Pedagogical Voice in an E-Learning System: Content Expert Versus Content Novice

Lincoln Sedlacek, Victor Kostyuk[✉], Matthew Labrum,
Kevin Mulqueeny, Georgina Petronella, and Maisie Wiltshire-Gordon

Reasoning Mind, Houston, TX, USA
lincolnsedlacek@gmail.com,
vkostyuk@reasoningmind.org, mlabrum@gmail.com,
mulqueeny@gmail.com, gpetronella@gmail.com,
maisiewg@gmail.com

Abstract. Studies have shown student performance in virtual learning environments is improved by the presence of a virtual pedagogical agent, particularly when the agent is voiced by a human voice. Furthermore, studies have shown students achieve higher learning outcomes when their teacher is a content expert in the material being taught. This study examines the question at the intersection of these two domains: Do students achieve higher learning outcomes in a virtual learning environment if the actor voicing the virtual agent is a content expert in the material being taught? The analysis found no evidence of such an effect, although more research should be conducted before firm conclusions are drawn.

Keywords: Pedagogical agents · e-Learning · Content expertise · Narration

1 Background

Over the years, a large body of research has accumulated concerning e-learning environments, virtual instruction, and online lessons. Studies have demonstrated positive learning outcomes for students who use e-learning and online instructional programs. At the same time, the computerized nature of these programs allows for easily replicated personalized learning experiences [3, 6].

In studies on the efficacy of e-learning programs, a great deal of attention has been given to the use of on-screen coaches—typically referred to as "pedagogical agents"—to improve learning outcomes. Pedagogical agents are defined in Clark and Mayer's *e-Learning and the Science of Instruction* [3] as "on-screen characters who help guide the learning process during an e-learning episode." The embodiment of a pedagogical agent varies greatly among e-learning programs—some are presented as high-quality talking-head videos, while others may appear as static, cartoon-like characters that "speak" only via printed text.

While research on the effectiveness of pedagogical agents is still in its early stages, multiple studies have shown correlations between the use of embodied pedagogical agents in e-learning programs and improved student learning outcomes [1, 2, 7, 10]. Additionally, studies have shown that, while student performance is not necessarily

improved by the agent having a realistic human appearance [10], student performance *is* improved when the agent exhibits human-like behavior, such as gesturing or gazing at relevant material, or speaking in a personalized fashion [7, 9]. Further research has shown that, to improve learning, agents should communicate through speech, rather than through on-screen text [1, 10]; that students perform better when such speech is polite, rather than more direct [12]; and that students perform better when the pedagogical agent has a human voice as opposed to a machine-generated voice [2].

Of course, giving a pedagogical agent a human voice requires a human voice actor. And while there have been several studies on the ideal appearance, sound, and demeanor of a pedagogical agent, there is a dearth of research on what qualities are important in the voice actor voicing a pedagogical agent. The developers of the Reasoning Mind blended learning program for middle school math, *Edifice*, hypothesized that student learning would increase among users of their e-learning software if the voice actors voicing pedagogical agents were content experts in the mathematical subject matter.

1.1 Theoretical Basis

There is some theoretical basis for believing a content-expert voice actor would have a more positive impact on student learning outcomes. Research has shown that, in a real-world mathematics classroom, instructor competency in classroom subject matter greatly affects student learning [5].

However, a pedagogical agent's voice actor is very different from a classroom instructor. Classroom instructors are typically in charge of curriculum construction, lesson content and structure, lesson delivery, one-on-one intervention, classroom management, and a host of other tasks. Voice actors for pedagogical agents typically control a much narrower set of features of their (virtual) classrooms. True, there are examples of virtual learning environments like Khan Academy [4], where Sal Khan originally planned the lesson, wrote the script, and then delivered it in his own voice as an unseen pedagogical agent. Yet many virtual learning environments (Reasoning Mind *Edifice* included) employ curriculum designers who plan, design, and write lessons; have programmers who create the lesson visuals; and recruit voice actors after the lesson is scripted to deliver the lines for the final product. In such settings, voice actors narrate, with limited deviation, pre-written instructional scripts—they do not compose those scripts.

Despite the limited role voice actors have when compared to a more typical course instructor, there are still reasons why their status as a content expert or a content novice may be important when it comes to voicing a pedagogical agent in a virtual or online math lesson:

1. A content novice may not understand the technical vocabulary used in the script, resulting in mispronunciations and disfluencies that are distracting to the student.
2. A content novice may speak without confidence in the delivery, making the student question their understanding and disengage.
3. In an attempt to maintain fluency and confidence, a content novice may deliver lines or modify the script in a way that impedes student understanding, or worse yet, promotes an incorrect understanding of the material.

This last effect may be the most important, and is one that Reasoning Mind *Edifice*'s creators worried might have been occurring within the *Edifice* program.

The following is an example of how line delivery may impede student understanding. During several *Edifice* lessons, students learned about, and completed exercises involving, inverse proportionality. Such lessons would include lines like the following:

> "So, if these quantities are *inversely* proportional, and one of them *increases*, then the other will *decrease*."

A mathematical content expert would understand the importance of distinguishing the words "increases" and "decreases" from each other, which can be done by emphasizing the syllables "in-" and "de-" (highlighting the parts of the words that are different). However, one of the *Edifice* voice actors would consistently emphasize the "-crease" syllable of both words instead, a practice that obscured the meaning of the line by highlighting the *similarity* between the two words.

This, of course, is just one example of line delivery being affected by a voice actor's status as a content novice or content expert. During the *Edifice* creation process, however, several more occurred, such as:

- Word mispronunciations (e.g., pronouncing "multiplicative" as "**muhl**-*tih*-play-*tihv*," which leaves out one full syllable)
- Emphases that obscured the meanings of lines (such as the "*in*crease/*de*crease" vs. "in*crease*/de*crease*" example given above)
- Carelessly paced speech patterns that actively communicated an incorrect meaning of a line (e.g., speaking the line, "The answers are twenty, two, and seven," too quickly, making it sound like, "The answers are twenty-two and seven")
- Emphases that actively communicated an incorrect meaning of a line (e.g., speaking a line like, "You drew a line parallel to line CD. You should have drawn a line perpendicular to line BC," while only emphasizing parallel and perpendicular; such a mistake might lead students to hear "CD" and "BC" as the same, and not recognize that mistake)

As Reasoning Mind curriculum creators, who are content experts in mathematics[1], reviewed lessons in their final form, many felt that these issues were occurring, and may affect student learning outcomes.

All three *Edifice* voice actors had professional experience in voice acting, with each having professional performance credits to their name. However, these voice actors also self-reported having average mathematical competency—they generally understood the material being covered, but wouldn't have considered themselves content experts. Further discussion made it clear that, in many cases where lines were delivered problematically, it was due to a lack of expert-level understanding of the material at hand.

---

[1] Most Reasoning Mind curriculum developers have PhDs or Master's degrees in mathematics or another STEM field.

To better understand the relationship between a pedagogical agent voice actor's mathematical competency and student learning outcomes, we ran an experiment to see if student learning outcomes would be higher with amateur voice actors who were content experts as opposed to professional voice actors who were content novices.

2   Methods

2.1   The Reasoning Mind *Edifice* Platform

Reasoning Mind *Edifice* is a blended learning platform for middle school (Grades 6, 7, and 8) mathematics in which students spend part of their class time learning online, and part learning from teacher-led instruction. Within the online program, students go through a series of over 100 lessons throughout the school year.

Students complete lessons in a "virtual classroom" (See Fig. 1), in which mathematical instruction is delivered through text, illustrations, animations, and voice. Most of the virtual classroom consists of a large "whiteboard" on which information appears in the form of text, formulas, figures, diagrams, problems, animations, and other relevant material.

Also visible in the virtual classroom are two virtual students and a virtual tutor, who go through the lesson with the user. The virtual students act much as a typical student would, asking questions, attempting problems, and contributing to the classroom
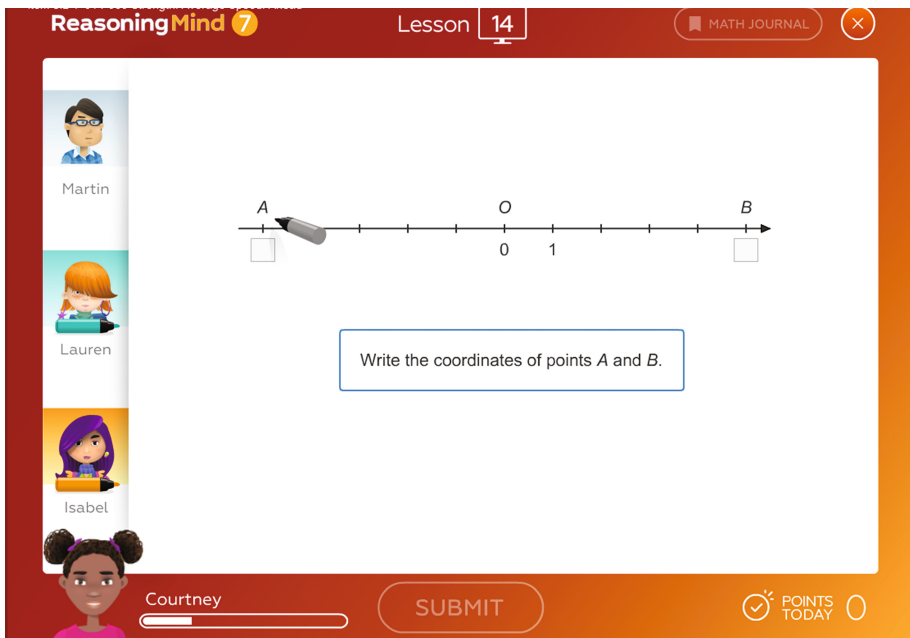

Fig. 1. The Reasoning Mind *Edifice* virtual classroom

discussion of mathematics. The virtual tutor—the pedagogical agent this experiment concerns itself with—leads the lesson much like a traditional teacher would, reviewing homework, introducing the students to new material, explaining concepts, walking students through example problems, leading virtual classroom discussions, asking questions, reacting to students' answers to questions, and reviewing lesson material.

All virtual characters can ask the student user questions; the virtual tutor does this frequently, and the virtual students do occasionally. Student users then respond via one of *Edifice*'s many response tools. A virtual character will give one of several possible reactions to the user's response, depending on what response they gave; a correct response will usually result in some type of congratulatory reaction, while an incorrect response will typically result in course correction coupled with reassuring motivation. If a student user fails to respond to a question for a certain amount of time, a virtual character will continue the lesson, typically with an explanation of the correct answer. Adaptivity occurs at the fine-grained level of individual student answers (e.g., the virtual tutor is programmed to adapt feedback and follow-up questions to answers indicative of particular misconceptions), and at a course-grained level, where stronger students get more challenging problems or less scaffolding during the problem-solving process while weaker students get simpler problems and more scaffolding.

The lessons in *Edifice*'s middle school curriculum house a cast of three virtual tutors and seven virtual students. Virtual tutors and students are rotated through lessons at a rate of two students and one tutor per lesson, providing variety in the virtual characters' genders, races, personalities, portrayed mathematical strengths, etc.

As Reasoning Mind *Edifice*'s pedagogical agents, the virtual characters—and the virtual tutor, in particular—are key features when it comes to promoting positive learning outcomes among student users of the program.

Edifice has shown promising results in improving student learning and engagement. Students using *Edifice* have shown greater growth in measures of achievement [8] and engagement [11] compared to similar students in traditional classrooms.

2.2   Experimental Design

This experiment required one student group to go through lessons voiced by *Edifice*'s professional voice actors (content novices), who had average general competency in school mathematics. Another student group went through lessons voiced by content experts who were not trained voice actors . Students groups were randomly assigned.

Selecting New Voice Actors. Content-expert voice actors were selected using an audition process similar to the process followed to select the original voice actors, moderated by a panel of judges, three of whom had PhDs in mathematics. However, this time, judges were instructed to place a heavy emphasis on how competent and knowledgeable the voice actors sounded while explaining mathematical material. The result was the selection of three new voice actors rated across the board as sounding: (1) like content experts in mathematics, and (2) generally better at performing the content for chosen lesson sections than at least two of our original three voice actors.

The judges also selected another new voice actor that they rated as sounding as close as possible to the original voice actors (a novel content novice). The inclusion of this new content novice voice actor in the experiment was done to investigate the possibility of a novelty effect, as explained in the "implementation" section below.

Implementing New Voice Actors in Lessons. Study designers chose three seventh-grade lessons in which they could implement new voice actors. Each lesson had a different one of the three original professional voice actors who were content novices.

Lesson 105 received two treatment conditions. The version recorded using the original content-novice voice actor was Condition A. Condition B was recorded using the novel content-novice voice actor. Condition C was recorded using the content-expert voice actor. These three conditions allowed for the investigation of a possible novelty effect: if students' performance improved in comparable measure in both Conditions B and C, it would likely be due to a novelty effect. If student performance improved due to the content-expert voice actor in Condition C, student learning outcomes should be significantly higher in Condition C than in Condition B.

Lessons 110 and 112 were each re-recorded once. The version recorded using the original content-novice voice actor was Condition A. Condition B was recorded using the content-expert voice actor. A third version of Lessons 110 and 112 was also created in which the lessons began with the content-expert-voiced virtual tutor being introduced in a short, introductory video. This version was related to another study, and whether students received video or non-video versions of these lessons was randomized for both lessons to mitigate any effects of the third condition.

For every condition with a novel voice actor, virtual tutor avatars were changed, and each new tutor introduced themselves as a new tutor so that students would not be confused by what appeared to be their old tutor suddenly having a different voice. Some lines were also given minor revisions to accommodate new personalities and speaking patterns. (For example, it might distract a student if a new virtual tutor used a catchphrase that was unique to one of the original virtual tutors).

2.3   Participants

Participants in the study were seventh-grade mathematics students from a mid-sized city in Central Texas. Using the experimental setup of the three lessons mentioned above, we conducted a randomized-controlled trial. In the first lesson (Lesson 105), 191 students were included in the experiment, with 65 included in Condition A (content-novice professional voice actor), 59 included in Condition B (content-expert amateur voice actor), and 67 in Condition C (novel content-novice amateur voice actor). The second and third lessons (Lessons 110 and 112) each had 168 students included in the experiment, with 54 students in Condition A and 114 in Condition B.

3   Results

We randomly assigned 191 students into groups that received different versions of Lessons 105, 110, and 112; however, not all students completed every lesson or every outcome measure, due to instances of student absences and other conflicts. For this reason, the sample sizes vary slightly for the homework, quiz, and final exam. In lesson 105, the original content novice was the tutor voice actor for 65 students, the content expert was the tutor voice actor for 59 students, and the novel content novice was the tutor voice actor for 67 students. To establish baseline equivalence of the three groups, we examined the average of scores on two quizzes halfway through the course. The groups did not differ significantly, averaging 54.1 (SD = 27.8), 52.2 (SD = 26.2), and 53.5 (SD = 27.4) for the original content novice, content expert, and novel content novice, respectively (pairwise *t*-tests between groups had *p*-values equal to 0.66 for the original content novice and the content expert, 0.89 for the original content novice and the novel content novice, and 0.76 for the content expert and the novel content novice).

Measured outcomes for Lesson 105 are shown in Table 1. Student learning outcomes were determined by measuring lesson accuracy (i.e., the percent of questions answered correctly within the given lesson) and homework accuracy (percent of questions answered correctly on the assigned homework following the lesson), in addition to accuracy on the quiz in Lesson 110 and the Final Exam, both measured as a percentage of questions answered correctly. Number of timeouts (i.e., the number of times a student did not answer a question posed by a virtual character) was used as a measure of student engagement and in-lesson understanding (with the belief being that students who didn't answer questions were less engaged or had trouble understanding the material). Lesson completion time (number of hours it took students to complete the lesson) was also used as a potential measurement for student engagement and in-lesson

Table 1. Lesson 105 ANOVA results

	Condition A: original con. novice M (SD)	Condition B: con. Expert M (SD)	Condition C: novel con. novice M (SD)	F (df)	p-value
Lesson accuracy	63.5 (20.1)	61.3 (20.6)	59.4 (22.5)	0.62 (2,188)	0.54
Lesson completion time	0.79 (.22)	0.83 (.20)	0.88 (.22)	2.66 (2,188)	0.07
Number of timeouts	8.4 (8.6)	6.4 (6.7)	10.5 (12.7)	2.79 (2,188)	0.06
Homework accuracy	41.1 (32.9)	48.5 (32.4)	37.5 (29.9)	1.76 (2,166)	0.18
Quiz accuracy	58.7 (42.8)	49.1 (41.4)	48.3 (45.5)	0.95 (2,163)	0.39
Final exam accuracy	57.0 (29.9)	45.9 (32.5)	44.9 (27.7)	2.08 (2,129)	0.13

understanding (with the assumption that engaged students who had an easier time understanding the material would finish the lesson more quickly). However, this measure should not be over-interpreted: it may also reflect differing voice actor speech speeds.

According to a one-way analysis of variance (ANOVA), the three groups of students did not differ significantly on lesson accuracy, lesson completion time, number of timeouts during the lesson, homework accuracy, Quiz accuracy (in Lesson 110), or Final Exam performance. The number of timeouts and lesson completion time approached significance, with p-values less than 0.1, but pairwise comparisons found no significant differences between any pair of groups. There also did not appear to be any effect of the novelty of the guest tutors.

In Lessons 110 and 112, one group of students received instruction from virtual tutors voiced by the content-novice voice actors (54 students), while the rest were instructed by virtual tutors voiced by content-expert voice actors (114 students). We checked baseline equivalency: the groups did not differ significantly on the average performance of two quizzes halfway through the course, with students who had the content novice averaging 54.1 (SD = 27.8) and those with content experts averaging 52.8 (SD = 26.6, p = 0.73). According to an independent-sample t-test, performance within Lesson 110 was not significantly different between the two groups on lesson accuracy, lesson completion time, number of timeouts, or subsequent homework accuracy. Students with the content-novice voice actors scored 10% points higher on the quiz in Lesson 110 than students with content-expert voice actors, an effect size of 0.23 standard deviations. A similar t-test found performance in Lesson 112 was not significantly different between the two groups, but the content-expert group took significantly longer to complete the lesson. The content-novice group scored significantly higher on the Final Exam in lesson 114, an effect size of 0.39 standard deviations (Tables 2 and 3).

Table 2. Lesson 110 t-test results

	Condition A: content novice M (SD)	Condition B: content expert M (SD)	t (df)	p-value
Lesson accuracy	63.0 (20.8)	60.8 (19.9)	0.66 (161)	0.51
Lesson completion time	0.70 (.14)	0.71 (.16)	–0.66 (161)	0.51
Number of timeouts	3.7 (4.4)	3.9 (3.9)	–0.32 (165)	0.75
Homework accuracy	47.6 (37.6)	44.3 (34.4)	0.51 (136)	0.61
Quiz accuracy	58.7 (42.8)	48.7 (43.3)	1.38 (164)	0.17

Table 3. Lesson 112 *t*-test results

	Condition A: content novice *M (SD)*	Condition B: content expert *M (SD)*	t *(df)*	p-*value*
Lesson accuracy	69.6 (19.3)	68.2 (18.6)	0.43 (147)	0.67
Lesson completion time	0.81 (.17)	0.89 (.20)	**−2.54 (147)**	**0.01***
Number of timeouts	7.7 (7.4)	6.1 (5.8)	1.46 (152)	0.15
Homework accuracy	47.3 (29.5)	47.7 (31.8)	−0.07 (128)	0.94
Final exam accuracy	57.0 (29.9)	45.4 (30.1)	**2.04 (130)**	**0.04***

*a t-test p value of less than 0.05

In all three lessons, there was no significant difference between the groups in each measure of student engagement and student performance after adjusting for multiple comparisons.

4 Discussion

Somewhat surprisingly, the results indicate that the use of content experts as voice actors for pedagogical agents provides no significant effect on either student engagement or student performance. For each group of students (those taking lessons in which the virtual tutor was voiced by the content novice, versus those taking lessons in which the virtual tutor was voiced by a content expert), in each lesson, baseline equivalency was determined by noting the average scores on two quizzes approximately halfway through the course were not significantly different. In all three lessons, there was no significant difference between the groups in each measure of student engagement and student performance after adjusting for multiple comparisons. Prior to the adjustment, two statistically significant (at the 0.05 level) differences were present, in the time taken to complete the last of the three experimental lessons and in the score on the final exam.

Our results may suggest that, unlike in a real-world mathematics classroom, in which instructor quality greatly affects student learning [5], the scripted environment of an e-learning system greatly reduces the effect a tutor voice actor's personal understanding of the material may have on student performance. Therefore, while having a human voice for a pedagogical agent in a virtual learning environment improves student learning outcomes [2], it may not matter whether or not that human voice actor is a content expert.

It is also possible that the mathematical competency of a pedagogical agent's voice actor does have an effect on student learning outcomes, and we simply did not detect such an effect. Such an effect could have gone undetected by us due to an insufficient sample size, or due to difficulties associated with completely isolating variable voice actor characteristics. (For example, voice actors did not differ solely in whether they

were a content novice or a content expert, but also in their voiceprints, recording equipment, personal vocal tics, etc.)

Another possibility is that professional voice actors—even ones who are content novices—have developed the skills necessary to deliver material for which they lack a content-expert-level understanding, while amateur voice actors—even content experts—aren't experienced enough in voice acting to deliver all material effectively. This could explain our results, although our lack of an effect on student learning outcomes is still a bit surprising since the presence of Conditions A (content novice, professional voice actor), B (content expert, amateur voice actor), and C (content novice, amateur voice actor) should have controlled for the possibility of this effect. That is, if the advantages of being a professional voice actor and being a content expert were equal, both should have resulted in significantly higher learning outcomes than the content novice, amateur voice actor condition.

It's also notable that, at the time our experimental design was implemented, seventh-grade *Edifice* students had already gone through over 100 lessons with their original seventh-grade virtual tutor voice actors. While three lessons with content expert voice actors may not have resulted in a significant improvement in students' learning outcomes, it is possible that a full year of such lessons would have made a large significant difference. Additionally, the distraction of new virtual tutors, along with the time it took students to get used to them, might have mitigated any benefit students could have gotten from the content expert voice actors' better delivery of mathematical content.

A final possibility is that, while mathematical content experts may have a better idea of how best to explain mathematical content *to another content expert*, voice actors with lower levels of mathematical competency might be more highly attuned to what parts of an explanation students might have trouble with. Therefore, voice actors with lower levels of mathematical competency might better know what parts of an explanation they should emphasize or otherwise set apart when delivering lines to a student. This would result in content novice voice actors fostering higher student learning outcomes, even as they would likely be rated as having poorer voice actor performance by mathematical content experts.

5   Conclusion

In a virtual learning environment, students tend to achieve higher learning outcomes when an embodied pedagogical agent is present in the environment [1, 2, 7, 10]—and they tend to achieve even higher learning outcomes when said virtual pedagogical agent has a natural, human voice [2]. While we theorized that students would achieve even higher learning outcomes if that virtual pedagogical agent was voiced by a content expert, our study found no evidence for this hypothesis. However, while our foray into examining this question may give some guidance to aspiring e-learning designers, more research should be conducted before arriving at firm conclusions.

References

1. Atkinson, R.K.: Optimizing learning from examples using animated pedagogical agents. J. Educ. Psychol. **94**, 416–427 (2002). doi:10.1037/0022-0663.94.2.416
2. Atkinson, R.K., Mayer, R.E., Merrill, M.M.: Fostering social agency in multimedia learning: examining the impact of an animated agent's voice. Contemp. Educ. Psychol. **30**(1), 117–139 (2005). doi:10.1016/j.cedpsych.2004.07.001
3. Clark, R.C., Mayer, R.E.: e-Learning and the Science of Instruction, 4th edn. Wiley, Hoboken (2016). doi:10.1002/9781119239086
4. Dreifus, C.: Salman Khan turned family tutoring into Khan Academy. The New York Times, 27 January 2014. https://www.nytimes.com/2014/01/28/science/salman-khan-turned-family-tutoring-into-khan-academy.html. Accessed 9 Feb 2017
5. Hanushek, E.A., Rivkin, S.G.: Chapter 18 teacher quality. In: Hanushek, E.A. (ed.) Handbook of the Economics of Education, pp. 1051–1078 (2006). doi:10.1016/s1574-0692 (06)02018-6
6. Horn, M.B., Staker, H.: The Rise of K-12 Blended Learning. Innosight Institute, Inc., Mountain View (2011)
7. Lusk, M.M., Atkinson, R.K.: Animated pedagogical agents: does their degree of embodiment impact learning from static or animated worked examples? Appl. Cogn. Psychol. **21**(6), 747–764 (2007). doi:10.1002/acp.1347
8. Mingle, L., Kostyuk, V., Khachatryan, G.: Adapting international math curricula using an online platform. In: Poster to be Presented at The Annual Meeting of the American Educational Research Association in San Antonio (2017)
9. Moreno, R., Mayer, R.E.: Personalized messages that promote science learning in virtual environments. J. Educ. Psychol. **96**(1), 165–173 (2004). doi:10.1037/0022-0663.96.1.165
10. Moreno, R., Mayer, R.E., Spires, H., Lester, J.: The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents? Cogn. Instr. **19**, 177–214 (2001). doi:10.1207/s1532690xci1902_02
11. Mulqueeny, K., Mingle, Leigh A., Kostyuk, V., Baker, Ryan S., Ocumpaugh, J.: Improving engagement in an e-Learning environment. In: Conati, C., Heffernan, N., Mitrovic, A., Verdejo, M.Felisa (eds.) AIED 2015. LNCS, vol. 9112, pp. 730–733. Springer, Cham (2015). doi:10.1007/978-3-319-19773-9_103
12. Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., Collins, H.: The politeness effect: pedagogical agents and learning outcomes. Int. J. Hum.-Comput. Stud. **66**(2), 98–112 (2008). doi:10.1016/j.ijhcs.2007.09.003