Coaching Intercultural Communication in a Serious Game

H. Chad Lane[a], Matthew Hays[b], Mark Core[a], Dave Gomboc[a], Eric Forbell[a], Milton Rosenberg[a]
[a]*Institute for Creative Technologies, University of Southern California, USA*
[b]*Department of Psychology, University of California Los Angeles, USA*
lane@ict.usc.edu

Abstract: We consider the use of virtual humans and an intelligent tutoring system (ITS) for the teaching of cultural social conventions. Learning occurs in a serious game that requires the learner to establish trust and reach agreements with virtual characters of a different culture. Our tutoring system provides culturally focused learning support during and after the meetings with these virtual characters. In a study intended to determine the effectiveness of the ITS, we found that guidance provided during meetings seemed to improve learners' understandings of culturally-related "phases" in meetings (e.g., when to talk about business) as well as greater success in an unsupported posttest meeting, but with no overall increase in cultural understanding when compared with learning in passive and unguided conditions.

Keywords: serious games, intelligent tutoring systems, cultural awareness, intercultural competence, intercultural communication, interpersonal skills

Introduction

Culture can play a significant role in the success or failure of face-to-face encounters. Many of the expectations we hold going into a conversation or meeting have cultural explanations, or at least are influenced by our cultural values and backgrounds. Similarly, meanings we infer from the communicative acts of others in conversation are influenced heavily by our own cultural "lenses." So, when we enter into conversations with people from cultures other than our own, the differences can quickly become a source for confusion, misunderstanding, and at worst, conflict. Awareness of such differences – and a willingness to assume different cultural perspectives – is key for growth towards intercultural competence [1]. In this paper, we consider the use of virtual humans, serious games, and an intelligent tutoring system (ITS) for the learning of cultural social conventions. The aim is to teach trust-building strategies and how to reach agreements with individuals from another culture.

1. Immersive Environments for Cultural Learning

Immersive learning environments provide new and unique ways in which to learn about a new culture [6]. High-fidelity graphics, sound, and first-person perspectives make it possible to simulate many tangible aspects of a specific culture (e.g., dress, gestures) and provide more authentic practice environments than may otherwise be feasible using more traditional live role-play and media-based approaches. Further, recent advances in artificial intelligence (AI) and cognitive modeling now permit rich modeling of emotions, language, tasks, and more [11]. When built with cultural accuracy, these models (and the resulting

Figure 1. The ELECT BiLAT meeting screen. Learners select communicative actions from the menu on the left and see and hear the character's animated response. The dialogue is recorded in the textbox in the lower right corner. Coaching utterances also appear in this box.

virtual humans) may open new avenues for teaching the cognitive and interpersonal aspects of learning about different cultures.

A number of cultural learning systems exist that take advantage of these capabilities. The Tactical Language and Culture Training System provides a mission practice environment that allows learners to explore a virtual town while speaking to locals in Arabic, make culturally appropriate gestures, and accomplish goals such as getting the names of contacts and directions [4]. VECTOR, also a cultural learning system, situates learners in a virtual foreign town, but uses English utterances via menu selections for interaction with locals. An example of a goal in VECTOR is to find a bomber and stop him from attacking his next target [8]. The Adaptive Thinking and Leadership system, a team-training system, uses human role players in intercultural scenarios. Learners take roles as people from different cultures and are given believable back-stories and goals [10]. We have conducted our research in another cultural learning system: ELECT BiLAT (Enhanced Learning Environments with Creative Technologies for Bi-lateral negotiations). BiLAT, a serious-game-based immersive learning environment, teaches the preparation, execution, and understanding of bi-lateral meetings in a cultural context [3]. Here, we focus on face-to-face meetings between learners and virtual characters, even though BiLAT's overall scope is much broader. This represents a fundamental skill necessary for intercultural communication, and we also perceive a scientific need to better understand the extent to which virtual humans can be used to teach intercultural skills in lieu of human role players.

A BiLAT meeting consists of two modes: dialogue and negotiation. A screenshot is shown in Figure 1. It shows a menu of conversational actions (e.g., questions and statements – lower left in the screenshot) as well as physical actions (e.g., removing sunglasses, giving a gift). The character responds to the learner with a synthesized voice and physical gestures; corresponding text is displayed in the dialogue window in the lower right corner. Although dozens of variables govern the actions of the character, the variable of primary importance is trust. While characters may display a variety of emotions in their responses, trust is the persistent record of how well players have used their interpersonal and intercultural skills. In the simulation, trust is a major factor in whether BiLAT characters will agree to negotiate and what deals they will accept. A mistrusting character may demand uneven deals or refuse to negotiate. Learners communicate with BiLAT characters by selecting from a large collection of hand-authored communicative actions. Characters' responses depend on a number of factors, including how well the learner prepared, the current meeting phase, the trust level, and a virtual dice roll. Twelve responses exist for each available user action, along with a set of generic responses (e.g., to display confusion). Each action entails a possible change to the trust variable of the character. The dice roll is intended to simulate uncertainty in human behavior – cognitive and emotional modeling techniques can be used to simulate these reactions in more principled ways [11]. This comes at a greater cost, however, in terms of knowledge engineering and maintenance.

Because of variability between cultures with respect to how time is treated [9], BiLAT also models distinct time spans corresponding to the expectations of characters during meetings. This includes an opening phase, a social period, a business period, and a closing social period. Scenario authors are required to indicate the times (or "phases") for which actions are appropriate. If a learner chooses an action that is not appropriate for the current phase of a meeting, the character will respond negatively, which is revealed in the content of the response, gestures, and (usually) loss of trust. One of the learning objectives (LOs) underlying BiLAT is that the learner should follow the lead of his or her host – this is one focus of the ITS and is discussed in the next section.

2. Coaching Culture

In BiLAT, the learner must repeatedly select conversational actions that simultaneously achieve game objectives and respect the targeted cultural norms. This can be a significant challenge for a learner who does not understand the new culture or the differences between it and his or her own. Our ITS, an instance of the Intelligent Guided Experiential Learning (IGEL) framework [2], provides learning support in two ways. During meetings, feedback and hints can be delivered by a coach, while after meetings, a reflective tutor reviews the meeting with the learner, gives more detailed feedback, and asks reflection questions. In the present study, we focus primarily on the details of the coach that deal with the timing of communicative actions. This is also highlighted in the evaluation described in Section 3.

2.1 Hinting and Feedback

When the coach decides to provide some form of guidance, a message appears in the dialogue window of BiLAT (lower right in Figure 1). These messages are intended to promote learning and reduce frustration. For example, if an inappropriate gift is given to the virtual character, the coach might explicitly state the gift is not an acceptable one for the virtual character's culture (e.g., "alcohol is not generally acceptable to Arabs"; [9]). Coaching messages are generally shorter given the context of a live meeting, while reflective tutoring sessions are used to get into the underlying cultural issues through interactive questioning and explanations. The coach is capable of abstract as well as concrete feedback. At both levels of specificity, three categories of advice are available:

- **hints**: a suggestion pointing the learner to an appropriate next action
- **negative feedback**: statement that an action was poor with a short explanation
- **positive feedback**: praise for a good action and (possibly) a short explanation

Because cultural rules vary across individuals and are generally regarded as ill-defined, actions often resist clear classification as "right" or "wrong." We therefore allow for "mixed" assessments, meaning that giving both positive and negative feedback is possible. Feedback messages in these situations can be concatenated or delivered individually based on other factors, such as if related errors (or related "good" actions) were selected earlier in the meeting. These decisions depend on the configuration of the coaching algorithm, which can be used to control the content and timing of coaching interventions. In the experiment described in Section 4, we implemented a model-scaffold-fade algorithm to control feedback timing.

The IGEL reflective tutor, which runs after meetings and is responsible for delayed feedback, presents the learner with a playback of salient moments from their meeting and provides support for review and reflection. Specifically, the reflective tutor uses the series of assessments made by the expert model and gives feedback that is more verbose than what is delivered during meetings by the coach. It also asks multiple choice questions that ask the learner to think about rules that may have been violated and if better actions could have been taken. The reflective tutor also highlights phase-related errors by including these as reasons for why certain actions might have been unsuccessful and discussing when business or smalltalk topics should be addressed (rather than when the learner chose to bring them up). More details of how the reflective tutor is implemented can be found in [2, 5].

2.2 Assessing Meeting Actions

Of course, it is critical for the coach and the reflective tutor to have assessments of actions available to decide whether to give feedback and if so, what to say. In BiLAT, each time an action is taken in a meeting or the learner responds to a character's question, IGEL's expert model is called to judge the action's quality. The correctness of an action is determined in two ways:.

1. The action is checked to see if it is phase-appropriate.
2. The active learning objectives are determined, along with their positive and negative association with the action.

As discussed above, meeting phases are windows of time during a meeting that define when certain categories of actions are appropriate or not. They are culture dependent and the expert module dynamically assigns negative assessments when an inappropriate action is chosen. This is implemented via a link into the domain knowledge.

Domain knowledge is represented in a primarily procedural form, with additional tags that allow for optional steps, rules-of-thumb, and commonly applied incorrect steps. We currently use the following set of cultural rules to classify phase mismatch errors:

1. Don't discuss business during social periods.
2. Don't stray too far from business in a business period.
3. Regarding the opening of meetings:
    a. Opening actions (e.g., greetings) are not appropriate later in the meeting.
    b. The opening is too early for some social actions.
4. Regarding the closing of meetings:
    a. Closing actions (e.g., leaving) are not appropriate earlier in the meeting.
    b. Some social actions are not appropriate when concluding the meeting.

The expert model considers in which phases an action is permitted and in which phase it is performed. For example, if an action is permitted only in the opening phase (e.g., greeting in Arabic) but is performed in the pre-business phase, the expert model will return a negative assessment of rule 3a. However, it will also find positive evidence of understanding the rule of thumb to always greet in the language of your counterpart. The final assessment will be therefore be *mixed*. Another example of a phase-related rule states that it can be harmful to rush into business [9, p.58]. When a business-phase action is taken prematurely, then, the expert model will register *incorrect* if the action is never appropriate. If, on the other hand, the action would be considered acceptable at a later time, a mixed assessment would be recorded. In short, learners receive credit for any "goodness" in the actions they take regardless of the outcomes in any specific situation.

3. Evaluation

BiLAT provides practice in developing negotiation strategies, trust building, and appropriate meeting behavior (i.e., choosing phase-appropriate actions that respect a partner's culture). Together, practice in BiLAT and extrinsic feedback from IGEL (as opposed to *intrinsic* feedback through character actions and responses), should allow the learner to gather a practical understanding of the learning objectives. In this section, we report on a small study intended to examine the learning contributions of BiLAT and IGEL.

3.1 Research Questions: Interactivity and Real-time Feedback

Our first question was whether actually playing BiLAT would be pedagogically beneficial. A common notion is that maximum gains can be produced by errorless learning, wherein the conditions of instruction are such that it is impossible for learners to make errors. In the present study, we used a video-only condition, in which participants watched videos of perfect gameplay. Optimal behavior is modeled for these learners. Nevertheless, we expected that the video-only condition would suffer because it would lose the benefit of interactivity. Participants who actually play BiLAT would select actions they believed to be correct. Perhaps of even more value: participants would also select *against* actions they believed to be incorrect. We expected these deliberations would produce general gains when meeting with new characters, and therefore believed the interactive conditions would be superior to the video-only condition.

Our second question was whether the coach adds any pedagogical value. We hypothesized that the coach would produce learning gains because of its ability to identify incorrect actions and the reasons those actions are incorrect. Ideally, this type of feedback could generalize to other actions – and the learner's ability to self-evaluate whether those actions would be advisable. To get at this question, we included two interactive conditions: one in which the coach was active (*yes-coach*) and one in which it was not (*no-coach*). In all three conditions, the reflective tutor was active to ensure that participants in all three conditions received extrinsic pedagogical support.

3.2 Method

Participants. Participants were thirty U.S. Citizens recruited by flyer from the campus of the University of Southern California. As compensation for their three hours of participation, they were paid $60.

Procedure. All participants received an instructions packet, watched a video on using BiLAT, took the pretest, conducted meetings with three different characters (including reading background information about each and reflective tutoring sessions), conducted a fourth meeting with no coaching or reflective tutor, and finally took the posttest.

Design. Three meetings occurred in each of three conditions: video-only, no-coach, and yes-coach. Video-only participants observed expert play with coaching and reflective tutoring active. Participants without coaching conducted meetings and received reflective tutoring support after each one. Finally, those in the yes-coach condition played with coaching and reflective tutoring support. Participants' experiences were otherwise identical.

Measures. We made two comparisons of the three between-subjects conditions. The first comparison was *success*, which was measured on an uncoached fourth meeting with a new character. This measurement was binary: whether the participant was able to achieve the mission objective in 25 minutes or less (repeated attempts are possible until time expires). A successful meeting required achieving given objectives and by choosing actions that increased the meeting partner's trust to the point where agreements were possible.

The second comparison was the pretest-posttest change. A situational judgment test (SJT) was used for the pre- and posttests. SJTs present a series of scenarios and ask the participant to rate the quality of a small set of potential responses on a Likert scale [7]. Each action is rated between 0 (never take this action) up to 10 (definitely a good action). The test was given to three subject-matter experts (SMEs) and the means of their answers were used as the gold standard. Two measurements were used to evaluate participants' mastery of the learning objectives. One is the correlation between participants' answers and SMEs'. The correlation represents the degree to which the two groups correspond in their overall assessment of the responses, but does not capture valence data. To illustrate: let us say that a SME rates an answer as 10 and that is the highest rating the SME gives. If a participant rates that answer as 4, but 4 is the highest rating that participant gives, the two may be perfectly correlated. To capture information about the valence of answers, we calculated what we have called a *ballpark* score. To illustrate: if an SME rates a response as 8, the participant would be said to be "in the ballpark" if the participant provides a rating of 7, 8, or 9. The change in each score was measured from pretest to posttest.

3.3 Results and Discussion

Fourth meeting. Three participants – one from each condition – were omitted from this analysis because of experimenter error. As a result, this analysis included a total of 27 participants (nine per condition). Figure 2 presents the success data from our experiment. As can be seen, the combination of interactive gameplay and feedback from the coach helped players to be successful in the fourth meeting (in which both the coach and reflective tutor were deactivated). 56% of the participants in the video-only condition were successful. With the addition of interactivity – and the ability to make errors – 67% of the participants in the no-coach condition were successful. Finally, those who had coaching during practice had the highest rate of success in the fourth meeting at 89%.The number of participants per cell (nine) in this analysis was too small for a reliable statistical analysis to be run. Nevertheless, we view these results as encouraging.

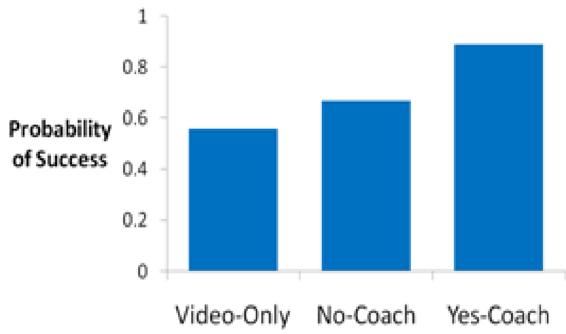

Figure 2. Probability of success by condition.

Situational Judgment Test. The uncoached meeting suggested learners who interacted and were coached improved their ability to play BiLAT. To check for learning beyond the game, we used the SJT to determine whether participants were mastering the cultural rules being taught. Overall, these data initially painted a different picture. We did not find that the conditions produced a differential increase in correlation with SMEs between conditions: $F(2, 27) = .805$, $p = .457$, nor on the ballpark measure: $F(2, 27) = .111$, $p = .896$.

Immediately, we wondered why there would be disparity between performance in the game and improvement on the SJT. One possibility is that the pretest SJT ratings affected posttest ratings (i.e., that learning occurred from the test itself). It is unlikely that participants were able to remember the exact numbers that they had provided hours before. Nevertheless, pretest ratings may have served as anchors for posttest responses. To the extent that this interference reduced variance in posttest responses, it reduced the ability for between-group differences to emerge.

Another, more likely possibility is that this particular experiment and the SJT were not perfectly aligned. The “correct” answers on the SJT were provided by subject matter experts, whose knowledge spans all of the elements of BiLAT and all of the broader domain knowledge. The SJT itself was not originally designed as a measure for this experiment. As a result, the SJT taps knowledge about many more issues of negotiation and culture than we could have expected to address in three hours (and only 90 minutes of practice) with our participants. The SJT may be too broad in scope to examine gains in knowledge in our participants as a result of our experimental procedure.

We therefore ran a second analysis of learning gains in the SJT in which we culled data from questions that did not address issues the coach directly supports. For example, we omitted responses to prompts about the reliability of information during the meeting preparation phase since participants did not use this part of BiLAT (to save time, they were instead provided with dossiers). By removing these responses from our analysis, we increased our power to detect learning in two ways. First, to the extent that participants’ responses to culled answers remained fixed from pretest to posttest, estimates of change would be deflated and would therefore underestimate the amount of learning in the subset of SME knowledge tapped by BiLAT and IGEL. Second, to the extent that participants’ responses to culled answers wandered randomly, pre- and posttest correlations would be attenuated and within-group variance in both of our measures would increase. This secondary analysis therefore reflected content most directly present in our experimental interventions. The domain of meeting-phase-specific behavior was highlighted for this analysis. It draws on cultural understanding and appropriate use of negotiation strategies.

Correlation data for phase-specific questions. Figure 3 presents the correlation data for the three conditions in our experiment. Interactivity with coaching produced modest (but not significant) gains in understanding of phase information: F(2, 27) = 2.062, p = .147. Post-hoc tests revealed a marginally significant difference between in favor of coaching over no-coaching: t(19) = 1.72, p = .054.

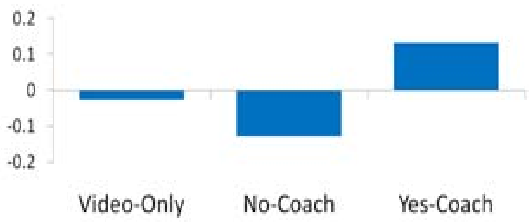

Figure 3. Changes in correlation with experts on meeting-phase-specific prompts (by condition)

Figure 4. Changes in “ballpark” score on meeting-phase-specific prompts (by condition)

This was expected: an action’s phase-appropriateness is the branching point from which the coach begins to decide to provide feedback. If the coach were to produce benefits in only one aspect, this would be it. We note that the no-coach condition is actually worse than the video-only condition. This pattern may reflect the inadequacy of discovery (i.e., trial-and-error) learning. Without guidance from the coach or a model of ideal gameplay, which was present in the video-only condition, players in the no-coach condition are left to wander through the action menus. It appears this unguided activity is unproductive.

Ballpark scores on phase-specific questions. Figure 4 presents the ballpark data for the three conditions in our experiment. Again, playing BiLAT and receiving real-time feedback from IGEL resulted in somewhat superior comprehension and retention of the LOs related to meeting phase: F(2, 27) = 1.681, p = .205. Post-hoc tests revealed a marginally significant difference between the superior yes-coach and the inferior video-only conditions: t(19) =

1.41, p = .086. This finding provides additional, albeit preliminary, support for our conclusions from the correlation data. On the other hand, the reversal of the video-only and no-coach relationship in the correlation and ballpark measures suggests some instability in our measures. This is likely due to our small sample size and to the relatively small number of data points per participant; there were only five phase-specific prompts on the SJT.

4. Conclusion

This paper described a serious-game-based approach to teaching intercultural communication skills with virtual human characters and with the support of an intelligent tutoring system. The preliminary results are suggestive that learning does occur and is likely due to the interactivity and from the intelligent tutor's support. Specifically, the results suggest that the intelligent support was most helpful in teaching cultural rules relating to meeting phases and the timing of communicative actions, which are critical elements of succeeding in ELECT BiLAT. Our future experimentation plans continue to revolve around the role and impact of feedback in cultural learning environments. Specifically, we have built versions of the coach that give exclusively abstract or concrete feedback. Our hypothesis is that game performance will increase with higher levels of concrete guidance, and that less directive help will reduce game performance (i.e., it will be more difficult since the student must reason about actions), but produce more robust 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] Bennett, M.J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In R.M. Paige (ed.) Education for the Intercultural Experience (pp. 27-71). Yarmouth: Intercultural Press.

[2] Gomboc, D., Lane, H.C., Core, M., Karnavat, A., & Rosenberg, M. (2008). An intelligent tutoring framework for simulation-based training. In *Proceedings of the 16th International Conference on Computers in Education (ICCE2008)*. Taipei, Taiwan.

[3] Hill, R.W., Belanich, J., Lane, H.C., Core, M.G., Dixon, M., Forbell, E., Kim, J., & Hart, J. (2006). Pedagogically structured game-based training: Development of the ELECT BiLAT simulation, *in the Proceedings of the 25th Army Science Conference (ASC 2006)*, Orlando, FL.

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

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

[6] Lane, H.C. (2007). Metacognition and the development of intercultural competence. *In Proceedings of the Workshop on Metacognition and Self-Regulated Learning in Intelligent Tutoring Systems at the 13th International Conference on Artificial Intelligence in Education* (pp. 23-32). Marina del Rey, CA.

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

[8] McCollum, C., Deaton, J., Barba, C., Santerelli, T., Singer, M.J., & Kerr, B.W. (2004). Developing an immersive, cultural training system. *In Proceedings of I/ITSEC: Interservice/Industry Training, Simulation, and Education Conference*. Orlando, FL.

[9] Nydall, M.K. (2006). *Understanding Arabs: A guide for modern times (4th ed)*. Boston, Intercultural Press.

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

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