avatar may have triggered a mismatch between appearance and voice. As Garau et al. (2003) found, increasing the photographic realism of an avatar requires behavioral realism to also increase, or the effect of social presence fades.

The third possibility may be due to people's tendency to evaluate their own appearance and performance higher than objectively warranted. Epley and Whitchurch (2008) found that adults have the tendency to perceive their own face as being more physically attractive than they actually are. Participants were more likely to recognize an attractively enhanced version of their face more quickly than their actual image. The look-alike avatars in our study were generated using participant's photos, but no enhancement was made to make them more physically attractive. This gap between expectancy and reality may have impeded participant's performance. Similarly, studies have shown that people evaluate their own abilities and performance levels higher than their actual achievement level (Dunning, 1999). People are also more willing to adopt flattering information about the self, but are more critical about unfavorable information, and may even derogate the credibility of the source if self-threatening (Chambers & Windschitl, 2004). In our study, Pupil X was careful in giving only minimal information during the interactions (teach session and observe session). Perhaps the mediocre performance of the look-alike avatar had not been good enough for participants to want to adopt or associate with Pupil X's performance. The participant may have perceived the look-alike avatar student as "doing poorly" or being "uncooperative" to the participant. Van Vugt, Bailenson, Hoorn, and Konijn (2010) found that facial similarity was not enough to drive positive impressions or favorable associations. Under certain circumstances, people can feel less favorable toward their look-alike avatar, especially when similarity is paired with negative characteristics. Their study revealed that when the look-alike agent was helpful, facial similarity increased participant's rating of interest and involvement. When the look-alike agent was unsuccessful in assisting the participant, male participants had more negative responses to the look-alike agent. Results suggested that users may avoid people with similarities, or even show some rejection, when they did not want to be associated, felt threatened by the similarity, made self-associations with negative traits, or maintained distance to protect themselves (Lerner & Agar, 1972).

This was an interesting outcome where technological features revealed results based on how humans constitute meaning from technological features (MacKenzie & Wajcman, 1985; Stahl & Hesse, 2006). Literature has revealed that facial similarity in avatars when unconscious, could draw people to favor voting behavior or persuade user's decision making. On the other hand, when facial similarities were conscious, people had higher expectations concerning appearance, and abilities than their actual state. Look-alike features may have potential, but careful design choices need to be made for positive outcomes on learning. For example, a design choice that helps participants identify more with the pupil may be as simple as making participants unaware of the facial similarity manipulation, using a computer-generated voice, or enhancing the appearance of the look-alike avatar beyond the participant's actual image, or even significantly simplifying the appearance. McCloud (1993) suggested that using the iconic form and simplicity of a cartoon instead of a realistic graphic image, was one way of helping people understand their own internal shapes. It would be interesting to also see whether participants actually learn more by interacting with look-alike avatars that outperform the participant. In any case, a larger sample size and further empirical research is needed to determine if this is the case.

In the literature on customization, studies have found that avatar customization influences not only the users' mental representations but also their behavior (Yee & Bailenson, 2007). In relation to learning, Langer (1975) found that giving individuals choices has led to better performance, more satisfaction, and more intrinsic motivation when performing tasks. Cordova and Lepper (1996) found that that even simple choices, such as naming a task, a player, or "space mission," elicited a positive reaction in learning, intrinsic motivation, and enjoyment. Our hypothesis was somewhat confirmed as the PTO (customization) group performed better than the PTO (look-alike) and other LBT control conditions (from experiment 1). However, two questions remain as to why the PTO customization group performed lower in comparison to the PTO control group, and why the amount of customization may have mattered.

One explanation may be in how customization was applied to our study. Previous studies with Teachable Agents have shown successful cases where customizing a pupil agent (computer controlled program), helped tutors develop ownership and responsibility, which assisted learning (Biswas et al., 2001; Chase et al., 2009; Schwartz, Blair, Biswas, Leelawong, & Davis, 2007). In our current study customization did not involve a computer program or the participant's own virtual representation. The study involved the customization of the appearance of a virtual pupil avatar that was eventually controlled by another person (confederate). Although transforming a pupil's knowledge may be favorable, transforming the pupil's physical appearance may have somehow violated the identity of the pupil. At the time when participants customized the pupil avatar, the situation may not have been much different with customizing a pupil agent (computer-controlled program). However, when the participants started teaching and observing the pupil that was remotely controlled by another person, and heard a live person's voice coming from that avatar, the tutor-tutee relation may have turned awkward in terms of who had ownership or responsibility. Peer tutoring is usually a positive relation, and pupils are considered autonomous, independent thinkers that deserve respect. Perhaps changing the physical appearance, body shape, ethnicity, and/or gender of a live independent pupil without their consent objectifies the pupil, and violates a social norm. Doing so may have led to the loss or prospect of ever establishing an equal peer relationship.

Another explanation may be found in the interesting finding of how the number of times participants customized the pupil avatar, influenced the posttest performance. The results showed that the number of times the participant customized the avatar influenced the learning outcome on the posttest, where too little (low level: 0–15 times) and too much (high level: 30+ times) customization seemed to hurt performance. One possibility of why the amount of customization may have mattered may be that the customizing process triggered stronger ownership and responsibility associated with the external features of Pupil X (e.g., appearance of the avatar) rather than the information exchanged during the tutor-tutee interaction (e.g., knowledge developed from the tutor-tutee session). Too little customization may not have been enough to develop any sense of responsibility or relation, and too much customization may have allowed external features to get in the way of focusing on other parts of the interaction with the tutee. Since the participants in the PTO control group only focused on developing the tutor-tutee relation (e.g., knowledge), perhaps they were able to focus on Pupil X's responses during the observation session. Customization features seem to have potential, but may be limited to the participant's own representation, or a computer program where ownership and responsibility are somewhat clear.

6. General discussion

Two studies examined whether the value of recursive feedback extends to an online virtual environment where the tutor and pupil communicate through virtual avatars, and never in person. The studies focused on the practical application of *recursive feedback* in an online environment, building on previous findings that tested the hypothesis that an important facet of learning by teaching is the opportunity to observe their pupils independently use what they have been taught in a relevant performance context. The first study isolated the effect of recursive feedback during LBT. The studies took place in Second Life where college students who taught and observed their pupil exhibited superior learning relative to several control conditions. All control conditions included elements of LBT but not recursive feedback. The value of recursive feedback seemed to extend to online virtual environments even if students never meet in person. The effectiveness of recursive feedback required that tutors maintain representations of their own understanding, of what they taught, and of the understanding of their pupils. Moreover, these representations need to be entertained simultaneously, so the tutors can sort out which aspects of the pupils' performance relates to which level of representation. The findings indicate that recursive feedback is effective and should be included in LBT instructional models for online learning environments.

The second experiment explored whether design choices to create one's own ideal social environment had implications for student learning. Results show that the recursive feedback control group outperformed both the customization group and look-alike group. However, customization still seemed to have positive implications for learning, as posttest performance was higher compared to other LBT control conditions. The findings showed that too little or too much customization might have influenced the participant to focus more on the external features of the pupil than on the pupil's internal knowledge or pupil's response. The look-alike features had the lowest performance among the recursive feedback conditions, but still showed better learning performance compared to the LBT control conditions where no teaching was involved (i.e., PPP and PPO).

The two studies naturally raise a number of important questions for subsequent researchers. Compared to previous findings (Okita & Schwartz, 2006), the current studies were heavier on the technical side. The studies focused more on practical applications to see if the value of recursive feedback extends to an online virtual environment where students do not physically meet. Regarding causality we feel that the recursive feedback effect was due to the felicitous confluence of many important conditions that favor learning. This would include a sense of social connection, familiarity, ownership, responsibility, motivation, and uptake of negative versus positive feedback that is, in addition to the technical features that may have influenced the social interaction. An approach may be to decompose each psychological and contextual variable with an appropriate research design. However, the intention of this study was not to isolate a specific variable or to find a strict causal attribution, but to examine how important learning conditions come together. We theorize that each of the aforementioned isolated variables were amplified when put into the context of recursive feedback. We do however acknowledge that the reader can benefit from drawing connections to the cognitive literature to further understand the underlying mechanism that drive the findings. The discussion section for each experiment addresses these connections, and we have described some possibilities. It is likely that numerous causes work together (e.g., sense of connection, responsibility, motivation) and that there is no single causality. This is, however, one reason why learning by teaching can be such a powerful instructional method: It is a single situation that naturally brings to bear many positive forces for learning.

While we are not in a position to address mechanism questions based on this study, one might entertain two classes of mechanisms: cognitive and affective, that seemed to drive our findings from both studies. One cognitive account that might come into play is the dual task of problem solving; this would require one "process" that applies a sequence of steps to solve a problem and a second "process" that evaluates the problem-solving process. In essence, people need to run and coordinate two processes that work to solve the problem, the one doing the "actual work" and the one that evaluates the work and self-reflects to see what changes may be needed to improve the results. If the answers are too far out of alignment, this should trigger a set of activities to resolve the gap and discrepancy. By hypothesis, the dual-task demands of both processes can be alleviated by observing somebody else problem solve. Instead of running two operations, observing someone else solving the problem may free up capacity and permit more effective reflection and debugging. A second, but not incompatible, cognitive account might be that observing another person provides new issues and questions that help overcome the inertia of one's own ideas (Azmitia, 1996). For example, reciprocal teaching (Palincsar & Brown, 1984) and peer tutoring can help one revise knowledge by providing information that one would never have considered, or listening to explanations from another point of view. Recursive feedback may slow down students to think so they become more reflective (Schön, 1983), a possibility that could be evaluated by an appropriate research design. The limitations of this cognitive mechanism are that participants who observed someone they had not taught did not learn as well as students who observed their own pupil. This suggests the possibility of an affective account, because the "alleviating of dual task demands" applies even if Pupil X is not one's pupil. The participants who observed their own pupil (Pupil X) answer questions may have been more attentive to their pupil's performance; possibly a sense of responsibility or ownership-type relationship, toward the student (as in an advisor/advisee or coach/team relationship) leads to more attention to the pupil's answers. Having some idea of what the pupil knows, the tutor will often realize that he or she has not taught relevant ideas, which may cause the tutor to think the process through more deeply.

Virtual representations such as agents and avatars incorporate natural human-like elements (e.g., appearance, movements) to technology allowing participants to make social attributions to virtual representations that trigger affection. As we have found in the work with teachable agents, working with computer programs (or virtual agents) or people can attribute social qualities to the programs (Biswas et al., 2001; Reeves & Nass, 1996; Turkle, 1984). Okita et al. (2007) found that people learn more when they believe they are interacting with a person than when they believe they are interacting with a computer program. The question however, was whether social attributions are a special catalyst to the value of recursive feedback. This was somewhat revealed in our two studies, as we found that social attributions made by tutoring a person through a virtual avatar and observing recursive feedback led to significant learning gains without meeting in person. However, the second study revealed that design choices, such as customizing one's pupil might differ according to whether the learner is making social attributes to a computer programs (agent) or a virtual representation of a real person (avatar). Customizing a pupil agent helped students learn, but excessively customizing a pupil avatar seemed to hinder learning as it violates a social norm and objectifies the pupil. The study also helps point out some irony in technological features, as certain choices may be favorable in one situation but unfavorable in another: customizing a pupil representation that is an agent (computer program) has positive effects on motivation and learning, but customizing a pupil representation that is controlled by another human may have unfavorable consequences for learning, although the amount of customization may influence what the tutor focuses on about their pupil (e.g., knowledge or appearance). Look-alike features in avatars may have favorable consequences if the learner is unconscious about the facial similarity, but conscious acknowledgment of similarity may trigger direct association with failure, and higher expectancy in avatar appearance and performance.

Other ways in which social attributions may trigger affection and catalyze recursive feedback may have to do with the recursive mapping between the tutor and pupil thoughts. In LBT, tutors can ask, "What is my pupil thinking?" "Can she answer that? Did she understand what I explained?" This "theory of mind" bridge may help students think about thinking, compared to a situation where they need to think about a specific content. There are several possible practical implications for designing effective instruction, especially for LBT, which should include the final phase of observing recursive feedback.

7. Future work

One way to make additional headway is to explore different variants of LBT and recursive feedback, since not all learning situations will produce favorable performance. To pursue this empirical tact, which requires tightly matched experimental conditions, it may be important to use a simpler research paradigm than the studies this paper presents. Given a simple design, a number of useful variations can be planned to help isolate causes of learning. For example, the study did not include a condition in which participants taught Pupil X and then observed a different Pupil Y answer questions. A condition that involves teaching one pupil and monitoring another would have helped to test the importance of the social connection between the tutor and the pupil. For example, participants who observe their own pupils may learn better than participants who observe another tutor's pupil, because they have a strong connection by which to make the affective and cognitive recursion. This would make a good follow-up line of research, now that the basic effect has been established. The benefits of watching someone else's student will be a function of the degree to which tutors project their original teaching into the pupil they subsequently observe. For example, if the tutor actively imagined how their pupil would answer the same questions, then they should gain the benefits of watching another tutor's pupil. However, if the tutors do not engage in that level of imagination, then they may not gain the benefit of recursive feedback. It would be worthwhile to identify specific mechanism that help tutors continue to think about their pupils even when they are out of sight. Such research may have useful implications for online asynchronous courses, where teachers presumably need to imagine how their students would respond since they do not have real-time interaction with their students.

Another limitation that needs to be addressed in future work is that even though all participants were given equal amount of time (and used the full time) to customize, there was still a difference in the amount of customization across participants. Possibly, some participants were more selective with their choices than others. As a result, the amount of customization may have influenced motivation for learning. In a future study a random assignment of participants to heavier and lighter customization conditions, may help further identify the relation between customization, learning, and motivation.

8. Summary

Two studies tested the hypothesis that an important facet of learning by teaching is the opportunity to watch one's student perform. The studies took place in Second Life where college students who taught and observed their pupil avatar exhibited superior learning relative to several control conditions. The first study isolated the potential value of recursive feedback through an ablation study that systematically removed various phases of the full LBT cycle. The measures (i.e., Question Sets A, B, C) were analytically introduced at different points in the study to gain some purchase on how the different LBT phases influenced learning. The value of recursive feedback seemed to extend to online virtual environments even if students never meet in person. The findings indicate that recursive feedback is effective and should be included in LBT instructional models for online learning environments. The second experiment explored whether design choices often used to create one's own ideal social environment had any additional value for recursive feedback during LBT. The customization condition involved tutoring a pupil avatar that the participant customized prior to the study and observing the pupil avatar answer questions. The doppelgänger look-alike condition involved tutoring a pupil avatar that looked like the participant and observing the pupil avatar answer questions. Results show that the recursive feedback control group outperformed both the customization group and look-alike group.

However, customization still seemed to have positive implications for learning, as the posttest performance was higher compared to other LBT control conditions. Findings show that too little or too much customization might have influenced the participant to focus more on the external features of the pupil than on the pupil's internal knowledge or pupil's response. The look-alike features had the lowest performance among the recursive feedback conditions, but still better compared to the LBT control conditions where no teaching occurred.

References

Ahn, S. J., & Bailenson, J. N. (2011). Self-endorsing versus other-endorsing in virtual environments: the effect on brand attitude and purchase intention. *Journal of Advertising, 40*(2), 93–106.

Andsager, J. L., Bemker, V., Choi, H.-L., & Torwel, V. (2006). Perceived similarity of exemplar traits and behavior: effects on message evaluation. *Communication Research, 33*, 3–18.

Arnseth, H. C., & Ludvigsen, S. (2006). Approaching institutional contexts: systemic versus dialogic research in CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*, 167–185.

Azmitia, M. (1996). Peer interactive minds: developmental, theoretical, and methodological issues. In P. B. Baltes, & U. M. Staudinger (Eds.), *Interactive minds: Life-span perspectives on the social foundation of cognition* (pp. 133–162). Cambridge, UK: Cambridge University Press.

Bailenson, J. N. (2012). Doppelgangers: a new form of self. *The Psychologist, 25*(1), 36–39.

Bailenson, J. N., Iyengar, S., Yee, N., & Collins, N. (2008). Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly, 72*(5), 935–961.

Bailenson, J. N., Patel, K., Nielsen, A., Bajscy, R., Jung, S., & Kurillo, G. (2008). The effect of interactivity on learning physical actions in virtual reality. *Media Psychology, 11*, 354–376.

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences, 17*, 102–141.

Bailey, R., Wise, K., & Bolls, P. (2009). How avatar customizability affects children's arousal and subjective presence during junk food-sponsored online video games. *CyberPsychology and Behavior, 12*(3), 277–283.

Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychology, 3*, 265–299.

Bandura, A., & Huston, A. C. (1961). Identification as a process of incidental learning. *Journal of Abnormal & Social Psychology, 63*, 311–318.

Bandura, A., & Jeffery, R. W. (1973). Role of symbolic coding and rehearsal processes in observational learning. *Journal of Personality and Social Psychology, 26*, 122–130.

Barab, S. (2006). Design-based research, a methodological toolkit for the learning scientist. In *The Cambridge handbook of the learning sciences* (pp. 153–169). New York: Cambridge University Press.

Bargh, J., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72*, 593–604.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 680–740). Boston: McGraw-Hill.

Beall, A. C., Bailenson, J. N., Loomis, J., Blascovich, J., & Rex, C. (2003). Non-zero-sum mutual gaze in immersive virtual environments. In *Proceedings of HCI international* (pp. 1108–1112). New York: ACM Press.

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

Biswas, G., Schwartz, D. L., Bransford, J. D., & , The Teachable Agents Group at Vanderbilt [TAG-V]. (2001). Technology support for complex problem solving: from SAD environments to AI. In K. Forbus, & P. Feltovich (Eds.), *Smart machines in education* (pp. 71–98). Menlo Park, CA: AAAI/MIT Press.

Blair, K., Schwartz, D., Biswas, G., & Leelawong, K. (2007). Pedagogical agents for learning by teaching: teachable agents. *Education Technology, 47*, 56–61.

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education, 23*, 194–222.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: the role of nonmotivated factors in above-average and comparative optimism effects. *Psychological Bulletin, 130*, 813–838.

Chase, C. S., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: increasing the effort towards learning. *Journal of Science and Educational Technology, 18*, 334–352.

Chi, M. T. H. (2000). Self-explaining expository texts: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.). *Advances in instructional psychology, Vol. 5* (pp. 161–237). Mahwah, NJ: Erlbaum.

Chi, M. T. H., deLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477.

Chi, M. T. H., Silver, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471–533.

Cohen, J. (1986). Theoretical considerations of peer tutoring. *Psychology in the Schools, 23*, 175–186.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*, 715–730.

Crook, D. (2002). Deferring to resources: collaborations and traditional versus computer-based notes. *Journal of Computer-Assisted Learning, 18*, 64–76.

De Bruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London B, 269*, 1307–1312.

Deci, E. L., & Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. In L. Berkowitz (Ed.). *Advances in experimental social psychology, Vol. 13* (pp. 39–80). New York: Academic Press.

Dede, C., Whitehouse, P., & Brown-L'Bahy, T. (2002). Designing and studying learning experiences that use multiple interactive media to bridge distance and time. In C. Vrasidas, & G. Glass (Eds.), *Current perspectives on applied information technologies. Distance education, Vol. 1* (pp. 1–30). Greenwich, CT: Information Age Press.

Dufrene, B., Noell, G., Gilbertson, D., & Duhan, G. (2005). Monitoring implementation of reciprocal peer tutoring: identifying and intervening with students who do not maintain accurate implementation. *School Psychology Review, 34*(1), 74–86.

Dunning, D. (1999). A newer look: motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry, 10*, 1–11.

Epley, N., & Whitchurch, E. (2008). Mirror, mirror on the wall: enhancement in self-recognition. *Personality and Social Psychology Bulletin, 34*, 1159–1170.

Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: a component analysis. *Journal of Educational Psychology, 81*, 173–177.

Fox, J., & Bailenson, J. N. (2009). Virtual self-modeling: the effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology, 12*, 1–25.

Fox, J., Bailenson, J. N., & Binney, J. (2009). Virtual experiences, physical behaviors: the effect of presence on imitation of an eating avatar. *PRESENCE: Teleoperators & Virtual Environments, 18*(4), 294–303.

Garau, M., Slater, M., Vinayagamoorhty, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 529–536). New York: ACM Press.

Garbarino, J. (1975). The impact of anticipated reward upon cross-age tutoring. *Journal of Personality and Social Psychology, 32*, 421–428.

Graesser, A. C., Person, N., & Magliano, J. (1995). Collaborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychologist, 9*, 359–387.

Greenwood, C. R., Carta, J. J., & Hall, R. V. (1989). The use of peer tutoring strategies in classroom management and educational instruction. *School Psychology Review, 77*(2), 258–275.

Hara, N., & Kling, R. (2001). Student distress in web-based distance education. *EDUCAUSE Quarterly, 3*, 68–69.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hodges, L. F., Anderson, P., Burdea, G. C., & Hoffmann, B. O. (2001). Treating psychological and physical disorders with VR. *IEEE Computer Graphics and Applications, 21*(6), 25–33.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the role of choice: a cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology, 76*, 349–366.

Juan, M. C., Alcaniz, M., Monserrat, C., Botella, C., Banos, R. M., & Guerrero, B. (2005). Using augmented reality to treat phobias. *IEEE Computer Graphics and Applications, 25*(6), 31–37.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772–775.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: toward the understanding of a double-edged sword. *Current Directions in Psychological Science, 7*(3), 67–72.

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proceedings of artificial intelligence in education* (pp. 383–390). IOS Press.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*, 311–328.
Lawton, D., Vye, N., Bransford, J., Sanders, E., Richey, M., French, D., et al. (2012). Online learning based on essential concepts and formative assessment. *Journal of Engineering Education, 101*(2), 1–44.
Lerner, M. J., & Agar, E. (1972). The consequence of perceived similarity: attraction and rejection, approach and avoidance. *Journal of Experimental Research in Personality, 6*, 69–75.
McCloud, S. (1993). *Understanding comics*. Amherst, MA: Kitchen Sink Press.
MacKenzie, D. A., & Wajcman, J. (1985). *The social shaping of technology: How the refrigerator got its hum*. Milton Keynes: Open University Press.
Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social Cognitive and Affective Neuroscience, 1*, 75–86.
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*(2), 186–207.
Morgan, R., & Toy, T. (1970). Learning by teaching: a student-to-student compensatory tutoring program in a rural school system and its relevance to the educational cooperative. *Psychological Record, 20*, 159–169.
Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Erlbaum.
Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed.). (pp. 125–143) Mahwah, NJ: Erlbaum.
Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218.
Okita, S. Y., Bailenson, J., & Schwartz, D. L. (2007). The mere belief of social interaction improves learning. In D. S. McNamara, & J. G. Trafton (Eds.), *The proceedings of the 29th meeting of the cognitive science society* (pp. 1355–1360). Nashville, USA: August.
Okita, S. Y., & Jamalian, A. (2012). Learning from the folly of others: learning to self-correct by monitoring the reasoning of projective pedagogical agents. In J. Aalst, K. Thompson, M. J. Jacobson, & P. Reimann (Eds.). *Proceedings for the 10th international conference of the learning sciences (ICLS), Vol. 2* (pp. 281–285). Sydney: University of Sydney: ISLS.
Okita, S. Y., & Schwartz, D. L. (2006). When observation beats doing: learning by teaching. In S. Barab, K. Hay, & D. Hickey (Eds.). *Proceedings of the 7th international conference of the learning sciences, Vol. 1* (pp. 509–515). New Jersey: Lawrence Erlbaum Associates, Bloomington, USA.
Okita, S. Y., & Schwartz, D. L. Learning by teaching: the recursive feedback that comes from observing a human pupil or agent, Journal of the Learning Sciences, submitted for publication.
Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension monitoring activities. *Cognition and Instruction, 1*, 117–175.
Ploetzner, R., Dillenbourg, P., Praier, M., & Traum, D. (1999). Learning by explaining to oneself and to others. In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 103–121). Oxford: Elsevier.
Ratan, R., & Bailenson, J. N. (2007). *Similarity and persuasion in immersive virtual reality*. Panel presentation to the Communication and Technology Commission of ICA.
Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television and new media like real people and places*. New York: Stanford University, Center for the Study of Language and Information.
Renkl, A. (1995). Learning for later teaching: an exploration of mediational links between teaching expectancy and learning results. *Learning and Instruction, 5*, 21–36.
Rittschof, K., & Griffin, B. (2001). Reciprocal peer tutoring: re-examining the value of a cooperative learning technique to college students and instructors. *Educational Psychology, 21*(3), 313–331.
Riva, G., Bacchetta, M., Baruffi, M., Rinaldi, S., & Molinari, E. (1999). Virtual reality based experiential cognitive treatment of anorexia nervosa. *Journal of Behavior Therapy and Experimental Psychiatry, 30*(3), 221–230.
Rizzo, A. A., Bowerly, T., Buckwalter, G., Klimchuk, D., Mitura, R., & Parsons, T. D. (2006). A virtual reality scenario for all seasons: the virtual classroom. *CNS Spectrums, 11*(1), 35–44.
Rizzo, A. A., Difede, J., Rothbaum, B. O., Johnston, S., McLay, R. N., Reger, G., et al. (2009). VR PTSD exposure therapy results with active duty OIF/OEF combatants. In J. D. Westwood, S. W. Westwood, R. S. Haluck, H. M. Hoffman, G. T. Mogel, R. Phillips, et al. (Eds.), *Medicine meets virtual reality, 17 – NextMed: Design for the well being*. Amsterdam: IOS Press.
Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: reflective knowledge building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*, 534–574.
Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: the role of explaining and responding to questions. *Instructional Science, 33*(4), 321–350.
Ryokai, K., Vaucelle, C., & Cassell, J. (2002). *Literacy learning by storytelling with a virtual peer*. Paper presented at the meeting of computer support for collaborative learning.
Säljö, R. (2004). Learning and technologies, people and tools in coordinated activities. *International Journal of Educational Research, 41*, 489–494.
Schön, D. A. (1983). *The reflective practitioner – How professionals think in action*. NY: Basic Books.
Schwartz, D. L., Blair, K. P., Biswas, G., Leelawong, K., & Davis, J. (2007). Animations of thought: interactivity in the teachable agents paradigm. In R. Lowe, & W. Schnotz (Eds.), *Learning with animation: Research and implications for design*. UK: Cambridge University Press.
Segovia, K. Y., & Bailenson, J. N. (2009). Virtually true: children's acquisition of false memories in virtual reality. *Media Psychology, 12*, 371–393.
Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction, 20*(2), 119–150.
Siegler, R. S. (1995). How does change occur: a microgenetic study of number conservation. *Cognitive Psychology, 28*, 225–273.
Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Garnott, & J. Parziale (Eds.), *Microdevelopment: A process-oriented perspective for studying development and learning* (pp. 31–58). Cambridge, MA: Cambridge University Press.
Sikora, A. C., & Carroll, C. D. (2002). *Postsecondary education descriptive analysis reports (NCES 2003-154)*. US Department of Education, National Center for Education Statistics. Washington, DC: US Government Printing Office.
Slamecka, N. J., & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 592–604.
Stahl, G., & Hesse, F. (2006). ijCSCL-a journal for research in CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*(1), 3–8.
Turkay, S., & Adinolf, S. (2010). Free to be me: a survey study on customization with World of Warcraft and City of Heroes/Villains players. *Procedia Social and Behavioral Sciences, 2*, 1840–1845.
Turkle, S. (1984). *The second self. Computers and the human spirit*. New York: Simon and Schuster.
Uretsi, J. A. R. (2000). Should I teach my computer peer? Some issues in teaching a learning companion. In G. Gautheir, C. Frasson, & K. VanLehn (Eds.), *Intelligent tutoring systems* (pp. 103–112). Berlin: Springer-Verlag.
Van den Bossche, P., Gijselaers, W., Segers, M., & Kirschner, P. (2006). Social and cognitive factors driving teamwork in collaborative learning environments. *Small Group Research, 37*, 490–521.
Van Vugt, H. C., Bailenson, J. N., Hoorn, J. F., & Konijn, E. A. (2010). Effects of facial similarity on user responses to embodied agents. *ACM Transactions on Computer-Human Interaction, 17*(2), 1–27, 7.
Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13*, 21–39.
Yee, N., & Bailenson, J. (2007). The Proteus effect: the effect of transformed self-representation on behavior. *Health Communication Research, 33*, 271–290.
Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269–274.