# Impact of Agent Role on Confusion Induction and Learning

Blair Lehman and Art Graesser

University of Memphis, Memphis, TN 38152, USA
[balehman, graesser]@memphis.edu

**Abstract.** The presentation of contradictory information to trigger deeper processing and increase learning has been investigated in a variety of ways (e.g., conversational agents, worked examples). However, the impact of information source (e.g., expertise, gender) and the relationship between the contradicting sources (e.g., status level) has not been investigated to the same degree. We previously reported that confusion can successfully be induced and learning increased when contradictory information was presented by two conversational agents (tutor, peer student). In the present experiment we investigated contradictions posed by two peer student agents. Self-reports of confusion and learner responses to embedded forced-choice questions revealed that the contradictions still successfully induced confusion. There were, however, differences in the nature of confusion induction based on the inter-agent relationship (i.e., student-student vs. tutor-student). Learners performed better on transfer tasks when presented with contradictions compared to a no-contradiction control, but only when they were successfully confused.

**Keywords:** confusion, contradiction, affect, tutoring, animated pedagogical agents, intelligent tutoring systems, learning

## 1 Introduction

To understand a concept it is important to learn why a particular strategy or explanation is correct and why alternatives are incorrect. However, it is often difficult for learners to understand both aspects. One method to help learners reach this level of understanding is the presentation of contradictory information [1-5]. Contradictory information has been presented in a variety of contexts, such as conversational agents [1,3], sources within a text [6], and worked examples [2,4,5], to create cognitive conflict (see Limón [7] for review), cognitive disequilibrium [8-10], and confusion [1,3]. In all instances, the contradictory information is expected to increase learning by causing learners to stop, think, and deliberate over which alternative is correct in an effort to resolve their current cognitive and affective conflict.

There are two important considerations when presenting contradictions to increase learning. First, the contradiction must be highlighted such that learners are aware that there is a contradiction and that the two alternatives are not compatible (i.e., cannot

adfa, p. 1, 2011.

both be correct) [7]. Unfortunately, learners often dismiss the contradiction and do not engage in the beneficial cognitive activities required to compare the competing alternatives and determine which one is correct. Learners can ignore the contradiction, reject or deny the validity of one alternative, exclude one alternative from the explanation of a concept, or reinterpret one alternative so that the two alternatives are no longer in conflict [11]. Thus, it is important to present contradictions that are salient to learners within a context that requires their resolution.

The second issue to consider when presenting contradictions to increase learning is the sources of the contradicting alternatives. Research on the presentation of contradictions within a text has found that contradictions actually draw more attention to the source of information [6]. Participants have been found to have more fixations and longer gaze times on the sources of information (e.g., person A vs. person B) while reading and increased citations of sources when writing summaries compared to when sources agreed. Attention to sources can lead to source evaluation, which has been found to increase comprehension [12-15]. In fact, learners who performed better on comprehension assessments were found to evaluate information sources more while reading than those who performed less well [13].

Contradictions have also been found to be an effective catalyst for deeper reasoning when presented by conversational agents. In a series of experiments, conversational agents presented contradictions during trialogues (i.e., three-party conversations) to induce confusion and promote learning [1,3]. One agent served as a tutor, whereas the other agent served as a peer student agent. Learners who were successfully confused by the contradictions performed significantly better on measures of learning and transfer tasks compared to when the agents agreed. However, the effectiveness of confusion induction was consistently found to differ depending on which agent was correct (i.e., tutor vs. student) when the agents disagreed. This finding raises the question as to how agent role (e.g., status, status differential, gender, etc.) impacts confusion induction and learning. Baylor and Kim [16] have indeed reported that agent roles in learning environments that do not pose contradictions can impact both motivation and learning.

The present research is an initial attempt to determine the impact of agent role when contradictions are presented. To completely address this question, research should examine confusion induction and learning when agent role differs (tutor, peer student) and is the same (peer student, tutor or expert) as well as when agent characteristics (e.g., gender, race, age) are varied. The present research replicates Lehman et al. [3], but with two peer student agents instead of a tutor agent and peer student agent. Three research questions are investigated in the present research. When contradictions are presented by two peer student agents, will confusion be successfully induced (question 1) and will learning increase (question 2)? Finally, the third research question will address the similarities and differences between confusion induction and learning outcomes when contradictions are presented by agents of different status (tutor, student, [3]) compared to agents of the same status (two students). The impact of agent role will be investigated within a learning environment that diagnoses flaws in research case studies to help learners better understand research methods concepts.

## 2 Methods

### 2.1 Manipulation

We experimentally induced confusion with a contradictory information manipulation over the course of learning research methods concepts (e.g., replication, control group, validity). This was achieved by having the two student agents (male student and female student, see Figure 1) stage a disagreement on an idea and eventually invite the human learner to intervene (note that student agent refers to the animated agents, the actual human learner is referred to as learner). This confusion induction method has been found to successfully induce confusion when contradictions were posed by tutor and student agents in previous experiments [1,3].

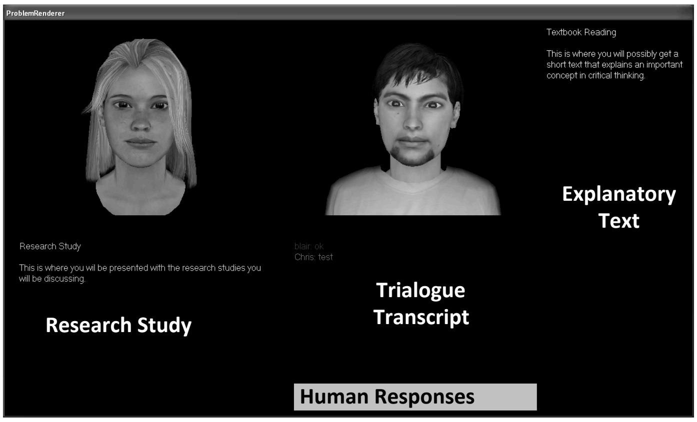


Figure 1. Screenshot of learning environment interface

Contradictions were introduced during trialgoues (three-party conversations) identifying flaws in sample research studies. Some studies had subtle flaws while others were flawless. There were four contradictory information conditions. In the *true-true* condition, both student agents agreed and presented correct opinions. In the *true-false* condition, the female student agent presented a correct opinion and the male student agent disagreed by presenting an incorrect opinion. In contrast, the male agent presented a correct opinion and it was the female agent that disagreed with an incorrect opinion in the *false-true* condition. Finally, in the *false-false* condition, both agents agreed but the opinions that they presented were incorrect. It should be noted that all misleading information was corrected after learners completed all four trialogues and posttests and that learners were fully debriefed at the end of the experiment.

### 2.2 Participants & Design

Participants were 32 undergraduate students from a mid-south university in the US and participated for course credit. The experiment had a within-subjects design with four conditions (*true-true*, *true-false*, *false-true*, *false-false*). Learners completed one trialogue in each of the four conditions with a different research methods topic in each session (4 in all). Order of conditions and topics and assignment of topics to conditions was counterbalanced across learners with a Graeco-Latin Square.

### 2.3 Procedure

The experiment occurred over two phases: (1) knowledge assessments and trialogues and (2) a retrospective affect judgment protocol.

**Knowledge Tests.** Research methods knowledge was assessed with flaw identification tasks before and after trialogues (pretest and posttest, respectively). The flaw identification tasks consisted of a description of a previously unseen study and learners were asked to identify flaw(s) in the study by selecting as many items as they wanted from a list of eight research methods topics. The list included four topics that could potentially be flawed (discussed in the trialogues) and four distractor topics (not discussed in the trialogues). Learners also had the option of selecting that there was no flaw, although each study contained one flaw. The pretest involved the presentation of four case studies that each contained one flaw. The flaw in each case study corresponded to one of the topics discussed in the trialogues.

The posttest consisted of both near and far transfer versions of the studies that were presented in the trialogues. The near transfer studies differed from the studies in the trialogues on surface features, whereas the far transfer studies differed on both surface and structural features. Each topic discussed during the trialogues had one near and one far transfer study, resulting in eight transfer studies in all on the posttest.

**Trialogues.** First, learners signed an informed consent and then completed the pretest. Learners then began the first of four trialogues. A webcam and a commercially available screen capture program (Camtasia Studio™) recorded learners' face and screen, respectively, during the trialogues.

Each trialogue began with a description of a study, which learners read and then began the discussion with the agents. The excerpt in Table 1 is an example trialogue. This is an excerpt from the *true-false* condition, where the female (Mary) and male (Chris) student agents are discussing a flawed study with Bob (learner). The discussion of each study involved five trials. For example, in Table 1 the dialogue turns 2 through 5 represent one trial. Each trial consisted of the student agents asserting their opinions (turns 2 and 3), prompting the learner to intervene with a forced-choice (FC) question (turn 4), and obtaining the learner's response (turn 5).

This cycle was repeated in each trial, with each trial becoming increasingly more specific about the scientific merits of the study. The trialogue in Table 1 discusses a study that uses an inappropriate control group. Trial 1 broadly asked if learners would change their behavior based on the results of the study ("Would you recommend the diet pill to a friend?"), while Trial 2 addressed whether or not there is a problem in the methodology of the study ("Do you think the methodology of the study was good or problematic?"). Trial 3 began to specifically address the problem in the study (turns 2-5). Trial 4 then directly addressed the appropriateness of the control group, (turns 6-9). After Trial 4, learners were presented with an explanatory text to potentially alleviate their confusion, which was adapted from the electronic textbook that accompanies the *Operation ARA!* ITS [17]. Finally, Trial 5 repeated the forced-choice question posed to learners in Trial 4 without the presentation of the agents' opinions. For the present paper only Trials 1-4 are relevant and will be included in subsequent analyses. Learners then completed the posttest after discussing the four studies.

Table 1. Excerpt of trialogue from *true-false* condition

| Turn | Speaker | Dialogue |
|---|---|---|
| | | Participants took this new diet pill and reported losing 10 pounds in the first month. None of the participants exercised or ate super healthy or anything, they just acted normally. The researchers even compared the pill to another group who didn't take it and just acted how they normally do. The group that didn't take the pill didn't lose any weight over a month. So the pill works about ten times as well as not doing anything. <*Case study*> |
| 1 | Mary | Chris, are these two groups similar in every way except taking the pill or not? <*Advance dialogue*> |
| 2 | Chris | Umm, I think they were the same in every way. <*Assert opinion*> |
| 3 | Mary | Well, the two groups don't seem the same to me. <*Disagree with Chris*> |
| 4 | Chris | Looks like we still disagree. Bob were these two groups exactly the same or different? <*Contradiction & forced-choice question*> |
| 5 | Bob | Different <*Response*> |
| 6 | Chris | I don't think they need to make any changes to the control group. <*Assert opinions*> |
| 7 | Mary | No, I don't think so, it needs to change. <*Disagree with Chris*> |
| 8 | Mary | It's too bad that we still disagree. Bob, should the control group have taken some kind of a pill or should it stay the same? <*Contradiction & forced-choice question*> |
| 9 | Bob | stay same <*Response*> |

**Retrospective Affect Judgment Protocol.** Learners completed a retrospective affect judgment protocol [18] after completing the posttest. Videos of learners' face and screen were synchronized and learners made affect ratings while viewing these videos. Learners were provided with a list of affective states (anxiety, boredom, confusion, curiosity, delight, engagement/flow, frustration, surprise, and neutral) with definitions. Affect judgments occurred at 14 pre-specified points (e.g., after contradiction presentation, after forced-choice question, after learner response) in each trialogue (56 in all). In addition to these pre-specified points, learners were able to manually pause the videos and provide judgments at any time.

## 3 Results & Discussion

The analyses were conducted in three phases: self-report confusion ratings, forced-choice (FC) question response accuracy, and transfer test performance. We conducted these analyses in order to determine the impact of agent role (tutor-student vs. student-student) on confusion induction and learning. The results from the current experiment were compared to previous findings from an experiment that involved trialogues with tutor and student agents (tutor-student experiment) [3]. Mixed-effects linear or logistic regression models were constructed for each dependent measure, with one exception, to compare the experimental conditions (*true-false, false-true, false-false*) to the no-contradiction control condition (*true-true*).

### 3.1 Self-Report Confusion Ratings

In the tutor-student experiment confusion was reported more often when learners were in the *true-false* and *false-false* conditions compared to the *true-true* condition. Confusion self-report ratings for the first four trials of each trialogue were investigated for the present experiment. A mixed-effects logistic regression revealed that in the student-student experiment, learners also reported more confusion in the *true-false* and *false-false* conditions than when in the *true-true* condition, $\chi^2(3) = 6.90$, $p = .038$. Table 2 shows the coefficients for the models along with the mean proportional occurrence of confusion. These findings suggest that confusion induction can still be successful when contradictions were presented by two peer student agents. It is interesting, however, that the same pattern of findings emerged when both agents had the same status level. Ostensibly, the contradiction in the *true-false* and *false-true* conditions should evoke the same degree of confusion, but this was not the case. This suggests that other characteristics of the agents may need to be taken into consideration (e.g., gender, perceived knowledge).

Table 2. Proportional occurrence of trialogue dependent measures

| | **Induction Condition** | | | | **Coefficient (B)** | | |
|---|---|---|---|---|---|---|---|
| | *Tr-Tr* | *Tr-Fl* | *Fl-Tr* | *Fl-Fl* | *Tr-Fl* | *Fl-Tr* | *Fl-Fl* |
| **Confusion Self-Report** | .113 | .159 | .118 | .150 | **.490** | .221 | **.421** |
| **FC Question** | | | | | | | |
| Trial 1 | .688 | .563 | .500 | .500 | -.530 | -.795 | -.787 |
| Trial 2 | .844 | .594 | .656 | .406 | **-1.28** | **-1.05** | **-2.06** |
| Trial 3 | .750 | .563 | .656 | .406 | **-.847** | -.452 | **-1.48** |
| Trial 4 | .656 | .688 | .719 | .500 | .126 | .298 | -.667 |

*Notes.* Tr: True; Fl: False; Tr-Tr was the reference group for each model, hence coefficients for this condition are not shown in the table. Bolded cells refer to significant effects at $p < .05$.

### 3.2 Forced-Choice Question Response Accuracy

Two analyses were conducted to investigate FC question response accuracy during trialogues. First, we constructed four mixed-effects logistic regressions to investigate response accuracy in each trial (see Table 2). In the tutor-student experiment, learners were less likely to respond correctly when in the experimental conditions as the trialogues became increasingly more specific (i.e., Trials 2-4) compared to the no-contradiction control condition. This reduction in correct responses is hypothesized to display confusion and uncertainty. A similar pattern emerged in the present student-student experiment with learners being less likely to respond correctly when in the experimental conditions compared to the no-contradiction control condition in Trials 2 ($\chi^2(3) = 13.6$, $p = .002$) and 3 ($\chi^2(3) = 8.66$, $p = .017$). The one exception was that the *false-true* condition did not differ from the *true-true* condition in Trial 3. Interestingly, when the trialogue specifically addressed the flaw in the study (Trial 4) in the present experiment, the experimental conditions did not differ from the no-contradiction control condition, $\chi^2(3) = 3.95$, $p = .134$. Performance in Trial 4 was then the primary difference between the two experiments.

Second, we investigated response accuracy compared to random guessing (or chance) in each condition with the hypothesis that responses similar to random guessing would display confusion and uncertainty. Since the questions adopted a two-alternative format, random guessing would yield a score of .5. In the tutor-student experiment this analysis revealed the general pattern that *true-true* performed above chance and *false-false* performed below chance, whereas *true-false* and *false-true* generally remained at chance level. One-sample t-tests comparing learner responses to .5 (chance) revealed the following overall pattern: *true-true* and *false-true* were significantly greater than chance and *true-false* and *false-false* were statistically indistinguishable from chance. There were two exceptions to this pattern: (a) *true-false* was greater than chance on Trial 4 and (b) *false-true* was at chance level in Trial 1.

There are two overall differences when the patterns from the tutor-student and student-student experiments are compared. First, learner responses in the *false-false* condition were found to remain at chance level in the present experiment, suggesting that learners may have been more skeptical of incorrect agent opinions, even when the agents agreed. Second, learners responded above chance levels in the *false-true* condition. This is a somewhat perplexing finding given that responses in the *true-false* condition were generally still at chance level. Even though the agents had the same status level, there may have been other agent characteristics (e.g., gender, perceived knowledge) or trialogue characteristics (e.g., which agent stated their opinion first) that influenced learner responses.

### 3.3 Transfer Task Performance

Learner performance on both transfer tasks was assessed with hits (correctly identifying the presence of a flaw) to investigate learning. In the previous tutor-student experiment, performance on multiple-choice knowledge assessments was used to measure learning. The results from that experiment revealed that learners only benefited from the presentation of contradictions when they were successfully confused during the trialogues. Two analyses were conducted to investigate learning in the present student-student experiment.

First, mixed-effects logistic regressions revealed that there were not significant condition differences on either transfer task (Near Transfer: $\chi^2(3) = 4.95$, $p = .176$, Far Transfer: $\chi^2(3) = 1.41$, $p = .703$). This finding was consistent with the previous tutor-student experiment and is likely due to the fact that confusion induction success was not taken into consideration. The second analysis then involved dividing learners into low- and high-confusion cases based on a median split of self-report confusion ratings. Mixed-effects logistic regression models were constructed to investigate the induction condition × confusion (low, high) interaction (see Table 3). A significant model was found for the near transfer task ($\chi^2(7) = 11.1$, $p = .067$), but not for the far transfer task ($\chi^2(7) = 6.26$, $p = .255$). The main effect for confusion was not significant for either model ($p$'s > .1).

The interaction was probed by regressing near transfer hits for the low- and high-confusion cases separately. The model for low-confusion cases was not significant, $\chi^2(3) = .435$, $p = .467$. However, the model for the high-confusion cases was significant, $\chi^2(3) = 10.5$, $p = .008$. When learners were in the *true-false* and *false-true* conditions, they performed significantly better on the near transfer task than in the *true-true* condition. It is possible that this increased performance was actually due to increased guessing. To address this issue, we investigated false alarms (incorrectly identifying the presence of a flaw) for the near transfer case studies. The induction condition × confusion model for false alarms was not significant, so the learning effect cannot be attributed to guessing.

Despite the fact that different types of assessments were used (multiple-choice questions vs. transfer tasks), the findings in the present experiment are very similar to those in the tutor-student experiment. It appears to be critical that learners are successfully confused to benefit from the presentation of contradictory information.

Table 3. Proportional occurrence of transfer test performance

| | **Induction Condition** | | | | **Coefficient (B)** | | |
|---|---|---|---|---|---|---|---|
| | *Tr-Tr* | *Tr-Fl* | *Fl-Tr* | *Fl-Fl* | *Tr-Fl* | *Fl-Tr* | *Fl-Fl* |
| **Near Transfer** | | | | | | | |
| Low Confusion | .571 | .533 | .471 | .500 | -.168 | -.414 | -.311 |
| High Confusion | .273 | .647 | .600 | .222 | **1.84** | **1.60** | -.273 |
| **Far Transfer** | | | | | | | |
| Low Confusion | .350 | .200 | .412 | .385 | -.800 | .433 | .092 |
| High Confusion | .455 | .500 | .214 | .500 | .171 | -.160 | .244 |

*Notes.* Tr: True; Fl: False; Tr-Tr was the reference group for each model, hence coefficients for this condition are not shown in the table. Bolded cells refer to significant effects at $p < .05$.

# 4 Conclusion

Contradictory information has been used to increase learning with different methods of presentation (e.g., [1-5,7]). This strategy is expected to be effective because it creates a state of mental discomfort through occurrences of cognitive conflict [7], cognitive disequilibrium [8-10], and confusion [1,3], which then trigger learners to engage in effortful cognitive activities (e.g., reflection, problem solving) that ultimately bring about deeper comprehension [19-20]. The present experiment continues this line of research, but also addresses the less researched issue of the sources of contradictions. We have conducted an experiment that, when compared with the findings of a previous experiment [3], allows for the impact of source to be investigated.

Overall we have found that the presentation of contradictory information by two peer student agents can still successfully induce confusion and had a positive impact on learning. Findings for self-reported confusion mirrored the pattern when contradictions were presented by tutor and student agents [3]. The patterns differed, however, when response accuracy was investigated. This more objective measure of uncertainty and confusion indicated that learners were influenced by agent role and the inter-agent relationship in the trialogue. Although, it was the case that similar learning patterns were found regardless of the inter-agent relationship. In both experiments learners performed better when in the contradictory information conditions (*true-false*, *false-true*) when they were successfully confused. This finding across both experiments is consistent with impasse-driven theories of learning [20] and also cognitive conflict research (e.g., [7,21]) in which learners must be triggered through awareness of the conflict to begin engaging in the cognitive activities that benefit learning.

It was not the case, however, that the two conditions in which the agents disagreed (*true-false*, *false-true*) were identical in all respects in the present experiment. In particular, the *true-false* and *false-true* conditions differed on self-reported confusion and forced-choice question response accuracy. Given that the two agents had the same status level (peer student), it could be expected that similar patterns would emerge for both conditions. The findings for the *true-false* condition adhere to the expected pattern with increased self-reported confusion and response accuracy at chance level, whereas the *false-true* condition did not differ from the *true-true* condition on self-reported confusion and generally responded above chance level. This suggests that status level is not the only agent characteristic that should be considered and trialogue characteristics (e.g., which agent stated their opinion first) may need to be considered as well. For example, in the present experiment the agents differed on gender. When there is no clear authority figure or expert learners may align with an agent based on other characteristics. Research has suggested that agent gender and ethnicity in relation to learner gender and ethnicity can impact the learning experience (e.g., [16,22-23]). There were too few participants in the present sample to investigate these differences, but future research will need to investigate additional characteristics and how they impact the effectiveness of the presentation of contradictions to trigger deeper processing and ultimately have a positive impact on learning.

**Acknowledgements.** This research was supported by the National Science Foundation (NSF) (ITR 0325428, HCC 0834847, DRL 1235958). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

# 5 References

1. D'Mello, S., Lehman, B., Pekrun, R., Graesser, A.: Confusion can be beneficial for learning. Learning and Instruction, 29, 153--170 (2014)
2. Grosse, C., Renkl, A.: Finding and fixing errors in worked examples: Can this foster learning outcomes? Learning and Instruction, 17, 612--634 (2007)
3. Lehman, B., D'Mello, S., Strain, A., Mills, C., Gross, M., Dobbins, A., et al.: Inducting and tracking confusion with contradictions during complex learning. International Journal of Artificial Intelligence in Education, 22, 71--93 (2013)
4. McLaren, B., Adams, D., Durkin, K., Goguadze, Mayer, R., Rittle-Johnson, B., et al.: To err is human, to explain and correct is divine: A study of interactive erroneous examples with middle school math students. In: Ravenscroft, A., Lindstaedt, S., Delgado Kloos, C.,

Hernandex-Leo, D. (eds.) Proceedings of EC-TEL 2012: Seventh European Conference on Technology Enhanced Learning, pp. 222--235. Springer, Berlin (2012)
5. Tsovaltzi, D., Melis, E., McLaren, B.: Erroneous examples: Effects on learning fractions in a web-based setting. International Journal of Technology Enhanced Learning, 4, 191--230. (2012)
6. Braasch, J., Rouet, J-F., Vibert, N., Britt, M.: Readers' use of source information in text comprehension. Memory & Cognition, 40, 450--465 (2012)
7. Limón, M.: On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. Learning and Instruction, 11, 357--380 (2001)
8. Festinger, L.: A theory of cognitive dissonance. Row Peterson, Evanston (1957)
9. Graesser, A., Lu, S., Olde, B., Cooper-Pye, E., Whitten, S.: Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when devices breakdown. Memory & Cognition, 33, 1235--1247 (2005)
10. Piaget, J.: The origins of intelligence. International University Press, New York (1952)
11. Chinn, C., Brewer, W.: An empirical test of a taxonomy of responses to anomalous data in science. Journal of Research in Science Teaching, 35, 623--654 (1998)
12. Bråten, I., Strømsø, H., Britt, M.: Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. Reading Research Quarterly, 44, 6--28 (2009)
13. Goldman, S., Braasch, J., Wiley, J., Graesser, A., Brodowinska, K.: Comprehending and learning from internet sources: Processing patterns of better and poorer learners. Reading Research Quarterly, 47, 356--381 (2012)
14. Strømsø, H., Bråten, I., Britt, M.: Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. Learning and Instruction, 20, 192--204 (2010)
15. Wiley, J., Goldman, S., Graesser, A., Sanchez, C., Ash, I., Hemmerich, J.: Source evaluation, comprehension, and learning in internet science inquiry tasks. American Educational Research Journal, 46, 1060--1106 (2009)
16. Baylor, A., Kim, Y.: Simulating instructional roles through pedagogical agents. International Journal of Artificial Intelligence in Education, 15, 95--115 (2005)
17. Halpern, D., Millis, K., Graesser, A., Butler, H., Forsyth, C., Cai, Z.: Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. Thinking Skills and Creativity, 7, 93--100 (2012)
18. Graesser, A., D'Mello, S.: Emotions during the learning of difficult material. In: Ross, B. (ed.) The Psychology of Learning and Motivation, vol. 57, pp. 183--225. Elsevier (2012)
19. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learning and Instruction, 22, 145--157 (2012)
20. VanLehn, K., Siler, S., Murray, C., Yamauchi, T., Baggett, W.: Why do only some events cause learning during human tutoring? Cognition & Instruction, 21, 209--249 (2003)
21. Chan, C., Burtis, J., Bereiter, C.: Knowledge building as a mediator of conflict in conceptual change. Cognition and Instruction, 15, 1--40 (1997)
22. Baylor, A., Kim, Y.: The role of gender and ethnicity in pedagogical agent perception. In Richards, G. (ed.) Proceedings of the World Conference on E-learning in Corporate, Government, Healthcare, and Higher Education, pp. 1503--1506. AACE, Chesapeake (2003)
23. Moreno, R., Flowerday, T.: Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. Contemporary Educational Psychology, 31, 186--207 (2005)