Feedback Specificity and the Learning of Intercultural Communication Skills

Matthew HAYS[a], H. Chad LANE[b,1], Daniel AUERBACH[b],
Mark G. CORE[b], Dave GOMBOC[b], & Milton ROSENBERG[b]
[a]*Dept. of Psychology, University of California Los Angeles, USA*
[b]*Institute for Creative Technologies, University of Southern California, USA*

Abstract. The role of explicit feedback in learning has been studied from a variety of perspectives and in many contexts. In this paper, we examine the impact of the specificity of feedback delivered by an intelligent tutoring system in a game-based environment for cultural learning. We compared two versions: one that provided only "bottom-out" hints and feedback versus one that provided only conceptual messages. We measured during-training performance, in-game transfer, and long-term retention. Consistent with our hypotheses, specific feedback utterances produced inferior learning on the in-game transfer task when compared to conceptual utterances. No differences were found on a web-based post-test. We discuss possible explanations for these findings, particularly as they relate to the learning of loosely defined skills and serious games.

Keywords. Feedback, serious games, intelligent tutoring systems, virtual humans, cultural awareness, intercultural communication, interpersonal skills

Introduction

The role of explicit feedback in learning has been studied from a variety of perspectives and in many contexts, including its positive and negative impacts on performance (e.g., [4]), how best to design feedback messages (e.g., [9,13]), and on its use in managing student affect (e.g., [7]). Carefully constructed, timely feedback can promote efficient and robust learning. Intuitively, it may seem that feedback that most rapidly improves performance would lead to better learning; however, work in cognitive psychology suggests that the opposite may often be the case. Multiple lines of work suggest that learning occurs when students reach (and overcome) *impasses* [15] and that the introduction of *desirable difficulties* during practice leads to better retention [1]. Practice that is too easy (or too difficult) can slow or even have a detrimental effect on learning; thus, the goal is to properly modulate difficulty during practice—not to minimize it. We view the specificity of feedback as one avenue for modulating difficulty. The immediate usefulness of feedback should affect both the efficiency of training and subsequent learning. In this paper, we describe a game-based environment for cultural learning and report an experiment in which we manipulated the specificity of feedback provided by an intelligent tutoring system (ITS) during training.

[1] Corresponding Author: H. Chad Lane, USC Institute for Creative Technologies, 13274 Fiji Way, Marina del Rey, CA 90292 USA, email: lane@ict.usc.edu

Figure 1. Screenshots of two meetings with BiLAT characters.

1. Serious games for cultural learning

Immersive learning environments provide new and unique ways in which to acquire culture-specific knowledge and skills. High-fidelity graphics, sound, and first-person perspectives permit simulation of many observable aspects of a culture (e.g., dress, gestures) and provide more authentic practice environments than may otherwise be feasible using more traditional live role-play and passive, film-based approaches [10]. Further, recent advances in artificial intelligence (AI) and cognitive science now permit rich modeling of emotions and language [14]. When built with cultural accuracy, these models (and the resulting virtual humans) may open new avenues for teaching the cognitive and interpersonal aspects of different cultures. A number of cultural learning systems have been developed that take advantage of these capabilities (e.g., [2,3]).

We have conducted the present research in ELECT BiLAT (Enhanced Learning Environments with Creative Technologies for Bilateral negotiations). BiLAT, a serious-game-based immersive learning environment, teaches the preparation, execution, and understanding of bilateral meetings and emphasizes culturally appropriate behavior [3,5,6]. Here, we focus on face-to-face meetings with virtual Iraqi characters, even though BiLAT's scope is much broader. These meetings allow trainees to practice the skills necessary for intercultural communication and allow researchers to evaluate the efficacy of using virtual humans to teach intercultural skills.

Learners communicate with a BiLAT character by selecting from a set of pre-authored communicative actions (see Figure 1 for a screenshot). These actions include conversational actions (e.g., questions, statements) as well as physical actions (e.g., removing helmet and sunglasses). The character responds with physical gestures and a synthesized voice. Characters' responses depend on several variables, including trust and a virtual dice roll. Trust is the persistent record of how well learners have used their interpersonal and intercultural skills and is an important factor in determining characters' emotional displays. The virtual dice roll is used to simulate the uncertainty in human behavior and plays a small role in the selection between several possible responses to each action [3].

Because of variability between cultures with respect to how business is conducted, BiLAT also models distinct time spans of meetings that are expected in the meeting partner's culture. With the Iraqi meeting partners in the present study, meetings include an opening phase, a social period, a business period, and a closing social period. If a

learner chooses an action that is not appropriate for the current meeting phase, the character will respond negatively and trust will usually decrease. Because of its importance in the Iraqi culture, following the host's lead (i.e., being aware of the meeting phase and when it changes) is a primary teaching objective of BiLAT and is a focus of our ITS and our investigations.

In BiLAT, actions taken by the student must simultaneously achieve game objectives and respect the targeted cultural norms. This can be a significant challenge for someone unfamiliar with the target culture, and so our ITS provides guidance during meetings in the form of a virtual coach. The coaching algorithm implements *fading*—that is, a greater level of support is provided early, and this is withdrawn as the learner demonstrates competence [5]. The algorithm also selects messages that are more conceptual in nature (i.e., more vague) before delivering those that are specific (i.e., concrete, or "bottom-out") in the hope that learners can reason from the more general assistance. Students' actions are assessed as *correct*, *mixed*, or *incorrect* by consulting links into a lightweight procedural representation of the domain knowledge that identifies necessary steps, optional steps, rules-of-thumb, and actions to avoid [5,6]. In the ITS, pedagogical models used for assessment and feedback adhere to general findings of the education literature (e.g., [13]); our present goal is to take a step towards empirically justifying using these models in the domain of cultural learning.

2. Experiment

As mentioned above, conditions of instruction that maximize learning gains are often *not* those that maximize performance gains. Introducing obstacles for the learner during training introduces difficulty, but overcoming these obstacles can recruit processes that support learning [15]. As a result, conditions that impair performance can result in superior long-term retention and transfer [1]. One such condition is reduced feedback during training. This may seem improbable; it is easy to define feedback simply as more information provided to the learner. It makes sense that, when learning is the goal, more information *must invariably* be better—but this is not the case (e.g., [11,12]).

We were interested in whether the specificity of feedback also affects learning and performance in different ways. If so, our goal as we refine BiLAT—and the goal of other simulation designers—should be to construct coach utterances that engender maximal *learning* gains, but not necessarily maximal *performance* gains. We therefore manipulated the specificity of the coach's feedback and hint utterances. We created two sets of coach utterances. One set was designed to be *specific*, providing "bottom-out" feedback at the action level (e.g., "You are still in full combat gear, including your helmet and sunglasses."). The other set was designed to be vague, providing feedback at the *conceptual* level (e.g., "Be sure to avoid appearing overly defensive or protective."). During gameplay, the coach provided one type of feedback or the other (rather than transitioning between the two, as is its default behavior). We believed that, specific feedback would allow learners to progress through training more rapidly (because the advice would be easier to follow), but conceptual feedback would provide superior long-term retention and the ability to transfer information to new situations.

2.1. Hypotheses

We expected superior performance during training in the specific-coach condition. We expected overall improvement on an explicit judgment test in both conditions. We expected greater improvement on the judgment test in the conceptual-coach condition than in the specific-coach condition. We expected that this improvement would be driven in part by improvements in understanding meeting phases. Finally, we expected superior transfer of learning to new situations in the conceptual-coach condition.

2.2. Method

Participants. Participants responded to flyers posted on university campuses in southern California. Of 100 respondents, 47 were able to travel to our Institute to participate in the experiment. As compensation for their three hours of participation, they were each paid $60.

Measures. A web-based situational judgment test (SJT) was used to examine participants' understanding of concepts needed for successful interactions with the characters in BiLAT [8]. The SJT presents a series of scenarios and asks participants to provide ratings for several actions related to each scenario. Participants rated the quality of these actions from 0 (never take this action) to 10 (definitely take this action). For each action, the mean ratings from three subject-matter experts (SMEs) were used as the "correct" response. We defined learning as the increase in a participant's correlation with the SMEs after using BiLAT. Further, Institute researchers had previously tagged each SJT action with one or more *learning objectives* (LOs), which are the basis of the student model in our ITS. By analyzing only SJT actions tagged with meeting-phase-related LOs, we were also able to determine whether participants learned about meeting phases in particular.

An in-game transfer task was used to determine participants' ability to apply the concepts they practiced during their meetings in BiLAT. This transfer task consisted of meeting with another character in a new scenario. During meetings with this character, the coach was disabled. A participant's mastery was defined with respect to the proportion of actions during this meeting that were culturally or interpersonally inappropriate. Because actions in BiLAT are also tagged with LOs, we were able to detect improvements in participants' ability to apply knowledge of meeting phases.

Procedure. People who responded to the flyer were emailed a link to the SJT. After completing the SJT, an appointment was made for participants to come to our Institute. Upon arriving at the Institute, participants received an instructions packet that gave them a brief overview of the purposes and functions of BiLAT and the ITS. Participants then watched a video on how to use BiLAT, which included indirect references to meeting phases and other culturally appropriate behavior, as well as the mechanics of executing actions within the game.

Next, participants read a scenario about a problem in a U.S.-constructed market in Iraq. Resolving the market problem required meeting successfully with three virtual characters in BiLAT. Participants were able to meet with each character as many times as was necessary in order to resolve the problem, but had to meet with the characters in a pre-determined order. Participants were allotted 100 minutes to resolve the problem.

Random assignment determined which coach—conceptual or specific—guided them during their meetings.

When time expired or the market problem was resolved, participants read another scenario about a problem with medical supply theft and doctor safety in an Iraqi hospital. Participants met for up to 30 minutes with one virtual character—the hospital administrator, Aziz—to resolve this problem. The coach was not available during meetings with Aziz so we could use them as a measure of applied knowledge transfer from what they had practiced in attempting to resolve the market problem.

One week after interacting with BiLAT, participants were emailed a link to the SJT, which they were asked to complete. They completed it after an average of 1.8 days. Finally, participants were thanked, compensated, and debriefed.

Design. The single independent variable—coach type (conceptual or specific)—was manipulated (between subjects) while participants attempted to resolve the problem in the market. As described in Section 1, the conceptual coach's utterances were more general in nature while the specific coach gave concrete, easy-to-understand hints and feedback. It is critical to note that, other than the content of their utterances, the coaches were identical. That is, the algorithm by which the coach decided whether to provide a hint was independent of the wording of that hint.

Table 1. Specific vs. conceptual feedback differences in practice, in-game transfer task, and on written test.

	ERRORS DURING COACHED MEETINGS		UNCOACHED MEETING ERRORS		SJT IMPROVEMENT	
	p(error)	p(phase error)	p(error)	p(phase error)	overall	phase
specific	.22	.13	.20	.13	.17	.48
conceptual	.20	.12	.15	.07	.18	.53

3. Results

Due to experimenter error, data from one participant were corrupted and were therefore excluded from all analyses. A summary of the results is shown in Table 1 and explained below.

3.1. Performance during training

First, we examined whether the two coaches produced different performance during training. We defined "performance" as the probability that an action selected by the participant was erroneous. Participants in the specific-coach condition committed erroneous actions 22% of the time (*SD* = 8.89%). Participants in the conceptual-coach condition committed errors 20.1% of the time (*SD* = 5.02%). This difference was not statistically significant: $t < 1$, *ns*. The lack of observed differences seems to contradict our hypothesis that the specific coach would be easier to obey than would the conceptual coach. However, what differences there were may have been overshadowed by the 100-minute-long training period. Another source of this overshadowing may have been the implicit feedback from the characters themselves (i.e., cultural errors produce negative character reactions).

3.2. Conceptual understanding (SJT)

Next, we examined whether experience with BiLAT promoted participants' understanding of the learning objectives. Before interacting with the virtual characters in BiLAT, participants' average correlation with the SME was $r = .56$ ($SD = .18$). One week after using BiLAT, this correlation increased to $r = .74$ ($SD = .10$). The apparent improvement in alignment with SMEs' understanding was statistically significant: $t(45) = 7.873$, $p < .001$.

When this analysis was restricted to meeting-phase-related items, a similar pattern emerged. Before using BiLAT, participants' average correlation with the SME on meeting-phase-related items was $r = .33$ ($SD = .49$). One week after using BiLAT, this correlation increased to $r = .85$ ($SD = .14$). This improvement was statistically significant: $t(45) = 7.601$, $p < .001$. Together, these results support our hypothesis that BiLAT is able to dramatically improve participants' understanding of LOs.

Next, we examined whether the conceptual coach promoted a differentially greater increase in understanding as measured by the SJT. We found no main effect of coach type. Improvement for participants in the conceptual-coach condition was $M = .18$ ($SD = .16$). Improvement for participants in the specific-coach condition was $M = .17$ ($SD = .14$). The difference in improvement due to coach type was not statistically significant: $F < 1$, *ns*. These results suggest that—as measured by the SJT—the specificity of the coach's utterances did not differentially affect learners' explicit knowledge of the LOs.

When this analysis was restricted to meeting-phase-related items, we again found non-meaningful differences: $F < 1$, *ns*. These results suggest that, even for meeting-phase-related information, the conceptual coach did not improve participants SJT scores any more than did the specific coach. Together, these results suggest that participants' declarative knowledge of the LOs did not differentially profit from experience with one coach versus the other.

3.3. Within-game transfer

We next turned to analyses of participants' ability to *apply* the knowledge they gained during the experiment. This analysis was based on their performance in their uncoached meeting(s) with Aziz. A software error corrupted data from four participants' meetings with Aziz; they were excluded from the following analyses.

We began by examining whether the conceptual coach differentially reduced the probability that the participant would select inappropriate actions while trying to solve the hospital problem. Unlike in the SJT analyses, we found a main effect of coach type. Participants in the conceptual-coach condition executed inappropriate actions 15.4% of the time ($SD = 6.43\%$) during their interactions with Aziz. Participants in the specific-coach condition executed inappropriate actions 19.91% of the time ($SD = 7.69\%$) during their interactions with Aziz. The apparent superiority of participants in the conceptual-coach condition was statistically significant: $F(1, 36) = 4.2$, $p = .047$. This result supports our hypothesis that the conceptual coach provided participants with a superior ability to transfer knowledge to uncoached meetings in a new context.

When this analysis was restricted to meeting-phase-related actions, a similar pattern emerged. Participants in the conceptual-coach condition committed meeting-phase-related errors 7.48% of the time ($SD = 12.60\%$) during their interactions with Aziz. Participants in the specific-coach condition committed meeting-phase-related errors 13.32% of the time ($SD = 15.43\%$) during their interactions with Aziz. The

apparent superiority of participants in the conceptual-coach condition was again statistically significant: $F(1, 33) = 8.675$, $p = .006$. This result supports our hypothesis that the conceptual coach provided participants with more ability to transfer meeting-phase-related knowledge to a new character in a new scenario.

4. Discussion and conclusion

The overall gains in SJT scores suggest that BiLAT is an effective pedagogical tool. A potential objection to this interpretation is that our participants may have learned from the 20-minute instructional video that preceded BiLAT but followed the pre-test. However, earlier research on BiLAT revealed that using BiLAT provided significant advantages over watching nearly an hour of perfect BiLAT gameplay [6]. It is therefore unlikely that the video could have been responsible for the substantial increases in SJT scores we observed, but this possibility is worth investigating directly.

We believe that the lack of a main effect of coach type on the SJT (the web-based post-test) may be due to the nine-day delay between manipulation and test. We observed differences in the application of knowledge in the meetings with Aziz, which occurred immediately after the coached meetings. Perhaps these differences attenuated over time—or the extended time to reflect allowed learners to generalize their understanding of the domain. An immediate post-test may have shown different results.

The transfer task results suggest that the conceptual coach further increased participants' ability to apply what they learned from BiLAT. This finding provides evidence that adjustment of feedback specificity is a valid approach to modulating difficulty faced by learners during training. This finding also demonstrates that properly modulated difficulty improves learning more than does minimizing obstacles to performance gains. Allowing difficulties to emerge and be overcome at the appropriate times is often superior to removing difficulties altogether. These results reinforce the fundamental principles of impasse-driven learning and desirable difficulties in training [1,15]. That is, in the design of feedback in serious games, intentionally providing *less help* can promote learning (but must be done judiciously).

The transfer task results also serve to illustrate two dissociations that are critical to the design of learning environments. First, although the conceptual coach produced superior transfer, the two coaches produced equivalent during-training performance. Although performance during training is an intuitive index of long-term learning and transfer, it is an unreliable one. Indeed, greater gains were observed in the condition designed to provide *less* support during training. Second, the differential effect of coaches on transfer performance was also not mirrored on the SJT, on which the two coaches yielded similar gains. This dissociation underscores the difference between knowledge and the ability to *apply* that knowledge; whereas the two coaches provided similar gains on the SJT, the general coach better enabled participants to transfer that knowledge to a new context.

Ultimately, our aim is to empirically determine the ideal balance of intrinsic feedback (from the characters' responses), explicit feedback (from the ITS), and other pedagogical support (such as through preparation materials and support for reflection) for learning cultural knowledge in BiLAT. Further experimentation is necessary to determine the relative impact of each of these elements on difficulty and learning in BiLAT. We therefore intend to continue evaluating the pedagogical value of different types and timings of feedback as well as the implicit difficulty of the game (e.g., by

adjusting how forgiving or irascible the characters are). We hope that an overall understanding of the proper level of difficulty and provision of support for learners in BiLAT can be achieved.

As we continue to extend and investigate BiLAT, our research will focus on examining the relevance of principles of learning and cognition to the design of unsolicited, explicit feedback in immersive simulations. Particularly when they are used in complex domains like cross-cultural negotiation, ITSs that are more informed about human cognition will be better able to provide feedback in a way that produces effective training and superior long-term learning.

Acknowledgements

The project described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

[1] R.A. Bjork, Memory and metamemory: Considerations in the training of human beings, In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing*, Cambridge, MA: MIT Press, 185-205.

[2] W.L. Johnson, Serious use of a serious game for language learning. In *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (2007), 67-74.

[3] J. Kim, R.W. Hill, P. Durlach, H.C. Lane, E. Forbell, M. Core, S. Marsella, D. Pynadeth, & J. Hart, BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education* (in-press).

[4] A.N. Kluger & A. DeNesi, Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science* **7**, 3 (1998), 67-72.

[5] H.C. Lane, M.G. Core, D. Gomboc, A. Karnavat, & M. Rosenberg, Intelligent tutoring for interpersonal and intercultural skills. In *Proceedings of Interservice/Industry Training, Simulation, and Education Conference* (2007), Orlando, FL.

[6] H.C. Lane, M. J. Hays, M.G. Core, D. Gomboc, E. Forbell, D. Auerbach, & M. Rosenberg, Coaching intercultural communication in a serious game. *Proceedings of the 16th International Conference on Computers in Education* (2008), 35-42.

[7] M.R. Lepper, M. Woolverton, D.L. Mumme, & J. Gurtner, Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S.P. Lajoie & S.J. Derry (Eds.), *Computers as cognitive tools* (1993), 75-105.

[8] P. Legree & J. Psotka, Refining situational judgment test methods. *In Proceedings of the 25th Army Science Conference* (2006), Orlando, FL.

[9] J. McKendree, Effective feedback content for tutoring complex skills, *Human Computer Interaction* **5** (1990), 381-413.

[10] E.M. Raybourn, Applying simulation experience design methods to create serious game-based adaptive training systems. *Interacting With Computers*, **19**, (2007), 206-214.

[11] R.A. Schmidt & R.A. Bjork, New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science* **3**, (1992), 207-217.

[12] R.W. Schulz & W.N. Runquist, Learning and retention of paired adjectives as a function of percentage occurrence of response members. *Journal of Experimental Psychology* **59** (1960), 409-413.

[13] V. Shute, Focus on formative feedback, *Review of Educational Research* **78**, (2008), 153-189.

[14] Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., & Rickel, J. Toward virtual humans. *AI Magazine* **27**, 2, (2006), 96–108.

[15] K. VanLehn, Rule learning events in the acquisition of a complex skill: An evaluation of Cascade, *Journal of the Learning Sciences* **8**, 1 (1999), 71-125.