International Journal of Artificial Intelligence in Education 19 (2009) xx-xx
IOS Press

Advancing development of intercultural competence through supporting predictions in narrative video

Amy Ogan, Vincent Aleven, *Human Computer Interaction Institute*
Christopher Jones, *Modern Languages Department*
Carnegie Mellon University, Pittsburgh, PA 15213, USA
aeo@andrew.cmu.edu, aleven@cs.cmu.edu, cjones@andrew.cmu.edu

Abstract. Most successes in intelligent tutoring systems have come in well-defined domains like algebra or physics. We investigate how to support students in acquiring ill-defined skills of intercultural competence using an online environment that employs clips of feature films from a target culture. To test the effectiveness of a set of attention-focusing techniques (*pause-predict-ponder*) we created ICCAT, a simple tutor that enhances an existing classroom model for the development of intercultural competence. We ran a study in two French Online classrooms with 34 participants, comparing ICCAT versions with and without these techniques. We found that the addition of pause-predict-ponder seemed to guide students in acquiring cultural knowledge and significantly increased students' ability to reason from an intercultural perspective. Further analysis of the posttest and the post-video viewing discussion found that students in the experimental condition were significantly assisted by the prediction, and were able to maintain a high quality of discussion over time. The research thus establishes that a simple model for intercultural competence activities, enhanced with the novel pause-predict-ponder techniques, is a viable approach to creating tutors in an ill-defined domain, and possibly better suited to the demands of the domain than the standard problem-solving approach embedded in intelligent tutoring systems.

Keywords. Ill-defined domains, intercultural competence, multimedia

INTRODUCTION

Intelligent tutoring systems (ITS) have been shown to be effective at increasing student learning in domains with well-structured tasks (e.g. Koedinger, Anderson, Hadley, & Mark, 1997; VanLehn et al., 2005). A recent review paper (VanLehn, 2006) defines intelligent tutoring systems as systems capable of guiding students step-by-step during complex problem-solving activities. These systems often follow the principles developed by Anderson, Corbett, Koedinger, and Pelletier (1995) for guiding the construction of intelligent tutors. As researchers move towards domains with ill-defined tasks, however, new or revised methodologies will be necessary for successful tutoring (Lynch, Ashley, Aleven, & Pinkwart, 2006). Reasoning in such domains is different because there is often no clear path to a solution (e.g., the number of valid interpretive responses to a given literary text is essentially infinite), and when a solution is reached, it may be hard to verify and thus requires a supporting argument for its validity. Because of these different reasoning processes, teaching and learning are different with ill-structured tasks. Whereas the main ITS paradigms provide feedback on student solutions (e.g., model tracing, as described in Anderson, Boyle, Corbett, & Lewis, 1990; constraint-based tutors, as described in Mitrovic, Mayo, Suraweera, & Martin, 2001; see also VanLehn, 2006),

similar definitive feedback is hard to give in ill-defined tasks because it may be difficult to tell whether an intermediate step is on the path to a valid solution. Also, classroom methods for teaching these domains, such as discussion with peers or Socratic dialogue, are not congruent with a typical tutoring system interface (Lynch et al., 2006).

In our current research, we investigate how a simple tutor can help students in a language learning setting acquire intercultural competence, that is, the ability to think and act in culturally appropriate ways. Students are expected not just to communicate in the given language, but also to "gain insight on native perspectives, opinions, and values" by learning to notice, analyze, and explain differences between the target culture and their home culture (Scarino, 2000). Cultural learning is an example of a domain in which solutions may depend on the learner's interpretation of events or language (Kramsch, 1993). While there may be rules that describe cultural norms or behaviors, they tend to have many exceptions depending on the circumstances or on individual traits of the actors in a situation (Edgerton, 1985). Additionally, traditional classroom tasks in this domain do not match the typical problem format of ITS (Kitajima & Lyman-Hager, 1998). For example, to discuss immigration issues in France, students could write an essay, role-play, or watch a film. Support from an intelligent system may be a good way to assist students in learning intercultural competence, but we will need to investigate how ITS research can be adapted to best support ill-defined domains.

In light of the challenges involved, researchers have tried various approaches for support in educational environments for ill-defined domains. Some systems rely on developing complex domain ontologies or artificial intelligence models, which are highly resource-intensive to create and often only cover a limited part of a domain. For example, Rashi (Woolf et al., 2005) supports inquiry learning by encouraging students to make hypotheses about phenomena. The work is done within domains described by large ontologies written by subject matter experts. An AI approach has also been tried in the domain of cultural learning. BiLAT (Kim et al., 2009) is an educational game with an intelligent tutoring system for acquiring negotiation skills in a cultural context. Similarly, Croquelandia (Sykes, Wendland, & Moore, 2008) teaches a limited set of cultural skills, namely communicative pragmatics such as making requests, through a simulated immersive environment. To investigate techniques that apply across many domains and to allow more rapid development of new tutors, however, it may be beneficial to investigate how to address these difficult domains without the need to build a complete model. LARGO is an example of a system with only a partial domain model, which uses authentic documents to support students in interpreting and explaining legal arguments (Pinkwart, Lynch, Ashley, & Aleven, 2008). One disadvantage of LARGO's partial model is that it contains knowledge about general patterns that may occur in an argument but does not know about the specific content of the task. In our work in the cultural domain, we have seen that if students are not supported with content knowledge, they may successfully display some cultural skills and still end up with incorrect factual knowledge (Ogan, Walker, Aleven, & Jones, 2008).

In our research project, we created ICCAT (InterCultural Competence Attention-focusing Tutor), a simple tutor that augments, in an online setting, an existing classroom model for the development of intercultural competence (Liddicoat & Crozet, 2001). This model comprises four steps for acquiring cultural knowledge from authentic sources such as films. These steps are facilitated by the instructor, who points out cultural elements that students miss or that are difficult to notice and helps them reflect on what they have seen. In the context of an online course, or while completing homework online, it is a much greater challenge to provide this support. Video content can now be easily distributed over the internet, but video viewing without direct guidance from an instructor can result in shallow or little learning (Mills, Herron, & Cole, 2004; Pusack & Otto, 1997). There is a need to provide students with

guidance to help them notice the salient features of the video, and to help them interpret and discuss the events depicted in the video from a new perspective. Our system provides this kind of guidance. ICCAT follows the phases of the Liddicoat model using short feature film clips from the target culture that are selected for their cultural content. To guide students in interpreting the events from intercultural perspectives, it augments these phases with attention-focusing and reflective scaffolding: (a) the clip *pauses* at moments of high cultural interest, (b) students *predict* what will happen next in a series of deep questions, and then (c) after they have viewed the rest of the clip, students *ponder* as they focus on the features that supported or contradicted their prediction and answer questions about the cultural content of the clip.

ICCAT thus diverges from a traditional ITS approach: there is no step-by-step feedback based on a full student model. Instead, ICCAT encourages open-ended reflection, which is not something ITSs tend to do well. However, it does take inspiration from the eight principles developed by Anderson et al. (1995) for guiding the construction of intelligent tutoring systems. As discussed in more detail below, ICCAT's pause-predict-ponder techniques are adaptations of some of these principles. The work thus represents a synthesis between an existing intercultural development model, tutoring technology, and cognitive tutoring principles.

In this paper, we describe the ICCAT system and use it as an example to describe how the intelligent tutoring systems principles might be adapted to support ill-defined domains. We then present a classroom experiment that evaluates whether the *pause-predict-ponder* technique employed by ICCAT leads to superior learning in the domain of intercultural competence, compared to video viewing with the unscaffolded Liddicoat model. We found that the addition of the pause-predict-ponder model seemed to guide students in acquiring knowledge about typical practices or behaviors in the target culture, as measured by multiple-choice questions. Additionally, the system significantly increased students' ability to reason from an intercultural perspective according to a validated scale (Steglitz, 1993), an ill-defined skill that is difficult to measure with a traditional posttest. We also present a further analysis of the posttest data and student discussions to better understand student behaviors and why these techniques might be effective. We discuss possible implications for using intelligent tutoring systems to support ill-structured tasks in other domains such as history or art criticism, where use of authentic materials and discussion are important teaching methods.

BACKGROUND

Cultural understanding, while established as a meaningful element of learning a second language, is often given a less than thorough treatment in lower-level language courses as instructors focus on more tangible speaking and writing skills (Maxim, 2000). In an attempt to include these higher-order skills, The American Council on the Teaching of Foreign Languages has set forth standards regarding what students should know and be able to do in *Standards for Foreign Language Learning in the 21st Century* (ACTFL, 1996). A significant number focus on cultural understanding, for example:

> *Standard 3.2:* Students acquire information and recognize the distinctive viewpoints that are only available through the foreign language and its cultures.
> *Standard 4.2:* Students demonstrate understanding of the concept of culture through comparisons of the cultures studied and their own.

These standards indicate a need for emphasis on developing intercultural competence as an integral part of the foreign language classroom. Furthermore, the document stresses the importance of going beyond a simple dissemination of knowledge of cultural practices to developing activities that cause students to reflect and gain insight on native perspectives, opinions, and values. For example, the Cultura project (Furstenberg, Levet, English, & Maillet, 2001) found that "The word *individualisme/ individualism,* is a prime example where highly positive connotations of words such as 'freedom,' 'creativity,' and 'personal expression' appear on the American side, while the French side is replete with such negative notions as '*égoïsme,*' '*égocentrisme,*' '*solitude.*'" Students do not make absolute conclusions about cultural issues (nor do we encourage them in our work to develop one single interpretation of a cultural phenomena), but rather use authentic documents to generate possible perspectives. The ability to notice, describe, and analyze these different perspectives is termed "intercultural competence."

In this work we focus on two of the main skills of intercultural competence, which have been described in Byram's *savoirs* (1997). S*avoir-comprendre* involves interpreting phenomena with a culturally appropriate perspective. Specifically, students should consider how observed behaviors or practices from the authentic input they receive might be indicative of different cultural values or customs, and reflect on or explain how those values and customs compare to those in one's own culture (Kramsch, 1993). S*avoir-s'engager* is the ability to make informed critical evaluations of aspects of one's own and other cultures. Practicing these skills differs from simply knowing a few facts about a culture (e.g., that the French standard criteria for completing high school is the Baccalauréat exam) in that students must be able to search for cultural explanations and take into account multiple points of view. These skills are difficult to acquire, in part because people often instinctively interpret events from their home culture's perspective, and much of their knowledge about culture is tacit. Bennett (1993) describes six stages through which people progress as they develop intercultural competence, and notes how rare it is to reach the final stage of complete competence.

A common approach in cultural teaching aimed at the ACTFL standards is to show multimedia in the form of videos, commercials, or other media, and then ask students to reflect and discuss in a classroom setting. Kitajima and Lyman-Hager (1998) suggest that "students can watch silent videos and discuss similarities and differences between, say, a train ride in their own country and in the target language community." In one such project at the University of Hawaii, students view Japanese commercials to identify cultural stereotypes and then role-play, in discussion groups, versions of these commercials created from their observations. This classroom discussion format creates synergy that can supersede the reflective capability of individual students. In another study (Herron & Dubreil, 2000), students watched French video as an "advance organizational tool" to gather information that was shown to be helpful in answering cultural assessment questions. In a different take on cultural perspectives, constructionist research such as the Cultura project (Furstenberg et al., 2001) invites students to construct their own knowledge of cultural values and attitudes. This construction is accomplished, largely without instructor intervention, by having students answer questionnaires about their own culture and then communicate with a class in France to assess the authentic cultural descriptions provided by the other class's questionnaires. This method of cross-cultural learning is motivational and deeply informative, but involves a great deal of overhead in linking classes across continents. None of these methods, unlike our project, attempt to scaffold the cultural learning using technology.

Our project focuses on the use of on-line multimedia technologies in language classes to support the development of intercultural competence. Much work has shown that multimedia instruction must

be well-designed to support student learning (Mayer, 2001). We help prepare students for an on-line discussion through video viewing, augmented with a set of attention-focusing techniques we call *pause-predict-ponder*. Previous work on language learning provides indirect support for the use of these kinds of attention-focusing techniques. Focusing students' attention on key components of language materials (e.g., by underlining or highlighting) has been shown to help students acquire new knowledge by noticing inconsistencies with their current knowledge, in domains like vocabulary or grammar (Chun & Plass, 1996; Dekeyser, 2005). More generally, Faraday and Sutcliffe (1997) found evidence that in order to correctly recall information, attention must be focused on appropriate features of multimedia, such as by revealing important information for emphasis. Prediction, a key attention-focusing technique in our work, has been used successfully in domains other than language learning. For example, it is one of the four components of reciprocal teaching (Palincsar & Brown, 1984), a program that supports students in learning strategies to comprehend text and read with understanding. Prediction-making has also been hypothesized to be a mechanism that helps students learn from viewing animations of computer algorithms (Byrne, Catrambone, & Stasko, 1999). In scientific inquiry, a domain where hypothesis forming is critical, it has been shown to successfully focus students' attention on the appropriate theory to test (Schunn & O'Malley, 2000; see also Smith & Reiser, 2005). Intercultural competence skills may also benefit from prediction scaffolding in a support system that helps students make explicit their beliefs about cultural phenomena.

SYSTEM DESCRIPTION

Building with the CTAT Tools

We built our simple tutors using the Cognitive Tutor Authoring Tools (CTAT), a set of tools from Carnegie Mellon University that facilitates rapid development of intelligent tutors. One advantage of CTAT is that it allows authors to build 'example-tracing tutors.' These systems behave like intelligent tutoring systems in that they have the capability to provide context-sensitive hints and feedback. However, unlike model-tracing tutors, they compare students' problem steps against examples of correct and incorrect problem-solving behaviors, rather than against output from an executable rule-based model. Example-tracing tutors are built by demonstration, and do not require complex AI programming (Aleven, McLaren, Sewall, & Koedinger, 2009; Koedinger, Aleven, Heffernan, McLaren, & Hockenberry, 2004). Thus, CTAT is attractive for use in ill-defined domains, in which individual problems or tasks are often unique and intricate, and developing a fully generalized domain model may not be desirable. While our current project does not exploit the full AI capabilities of the CTAT toolkit, the technical infrastructure would support extensions in that direction as described in our future research.

Selection of Video Content

The first step in creating the instructional materials was to collect a corpus of video clips. Based on the current curriculum in the French Online course in which we ran our study, we selected three themes for the cultural material, namely immigration, education, and unemployment issues in France. Then, we requested suggestions of feature films demonstrating cultural attitudes or behaviors from three French instructors who utilize such material in their current classrooms. Instructors were asked to

provide brief descriptions of appropriate scenes that fell under the chosen themes that they believed would be beneficial in teaching about cultural issues. We then viewed and annotated the films to extract 30-second to 2-minute video clips based on the guidance from instructors. The requirements for scenes to be chosen were that they a) afforded a natural moment to pause and ask students to make a prediction about the events of the second half of the clip, b) contained a "teachable moment" at the pause point such that the prediction was dependent on cultural knowledge, specifically a learning objective from the course, and not the plot of the film, and c) contained enough context to be comprehensible with only a brief introduction. Six video clips in our corpus were chosen for use in the study, which included two for each cultural theme.

Creating an ICCAT Tutor

Using these video clips, we developed the simple tutors for the experimental condition. ICCAT leads students through the four stages of the Liddicoat model:

1. Students first receive some form of *authentic input* – this step is often instantiated in the classroom by watching a film or other media that portrays realistic elements of the culture to be learned.
2. Students engage in *noticing,* where they detect and report unique cultural elements that they observe in the cultural input, possibly with guidance from an instructor.
3. Students then *reflect* on what has been noticed, again with instructor guidance.
4. Finally, students produce *verbal output* that is evaluated by the instructor or their peers. This output may take the form of an essay or a classroom discussion.

Liddicoat has shown that the model has had success in assisting students in acquiring cultural skills in classroom contexts. In order to scaffold the instruction in an online context, we developed a tutor based on this model and introduced the *pause-predict-ponder* approach described above.

Authentic Input

On ICCAT's first screen, students review a summary of the film and a brief introduction to the material in the clip. To illustrate, one clip is from *Monsieur Ibrahim*, a recent film dealing with issues of cultures clashing among immigrants in Paris (Petin, Petin, & Duperyeron, 2003). On the next screen students view the first half of the clip, which begins the first step of the Liddicoat model, *authentic input*. In this scene from *Monsieur Ibrahim*, a boy named Moses walks into a neighborhood convenience store and has a conversation with the elderly proprietor about the etymology of their names. The shopkeeper gently explains to Moses that while he is Muslim, he is not Arab. The boy asks, "Then why does my father say 'Go to the Arab's'?"

Notice

This places students at the second step in the Liddicoat model, *noticing,* where they detect and report unique cultural elements with the assistance of ICCAT's attention-focusing techniques. After the boy's question to the shopkeeper, the video *pauses* and students respond to a set of questions (see Figure 1) in which they: 1) *predict* the next event that will occur in the clip from a drop-down menu,

2) provide a more extensive natural language explanation of their choice, and 3) state what they believe an appropriate response to the situation might be in their own culture. In general, a correct prediction requires cultural knowledge, not just an understanding of the narrative content of the clip. The prediction, presented in a menu format, constrains students to several actual cultural possibilities. For this video clip, two possible responses include "The neighbors don't take the time to get to know me," and, "Anyone in this profession is labeled an Arab." The menu selection both scaffolds students in anticipating cultural reactions, and also helps to make them accountable for their response to the subsequent free response question because they have now made a choice that they must defend. To enable students to self-reflect, the system does not provide immediate feedback on these responses. This allows them to explore their own interpretation of the events prior to receiving tacit feedback from watching the events unfold in the film. Using the multimedia controls, students may rewind and review the video up to this point as often as they like, but are constrained from moving forward in the film until they have made their responses.

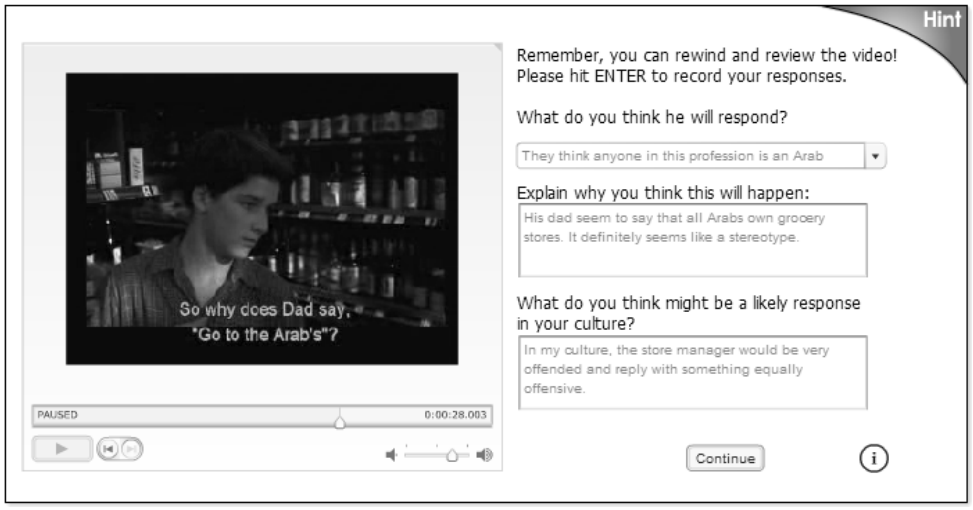

Fig. 1. Students watch the first half of the film and make their prediction.

Reflect

Once students have responded, the succeeding portion of the video clip plays and the ensuing cultural event is revealed. In *Monsieur Ibrahim*, the shopkeeper describes to the boy how his profession is often stereotyped by non-immigrant Parisians. As the story unfolds, students *reflect* on what has been noticed as they focus on the features that supported or contradicted their prediction, with help from ICCAT's *ponder* questions. These questions first ask students to do a self-assessment of whether they were correct in their prediction or not (see Figure 2). They are asked either to revise their prediction or to explicate elements that surprised them in free text, which may help them notice gaps between their prior knowledge and desired knowledge state. To do this, they may review the whole video clip as many times as they like. Finally, students are given a set of characterization and general comprehension questions that may be answered with "true," "false," or "maybe." That some of these questions do not have a definitive answer reinforces the idea that there are often unknowns in this ill-

defined domain, which may depend on interpretation or context. These questions are tutored with hints and error messages, as well as success messages that provide a summary of the evidence for a correct answer (see Figure 3).

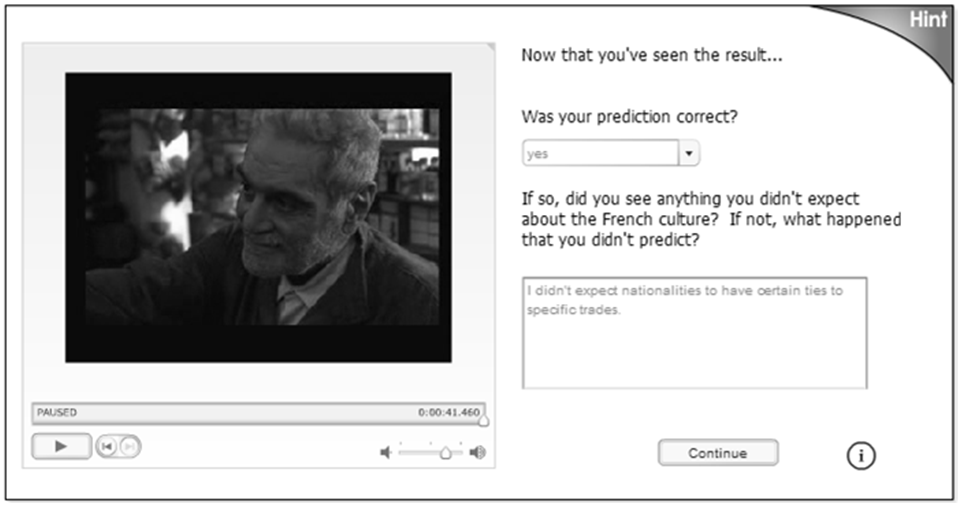

Fig. 2. Students view the result of their prediction and reflect.

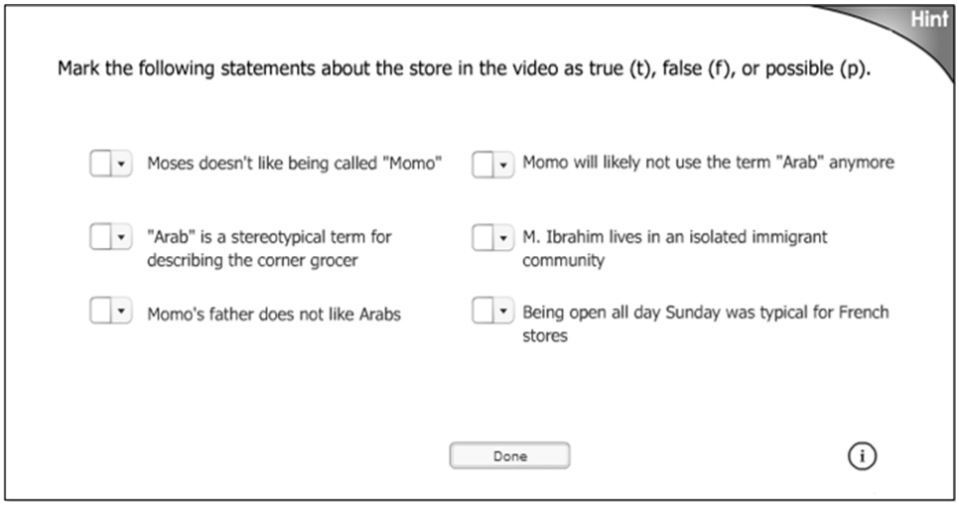

Fig. 3. Characterization questions with hints and feedback.

Create Verbal Output

In the final step of the Liddicoat model students *create verbal output* to be reviewed by the instructor and peers (see Figure 4). They participate in an asynchronous online class discussion where they use what they have seen and reflected on to discuss cultural differences, similarities or assumptions about

the French culture, and ask questions about the meaning of behaviors they observed in the clips. To complete the cycle of the developmental model, feedback on the output students create (in our system, peer feedback rather than instructor-driven) should assist them in noticing relevant features of the next authentic input.

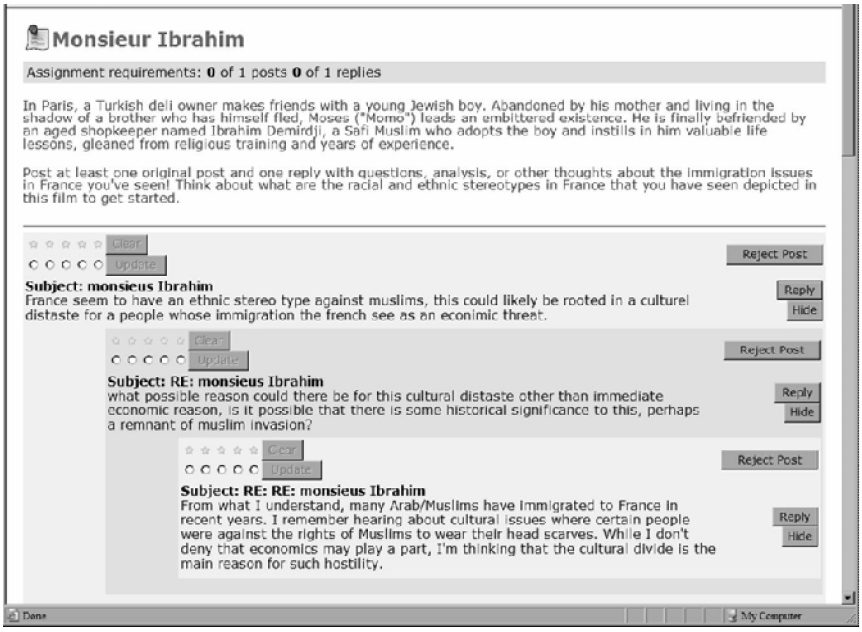

Fig. 4. Image of the online discussion board.

Figure 5 shows a sequence diagram of all the steps in the tutor. For each of the three cultural themes we selected, we created an instructional package with two ICCAT tutors using video clips that show slightly different perspectives on the issue in the theme.

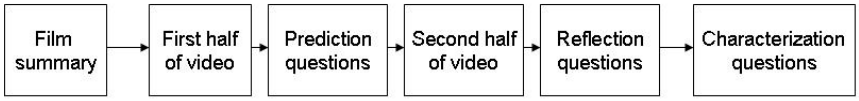

Fig. 5. Sequence in an ICCAT tutor

COGNITIVE TUTOR PRINCIPLES

So far we have emphasized that ICCAT is different from intelligent tutoring systems because it is based on the Liddicoat model, not a standard problem-solving approach. However, ICCAT does share many traits with standard ITS. As researchers begin to develop systems for ill-defined domains and search for principled design guidelines, it may be beneficial to consider a widely-cited set of principles that underlie cognitive tutors in well-defined domains (Anderson et al., 1995). Here, we use ICCAT

and the domain of intercultural competence to examine several of these principles that may be appropriate for teaching open-ended problems, and how they might guide the design of future systems. *Principle 2: Communicate the goal structure underlying the problem solving.* In this principle, Anderson et al. describe an approach Brown (1985) called *reification*, or loosely interpreted, a tutor interface that helps students "make thinking visible." In a well-defined domain, this typically means having students perform every step of a solution one at a time, and laying out these steps ahead of time in the interface. In ill-defined domains where students may take widely different approaches to crafting a solution, this principle might be best applied in making students' implicit assumptions and reasoning explicit in the interface. This idea is critical for the skills of intercultural competence, because it is a domain in which much of the knowledge is tacit. People tend to naturally adopt an "ethnocentric" perspective by adopting a cultural worldview that regards their home culture as central (Bennett, 1993), yet many are not aware of the assumptions they make that are based upon this perspective. Thus, communicating the tacit knowledge that underlies cultural behaviors may support the goal of making cultural learners aware of their own preconceptions. In the ICCAT interface, students make a prediction about what will happen next in the film at a point in which a particular learning objective occurs. As they make a prediction, students not only exercise their knowledge of French culture, but they also reify implicit beliefs. Only when students make these beliefs explicit in an accountable form such as a tutoring interface must they confront them as biased or inaccurate.

Principle 3: Provide instruction in the problem solving context. In well-defined domains, this means instruction should be contiguous with problem-solving practice, but in ill-defined domains, the idea of problem solving context may be taken even further. Situated learning has long been a mainstay of these domains, from legal reasoning with use of real courtroom case studies (Aleven & Ashley, 1993) to LOGO which teaches programming by directing a turtle to walk around an environment (Papert, 1980). As noted above, the literature on teaching intercultural competence also suggests using authentic input as the driver behind the analysis and explanation of culture. Therefore all the learning in ICCAT is contextualized both within a particular issue (e.g., immigration issues in France) and then with specific examples (e.g., a clip from *M. Ibrahim*, a movie based on this cultural issue).

Principle 4: Promote an abstract understanding of the problem-solving knowledge. This principle may seem to be in conflict with the idea of situated learning, since ill-defined tasks may have little meaning outside their context (Lynch et al., 2006). However, there may be ways of generalizing across tasks that will promote a greater abstract understanding. In our work, students watch two video clips for each issue which demonstrate slightly different viewpoints, and then they must reconcile the two in the discussion and posttest. This may help promote a more abstract understanding of the cultural issue.

Principle 8: Facilitate successive approximations to the target skill. While in cognitive tutors this principle often means that the tutor will fade problem scaffolding until the student is able to perform the whole skill independently, it may be manifested differently in ill-defined domains where skills are difficult to divide into component parts. In intercultural competence, students work their way along a spectrum of development which brings them closer and closer to full intercultural competence (Bennett, 1993). A system, rather than performing intermediate problem steps for a student, may be designed to nurture students in making steps along this spectrum. A beginning step in this developmental sequence is simply to realize that cultural differences exist and that they can be explained. In ICCAT, students are encouraged to attempt these explanations even when they may not

be able to give specific, elaborated reasons. They also give explanations from their home culture, a step which may develop understanding that cultural explanations exist in a foreign culture. Cultural knowledge may also have successive approximations; as in Principle 4, viewing multiple examples may help students get closer to a full, multi-faceted understanding of a cultural issue.

Principle 6: Provide immediate feedback on errors. Anderson et al. describe the benefits to learning of receiving immediate feedback on errors. On the other hand, in ill-defined domains, the notions of "desirable difficulties" (Bjork, 1994), learning from errors (Ohlsson, 1996), and delayed feedback (Schmidt & Bjork, 1992) may be even more applicable than in well-defined domains. Immediate feedback has been shown to hinder error detection and correction (Schooler, 1990), which are important skills in a domain like intercultural competence where learners need to react to rapidly changing conditions. ICCAT takes what may be an appropriate middle road and allows students to immediately view the result of their prediction and self-reflect on their understanding, before providing non-immediate feedback through questions and subsequent online discussion.

The applicability of the remaining principles to building tutors for ill-defined domains is less clear. Some may be intractably hard to fulfill for many domains – for example, building a rule-based representation of the skills may be impossible given disagreement between experts. However, significant elements of the principles apply in many ways in a tutor based on the Liddicoat model, and developing a generalized set of principles for ill-defined tutors is an important goal for researchers.

EXPERIMENT

Hypothesis

A study was completed in two introductory French classes to test the added value of the *pause-predict-ponder* techniques. The hypothesis was that focusing students' attention on moments that highlight a noteworthy cultural attribute through prediction, and encouraging them to reflect on their predictions, helps students acquire more cultural knowledge and a greater ability to reason from an intercultural perspective, compared to video viewing without the *pause-predict-ponder* techniques.

Participants

The study was conducted in the French Online class being taught at Carnegie Mellon University (CMU) and the University of Pittsburgh. This course was an introductory French course that formed part of the Pittsburgh Science of Learning Center and was taught online, except for an in-class session with all students once a week. In total, 38 students were enrolled between the two classes, but only 34 completed at least one of the assignments during the semester. Since each assignment covered novel cultural material, all students who completed at least one of the assignments were included in the analysis. There were 21 students who completed at least one assignment at the University of Pittsburgh. Because the course was offered in the evening at Pitt, many of these students were continuing education students. At CMU, 13 students completed at least one assignment. These students were generally freshmen or sophomores who enrolled in college directly following high

school. English was the native language for 29 students. Only three students had ever been to France, but 28 students had travelled outside of the United States.

Within each classroom, students were randomly assigned to either the experimental group that used the ICCAT system or a control group that viewed the same video clips without the attention-focusing techniques. Therefore conditions were balanced in each classroom. In our final analysis, 17 students were included in each condition.

Procedure

The study was comprised of three assignments spaced throughout the semester, which took the place of typical cultural reading and writing assignments for the class. The materials for each assignment were linked from the course webpage that provided related knowledge about the theme. Although the course was taught by a different instructor at each university, the material was accessed from the same online resource, and thus kept constant. Students received participation credit for each assignment.

Students first took a demographic survey. Then, in each assignment, students first took a short pretest that explored their knowledge of basic information about the assignment's cultural theme. Because the manipulation relies on focusing students' attention on key moments in the film, the pretest was designed with open-ended questions that allowed students to express their knowledge without priming them to identify particular answers in the video. Each assignment then consisted of two film clips to watch and discuss. To view the clips, the experimental condition used the tutor version described above, which includes the *pause-predict-ponder* method. The control condition on the other hand watched the video clips in a version of ICCAT from which the *pause-predict-ponder* steps had been removed. This tutor version did not pause the video or prompt for predictions or reflection, but simply provided an open space where students were encouraged to take notes. After each ICCAT session, students accessed the discussion board for that video clip. They were required to post at least twice over the course of a week and encouraged to return to the discussion board to read other replies and to post again. Finally, students took a posttest that covered cultural knowledge from both films in the theme. After the third assignment, students watched one final video and were asked to write a short essay response about the cultural issues they saw in the film.

Measures

We assessed how the interventions affected students' learning both in terms of their specific knowledge about French culture as it relates to immigration, education, and unemployment, and the level of intercultural competence displayed in the discussion board posts. First, following each assignment, students completed an online post-test that included 7 or 8 analytical questions covering the theme addressed by the assignment. The questions measured knowledge of French culture in multiple choice and true/false format questions, which were developed with French language instructors through an analysis of the cultural elements in video clips. They covered key knowledge related to the high-level cultural objectives of the course, and were situated within the context of the film. In general, questions asked students to analyze the events in the film (e.g., "From the factory scene in *Ressources humaines*, we know that social mobility is possible in French society.") or compare between the two films in the assignment (e.g., "Based on the clips you just saw, *Être et avoir* and *Le Péril jeune* are similar in EVERYTHING BUT: …"). Comparison questions focused on conceptual issues rather than superficial differences in setting or visual details. In particular, the

cultural attribute that was exemplified at the moment of each prediction made in the experimental condition had a related question on the posttest.

Although these questions enabled us to get a concrete glimpse at how students analyzed and synthesized the material in the instruction, these types of questions cannot hope to cover all of the learning in an ill-defined domain. Nor can they give an understanding of how students might apply their knowledge in an open-structured assessment or real life situations. To that end, we used students' discussion board posts as a measure of whether the intervention led students to reason more often from an intercultural perspective. To analyze their posts, we used a coding scheme developed and validated by Steglitz (1993), which distinguishes three main levels of desirable and undesirable cultural writing on a similar intercultural competence writing task, described here in simplified form:

Level 1: No awareness of cultural influences on behavior in the writing sample (e.g., "This is the wrong way to do things!")

Level 2: Students show awareness of multiple perspectives, but are unspecific about cultural causation (e.g., "I think the French do things differently in the education system")

Level 3: Shows clear cultural awareness with specific, elaborated cultural explanations (e.g., "French students seem to see education differently, perhaps because they take more personal responsibility in the education system")

Level three (specific, elaborated cultural explanations) is the desired target for students discussing cultural issues. To evaluate student responses, two human coders were trained using a coding manual. Then a subset of 20% of the data was coded by both raters, who were blind to condition. While there was 75.6% agreement between raters, this resulted in a Cohen's kappa value of .56 (Cohen, 1960). This kappa value, which indicates a medium level of agreement between raters, may be indicative of the ill-defined aspects of the domain in which experts do not always agree on correct responses. After discussion, one rater then coded the rest of the posts and singled out those that he believed to be difficult to categorize. The second rater also rated these "difficult" posts, and on this round of ratings a .67 kappa value was reached, which is in the "substantial agreement" range.

RESULTS

Pretest questions were evaluated by coders who rated the number of appropriate answers each student gave. Because of the differences observed in the type of students who were enrolled in the course at Carnegie Mellon University and the University of Pittsburgh, we expected that there might be a disparity in their performance. However, a *t*-test showed no significant difference on pretest scores between students in the two universities. Also, although we anticipated a disparity in cultural abilities between international students and U.S. citizens, a *t*-test on pretest scores found no significant differences, perhaps due to a small sample of five international students. Therefore, all students were analyzed in two groups, the experimental condition and the control.

Assessment of Posttest Responses

On the cultural knowledge posttests, students were given a point for each question answered correctly, and then these scores were averaged across the three assignments (see Figure 6). While students in the

control condition scored an average of only 62% (*SD* = .12), students in the experimental condition scored an average of 71% (*SD* = .15*)*. A one-way ANOVA revealed that this difference was marginally significant, $F(1, 32) = 3.47$, $p = .071$. Cohen's *d* (1988) was calculated to measure the relative size of the observed effect. The treatment had an overall effect size of .67 standard deviations over the control condition (generally accepted as a large size effect). When we explored each assignment independently, a disparity between the assignments emerged. The difference in means (Experimental: *M* = 72%, *SD* = .12; Control: *M* = 59%, *SD* = .21) on the first assignment was marginally significant ($F(1, 27) = 3.78$, $p = .058$), and Cohen's *d* showed a large effect size of .75. However, on the subsequent two assignments, the scores of the conditions seemed to converge. This difference may be explained by the fact that three students in the experimental condition with a score over 80% dropped out by the third assignment, while three students with scores under 60% dropped out of the control condition by the third assignment.

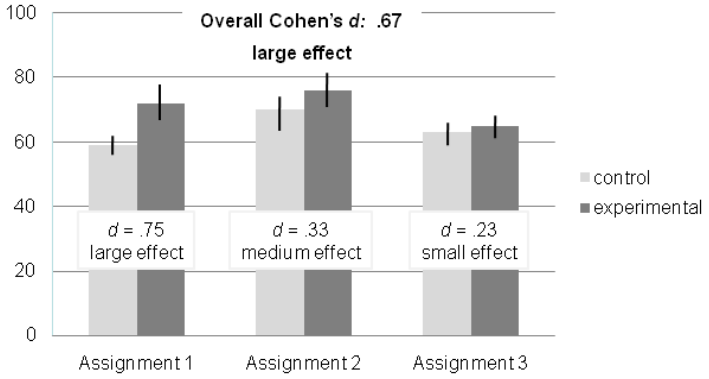

Fig. 6. Means for post-test by assignment.

While all of the questions on the posttest related to course objectives that were observable in the video, each test had one particular question whose answer was best exemplified in the video at the moment of the pause. Looking at these six questions only, students in the experimental group (*M* = .72, *SD* = .22) made significantly more correct responses on these particular knowledge components than the students in the control condition (*M* = .54, *SD* = .17), $F(1, 21) = 4.78$, $p = .04$). Examining the tutor data logs, students made a correct prediction 40% of the time at the moment of the pause, which is better than the chance rate of 25%.

Assessment of Student Discussion

In our second assessment, each post on the discussion boards was given a score according to the three main categories developed by Steglitz (1993), described above. Table 1 presents an example of a post classified at each level. Because the data are frequencies, we performed a Mann-Whitney test for non-parametric data to examine the differences between the conditions. Posts rated at level one, the lowest level of intercultural competence, were significantly more likely to occur in the control condition. Posts rated at level two were significantly more likely in the experimental condition. Posts rated at level three, the highest level of intercultural competence, were marginally more frequent in the

experimental condition, a result that may reflect the low number of level 3 posts overall (Mann-Whitney *U* tests, $U = 75.50$, $p = .01$; $U = 94.00$, $p = .05$; $U = 107.50$, $p = .08$ respectively). Figure 7 shows the frequencies of each level by condition.

Table 1
Examples of posts from each category

Level 1 no cultural explanation	I think it might be better to lay people off instead of reducing their working hours and thus keeping them around. These people would have more time and move on with their career. Although there would be no job opening, they might be able to get together and do a small business like making handcrafts that doesn't require large investment.
Level 2 unspecific cultural explanation	I think the living condition of these people had to do with racial segregation. As we watch on Tv recently about the riots in france, those were some of the issues that the immagrants were trying to address. Hey but violent is not the way to solve the problem thou. Their living condition is not on the positive side at all. Is there any movies out there that address the raacil classification of the immigarnts in France? I want to see that. I think there is more to the story than what this one minute clip is telling us.
Level 3 specific & elaborated cultural explanation	I was suprised by the level of comfort between Moses and Ibrahim. I think Ibrahim is comfortable with his informality with Moses because he gives him a nickname "momo" which is "less impressive." This is a good example of distinction between youth and adult in French culture. Remember, when talking to children they use "Tu" instead of "Vous." This distinction indicates that children are seen as somewhat inferior. Which can be taken as age discrimination. In addition, when Moses reveals that his father refers to Ibrahim as "The arab," it shows that Moses has been exposed to stereotypes and Ibrahim fuel his ignorance by explaining what is "arab" in terms of store hours.

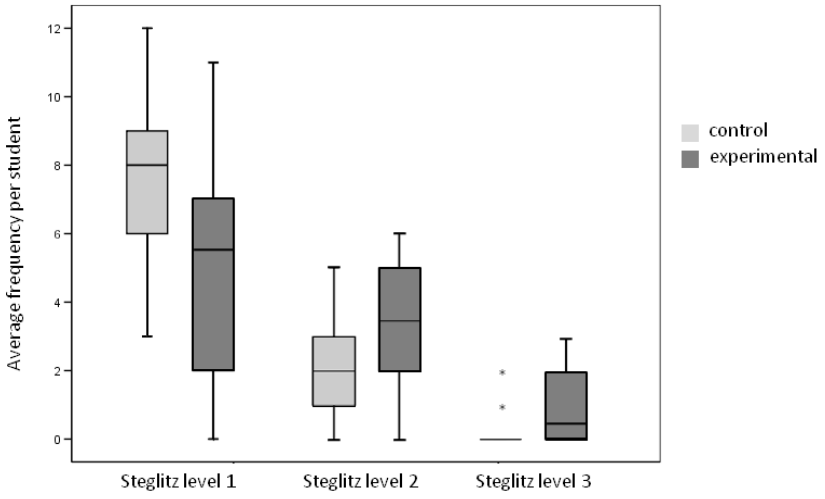

Fig. 7. Frequency of posts categorized at each Steglitz level.

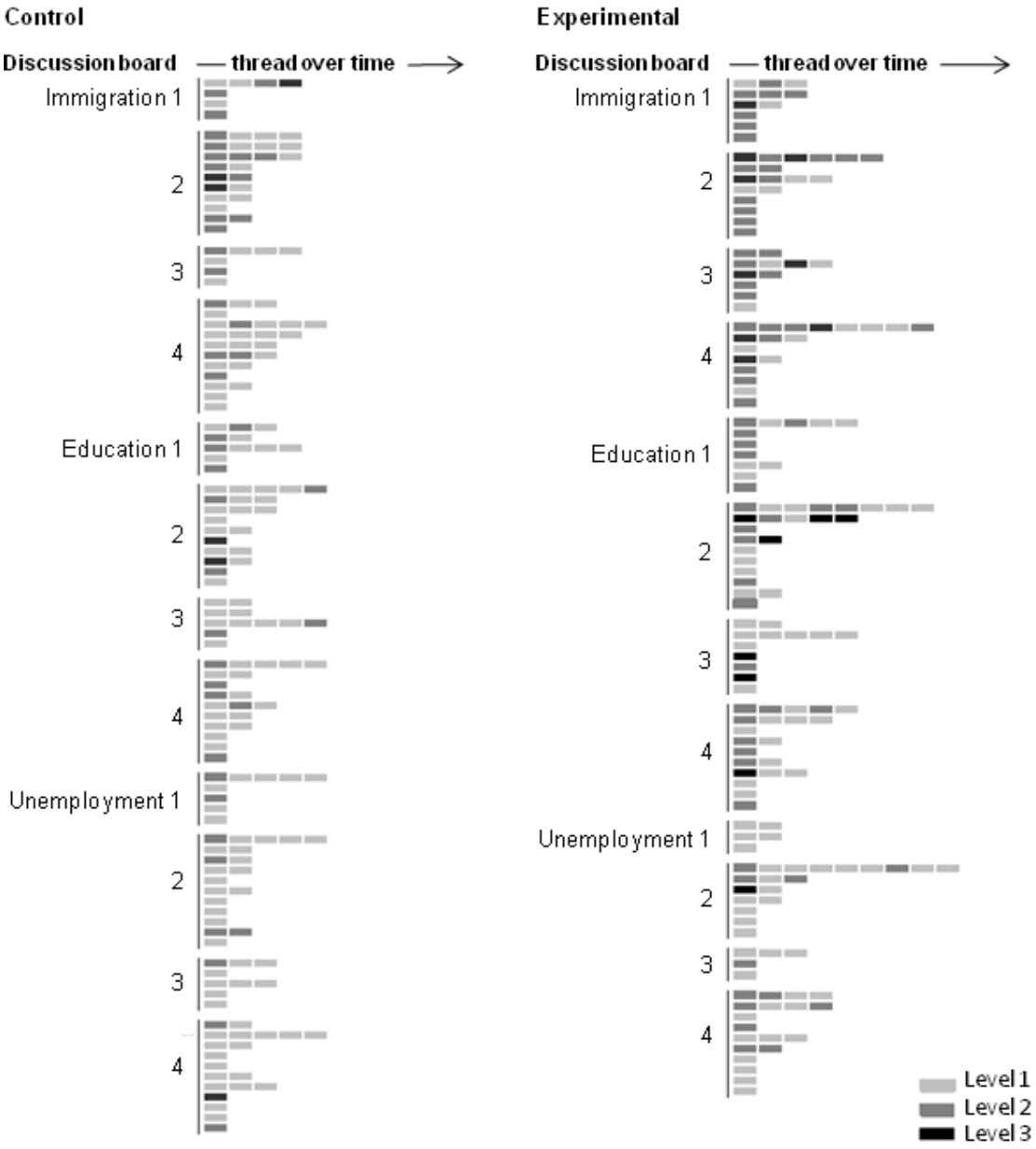

Fig. 8. Visualization of all the discussion board threads by condition. Each condition has 3 themes with 4 boards each (2 video clips per theme, with each classroom on a separate board).

These statistics showed the overall differences between the conditions. Next, we explored in more detail how the discussion progressed in each condition, collapsing over all assignments. To assist in

visualizing the flow of discussion, Figure 8 shows the discussion boards for each condition, where each row represents one thread in a discussion over time. Each bar is a single post in the thread, whose color indicates the level it was rated on the Steglitz scale.

To determine whether the discussions tended to improve over time, we developed a measure called upFrequency, which is intended to reveal whether the level of posts in the discussion went up over time. It was calculated by giving the thread a point for every time that a post was followed by a post of a higher level. For example, a thread that went: level 1, level 2, level 2, level 3 would have an upFrequency score of 2. Only threads of length greater than one were included in the analysis. An ANOVA, using condition as the between-subjects factor, revealed that the experimental condition (M = .35, SD = .54) was much more likely to make an improvement in the discussion than the control condition (M = .15, SD = .41), $F(1, 83) = 3.98$, $p < .05$). Additionally, we calculated downFrequency, the frequency at which discussion posts dropped in level over time. For example, a thread that went: level 2, level 1, level 3, level 1 would have a downFrequency of 2. Again, only threads of length greater than one were included in the analysis. In an ANOVA comparing conditions we found that the experimental condition (M = .83, SD = .80) was also much more likely to have a discussion decrease in level than the control condition (M = .46, SD = .50), $F(1, 83) = 7.15$, $p = .009$). This result is not surprising given that upFrequency was also greater in the experimental condition; the visualization in Figure 10 shows that the experimental condition had more "peaks" in the middle of threads. We additionally looked at how well the discussions began by examining the level of the first post in each thread. A Mann-Whitney test showed that the experimental condition (Mean Rank = 98.76) had a significantly higher rated first post than the control condition (Mean Rank = 77.83), and yet students were still better able to make improvements over the course of the thread ($U = 2910.00$, $p = .002$). In the control condition, a Mann-Whitney test reveals that the first posts (Mean Rank = 74.81) were likely to be significantly higher in level than the posts in the rest of the discussion (Mean Rank = 59.55), ($U = 1682.50$, $p = .01$), another indication that the control condition discussion did not improve over time.

Finally, in a comparison made between the posttest assessments and the discussion scores for each student, no significant correlations were found for any assignment ($p > .5$ for all correlations), which indicates that students were drawing on different skills in each assessment.

DISCUSSION

We found that students in the experimental condition exhibited significantly greater intercultural competence in their cultural writing, and also a marginally significant difference in their cultural knowledge. In an attempt to further understand how the *pause-predict-ponder* techniques beneficially influenced students' learning trajectories, we investigated the predictions made in ICCAT and their effect on individual posttest questions. Students made a correct prediction only 40% of the time before viewing the second half of a clip, meaning it is unlikely that they already had a solid grasp of this knowledge. However, they were better than the chance rate of 25%. This may be an indication that they were able to activate some of their prior cultural knowledge and relate it to appropriate information from the clips. On posttest questions that assessed knowledge related to the moment of these predictions, the experimental group was significantly better than the control. Students in the experimental condition were focused on the moment of the pause and were given the opportunity to reflect on the predictions they made about this knowledge; and the effect also marginally spills over

into other cultural knowledge on the posttest. So while it may not be feasible to *pause and predict* for every moment, this result suggests that we can help students understand many declarative knowledge components by focusing their attention through pauses in the video.

The prediction component of ICCAT also helps students acquire intercultural competence, by prompting them to search for cultural explanations in the clip and make a comparison to their home culture. Often they were confronted with incorrect choices they made, and thus had to work on integrating this new information into their understanding. We saw a significant difference in the ability to apply these intercultural skills between experimental and control students. Students in the experimental condition came much closer to including a reference to cultural explanations in every discussion post (i.e., level 2 in Steglitz's coding scheme). Because these perspective-taking skills are so difficult to acquire, this result represents a very positive sign of progress along the developmental scale of intercultural sensitivity (Bennett, 1993). Students acquire desirable skills for intercultural competence (e.g., if there is a difference from my culture, then explore whether or why it might be related to culture), whereas without ICCAT's techniques, students more frequently display shallow or undesirable intercultural competence skills (e.g., if there is a difference from my culture, the behavior or value is wrong). Further, the experimental group showed a better ability to maintain a productive discussion compared to the control group. We looked not only at isolated post quality, but also at how the discussion was influenced over time. In the control condition, the discussion often started out at a reasonable level (although generally not as good as the experimental group), but only seemed to devolve over time. The level of the discussion in the experimental condition increased and decreased as new students posted. Students showed the ability to ignore an off-topic comment or introduce new cultural knowledge into a thread, while in the control condition that was unlikely to happen. Additionally, inspection of the length of discussions suggests the experimental group was more prone to engage in extended discussions. The experimental group had a number of threads with over 5 posts, which did not occur in the control condition. Interestingly, they also had more threads of a single post with no replies. It may be that students in the experimental condition were more careful in selecting threads in which to reply – they chose to put effort into only the threads that were fruitful for learning.

On a final note, although both cultural knowledge and cultural discussion are important targets of a language classroom, we found that measures of the two competences are not well correlated. This implies that some students are able to acquire the declarative components of the domain without gaining any perspective-taking skills. A test that might be considered a normal posttest for cultural learning apparently does not capture these intercultural knowledge components. The results from this study strongly suggest that in assessing cultural learning, and perhaps in other similar ill-defined domains, it is important to include deeper, more comprehensive measures of competence.

CONCLUSION

The research presented in this paper investigates how a simple tutoring system for an introductory language course can be effective in helping students acquire a first level of intercultural competence. Instead of standard problem-solving, in our approach to this ill-defined domain we took an existing classroom model for the development of these skills and inserted authentic input in the form of culturally relevant clips from a feature film. We then augmented two of the steps to structure the support in an on-line environment in ways which are in alignment with principles of cognitive tutoring. We conducted a study to evaluate the added value of our augmented *pause-predict-ponder*

model in ICCAT. The results of this study indicate that these attention-focusing techniques help students to reason better from an intercultural perspective, while potentially increasing learning of declarative knowledge components. We also analyzed the use of the principles for cognitive tutor development in our system. Overall, incorporating the cognitive tutor principles into our system appeared to have positive benefits for learning.

While we saw a significant improvement in intercultural competence in students' writing due to ICCAT, even the improved scores of student contributions to the discussion board are lower than one would hope. In our current work, we explore two approaches to supporting cultural discussion (Ogan, et al., 2008). First, we have developed an extension to ICCAT that provides adaptive, guiding feedback on the discussion board to improve cultural writing. This extension has a simple model of a good discussion post, compares a given student post to the model, and provides private feedback to individual students based on the model. Second, we use a moderator support system to deliver adaptive support to a student from the class who moderates the discussion forum, based on this same model. These new components use relatively simple machine learning to support students in what we have discovered is a very difficult part of this domain. With this machine interpretation, we have the goal of creating a tutoring system for cultural knowledge and discussion that does not need a complete domain model, but does successfully incorporate adapted cognitive tutoring principles.

Designing interactions that lead students to engage in productive cultural reflection, and demonstrating their effectiveness in enhancing multiple measures of cultural learning, are novel contributions to the field of technology-enhanced learning. While the majority of the successes in this field have come in well-structured domains such as mathematics and physics, the current research demonstrates headway in an ill-defined, open-ended task. The work involves exploration of a novel approach to a type of learning that is increasingly used in courses at every level of education: learners working independently, studying resources at their own pace, and relying heavily on multimedia and asynchronous discussion. This approach has the potential to be applicable to a range of areas, such as film studies, history and anthropology, but more generally to any domain where learners make use of resources like film and audio to support deep learning of sophisticated concepts. Applying the cognitive tutor principles incorporated into ICCAT to these domains may have a significant impact on the creation of effective technology-based learning systems. This work represents a synthesis between an existing model of development, tutoring technologies, and cognitive tutoring principles adapted for the scaffolding of ill-defined domains.

ACKNOWLEDGEMENTS

Thanks to Robert Hausmann who provided significant statistical support. Also to Michael West and Cary Campbell, our French Online instructors, and OLI and the CTAT team for their technical assistance. This research is supported by NSF Grant # 0354420 to the Pittsburgh Science of Learning Center. The contents of the paper are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

REFERENCES

American Council on the Teaching of Foreign Languages, (1996). *Standards for Foreign Language Learning: Preparing for the 21st Century*. New York: ACTFL.

Aleven, V., & Ashley, K. (1993). What law students need to know to WIN. In A. Oskamp & K. Ashley (Eds.) *Proceedings of the 4th International Conference on Artificial Intelligence and Law* (pp. 152-161). New York, NY: ACM.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education*, 19, 105-154.

Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. In W. J. Clancey & E. Soloway (Eds.). *Artificial Intelligence and Learning Environments* (pp. 7-49). Scituate, MA: Bradford Company.

Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of Learning Sciences*, 4, 167-207.

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.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.) *Metacognition: Knowing About Knowing* (pp.185- 205). Cambridge, MA: MIT Press.

Brown, J. S. (1985). Idea amplifiers: new kinds of electronic learning. *Educational Horizons,* 63, 108-112.

Byram, M. (1997). *Teaching and Assessing Intercultural Communicative Competence.* Clevedon: Multilingual Matters Limited.

Byrne, M. D., Catrambone, R. & Stasko, J. T. (1999). Evaluating animations as student aids in learning computer algorithms. *Computers & Education*, 33, 253-278.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80, 183-198.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20(1), 37-46.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Ed.).* Hillsdale, NJ: Erlbaum.

Dekeyser, R. (2005). What makes learning second language grammar difficult? A review of issues. *Language Learning*, 55(S1), 1-25.

Edgerton, R. B. (1985). *Rules, Exceptions and Social Order.* Berkeley: University of California Press.

Faraday, P., & Sutcliffe, A. (1997). Designing effective multimedia presentations. In Pemberton, S. (Ed.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 272-278). New York: ACM Press.

Furstenberg, G., Levet, S., English, K., & Maillet, K. (2001). Giving a virtual voice to the silent language of culture: The Cultura project. *Language Learning & Technology*, 5(1), 55-102.

Herron, C., & Dubreil S. (2000). Using instructional video to teach culture to beginning foreign language students. *CALICO*, 17(3), 395-429.

Kim, J., Hill, R., Durlach, P., Lane, H.C., Forbell, E., Core, M., Marsella, S., Pynadath, D., & Hart, J. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education* 19, .

Kitajima, R., & Lyman-Hager, M. (1998). Theory-driven use of digital video in foreign language instruction. *CALICO*, 16(1), 37-48.

Koedinger, K., Anderson, J., Hadley, W., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30-43.

Koedinger, K., Aleven, V., Heffernan, N., McLaren, B. M., and Hockenberry, M. (2004). Opening the door to non-programmers: Authoring intelligent tutor behavior by demonstration. In J. Lester, R. M. Vicari, & F. Paraguacu (Eds.) *Proceedings of the 7th International Conference on Intelligent Tutoring Systems* (pp. 162-174). Berlin: Springer-Verlag.

Kramsch, C. (1993). *Context and Culture in Language Teaching*. Oxford: Oxford University Press.

Kramsch, C., & Anderson, R. (1999). Teaching text and context through multimedia. *Language Learning & Technology*, 2(2), 31-42.

Liddicoat, A. J., & Crozet, C. (2001). Acquiring French interactional norms through instruction. In K. R. Rose & G. Kasper (Eds.) *Pragmatic Development in Instructional Contexts*. Cambridge: Cambridge University Press.

Lynch, C., Ashley, K., Aleven, V., & Pinkwart, N. (2006). Defining ill-defined domains; A literature survey. In V. Aleven, K. Ashley, C. Lynch, & N. Pinkwart, (Eds.) *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 8th International Conference on Intelligent Tutoring Systems* (pp. 1-10). Berlin: Springer-Verlag.

Mayer, R. (2001). *Multimedia Learning*. New York: Cambridge University Press.

Maxim, H. (2000). Integrating language learning and cultural inquiry in the beginning foreign language classroom. *ADFL Bulletin*, 32**(**1). 12-17.

Mills, N. Herron, C., & Cole, S. P. (2004). Teacher-assisted versus individual viewing of foreign language video: Relation to comprehension, self-efficacy, and engagement. *CALICO*, 21(2), 291-316.

Mitrovic, A., Mayo, M., Suraweera, P., & Martin, B. (2001). Constraint-based tutors: A success story. In L. Monostori, J.V. Ancza & M. Ali (Eds.) *Proceedings of the 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems* (pp. 931-940). Berlin: Springer-Verlag.

Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, 3(2), 241-262.

Ogan, A., Walker, E., Aleven, V., & Jones, C. (2008). Two approaches for providing adaptive support for discussion in an ill-defined domain. In In V. Aleven, K. Ashley, C. Lynch, & N. Pinkwart, (Eds.) *Proceedings of the Workshop on Intelligent Assessment and Feedback in Ill-Defined Domains at the 9th International Conference on Intelligent Tutoring Systems* (pp. 1-12). Berlin: Springer-Verlag.

Palincsar, A.S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction,* 1, 117-175.

Papert,S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books, Inc.

Petin, M., & Petin, L. (Producers), & Dupeyeron, F. (Director) (2003). Monsieur Ibrahim [Motion picture]. Sony Pictures Classics.

Pinkwart, N., Lynch, C., Ashley, K., & Aleven, V. (2008). Re-evaluating LARGO in the classroom: Are diagrams better than text for teaching argumentation skill? In E. Aimeur, & B. Woolf (Eds.) *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (pp. 90-100). Berlin: Springer Verlag.

Pusack, J. P., & Otto, S. K. (1997). Taking control of multimedia. In M. D. Bush & R. M. Terry (Eds.) *Technology Enhanced Language Learning*, (pp. 1-46). Lincolnwood, IL: National Textbook Co.

Scarino, A. (2000). The Neglected Goals of Language Learning. *Babel*, 3(34) Summer, 4-11.

Schmidt, R.A., & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4), 207-217.

Schooler, L. J., & Anderson, J. R. (1990). The disruptive potential of immediate feedback. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 702-708) Hillsdale, NJ: Erlbaum.

Schunn, C. D., & O'Malley, C. J. (2000). Now they see the point: Improving science reasoning through making predictions. In L. Gleitman & A. Joshi (Eds.) *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp.889-894) Mahwah, NJ: Erlbaum.

Smith, B. K., & Reiser, B. J. (2005). Explaining behavior through observational investigation and theory articulation. *Journal of the Learning Sciences*, 14 (3), 315-360.

Steglitz, I.E. (2003). Intercultural perspective-taking: The impact of study abroad. (Doctoral dissertation, University of Minnesota, 1993) *Dissertation Abstracts International*, 54 (08), 2799A.

Sykes, J.M., Wendland, L., & Moore, P. (2008). CROQUELANDIA: Helping learners develop authentic intercultural communication skills in a synthetic world. Paper presented at EDUCAUSE Learning Initiative Annual Meeting, San Antonio, TX.

VanLehn, K. (2006) The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education.* 16, 3, 227-265.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R. H., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education*, 15(3), 147-204.

Woolf, B.P., Murray, T., Marshall, D., Dragon, T., Kohler, K., Mattingly, M., Bruno, M., Murray, D., & Sammons, J. (2005). Critical thinking environments for science education. In C. K. Looi, G. McCalla, B. Bredeweg & J. Breuker (Eds.)*Proceedings of the 12th International Conference on Artificial Intelligence in Education* (pp. 702-709). Amsterdam: IOS Press.