Fifth International Workshop on Culturally-Aware Tutoring Systems (CATS2014)

Workshop Co-Chairs:

Emmanuel G. Blanchard, IDÛ Interactive Inc., Canada
Isabela Gasparini, University of Santa Catarina State, Brazil
Amy Ogan, Carnegie Mellon University, USA
Ma. Mercedes T. Rodrigo, Ateneo de Manila University, The Philippines
http://www.cats-ws.org
https://www.facebook.com/SIGoCAT
Preface

The 5th International Workshop on Culturally-Aware Tutoring Systems (CATS2014) is a follow-up to a series of successful CATS workshop editions, organized in conjunction to ITS2008, AIED2009, ITS2010, and AIED2013. The present edition is organized in conjunction with ITS2014 and takes place at the University of Hawaii at Manoa.

The CATS workshop series has clearly established itself as the premier venue for presentations and discussions of research bridging culture and educational technologies. Indeed, in the past few years, dedicated culturally-aware technologies, guidelines and methods have been proposed, and promising results are now emerging along with new exciting challenges for our community. CATS2014 is thus a great occasion to continue investigating the place of culture in the ITS research field, which it considers from the three following perspectives:

- Designing ITS systems to teach cultural knowledge and intercultural skills,
- Enculturating ITS systems (i.e. developing both pedagogical strategies and system infrastructure mechanisms that incorporate cultural features),
- Considering cultural challenges in the ITS research cycle, and ways to address them.

The scientific quality of CATS2014 was ensured by an interdisciplinary program committee of 31 members with strong expertise in relevant fields of interest (ITS, culture and technology). Special care was also put in ensuring geographic diversity within the program committee: PC members were from institutions distributed across 17 countries with equal representations of Asia, Europe, North America, and South America.

5 full papers have been selected by the program committee for inclusion in these proceedings. But CATS also intends to be a space for experience sharing, communication, and networking. Consequently, two discussion sessions are also organized in the context of the workshop. The first will encourage participants to move beyond the student model to discuss the influence culture may have on every one of the commonly agreed-upon parts of ITS systems, whereas the second will focus on investigating strategies to allow ITS to deal with cultural groups beyond a national definition (for example, those resulting from differences in socio-economic status or from a digital divide).

We are very grateful to all the individuals who made CATS2014 possible. We would like to thank the members of the program committee for their efforts, as well as the Steering Committee, the Workshop Chairs, and the Organization Committee of the ITS2014 conference.

June 2014

Emmanuel G. Blanchard, Isabela Gasparini, Amy Ogan, and Mercedes T. Rodrigo
The CATS2014 Co-Chairs
Organization

Organizers
Emmanuel G. Blanchard, IDÛ Interactive Inc., Canada
Isabela Gasparini, University of Santa Catarina State, Brazil
Amy Ogan, Carnegie Mellon University, USA
Ma. Mercedes T. Rodrigo, Ateneo de Manila University, The Philippines

Program Committee
Mohammed Abdelrazek, King Abdulaziz University, Saudi Arabia
Ivon Arroyo, University of Massachusetts at Amherst, USA
Ryan SjD Baker, Teachers College Columbia University, USA
Jacqueline Bourdeau, Téléuniversity of Québec, Canada
Ignacio Casas Raposo, Pontifical Catholic University of Chile, Chile
Júnia Coutinho Anacleto, Federal University of Sao Paulo, Brazil
Vania Dimitrova, University of Leeds, UK
Benedict DuBoulay, University of Sussex, UK
Colette Faucher, Aix-Marseille University, France
Monique Grandbastien, University of Lorraine, France
Rudiger Heimgartner, Intercultural User Interface Consulting, Germany
Tsukasa Hirashima, Hiroshima University, Japan
Seiji Isotani, University of Sao Paulo, Brazil
W. Lewis Johnson, Alelo Inc., USA
Stan Karanasios, University of Leeds, UK
Cher Ping Lim, Hong Kong Institute of Education, Hong Kong
Chee Kit Looi, National Institute of Education, Singapore
Phaedra Mohammed, University of the West Indies, Trinidad and Tobago
Roger Nkambou, University of Québec at Montréal, Canada
José Palazzo Moreira de Oliveira, Federal University of Rio Grande do Sul, Brazil
Elaine M. Raybourn, Sandia National Laboratories, USA
Genaro Rebolledo Mendez, University of Veracruz, Mexico
Matthias Rehm, Aalborg University, Denmark
Benjamin Nye, University of Memphis, USA
Marcelo Soares Pimenta, Federal University of Rio Grande do Sul, Brazil
Silvia Schiaffino, ISISTAN, Argentina
Riichiro Mizoguchi, Japan Advanced Institute of Science and Technology, Japan
Mei Si, Rensselaer Polytechnic Institute, USA
Bao Hui Zhang, School of Education, Shaanxi Normal University, China
Imran Zualkernan, American University of Sharjah, Unified Arab Emirates
Beverly Woolf, University of Massachusetts at Amherst, USA
# Table of Contents

The Role of Educational Technologies in Supporting Students across Competing Cultural Frames  
_Samantha Finkelstein and Justine Cassell_  
6

An Advisor System for Cultural Adaptation in Instructional Design  
_Isabelle Savard, Gilbert Paquette, Jacqueline Bourdeau_  
15

Adapting a Cognitive Tutoring Strategy for Mathematics in Latin America  
_Ignacio Casas, Jason Imbrogno, Sergio F. Ochoa, Adriana Vergara_  
27

Culturally Aligned Pedagogical Agents for Mathematics Education  
_Melissa-Sue John, Ivon Arroyo, Imran Zaalkernan, Beverly P. Woolf_  
38

A theoretical approach to the development of critical incidents for cultural training  
_David Gerritsen, John Zimmerman, Amy Ogan_  
51
The Role of Educational Technologies in Supporting Students across Competing Cultural Frames

Samantha Finkelstein and Justine Cassell

Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA
{slfink, justine, etc.}@cs.cmu.edu

Abstract. The design of educational technologies involves making decisions based in cultural assumptions and ideologies, and student learning may be improved when these design choices are made intentionally with the students’ own cultural context in mind. Despite this, the classroom systems in which these technologies are deployed have their own culturally-based contexts which are often at odds with the expectations from some culturally-underrepresented groups. We believe that educational technologies are uniquely positioned to be able to both leverage student culture to support learning, while also supporting students in applying the learned domain content across relevant cultural contexts, such as the classroom. In this paper, we present the motivation and design of a reflective gaming system within the multiple-cultures framework called CoDES which demonstrates one way educational technologies can be designed to help students and teachers mediate competing culturally-based expectations.

Keywords: culture, dialect, AAVE, reflective gaming, educational technologies

1 Introduction

Educational technologies have demonstrated substantial success at supporting students’ learning within specific well-defined domain areas (Koedinger, Anderson, Hadley, & Mark, 1997) though much of what students are expected to do in the classroom is heavily dependent on context, and is not nearly so well-defined. Throughout the school day, for example, students are required to be sensitive to a variety of spoken and unspoken classroom and societal contexts, and are consequently required to navigate between different cultural lenses which inform and inspire their language and behavior. Demonstrating the appropriate behavior at the appropriate time is critical to students’ academic and social success. To make this more difficult, in many cases, these different expectations for students’ behavior may be directly at odds. For example, the language valued among peers
may not be the language valued among teachers (Denny, 2012), and the behaviors most likely to promote student learning may not be the behaviors most frequently elicited within the classroom (Lemke, 1990). One grounded example which exemplifies these tensions is dialect, where many students may start school not fluent in the Mainstream American English (MAE) dialect which is associated with power and success in contemporary society (Lippi-Green, 1997), and often required in schools (Fogel & Ehri, 2000). For students who instead speak other stigmatized dialects of English (such as African American Vernacular English (AAVE)), there are additional learning challenges these students face as they work to make sure they are meeting their teachers’ expectations to use the right dialect at the right time (Champion, Cobb-roberts, & Stewart, 2012; Craig, Hensel, & Quinn, 2009; Denny, 2012; Lee & Fradd, 1996).

While there are successful examples of technologies that demonstrate alignment with either the cultural expectations of mainstream society (e.g., Koedinger et al., 1997) or the cultural expectations of underrepresented student users (Eglash, Bennett, O’Donnell, Jennings, & Cintorino, 2006; Gilbert et al., 2006), there are very few that allow students to explicitly navigate through and reflect on realistic representations of these multiple competing cultural frames within the same learning environment (see discussion in Henderson, 2013). Despite this, we argue that educational technologies are actually particularly well-suited to mediate competing culturally-based expectations head-on by providing teachers and students with a shared platform through which they can openly reference and discuss hard politically- and culturally-charged themes. We discuss the motivation for and design of a reflective game called CoDES Contextual Dialect use in Education Spaces designed in line with Henderson’s (2013) multiple cultures framework which we believe demonstrates one way that educational technology designers can support students in navigating across cultural frames. Through reflective gaming principles such as process over product and questions over answers (Khaled, 2014), CoDES is designed to support both teachers and students in developing more positive language ideologies, think about the places in which people code-switch between dialects, and give students an opportunity to iteratively try out different dialects for different contexts.

2 Related work: ideologies and interventions

Siegel (2009) accurately opens his review of the issues surrounding the presence of stigmatized dialects such as AAVE in the classroom by pointing out that nearly all teachers and educational administrators have the interest of their students at heart, and want to do what’s best for them (p. 36). For many teachers, this involves ensuring students are fluent in Mainstream American English (MAE) (Champion et al., 2012; Denny, 2012; Siegel, 2006), which is a dialect of English associated with power and success in contemporary society (Lippi-Green, 1997). Unfortunately, to reach this end, many teachers employ classroom practices which treat dialect as an obstacle to avoid, rather than as a potential
bridge to success, using techniques such as correction and public denigration in response to student dialect use without ever having explicit discussion about dialect directly (see review, Siegel, 2009). In some cases, students will be banned from using this dialect at all in the classroom, with teachers reporting that a correct science answer presented in AAVE is a wrong answer (Ogan et al., 2014, under review). Despite this, there is evidence that when students are given classroom materials in their native dialect, such as reading comprehension texts (Rickford & Rickford, 1995) or pre-recorded voice samples demonstrating science talk structures (Finkelstein, Yarzebinski, Vaughn, & Ogan, 2013), they result in improved student performance. More broadly, longitudinal studies show that classrooms which support students using their native dialect in addition to the standard demonstrate the greatest academic and social success (Wolfram, 1998; Thomas, 2002).

2.1 Culture in Educational Technologies

Within the field of Intelligent Tutoring Systems, educational technologies have increasingly been bravely taking up the challenge of considering cultural-sensitivity or cultural-congruence in the design of systems. In many cases, however, these systems focus on what Henderson (2013) would call soft-multiculturalism – references to a students’ culture in surface-level, often-stereotyped ways, such as by designing “token characters” who share the target demographics’ appearance or reference stereotyped cultural preferences like hair style or music (Eglash et al., 2006; Gilbert et al., 2006; Pinkard, 2001). Even notable examples which work to include deeper cultural references such as non-standard dialect use along with the more surface references such as CRITS (Mohammed & Mohan, 2010) do not offer students a space to reflect on these different cultural frames. While these systems may result in improved student performance, they risk stereotyping or trivializing a students’ culture (Cassell, 2009) and miss out on a strong opportunity to help the student reflect on how their culture is or is not distinct from what’s expected in the classroom and in society (Henderson, 2013).

Despite the concerns regarding these soft multiculturalist systems, however, our own work discussing these and other technologies with elementary school teachers indicate that while teachers are generally on board with soft-multiculturalistic surface-level references to a students’ culture meant to more strongly engage them in the domain, they feel much more negatively about inclusion of deep references to aspects of culture which are politically- or culturally-charged, such as modeling an African American character speaking AAVE during science class (full study described in Ogan et al., 2014, under review). Teachers were explicit with their concern that even though students might like interacting with a character who speaks like they do, it would be unacceptable to demonstrate that it was okay to speak that way in school.
The dangers of ignoring the classroom system: Teachers, not technologies, are the agents of change

Though culture is a strong contemporary example of a tension that exists between the design of educational technologies and the classroom system, it is not the first; indeed, only two decades ago the very inclusion of these technologies within the classroom were perceived by many teachers as challenging their role as distributor of knowledge (Bruenjes, 2002). Early analyses of how teachers responded to these systems found that while those who were optimistic about the benefits of these systems and believed they had efficacy to facilitate student learning saw these benefits come to light within their classroom, those which were skeptical or dismissive of the systems did not see the same benefits (Demetriadis et al., 2003; Ertmer & Ottenbreit-Leftwich, 2010; Harris, Mishra, & Koehler, 2009; Levin & Wadmany, 2005). In line with these results, and in response to the trend within educational technologies to “teacher-proof” systems such that they can’t interfere with how they are “supposed to be deployed,” (Robinson, 1990), Fisher warns that it is teachers, not technologies, who are the agents of change (Fisher & Frey, 2007).

In other words, while technology can be designed with the goals of social or pedagogical good, the support of the classroom system (e.g., teachers and administrators) is critical for the system to be able to reach these goals. We must work with, and not against, the classroom system to best support the students who exist within it. That being said, there is additional research which demonstrates that the very inclusion of technology within the classroom has the ability to help influence teachers’ perceptions of their roles in productive ways (Ertmer, 2010; Levin & Wadmany, 2005). The goal as we see it, then, is for designers of educational technologies to first be able to identify and understand some of the competing cultural frames which are relevant to the technology’s targeted learning goals and demographic. This understanding can inform the design of a system that can appeal to all relevant education stakeholders (e.g., students, teachers, parents, and administrators) by representing these learning challenges in a multiple cultures perspective to which each stakeholder can relate.

CoDES: Contextual Dialect use in Education Spaces

This call for situating educational technologies more cleanly within the classroom system is a substantial one, and we do not claim to have figured out the ideal way to have reached these goals. However, we present here a new approach to bring these goals to light which we hope can promote discussion about next steps in the field of ITS. We’ve worked alongside a group of videogame designers at Carnegie Mellon’s Entertainment Technology Center (LaForce, Pausch, & Marinelli, 2000) to create a game environment that can contextualize more traditional ITS approaches to introduce cultural reflection around other educational domains. The game environment is designed to support 3rd – 5th grade ele-
mentary school students in becoming more reflective about dialect expectations within different contexts.

In CoDES, students work in groups of two to play as the partners to Alex, a gender-ambiguous (female pronouns used here for clarity) African American child who lives on a research space base along with her mother (the captain) and other families of astronaut scientists. As Alex explains to the players, this means that sometimes she gets to talk informally with her mom, like when they’re in their apartment, but when they’re out in the control center, she has to speak formally. Everyone on the ship, including the captain, is shown speaking both AAVE and MAE based on the context, such as where they are or what they’re talking about, to demonstrate how different dialects may be appropriate in different situations. The characters in the system talk about language style openly, such as Alex saying she has to remember to speak formally when she records her final ideas at the end of the day “for the captain’s official log” so that all of her ideas can “sound professional,” but that she should “chat informally with the rest of the crew” so it’s easier for everyone to “work through their ideas.” As the students work through the ITS within CoDES, they are prompted to record their working ideas informally for the crew so that the crew members on the ship can keep working on the problem. When the student reaches the correct answer, they are prompted to record their final answer formally to be sent off for the captain’s official records. This results in a space where the two are able to record their scientific ideas in both informal and formal contexts, listen to those recordings, reflect on whether or not they believe them to sound appropriate for the situation, potentially re-record the sound file, and perhaps even talk with each other openly about dialect use within this process.

Khaled’s (2014) reflective gaming principles include a call for both process over product and questions over answers. In other words, to promote an environment where children are not told exactly what to do, but given an environment where they are expected to make choices and directly reflect on the outcome of these choices. We focused on producing an environment where children are explicitly asked to record their ideas “informally” with crew members and “formally” for the captain’s official log, which, for many children, is a translation experience they are expected to do every day in the classroom, but one which in the majority of cases there is no explicit environment in which to practice this skill by producing language and reflecting on the outcome. We made the design choice to not include direct feedback to students within the game about their dialect use within their recordings for a number of reasons. First, we are primarily interested in supporting the process of students learning to be reflective about code-switching than the product of fluent MAE. Additionally, there are existing difficulties of accurate speech recognition for children who speak non-standard English, and dialect use can be an incredibly sensitive topic for children. Instead, we chose to design this system to support two children playing at once, thus aiming to promote human peers as the mutual providers of feedback. We are enthusiastic about this idea because of the research supporting the effectiveness of peers at providing feedback to each other (Kowal & Swain, 1994; Swain,
Brooks, & Tocalli-Beller, 2002) particularly when it comes to sensitive topics which may make students feel defensive when coming from an *othered* source (Ogbu, 1999).

5 Discussion and conclusion

CoDES was designed because of the competing realities that students may demonstrate improved academic performance when allowed to talk in their native dialect or language, but that the majority of US classrooms require their students to speak using Mainstream American English. In CoDES, we have designed a system where students are given opportunities to engage with educational material and practice the act of reasoning through their answer in AAVE and producing a final answer in MAE. CoDES also exposes this classroom / societal expectations of MAE through a system that demonstrates everyone code-switching based on context, which we hope will reinforce to both teachers and students that both AAVE and MAE are valid ways of communicating. Finally, it includes a reflection opportunity in which students are prompted to consider whether the recording they’ve produced is appropriate for the given situation, which we believe may promote explicit discussion of dialect and code-switching within the classroom. Our next steps for this game are to include additional learning tasks within the game environment, and then perform a study evaluating how children record their science talk formally and informally within the game, compared to a control condition where children are just given the learning components without the context of code-switching.

There are, however, key challenges to systems that address topics such as dialect head-on. Our own work has demonstrated that children who are not speakers of non-standard dialects may mock technologies which speak in AAVE, even when other AAVE-speakers are also in the room. This in-vivo racism reflects a painful reality of what happens within the classroom, and it is critical that the systems we design undergo rigorous analysis before they are fully deployed in classrooms to be sure we are not promoting or enabling this type of unwanted behavior. Additionally, technologies that exclusively cater to the minority culture without acknowledging the greater classroom context may be harder for teachers to accept in the classroom (Ogan et al., under review), and may promote confusion among students who don’t understand why the game characters are able to speak in the forbidden dialect.

By designing educational technologies that demonstrate a multiple cultures framework and directly demonstrate values and expectations privileged in different cultural contexts, we may be able to best support student learning and transfer, and help to promote positive language ideologies for both students and teachers. We are inspired by the literature which demonstrates that we have the ability to change negative ideologies through deploying carefully-designed technologies which challenge the existing roles of classroom stakeholders, and believe that educational technologies designed within this framework
are in a position to ideally promote in- and out-of-classroom student learning and social good.

Acknowledgments. Enthusiastic thanks to all of our supporters, and our hard-working ArticuLites! This work is supported in part by Carnegie Mellon University’s Program in Interdisciplinary Education Research (PIER) funded by grant number R305B090023 from the US Department of Education, the NSF (# 0946825), the Grable Foundation, and Achievement Rewards for Collegiate Scholars (ARCS).

References


An Advisor System for Cultural Adaptation in Instructional Design

Isabelle Savard1,2, Gilbert Paquette2, Jacqueline Bourdeau2

1. Faculté de médecine, Université Laval, Québec (Québec-Canada);
   Isabelle.savard@fmed.ulaval.ca
2. TELUQ, Centre de recherche LICEF, Montréal (Québec-Canada)
   {Jacqueline.bourdeau, Gilbert.paquette}@licef.ca

Abstract. In an age of globalization, it is important to consider cultural variables in the instructional design process, but as [2], [3], [7] and [16] point out, too few tools and guidelines exist to assist instructional designers in this task. Also, as pointed out by [13], professors are not always aware or even informed of the existence of cultural variables in pedagogical practices. We used a Design-based Research (DBR) iterative approach to identify cultural variables in the instructional design professional culture and we modeled knowledge regarding these variables via a formal ontology on the basis of which we created a “Cultural Diversity” knowledge base. Our “Cultural Diversity” knowledge base brings together knowledge regarding five cultures. Our advisor system, through an executable assistance process for cultural adaptation, uses this knowledge to advise the instructional designer, who then proceeds to adapt a pedagogical scenario to a culture other than his or her own. In this article, we present the executable assistance process model which enables the staging of different software agents that advise the designer, who must carry out the cultural adaptation of his or her pedagogical scenario.

Keywords: Instructional Design, Cultural Variables, Advisor System, Ontology

1 Introduction

Distance education continues to this day to gain in popularity. A number of factors can account for this reality: 1) the growing need for continuing education; 2) the exponential increase in the number of learners with Internet access; and 3) the desire on the part of universities to reach a distant clientele with time constraints and to attract an international clientele. Students no longer limit themselves to the institutions in their countries and do not hesitate to study “abroad” while remaining in the comfort of their homes. Professors also have a greater number of opportunities to give their courses to learners from a culture
other than their own or than that of learners for whom their courses were originally designed.

These new contexts have given rise to new challenges. The resources used are not always adapted to the cultural realities of the learners for whom they are intended. As pointed out by [13], professors are not always aware or even informed of the existence of cultural variables in educational practices. It is therefore important to consider cultural variables in the instructional design process, but as [2], [3], [7] and [16] point out, too few tools and guidelines exist to assist instructional designers in this task. We have identified seventeen variables that need to be considered in instructional design decisions when adapting a course (or pedagogical scenario) so that it may be offered to learners of a culture other than the one for which it was first planned.

2 Cultural Variables

In order to identify the cultural variables to consider when adapting a pedagogical scenario, we reviewed the literature and designed a web-based questionnaire that we had instructional designers from various countries complete. In all, sixty-six respondents from eleven countries have begun the questionnaire; fifty-five of them have completed it.

The literature review, conceptualization process and questionnaire response analysis allowed us to target variables that we have grouped into three major categories: Values, Common Practices and Human Interactions.

The Values category was inspired by the work of [4], [6] and [10] and consists of the following variables: relationship with authority, tolerance for uncertainty, individualism/collectivism, approach towards time. The other two categories were inspired by the work of [12]. The Common Practices category consists of the following variables: learning aims, lesson plan, rhythm of learning activities, learning situations, pedagogical communication, cooperation-collaboration, detailed feedback, summative evaluation methods, results interpretation. The Human Interactions category consists of the following: teacher’s role, learner’s role, reaching learning goals, available learning resources.

This work allowed us to develop a formal cultural variables ontology and a “Cultural Diversity” knowledge base, which brings together knowledge regarding five cultures: Quebec, Mauritius, France, Belgium and Gabon. To populate the knowledge base, we used the cultures for which a minimum of five respondents had completed the questionnaire.

The ontology, knowledge base and executable process were developed with the Tele-Learning Operating System (TELOS), presented by [8] and [9].
3 Tools Developed in the TELOS Environment

TELOS is a multi-agent system allowing various actors (computer engineer, technician, instructional designer, professor and learners) to interact in a homogenous environment adapted to the needs of each and covering the distance learning tools production chain. Within this chain, the designer uses the platform to design a formalized and executable pedagogical scenario and to implement all the resources that will allow the learners and the professor to apply the aforesaid scenario and to leave a product or learning trace. The central principle of TELOS is therefore the aggregation of resources in an executable multi-agent scenario where the participants accomplish tasks by consulting and producing different types of resources.

TELOS consists of graphics or text editors developed in order to construct particular models: knowledge model, ontological model, scenario model (executable processes), competence editor. These different editors are based on a web interface and on the use and production of XML files. In this way, interoperability between the editors is ensured, and it is possible to make reference to an ontology or a competence referential in a scenario (executable processes) edited in TELOS. The TELOS ontology editor allows for XML files that meet OWL-DL international specifications to be imported and exported.

In TELOS, we first developed the cultural variables ontology on the basis of the variables that we identified and presented in Savard, Bourdeau and Paquette (2013). We then developed a formalized and executable assistance process for cultural adaptation (a TELOS scenario) intended for instructional designers who need to adapt their pedagogical scenario for learners from a new culture. This assistance process offers instructional designers various tasks that allow them to produce a pedagogical scenario adapted to the new culture. It encompasses different software agents that provide adaptation advice to the designer at predetermined times. These agents use the knowledge represented in the “Cultural Diversity” knowledge base, which is constructed on the basis of the cultural variables ontology. Figure 1 illustrates these tasks and the relationships between them.
In the following sections, we describe in greater detail each of these developments: the cultural variables ontology, the “Cultural Diversity” knowledge base, the assistance process for cultural adaptation tasks and the software agents.

4 Cultural Variables Ontology and “Cultural Diversity” Knowledge Base

The engineering work on the formal ontology alternated with the questionnaire design, the questionnaire response analysis and the variable identification. Following the identification and stabilization of the various concepts and their relationships, a more technical task consisted in making the ontology computer-interpretable by formalizing it as an OWL-DL domain ontology. This modeling work was validated on three occasions by a modeling expert.

All of the seventeen variables enumerated in Section 2 are represented in the formal ontology. As illustrated in Figure 1, it is on the basis of this ontology that the knowledge about each of the cultures is organized. All of the knowledge represented in this manner constitutes the “Cultural Diversity” knowledge base.

For each of the cultural variables rendered in the ontology, we depicted the possible values via instances, and we established a relationship with each of the cultures represent-
ed on the basis of the responses obtained in the web-based questionnaire. Figure 2 presents an example for the variable “learner’s role”, which we have limited to a comparison between two cultures: Mauritius and France. Of course, the data for the five cultures are always represented in the knowledge base.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Object</th>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner's Role</td>
<td>Mauritius Learner</td>
<td>Has role</td>
<td>Work in class, Ask questions, Discuss, Suggest resources</td>
</tr>
<tr>
<td></td>
<td>France Learner</td>
<td>Has role</td>
<td>Listen passively, Work in class, Discuss</td>
</tr>
</tbody>
</table>

Fig. 2. Example of an Object, Attribute and Value for the Variable “Learner’s Role” from the Human Interactions Category

Figure 2 shows that differences exist between the roles of learners in Mauritius and those of learners in France. In Mauritius, learners are in the habit of asking questions and suggesting resources, whereas this is not the case in France. Furthermore, in France, learners are used to listening passively to their professor’s lecture, unlike learners in Mauritius. The following sections explain how software agents are put to use in the cultural adaptation assistance process in order to highlight these differences and advise the designer on the adaptation strategies to adopt.

5 The Executable Assistance Process for Cultural Adaptation

In the assistance process for cultural adaptation that we have developed, the designer interacts with different software agents, including one specialized agent for each of the variables identified, which are the concepts of the ontology. The specialized agent advises the designer who needs to adapt to a new culture.

This process involves steps 2 to 5 in the method for cultural variable processing presented by [14], which consists of a total of seven steps:
Step 1 is considered as preliminary to the treatment of cultural variables and is completed for the time being with a questionnaire (external to the assistance process), adapted from the work of [11]. Steps 6 and 7 may be added to the system once we have implemented the prototype with a clientele of instructional designers in authentic situations. For the time being, the assistance process has been implemented in TELOS and has been tested on numerous occasions for steps 2 to 5.

The assistance process begins with an analysis phase during which the designer establishes a case of cultural crossing, compares the portraits of crossed cultures, outlines the scenario, if a pedagogical scenario is to be reused, and finally, evaluates the adaptation complexity of the scenario to be reused. By the end of the analysis, the designer will have received advice on the technical complexity and the pedagogical adaptation complexity from the software agents. For each of the two completed evaluation grids, technical and pedagogical, one of the following global adaptation strategies is suggested to the designer: translation-localization, contextualization, modularization or creation of a new pedagogical scenario (adapted from [2]). On the basis of this advice and the comparison of the portraits of the crossed cultures, the designer must decide whether to reuse the pedagogical scenario as is, adapt it or create a new one. Reusing a scenario as is represents the simplest use case, since it leads the designer directly to the end of the assistance process. Adaptation (Adapt) is the most complex case (Step 5 in the method presented in Figure 3). The creation (Create) of a new scenario may also prove to be complex if major differences exist between the two cultures. Creation and adaptation may both lead to the indexation (Index) of a scenario, which consists in describing the properties of the new scenario or the new version of the scenario modified on the basis of the cultural characteristics. Indexation is tied to the seventh and final step of the method illustrated in Figure 3.

All the comparisons (between the designer’s culture and that of the learners whom the designer wishes to address, and between the portrait of the pedagogical scenario to be reused and the culture of the learners to be addressed) are based on the concepts of the ontology (identified variables).

Figure 4 presents the interface offered by TELOS to the instructional designer or the professor (who is in the process of adapting a pedagogical scenario and who follows the designed executable assistance process).
The list of tasks that make up the assistance process and the progress bar for each of the tasks are given on the left-hand side of the screen. The bar indicates whether a task has been started, has been completed or is in the process of being competed. An example of a task that leads the designer to compare the portraits of the cultures involved is given on the right. The system uses the knowledge from the “Cultural Diversity” knowledge base to present the two portraits to the designer (Portr A – designer’s culture and Portrait B – learners’ culture) so that he or she may compare them. The designer may highlight the similarities and differences in a comparison template provided to him or her. This information makes it possible for the designer to evaluate the adaptation complexity.

6 The Software Agents

Software agents capable of advising the designer are placed in strategic locations in the assistance process. All the software agents have specific responsibilities tied to each of the identified variables or to the evaluation of the pedagogical or technical adaptation complexity. Figure 5 illustrates one part of the assistance process for cultural adaptation in which collaboration between the designer and a software agent (here, the one who is responsible for adapting the summative evaluation methods) takes place in order to accomplish the task “Develop the adaptation strategy for the summative evaluation methods”. This results in the associated adaptation strategy as the product, which marks the end of this part of the assistance process.
Rules such as the ones presented in the table below, which concern the variable “responsibility for available resources”, define each of these software agents. This variable is interested in the fact that in some cultures, learners are responsible for providing relevant resources (e.g., articles, videos, images, etc.), whereas in other cultures, this responsibility lies entirely with the professor, and the learners may be considered as relatively passive consumers.

---

1 Figure 5 was designed using the TELOS scenario graphics language editor. The suitcase with the “abc” icon symbolizes a string (here, strategies adopted for the adaptation of the summative evaluation methods). The human figure icon symbolizes a human intervener (here, the designer) and the computer icon symbolizes a software agent (here, the one who is responsible for the summative evaluation methods). The oval shape symbolizes a process. An oval shape marked with small feet symbolizes an elementary human action.
Table 1. Example of Rules behind the Software Agent “Responsibility-Available Resources”

<table>
<thead>
<tr>
<th>Use Case: Adapt: Adapting Human Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agent:</strong> AdaptResponsibilities AvailableResources</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of the Rule</th>
<th>If…</th>
<th>Then display…:</th>
</tr>
</thead>
<tbody>
<tr>
<td>AddResponsibilitiesResourcesLearner</td>
<td>ScenarioResponsibilitiesForResources==Professor imposes a selection AND CultureBResponsibilitiesForResources==Learner suggests resources</td>
<td>The learners whom you wish to address are in the habit of suggesting resources that they deem interesting or useful. We advise you to:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) find out about the roles and responsibilities of professors and learners from colleagues who work there, 2) reflect on the advantages and disadvantages of allowing the learners to suggest resources and try to reach a compromise, 3) try allowing the learners to suggest resources that you can validate. You can let the learners know that you are not used to this practice, but that you are prepared to give it a try.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If you insist on imposing your selection, clearly explain why you deem it to be indispensable.</td>
</tr>
<tr>
<td>DecreaseResponsibilitiesResourcesLearner</td>
<td>ScenarioResponsibilitiesForResources==Learner must contribute AND Role1LearnerCultureB!=Suggest resources AND Role2LearnerCultureB!=Suggest resources AND Role3LearnerCultureB!=Suggest resources AND Role4LearnerCultureB!=Suggest resources AND Role5LearnerCultureB</td>
<td>The learners whom you wish to address are not normally required to suggest resources, so you propose a scenario where they are required to do so. We advise you to make sure that they know how and where to look. If they do not, it is preferable to teach this first. If you do not have time to teach the basics of research, direct the students to resources that will assist them in developing these skills. You can go slowly and ask for fewer resources in the beginning and later increase the level of involvement on the part of the students. You can also begin by suggesting a variety of resources to them and asking them to select the most useful ones for them from your suggestions.</td>
</tr>
</tbody>
</table>

In all, there are nineteen software agents totaling over one hundred rules, such as those given in Table 1. Various pieces of advice may be displayed to the instructional designer according to the rule or rules applied and on the basis of the variables identified as requiring adaptation during the evaluation phase of the adaptation process. Figure 6 presents an example of advice displayed for the designer.
The designer must then decide whether to accept the advice, reject it or complete it. This decision must take the form of a character string placed in the space provided for this purpose in the TELOS environment. All the adopted adaptation strategies (characters strings) are compiled by the system to form a collection: the detailed and adopted strategies. This collection is then made available to the designer who uses it to modify his or her scenario.

Once the adapted scenario has been completed, the system requests that the designers store their adapted scenario in a resource repository. Finally, the designers must decide whether or not to index their adapted pedagogical scenario on the basis of the new cultural parameters. Indexing the scenarios on the basis of the cultural variables could facilitate the reuse of teaching and learning resources.

7 Conclusion

The variable identification is the outcome of a long and thorough reflective process that is well anchored in the literature, our review of which served for both the questionnaire design and the final selection of the variables. Countries were selected where the French language is spoken to better isolate cultural factors without adding linguistic differences. Genericity by superimposing the two factors would be the subject of another project. The limited number of questionnaire respondents made it impossible for us to establish clear and reliable portraits of the five represented cultures, but our response analysis allowed us to select certain variables and leave others out. We have therefore successfully targeted important variables. In a DBR approach, there is an assessment loop at each stage, not only at the end as in a sequential process. Thus, the different results (ontology, assistance process, etc.) of the thesis presented by [15] have been assessed. Other assessments may be considered with designers in authentic situations since we consider this prototype of an advisor system to be promising with modeling that could prove to be very useful to
the designer who must adapt a pedagogical scenario to the needs of learners from another culture.

It could also be improved by relating it to work such as that of [1], which deals with a high-level ontology of culture, as well as that of [5], which examines an ontology of teaching and learning theories. Indeed, it could be interesting to develop specialized agents with the task of evaluating coherence in the scenarios. At this time, the agents that have been developed concentrate on cultural variables in instructional design and do not verify pedagogical coherence. As such, an entirely incoherent scenario could be developed without the designer’s being warned. For example, an agent could intervene and comment on the incoherence of a scenario where the learner is required to be constantly active, but where the sole teaching method planned is the lecture.

8 References


Adapting a Cognitive Tutoring Strategy for Mathematics in Latin America

Ignacio Casas\textsuperscript{1}, Jason Imbrogno\textsuperscript{2}, Sergio F. Ochoa\textsuperscript{3}, Adriana Vergara\textsuperscript{4}

\textsuperscript{1}School of Engineering, Pontificia Universidad Católica de Chile, Santiago, CHILE icasas@ing.puc.cl
\textsuperscript{2}Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA jpi@cmu.edu
\textsuperscript{3}Computer Science Department, Universidad de Chile, Santiago, CHILE sochoa@dcc.uchile.cl
\textsuperscript{4}School of Education, Pontificia Universidad Católica de Chile, Santiago, CHILE adrianavergara@uc.cl

Abstract. Cognitive tutoring technology has been extensively used in the United States, and it has proven effective for increasing students’ mathematics skills. However, reports of the usage of this technology in other regions are scarce. Given the urgent need to improve public education in Latin America, we have adapted and experimented with this technology in three countries (Chile, Mexico, and Ecuador) to learn if it is effective for mathematics learning at the middle school level (5th - 8th grade). We discuss difficulties, assessment, and lessons learned in the process of implementation.

Keywords: Intelligent tutoring systems; interactive learning environments; teaching/learning strategies; country-specific developments; evaluation of CAL systems.

1 Introduction

The purpose of this study\textsuperscript{2} is to explore strategies needed to successfully deploy and use cognitive tutoring technology in public schools of Latin American (LA) countries. We develop and experiment with a pedagogical framework that, considering scarce technological resources, takes advantage of personalized learning activities in the computer lab and collaborative (paper and pencil) strategies in the classroom. Even though the ultimate goal is to improve math learning among students, our methodology is centered on teacher training where the new technology-supported strategies are socialized and adapted to local idiosyncrasies. We want to make sure teachers feel motivated and are willing partici-
pants-leaders of the required change process. After training, we provide constant support and follow-up of the implementation in the classroom and lab.

Our research focuses on the teaching-learning strategies that can be enhanced by the technology (assuming that the technology has been correctly deployed\(^3\)). This involves substantial change in the teacher’s attitude, motivations, activities, and plans. The teachers need training and time for planning the new classroom-lab strategies. It involves major changes and it is a complex task.

1.1 Cognitive Tutor Technology

Following the theoretical principles developed by Anderson [1], [4], cognitive tutoring software (CT) was built at Carnegie Mellon University and is maintained and operated by Carnegie Learning Inc.\(^4\) Each student has a personalized “problem-solving” space (account), with just-in-time feedback and detailed tracking of his progress [11]. CT follows a personalized self-paced approach, allowing students to sequentially tackle progressively more difficult tasks. It tracks students’ progress in real time as they answer questions, ask for help and solve problems. It provides personalized feedback and hints when errors are made in key points [6].

Cognitive tutors have shown considerable potential, and evidence in the open literature indicates that they are effective in improving mathematics and science problem-solving skills [5], [10]. Specific mathematics cognitive tutors have been used in large school systems (primary/secondary level) in the United States, including Los Angeles and Chicago, as well as in rural areas [2].

1.2 Cognitive Tutor Strategies

The main objective of the CT software is to provide each student a unique, enriched environment where he/she can interact with the system in the solution of specific problems. Multiple graphical representations can be explored by the student for creative thinking practice [3], [9], [13].

A problem is presented and a solution is requested. Instead of jumping to the final solution, the software provides step-by-step scaffolding [7]. This divide-&-conquer strategy asks specific questions, from more easy to more complex, so the student can advance at his/her own pace in the solution of the problem.

The first question in each problem presented to the student is always related to the appropriate reading of the problem narrative. This can be particularly helpful in LA public schools, where student’s problem solving abilities are many times hindered by their lack of reading skills. The next questions (posed by the software) guide the student in the solution of the problem\(^5\).

\(^3\) We will see later on that this assumption is, regretfully, not correct in some cases of our experimentation.

\(^4\) Cognitive tutoring technology is a trademark property of Carnegie Learning Inc.

\(^5\) There is extensive literature with thorough description of CT technology ([2], [3], [4], [6], [10]).
The student gets feedback (positive or negative points in a roster of skills to be achieved) whenever he/she answers questions within a problem. The student gets immediate feedback on the “skill-o-meter” at any time during his work session [8]. Based on the skill-o-meter we have developed a web-based tool that provides teachers a complete view of student’s progress, both at an individual and full class scale. The teacher knows at any time where individual students are standing and thus can give them reinforcement on the topics of struggle [12].

1.3 Purpose of the Study

The broad objective is to understand how an educational technology that has been successfully used in developed countries can be adapted, contextualized, and integrated in a human and social system of a developing region, specifically the public educational sector of LA countries. In this case, the teacher, classroom, school, district authorities, and parents are part of the social system. The challenge is to both redesign these social arrangements (e.g. how teachers teach mathematics) and the technology system to fit and improve students’ capability in mathematics.

The key questions are as follows: can we motivate teachers in LA to change their teaching practices and take full advantage of the potentials of the CT technology? Once the teacher is motivated, does he/she get the resources and support from the school authorities to change his/her practice? Can they induce students to work with the CT technology and improve their math skills?

This paper describes the MCT (math cognitive tutor) initiative: an experimental study performed in public middle schools (5th to 8th grades in a K-12 system) of several LA countries with a teaching strategy based on the CT technology for math learning. A two-face implementation was performed during 2009-2010 in a number of public schools in Chile, Ecuador, and Mexico. Additional extended implementation continues in Chile (2011-2014).

2 Methodology

Building from the experiences in the United States, the MCT initiative seeks an important innovation: the definition and application of new teaching strategies that, based on the CT technology, are adapted to the LA educational idiosyncrasy. This starts with the negotiation of change strategies with the district and school authorities. It follows with the involvement of teachers on training and instructional design (b-learning) courses based on the CT. It culminates with the implementation of the technology-supported strategies in the math classroom.

---

6 The plan was initially designed for experimentation in four LA countries, but one of them could not fully implement it mainly due to lack of technological resources and support from authorities.
Once the school authorities have provided their support and there is sufficient infrastructure and resources in the school (compliance of “entrance” requirements), teacher involvement is the most critical issue in the implementation plan. The goal is to achieve high motivation and strong commitment of the teachers towards the new technology-based strategies.

In addition to the definition of the pedagogical strategies, we took an English version of the software contents and, considering cultural and idiosyncratic differences, transformed it into a Spanish version, coming up with a common structure for all the participating LA countries at the middle school level.

Even though the underlying theory and structure of the software tool remains the same as in the English version, contents and exercises were localized to the local cultures.

**General Implementation Plan.** The project’s plan focuses on training, motivating and empowering teachers (as leaders) in the use of a CT technology that has been adapted and localized for LA schools. The principal components of the plan are:

- Design of the strategies and implementation, negotiated with corresponding educational authorities.
- Adaptation of CT software modules and contents.
- Specialized training for math teachers and design of instructional units.
- Local implementation in selected public schools of different LA countries within a two stage pilot program.
- Creation of local learning communities and a repository of technology enhanced learning materials for mathematics.
- Assessment data collection and evaluation study.

An initial planning-design stage was accomplished during 2009 and training-implementation activities were performed in 2009-2010 in primary/secondary schools of Chile, Ecuador, and Mexico: over 730 students and 24 teachers participated in the study. A second experimental study is currently underway on 30 public schools in Chile.

**Design.** Public schools (as opposed to private) were chosen because of the enormous quality gap between the private and public educational system in the LA region. Private schools have many more resources and offer little or no opportunities to the socially vulnerable population. In each country, a district/municipality with a vulnerable population was selected which was willing and able to participate in the study. Within that district, treatment schools were randomly selected among those that did have the resources (computer labs, time for teacher training, and support from their authorities). (Some of the selected schools did drop off during the implementation due to lack of motivation,

---

7 Control schools and drop outs (including teachers and students) are not considered in these figures.
amongst other reasons.) Control schools were then selected within the chosen district/municipality.

**Software Adaptation.** This component focuses on the development of 14 lessons (comprising 54 instruction modules in the MCT software) of pre-algebra fundamentals used over a five month period of implementation in the 5th–8th grades. Activities at this stage included:

- Revision of curricula and assessment of teaching strategies used in math in each country.
- Development of CT modules and supporting documentation (teacher’s guide, student book, and exercises).
- Adjustment of cognitive tutoring modules to specific school contexts. This included changes in examples presented or language used, localized to the LA countries.

**Teachers’ Training.** Training teachers on the new pedagogical strategies and technology was of paramount importance. This included negotiating the integration of CT modules within the existing curricula, defining corresponding learning activities, learning self-paced pedagogies and methodologies, and monitoring and assessing students. Part of the training endeavor was devoted to working with teachers in the design of instructional units. Specialized seminars in technology management and planning were also provided to school directors. Additional training was given to technical support staff in schools.

**Pilot Implementation.** With the purpose of revising the strategies, class/lab methodologies, and contents, a small scale pilot with one or two schools per country and limited MCT lessons was first conducted during 2009 while other MCT modules were being developed. Afterwards, a full scale study was conducted during 2010. The following activities were performed during the extended study (second stage):

- Revision of class and lab infrastructure at each treatment school, to make sure there was one computer per student, appropriate local connectivity, and Internet access.
- Teachers were provided with the corresponding support documentation.
- Email, chat, and phone support was provided to teachers during implementation.
- On a weekly basis, the CT reports (with student and course progress) and suggested remedial actions were discussed with the teachers.
- Random monitoring of classes and labs was performed periodically.

**Local Learning Communities.** This component sought to increase the capacity of math teachers to use new forms of ICT technology to motivate students to pursue scientifically oriented careers. To do so, the project included the creation of local learning communities within and between the schools, thus stimulating the exchange of experiences and best
practices. A repository of best practices was created and used to promote knowledge exchange among teachers.

**Assessment Study.** This component focuses on the assessment of project-induced changes in schools, teacher practices, and student performance in math. Pre-tests and post-tests (pencil-paper based, similar format to national tests) were conducted at the beginning and at the end of the implementation, both at treatment and control schools. Additionally, surveys and interviews were conducted to learn about motivations and attitudes (towards the MCT system) of students, teachers, and school authorities.

### 3 Results and Discussion

The objective of the study’s evaluation process was to determine the effectiveness of the use of a CT system to improve skills in mathematics for 5th to 8th graders in public schools in selected LA countries. However, we also wanted to understand the motivations (and lack thereof) of teachers, students, and authorities to change their teaching-learning strategies. Before discussing the assessment results, we want to mention some culturally-oriented issues that impact in the implementation of the technology.

#### 3.1 Culturally-Oriented Issues

There were several cultural challenges that greatly increased the complexity of our endeavor. These issues, common among the participating LA countries, can be grouped in the following categories: authorities, teachers’ attitudes, and technological deployment.

With respect to educational authorities, the main problem refers to the complexity of the relationship between the high level (state or municipality) authority and the school principals. In our case, most municipality authorities would gladly support the MCT initiative, but would not provide additional resources for the schools. Municipality authorities could change very rapidly for political reasons, providing several discontinuities to the project. In some cases, school authorities would willingly participate but could not motivate their teachers to do it.

When asked, all teachers would agree about the need of change in their teaching strategies. However, it was very difficult to really implement this change in the classroom. Most young teachers would devote plenty of time to training in the new technologies and for change in their methodologies, as opposed to more experienced teachers. In general, we noticed that there is a widespread lack of confidence towards the real value of new technology-based strategies.

With respect to technology deployment, even though municipality authorities would claim they provided a sound infrastructure, there were many cases were we had to provide local servers and networks because of poor Internet bandwidth and lack of technical assistance for the setup of computer labs.
3.2 The Sample

Table I shows the number of participating (treatment and control) public schools, students, and teachers per country. As mentioned before, districts were selected by their willingness to participate and fulfillment of basic “entrance” requirements.

<table>
<thead>
<tr>
<th></th>
<th>Treatment (no attrition)</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Schools</td>
<td>Classrooms</td>
</tr>
<tr>
<td>Chile</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>Ecuador</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Mexico</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>17</td>
<td>24</td>
</tr>
</tbody>
</table>

In general, the selection of the participating districts was a somewhat difficult process. It is obvious that without full support and involvement of the district authorities, the implementation was impracticable. There were some initially invited districts that were necessary to discard due to their lack of real involvement. Finally, only one district per country participated in the project. All schools within a district were invited to participate, but only a few of them decided to experiment with the MCT system.

For instance, the selected Chilean district had 25 schools and only 17 of them were willing to participate. Of those 17, ten were assigned to treatment and seven were assigned to control. During the implementation process, three of the treatment schools dropped out for different reasons: problems with infrastructure, lack of involvement in training, reluctance toward teaching changes, and lack of support from school authorities.

Due to the training process - conducted locally in each country following the guidelines of a global strategy - most participating teachers were enthusiastic and willing to adopt the new strategies and technology. A few teachers (about 20% of initial participants) didn’t have enough time to complete the training. The later ones constituted dropouts from the implementation and in some cases the school as a whole could not participate. Attrition schools and teachers were not included in the assessment study.

3.3 Improvement of Math Skills

A large amount of collected data has been extensively analyzed. We present here solely observations from the data collected in Chile\textsuperscript{8}. An analysis of quantitative data (pre and post tests) shows the following:

- Students in treatment and control schools have statistically equivalent pre-test scores.
- Post-test differences between the treatment and control schools indicate a statistical improvement in math scores for those who used the MCT.

\textsuperscript{8} Data from Mexico and Ecuador shows similar tendencies as the Chilean data.
There is a variance in math performance among treatment schools.

The primary student achievement outcome measure used in this study comes from two comprehensive, grade-level pre-algebra tests given to all the students (treatment and control). Both instruments were measured with the Alfa-Cronbach test, giving a reliability of 0.77 for the pre-test and 0.80 for the post-test.

One test is given near the beginning of the implementation. The other is given five months later, after the implementation with the MCT system in the classroom and lab. Both exams consist of 44 multiple choice questions. The math material focuses on pre-algebra concepts, as does the software used by the treated students.

Our research team compiled the exam scores and difference scores from student level results on every question from both exams. We know whether a question was answered correctly (1), incorrectly (0), or left unanswered (blank). Adding up the correct responses for each student for each exam yields the total score for each separate exam. There was no added penalty for incorrect answers. The test outcome measure considered is the difference score from the exams. The difference score is measured at the student level by subtracting the initial exam score (pre-test) from the final exam score (post-test). A positive difference score means that a student scored higher on the second exam, while a negative difference score denotes a higher score on the initial exam. Unanswered questions were considered incorrect.

Comprehensive statistics on the difference scores are shown in Table II when we aggregate across the schools of similar type (treatment vs. control). Across nearly every measure shown in the tables, treatment students have a greater difference score value than control students. In short, a quick glance at Table II supports the notion of a positive overall treatment effect.

<table>
<thead>
<tr>
<th>Sample</th>
<th>All Students</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>788</td>
<td>388</td>
<td>400</td>
</tr>
<tr>
<td>Average</td>
<td>0.28</td>
<td>0.95</td>
<td>-0.37</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>5.17</td>
<td>4.99</td>
<td>5.27</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.18</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td>Maximum</td>
<td>19</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>Minimum</td>
<td>-22</td>
<td>-15</td>
<td>-22</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mode</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>-3</td>
<td>-2</td>
<td>-4</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

An ANOVA test reveals significant difference in difference scores\(^9\) (Chile: \(F(1,786) = 12.99, p = 0.0003\)) between the treated and control populations. The 95% confidence

\(^9\)“Difference in difference scores” measures the improvements of treated students over control students, using the difference score (post-test minus pre-test) as the outcome variable of interest.
interval from a two-sided t-test is (0.60, 2.03) in Chile. The point estimate of the difference-in-difference scores (1.32 in Chile) is approximately 25-30% of the difference score standard deviation across all students within the country, a very substantial improvement. The confidence interval (CI) on the test of whether the sample average from the treated and control groups are the same does not cover zero at the 95% level.

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.01</td>
<td>1.29</td>
<td>3.29</td>
<td>0.85</td>
<td>0.37</td>
<td>-0.25</td>
<td>2.23</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>5.09</td>
<td>5.40</td>
<td>5.52</td>
<td>5.27</td>
<td>4.54</td>
<td>5.08</td>
<td>4.18</td>
</tr>
<tr>
<td>Std Error</td>
<td>0.53</td>
<td>0.81</td>
<td>1.20</td>
<td>0.65</td>
<td>0.57</td>
<td>1.04</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Table III shows the difference scores by school (treatment only). There is significant variation even within treatment schools. Some even show negative difference scores, meaning that, on average, students performed better on the pre-test than the post-test of similar material. We hope to explain these large differences in subsequent implementations by considering MCT data and student, teacher, and school characteristics in the evaluation.

With these quantitative results, we believe we have answered the initial question about the feasibility of building practical cases where, with an innovative technology-supported strategy, we can improve student’s math skills in spite of shortcomings such as lack of resources, teacher time, teaching skills, math expertise, and initial students’ reluctance towards mathematics.

### 3.4 Motivations and Attitudes of Students and Teachers

As we mentioned in the introductory section, in this study we wanted to learn if we could motivate teachers in LA to change their teaching practices and take full advantage of the potentials of the CT technology. With respect to the students, we wanted to learn about their motivations and attitudes towards mathematics before and after using the MCT system. Some observations gathered from the Chilean schools’ surveys show that:

- A high percentage (78%) of treatment students was satisfied with their improvement of math performance due to the use of the MCT system.
- After implementation, a high percentage (67%) of treatment students increased their motivation toward learning math.
- After implementation, 68% of students felt more certain about their abilities to solve math problems and decreased their perception of math as a very complex discipline.
- 81% of students view the MCT system as a useful tool that substantially helps their learning process.

A positive difference in difference score means that treated students will improve their test scores more than control group peers.
• 85% of students were satisfied with the usability of the MCT software; 78% of students want to continue using the software tool in their math classes; 88% of students are motivated by the use of computers in their classes.

• 87% of students regarded the teacher as a helpful guide in the learning process using the MCT system.

• 82% of students would like to use the cognitive tutoring system methodology in other subject areas.

• A large percentage (75%) of teachers views the MCT system as a useful and effective mechanism that empowers their teaching process.

• 100% of participating school authorities (principals and area directors) considered the MCT system a useful tool for the teaching and learning of math.

• 100% of participating teachers and authorities would like to continue using the MCT system in the future.

All these qualitative observations from the Chile surveys are very similar to the ones collected for Ecuador and Mexico.

4 Conclusions and Future Work

This paper describes an innovative educational implementation based on cognitive tutoring technology. The main objective of the initiative has been the engagement of teachers in a change process that brings improvement of math learning in public schools of LA countries, specifically Chile, Ecuador, and Mexico.

Assessment results are encouraging and show that math learning in LA can be positively impacted with the use of cognitive tutoring technology. A necessary requirement is the involvement of motivated teachers that change the traditional passive classroom methodologies. With support from their teachers, students can get highly motivated to work with the software tools and the math contents. Students are encouraged by the personalized feedback and scaffolding aids provided by the CT. They perceive sustained progress in their learning. Teachers feel empowered with a richer teaching environment and use the feedback tools provided by the system as a guide for remedial actions.

According to our surveys and interviews, 100% of drop-outs were due to lack of resources (preparatory courses demand a significant amount of time and effort from teachers and authorities). The fact that 100% of participating teachers and authorities (not including drop-outs) want to continue using the MCT system in the future is an encouraging result that shows motivation and willingness to change.

The MCT system shows promising results. Currently, a new initiative is underway in 30 public schools of Chile, in urban and rural settings.
References

Culturally Aligned Pedagogical Agents for Mathematics Education

Melissa-Sue John¹, Ivon Arroyo¹, Imran Zualkernan², Beverly P. Woolf³

¹ Social Sciences and Policy Studies, Worcester Polytechnic Institute
  melissasuejohn@gmail.com, iarroyo@wpi.edu
² American University of Sharjah, United Arab Emirates
  izualkernan@aus.edu
³ School of Computer Science, University of Massachusetts-Amherst
  bev@cs.umass.edu

Abstract. This research investigates the impact of matching pedagogical agents to students’ social identities and questions whether improved learning and motivation can be produced. Pilot testing of children from three different countries who designed math avatars revealed that children overwhelmingly drew same gendered avatars, varied in the level of details of characters, and girls were significantly more likely to provide happy faces, and boys were evenly split between happy and neutral faces. We intend to use children’s depictions to design avatars to improve learning and reduce the achievement gap. Through iterative development and evaluation of cultural companions with dialectic and communication features, we plan to study Hispanic and Black students’ preferences to work with culturally and gender aligned characters, and to examine affective and motivational outcomes. This research moves towards generation of highly individualized, acculturated, pedagogically sound and accessible instruction, and will provide prescriptive principles about matching individuals’ needs.

Keywords. affect; culture; gender; mathematics education; minority students; motivation; pedagogical agents

1 Motivation and Prior Research

A wide gap exists in science, technology, engineering and mathematics literacy between low income and ethnic minority American students and their Caucasian-American peers, as highlighted by American reports, see Figure 1. The achievement gaps widens as students advance in school [1][2][3]. Our research investigates whether culturally-aligned learning companions, embedded in digital environments, can engage students, capture their interest, and be effective at transmitting motivational and cognitive support, and
thereby assist in reducing the achievement gap for underrepresented students. We are investigating the role of culture in learning and aim to go beyond modeling surface level traits such as color [4]. In doing so, we provide tutors that deliver adaptive instruction to students, an approach derived from decades of research indicating that students learn best with individualized instruction, tailored to their cognitive skill, affect, and gender [5][6][7][8][9]. More concretely, this research involves iterative development and evaluation of cultural and affective learning companions that have shown promise. Specifically, Caucasian versions of these characters are beneficial at improving affective experiences and motivation for mathematics problem solving and learning [10]. With these companions, students have reported lower levels of frustration, anxiety and boredom, at least when gender-matched with the companion. Previous studies show that students learn best from teachers whose race/ethnicity is matched to themselves [11][12], from teachers with similar accents to their own [13], and when they work with a partner who speaks their native language [14].

African-Americans, Latinos and Math Performance. A considerable gap exists between performance of Caucasian and African-American students at all ages (see Figure 1

![Figure 1: Racial/Ethnic gap in mathematics learning in the USA. Caucasian students outperform African-American students in mathematics K-12 standardized tests (top) and also outperform Hispanic students (bottom) for middle school students.](image-url)
and NAEP\(^{10}\)). This gap existed between 1973-2004 at all age levels (9, 13 and 17-year-olds), and the gap between these groups has not changed at all according to a 2009 report [15]. The effect size gap between Caucasian and African-American students at each of the ages approximates 1 sigma. Similar patterns are recorded for Latino students, see Figure 1. Analyses of grade 8 math achievement test data indicate that Caucasian students scored higher than both African-American and Latino students and Latino students score slightly higher than African-American students [16]. Research towards finding the underlying causes of these achievement gaps (e.g., [17][18][19][20]) has suggested multifaceted causes: personal (e.g., lack of motivation), social (e.g., parental involvement) and structural (e.g., curricula, racial discrimination). One explanation for the structural factors is that the socio-cultural aspects are not aligned to the curriculum. Teachers try to align to social realities of students. For instance, [21] explains ways a variety of teachers adapt to socio-affective realities of Hispanic students, involving their traditions as a way to catch and hold their students’ interest.

**Culture.** Culture refers to a complex construct that includes a full range of learned human behavior patterns including a person’s knowledge, language, belief, art, law, morals, custom, and any other capabilities or habits acquired as a member of society. It is a powerful human tool for survival, but a fragile phenomenon, as it is constantly changing [22]. When people speak of Italian, Samoan, or Japanese culture, they refer to the shared language, traditions, and beliefs that set people apart from others. In most cases, people who share a culture were raised by parents or caretakers from that culture, i.e., it was passed on from one generation to the next.

Being raised in a culture does not necessarily mean that a person completely adhere to that culture. Students might resist, rebel, etc. Therefore, we seek to discover an equation between students’ original culture and their preferences for subsequent cultural identities. We realize that preferences of majority populations in various countries do not necessarily apply to minority populations in the same country. A culture is rarely homogeneous, rather often mitigated. For example, while rooted in tradition, due to its multi-ethnic and multi-lingual character, and because the Indus valley has served as a gateway to India, people living in what is now Pakistan have constantly dealt with and absorbed cultures of their invaders; the last being the British. Consequently, it is not surprising that girls being brought up in even in conservative environment are exposed to western movies and T.V. and that this represents one dimension of their cultural identity. It would, however, be interesting to see whether they respond differently to avatars that are purely from the Pakistani vernacular as opposed to foreign ones.

**Communication and Dialect.** Culturally-sensitive digital companions are often based on modeling surface level traits such as skin color, ignoring deeper cultural phenomena associated with communication [4][23][24][25]. An exception is Finkelstein and col-

---

\(^{10}\) The National Assessment of Educational Progress (NAEP) is the largest nationally representative and continuing assessment of what America’s students know and can do in various subject areas.
leagues [26] who manipulated different dialect patterns in digital, including African-American Vernacular English (AAVE) and Mainstream American English (MAE). The results with 3rd grade students showed that: (1) students demonstrated improved science talk after listening to science explanation from a peer digital companion, and (2) this improvement was greatest among AAVE-speaking children with digital companions that spoke in AAVE. More specifically, this research studied whether listening to a digital companion’s science talk recordings increased students’ likelihood of producing on-task science contributions and scientific integration (ecological analogies, functionality, and prior knowledge) between their first and second science recordings, regardless of condition. Across all students the number of on-task science contributions and the amount of reasoning significantly increased from the first to second science recording. Results showed a significant main effect of science recording, with students increasing their production of strong arguments after hearing a digital companion in AAVE. This research motivated other researchers to experimentally compare the effects of cultural differences, including dialect-switching patterns in companions, an appearance that is consistent with students’ expectations for characters, and design of other cultural elements, including objects and familiar places. Finkelstein and colleagues [26] suggested several explanations for their result including: 1) a reduction of cognitive load for students when communicating in the students’ native dialect, since students fluent in the mainstream dialect may expend less effort during a learning task or are better able to demonstrate learning if they feel comfortable producing it in their native dialect; and 2) students felt an increased rapport, or sameness, with the agent who spoke in their own dialect, especially given that students who spoke with an AAVE-speaking agent spoke more loudly, more quickly, and with more pitch.

2 Intelligent Tutoring Systems and Pedagogical Companions

If computers are to interact naturally with humans to support learning, they should express social competencies and act like efficient and empathetic teachers and mentors. Some researchers have begun to explore this possibility in advanced learning technologies communities (e.g., [27]). It is understood that a localized solution (e.g., adapting software to a specific culture) runs the risk of being as offensive or irrelevant as the original version if a student does not belong to the represented cultural group. More than just cultural knowledge is required to relate to students; a sound understanding of environmental context is critical [28] [29].

For instance, gendered pedagogical companions have produced large differences in student motivation, affective experiences and degree of learning, depending on matching/mismatching the gender of the companion and student [27]. Likewise, recent research has explored how social factors play a role in learning technologies, e.g., a simple manipulation of the interface (changing the help button to say “working together”) and a brief
introduction, suggesting that students think of the software as a “teammate,” produced significant differences in student engagement with the system and improved help seeking behaviors [30]. These results are consistent with the idea that people seem to relate to computers in the same way they relate to humans and some of these relationships are very similar to actual social connections [31]. For example, students continue to engage in frustrating tasks on a computer significantly longer after an empathic digital response and they have lowered stress levels after receiving an empathetic message from digital characters [32]. In terms of actual learning outcomes the type of learning companion matters significantly. In one study students recalled more information when interacting with an artist agent compared to scientist agent while using a nanotechnology tutorial [33]. Moreover, students report reduced frustration and more general interest towards problem solving when working with gender-matched characters that emphasize the importance of perseverance/effort [34]. If such manipulations can influence students, it is very likely that more important cultural features would have a similar impact on students.

We believe that a danger exists of introducing stereotypes when researchers begin with a top-down theory, or from a researcher’s own understanding of what race, ethnicity and cultural background represent. On the other hand, we suggest a bottom-up approach by first identifying students’ own beliefs and depictions of their cultural experience. Having children draw as a way to mirror their preferred companions is a common technique used in psychology. Research into children’s drawings has focused on three main areas: (a) the internal structure and visual realism of children’s depictions (e.g., [35]); (b) the perceptual, cognitive, and motor processes involved in producing a drawing (e.g., [36]); and (c) the reliability and validity of the interpretation of children’s drawings (e.g., [37]). Very young children produce simple scribbles,
and later demonstrate representational intentions. With maturation and increased dexterity, children draw objects as they envision them rather than as they are actually perceived. Drawings of the human figure can also reflect a child’s social world. La Voy and colleagues [38] explored the idea that children from different cultural backgrounds may represent cultural differences in their drawings, because culture permeates a child’s representations of people. Differences across nations indicated that American children drew more smiles than Japanese children, who in turn drew more details as well as larger figures [38]. Similarly, Case and Okamoto [39] showed cultural differences between Chinese and Canadian children’s drawings. These findings suggest that children’s drawings not only reflect representational development but a child’s understanding of self and culture as well.

Having students draw characters as a way to tap into their minds and establish their expectations of pedagogical characters is an increasingly common technique, and has particularly been implemented for learning systems for mathematics education. For instance, Grawemeyer and colleagues [40] had participants within the autism spectrum express and externalize their individual ideas for an educational pedagogical agent for a mathematics educational game, and to combine their individual ideas with the ideas of others in a small group. Students created their own designs and also studied other students’ drawings, eventually creating a common prototype.

3 The Pilot Study

Our pilot study invited children from different countries to draw avatars (or pedagogical learning companions) for a mathematics game, see Figure 2. The goal was not to ask for complex representations, but instead, and similar to La Voy and colleagues [38], to explore cultural differences that are important factors to consider in the authors’ development of pedagogical agents. We understand that preferences of majority populations in various countries may or may not apply to minorities in the same country. However, we recognize that minority populations are often first or second generation immigrants from other countries. We began with several hypotheses based on gender and nationality for dress code, language expressivity, and affect. Specifically, we predicted that children would draw avatars of their own gender (H1) and dressed similarly to their national customs (i.e. clothing appropriate for their religion or geographical location) (H2). We expected girls to have more positive affect in tone and appearance than boys (H3), replicating findings from Arroyo and colleagues [34] and we expected that formality of a proposed software response would match the formality of each student’s cultural norms (H4).
Students were asked to help to design this math game. They were asked to supply a digital companion (left) and textual response that the system might provide to a child who got the answer wrong.

The children were not asked about the types of math games that had experienced earlier. However, Western and American television programs and movies are widely viewed in Pakistan and Argentina; children’s books often include western and American stories and characters. In addition, computer games are widely available in Pakistan and are played by children who are able to afford the computers and gaming consoles.

**Participants.** Seventy one children from North America, Argentina, Pakistan, and Jamaica, aged 7-10, were recruited from schools and neighborhoods which we had conveniently sampled. We wanted at least a few different racial backgrounds involved. Socio-economic status was middle and middle-high, in each country, with the exception of Pakistan students that came from three different schools, one low, one middle and one middle-high socio-economic status, in approximately equal amounts each Jamaican sample includes middle class students who pay to go to school. The Pakistan students come from three schools: 1) Balakot a semi-rural community located on the Himalayan earthquake fault line; the children represent the lowest socio-economic background in Pakistan and the parents are day laborers or migrant workers; 2) Lahore is a large urban center and the children belong to the lower middle-class urban classes of Pakistan; 3) Peshawar (Pushto), an urban setting and the children come from relatively better socio-economic background; they attend a private school and their parents typically are business people or high salaried employees of banks etc. We surveyed 15 students from Argentina, 14 students from the United States, 25 Pakistani, and 17 Jamaican students (~50% male). Students participated in school-like environments in general, with the exception of the USA, where students were recruited from individuals and neighborhoods in relation to the authors.

**Method.** Children were provided a printed six-page packet and asked to help us design avatars by drawing characters they thought would help younger children as they played a mathematics game, see Figure 3. On page 1, students were told “We are designing computer based math games for younger children. Can you help us?” On the second page, a
screenshot of a simple addition math game where children click on the moving fruit that indicates the correct answer. Students were asked to provide feedback that the digital character might provide for correct and incorrect responses of the participants.

Finally, teachers or parents were instructed to complete the student demographics (age, ethnicity, nationality and gender). In the case of Argentina (Spanish is the official language) the school was an English as a Second Language school, thus, the materials were left in English, and the teacher helped students to understand, reading aloud and translating when necessary. In the case of Pakistani schools, all materials were translated to Urdu, the official language. Coders translated student responses back to English afterwards, for both Pakistan and Argentina, whenever necessary.

**Results.** Although children were asked to create math avatars that looked like people, children came up with humanoid and non-humanoid images. In the Pakistani study, which involved a translation of instructions to the Urdu language, it was not clear that students had understood that we meant by “characters that look like humans”. We only coded humanoid images, see Figure 2. Two different human coders analyzed the pictures and messages that would be used to respond to a child’s correct/incorrect answers. They coded the variables, such as gender of the avatar, facial hair, clothing or emotion of the avatar. Because many of these metrics might be subjective, we had two independent coders. After coding was completed, we computed Cohen’s Kappa to analyze agreement between coders. Whenever a variable had a Kappa value less than 0.5, we reconsidered the variable and came up with a new coding scheme. The variable was recoded and the process repeated. Variables with very low Kappas were dropped from the analysis (e.g., age of the avatar). We then carried out Analysis of Variance on the variables of interest, and nationality, gender-child as fixed factors, with age of child as a covariate. In the case of discrete variables, we ran cross-tabulations and Chi-square tests.

As hypothesized, there were significant differences found for child’s gender ($\chi^2 = 38.9, p < 0.001$) and gender of the avatar, showing that most children drew characters of their same gender. Only a minority of children drew characters of unidentifiable gender. No significant differences were found for nationality. A significant interaction effect between gender of the student and nationality ($F = 3.9, p < 0.015$) showed that the Pakistani girls drew more fantasy characters than did children from other countries, or than did the boys of the same country, see Figure 2. Significant main effect was found for a child’s nationality in level of detail of the characters ($F = 3.6, p < 0.02$), showing that the students from Pakistan and Argentina drew more details than did the students from the United States or Jamaica, regardless of their gender. Further analyses showed differences in the amount of head-coverings, particularly drawn by boys in general ($F = 13.6, p < 0.001$), and for the Pakistani and the Argentine boys in particular. Contrary to our hypotheses that girls from Pakistan would draw headscarves to represent their cultural norms, they did not. In fact, they tended to not draw images of real people but drew fantasy figures from other cultures, e.g. Disney princesses. The head accessories that boys drew were actually hats. Students from the United States drew the least detailed clothing on their avatars ($F =$
The students from Jamaica drew more hair on their characters’ heads, and the students from Argentina drew the most facial hair. Meanwhile, girls in general drew significantly more hair on their avatars, both on the characters head ($F = 11.17, p < 0.02$) and more facial hair details (i.e. eye-brows, eye-lashes, moustache, etc.) ($F = 8.2, p < 0.001$). Girls also drew more details on the skin (e.g., freckles, dimples, tattoos, etc.) ($F = 4.5, p < 0.04$). No significant differences were found in the amount of accessories used, the kind of accessories, nor in the presence of shoes, facial features or limbs. Most students drew all of these, mostly full-bodied avatars and the amount of accessories did not have a consistent differential pattern across nations or genders. Perhaps the results of the Pakistanis young girls, can be interpreted as a mix of traditional culture and of TV-movie effect. A significant main effect was found for emotions expressed by avatars as drawn by girls and boys. As hypothesized, the girls across nations were more likely to draw avatars with happy faces, with boys evenly split between happy and neutral faces (gender effect, $F = 9.8, p < 0.003$) and a minority of students drew what was coded as angry in their characters, 5% of all students, all three were boys. As expected, a significant main effect was found for a child’s nationality predicting the formality of language (e.g., “nope” to say “no”, “awesome” to say “yes”) used by students for the avatars in response to incorrect or correct answers from players ($F = 9.7, p < 0.001$). Students from the USA used more informal language than did students from Jamaica and Argentina and the children from Argentina used the most formal language (i.e., least informal language) compared to Jamaica, United States and Pakistan. Significant effects were found across countries for the avatars’ answer after an incorrect answer from a student player, whereas no significant differences existed for having the character express a response after a player’s correct answer. A significant main effect for nationality ($F = 3.3, p < 0.03$) showed that students from Argentina used the least “polite” language as compared with students from other countries (e.g., least use of words such as sorry, please, thank you etc.), with children from the United States and Pakistan using the most polite language. Interestingly, there was an interaction effect between gender of the child and nationality ($F = 2.9, p < 0.05$), which indicated gender differences in the tone of the avatar’s response to incorrect answer for children of different genders. Actually, boys’ avatars from the United States used more polite language than girls’ avatars from the same country, despite the fact that the appearances of U.S. boys’ tended to be more aggressive than girls’ (see examples in Figure 3 drawings); the reverse happened for Pakistan, where girls’ avatars used more polite language than boys’.

**Future methods.** We intend to systematically explore the impact of culturally-aligned interventions on underrepresented students, specifically on their learning, affect and attitudes. We intend to advance a deeper understanding of these interventions to address social factors such as identity and bonding with students that impact affective outcomes while students learn, and how interventions can impact different minority groups. We will generate detailed principles regarding the influences that student backgrounds (the socio-cultural and affective baggage) bring to the design and effectiveness of digital educational
tutors. We will use the Wayang Tutoring System with characters of different genders and ethnicities/races as part of students’ regular math class. Students will be given a choice of characters that have clearly defined genders and ethnic characterizations; they will also choose to switch or remain with a character and their choices will be recorded in our database logs. Agent companions will be described as a friend and represented by several subcultures (e.g., Mexican-American, Puerto-Rican, Jamaican-American, Caucasian, African-American, Navajo, etc.). After finishing using the software with characters, students will be given a survey about their general perception of the characters, and open questions in focus groups about how to make characters more interesting/appealing, how to improve what the characters should say, what they think about how they speak, and corrections to their speech. Small focus groups (per cultural cluster) will be run, asking students for further thoughts about learning companions, with special emphasis on those that are culture-matched.

4 Discussion

This paper describes a first iteration of Design-Based Research, since we intend to both understand the effects of culture on learning and to design culture-sensitive learning principles. In the future we may reconsider our methodological framework using the iterative approach of DBR, to provide a better structure and direction for this research. If we consider that children draw what they expect, value, and desire, findings from our pilot study suggest that children, regardless of country, expect characters to be of their same gender. This is consistent with prior findings that indicate that matching the gender of the student with the character’s gender led to improved affective, behavioral and learning outcomes, such as engagement and reduced frustration (e.g., [34]). Girls also expect more details in their character’s hair, skin and facial hair. Boys might want to have more head coverings, particularly hats. Also, girls from Pakistan might prefer fantasy figures instead of figures that depict themselves. Lastly, the fact that girls in general drew more positive emotions on their characters could suggest an expectation of girl’s avatars to emote and act affectively – however, this needs to be examined further.

The level of detail expected in the characters will vary by country. Children from Argentina and Pakistan might expect more detail than do students of the USA or Jamaica, e.g., in clothes and hair. Meanwhile, differences across countries are especially marked in the kind of language to be used when the characters talk and respond to incorrect children’s answers; Argentine children apparently expect the least politeness. Expectations of politeness and amiability of the language can be explained by cultural differences. Argentines can be straightforward in their dialog (similar to European countries, e.g., France, Italy or Spain) and do not excuse themselves as much as do people from the United States (see [41]). Significant differences were present across countries, gender, and across country and gender. Main differences were in the language used to respond to incorrect an-
swers and in the look of characters, both across countries and gender. These differences span across the visual appearance of pedagogical agents as well as in the language used to communicate to players.

In future work we intend to develop a research plan that explains how the findings from this study will aid our design of pedagogical agents for underrepresented groups in the USA. We also will examine the relation between a child’s previous experience with math games and the resulting avatar, examining, for instance, the cultural background of existing game avatars and whether these set expectations for what would be culturally appropriate in a math game. Possible enhancements to this methodology include giving students exactly the same materials and time within each class to create new avatars. This will help control for effects that are connected with coloring, or time constraints that may impact the final outcome of drawings. In addition, recording students’ voices could lead to important speech features and differences in intonation that would be important when designing multimedia-based characters. This research has large implications for education in general and provides ways to understand cross-cultural differences especially at the subcultural level. One goal is to understand digital companions from the student’s perspective and to acquire a culture-based approach to educational environments.

Acknowledgement. This research was funded by the National Science Foundation, # 1324385, Cyberlearning DIP, Impact of Adaptive Interventions on Student Affect, Performance, and Learning, Arroyo, Woolf and Burleson, PIs and # 1109642, REESE, Personalized Learning: strategies to respond to distress and promote success, Arroyo, Woolf and Burleson. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

References


A theoretical approach to the development of critical incidents for cultural training

David Gerritsen, John Zimmerman, Amy Ogan

Human-Computer Interaction Institute,
Carnegie Mellon University, Pittsburgh, USA
{dgerrits, johnz, aeo}@cs.cmu.edu

Abstract. Researchers and developers working in cross-cultural training make frequent use of scenario-based instruction, the fundamental component of which is the critical incident. The traditional approach to developing critical incidents uses open-ended questions without theoretical grounding. We describe a new approach that starts from a theoretical and practical perspective in order to make the process more efficient, and more useful for transfer and assessment of cultural learning. We describe an implementation of the approach with an example from our research with international teaching assistants at American universities.

Keywords: Scenario-based training, critical incidents, international teaching assistants

1 Introduction

Researchers and managers interested in cross-cultural sensitivity and communication have been using scenario-based instruction as a training tool for several decades [1]. Scenarios take many forms ranging from simple scripts for role-playing to high-tech virtual-reality simulations with integrated artificial intelligence or intelligent tutoring systems [42.7.8, 13]. Scenarios work well because they allow cultural novices to willingly suspend disbelief and to become a character in a concrete, novel situation, learning as they act and react to the unfolding situation. Repeated practice of unexpected events is a critical part of learning how to navigate unfamiliar cultures [10], which is what makes simulations powerful.

Scenario-based instruction gives trainees exposure to content, i.e., incidents that actually happen in the target culture, and context, i.e., environmental cues that often accompany the incident. There are many scenario-based instructional methods for improving intercultural competence. The culture assimilator has been studied the most [42.7.1.4], and is implemented in many ITS-based systems. This technique exposes cultural novices to real-world scenarios that are known to be hard for non-natives to interpret. “Critical incidents” are a key feature of the culture assimilator. These are text-based vignettes about plausible
characters in culturally uncertain circumstances, with specific behavioral lessons embedded in environmentally rich narratives. What makes an incident critical is that the event appears confusing or disruptive to a cultural novice, but appears obvious to a native of the target culture. Critical incidents are used in many types of cultural training [1], and they are designed to include the attitudes, norms, values, and social contexts common to the target culture [4].

Critical incidents traditionally come from interviews of people who have experienced a foreign culture and can relate memories of confounding situations, and their impressions of why the events unfolded as they did [4]. Researchers collect and anonymize these stories from a wide sample of subjects before then presenting them to natives and experts from the target culture. They ask the experts to provide their own rationales for the unexpected behavior, and incidents are considered “critical” when the explanations from novices and experts fail to match. Although this approach has proven effective, it also invites large variance in the data due to the use of open-ended questions the initial interviews. Researchers ask subjects to describe “some specific intercultural occurrences or events that made a major difference in their attitudes or behavior toward the members of the other culture” [4, p. 97]. Such questions generate many interesting answers, but there is no coherent theme to the memories, and no hint of the underlying values that might explain entire classes of events within a society [6].

When it comes to building technologies for supporting cultural competence, we have found that open-ended interview tactics are insufficient instruments of knowledge acquisition. Many hours are lost to post-hoc categorization of incidents that may or may not fit specific environments. Furthermore, without a theoretical grounding to connect the scenarios, it is difficult to assess what the cultural novice might be learning. We therefore expanded the traditional methods of developing critical incidents by including useful features from knowledge engineering [18], the advantages of which we describe next. We will then explain our process in both abstract and concrete terms. We believe this approach is particularly relevant to the ITS community, and we conclude with suggestions for further implementation.

2 Potential advantages of a theory-based approach

In the world of culture training, every critical incident provides an inductive, experiential learning opportunity (and formative assessment) through exposure to specific situations with known solutions. We argue that our approach to developing critical incidents streamlines the process of discovering relevant situations and their known solutions, meaning that it strengthens the likelihood that trainees learn about underlying, latent cultural values, as opposed to visible yet disconnected cultural practices. By organizing scenarios around theoretical constructs, we attempt to limit not only the number of incidents we must collect through interviews, but also the number of scenarios a trainee must encounter before gaining insights on underlying principles they can apply to many situations. Unlike
traditional culture assimilators, this approach promotes *summative* assessment, as well as measures of transfer, because a trainee who has viewed some sufficient number of examples in a theoretical domain should be able to successfully select an appropriate response to previously unseen examples in the same domain.

To help clarify our process, we first provide a set of term definitions. By *target culture* we mean the unfamiliar cultural environment in which a trainee hopes to function without incident. A *cultural environment* may be a school or workplace in a new country, or even a local collaboration with international colleagues. *Trainees* or *cultural novices* (we use the terms interchangeably) are individuals who attempt to learn about the practices of the target culture. A *former cultural novice* is someone who has gained some expertise in the new environment after having lived or worked in that environment long enough to have noticed, reflected upon, and adapted to the target culture (stages known to describe the process of cultural adaptation [10]). It is important to note that a former cultural novice is not necessarily a cultural expert, and may at times maintain inaccurate or incomplete attributions about the culture.

**3 Critical incidents: A new approach**

Culture assimilators have been used extensively in training and assessment of cultural sensitivity, e.g., in industry [1], military [14], and research settings [4]. We are not the first to use theory as a guide in the development of culture assimilators [1, 2, 17]. The most common example of such work is with the *culture-general assimilator* [2], which attempts to train novices from *any* culture to operate in *any other* cultural environment. This culture-general method, however, focuses narrowly on the experiences of experts, and it abandons the aggregation method of the *culture-specific* assimilator (the traditional form that is described above). Our approach provides an advance by drawing on both traditions and aggregating critical incidents around theoretical principles that experts have helped pinpoint. We employ experts at the beginning of the process, rather than the end, which provides specific cultural dimensions to focus our interview questions. In addition, it brings to light incidents that former novices may not have yet encountered. In terms of knowledge engineering, the way we construct critical incidents can be thought of as preparation for modeling the *organization* and the *application* when those constructs have yet to be fully defined [18].

This section outlines our steps, provides explanations, and illustrates the approach with examples from our current research.

1. **Interview people with expertise in the research domain and cultural theory. Perform literature review.**

The initial step involves focused interviews [11] with experts in the field of interest, generally people familiar with the target culture as well as with the trainee population. It also
includes a review of the applicable literature, i.e., that which covers culture theory and/or the specific domain. Synthesizing these two sources of information provides direction for forward scenario simulation development (described in step two), which we use during interviews with former novices (step four).

In our current research we have interviewed experts in culture and educational systems. They are professors and researchers in the field, including the directors of international offices from two American universities, whose jobs include the evaluation and training of international teaching assistants (ITAs). Our trainees are ITAs living and working in the United States. The target culture is that of American college students in STEM courses. From this we can summarize the practical domain and its theoretical framework as the culture of learning in American college classrooms.

Our literature review included a wide selection of authors in culture theory and ITA training [e.g., 3, 5, 6, 8, 7,12, 16]. From this and the expert interviews we identified the Cultural Dimensions of Learning Framework (CDLF) [15] as the most relevant resource. It comprises eight (non-exhaustive) dimensions of culture and learning that may have an appreciable impact on how an ITA in America views the average college classroom. Two easily identifiable dimensions that connect to many cultural theories include equality/authority [15, cf. power-distance, 6, 7] and clock/event time [8]. Equality/authority is a dimension that, among several defining characteristics, describes how cultures differentially view and behave toward members of various social ranks. There are many ways this dimension plays out in the classroom, such as governing appropriate forms of address between professors and students. Clock/event time describes how closely a culture tends to follow a strict measure of time as an indication to start and stop activities. This dimension can have an impact on how students and TAs interact with respect to starting and finishing classes, office hours, and meetings, or how rigorously one should attend to due dates.

2. Create forward scenarios based on this top-down view.

Forward scenario simulation [11] is an interview technique that prompts experts with vignettes of common problem situations (i.e., forward scenarios) and asks how they might solve them. From the synthesis of the expert interviews and literature search, we create a small number of hypothetical (i.e., forward) scenarios that will prompt respondents with a familiar experience that evokes their own memories of being new to the cultural environment (step four). These scenarios should focus the interview on a subset of experiences that connect to a specific theoretical construct.

In our current work, one of the scenarios draws from equality/authority and describes an incident where students regularly “interrupt” the TA, displaying what could be interpreted as a lack of respect for the TA’s authority. Another scenario, drawn from clock/event time, describes a student requesting an extension on an assignment in order to spend more time considering the lesson.
3. Survey natives and former cultural novices of target population.

Before initiating interviews with former novices (i.e., the process of gathering personal stories and beliefs for later critical incident development), we create or find a survey instrument that can help us compare and contrast the cultural orientations of the novice and target populations. This will later support the introspection format phase of the interview [11] (described in step five). The survey need not be rigorously tested and validated, as it will be a qualitative starting point for a discussion with the former novices, and they will have an opportunity to present their impressions of the results. But it should at least have face validity as a measure of cultural values for natives and novices within the domain of interest. We distribute the survey to members of each population, and then create a statistical summary of the target population’s responses as an anchor for contrasting with former novices. This contrast may help identify interesting interview subjects from a large pool of former novices, but primarily it supports introspection during later interviews.

In our work, we administered a survey based on the CDLF [15] to 57 students from two different university classes in the target culture and 23 ITAs from the same institution. We compiled the means and standard deviations of the native population along all eight dimensions to create a comparative sample for the interview prompts in step five.

4. Interview former novices on survey responses and forward scenarios.

Steps four and five both involve gathering data through interviews with former novices, but each step has unique goals and techniques. Step four interviews are meant to assess the validity of the survey instrument and gather responses to the forward scenarios from step two. Both sets of interviews incorporate open-ended questions derived from traditional culture assimilator development [4], as described in the introduction.

We began all of our interviews with general questions about where the interviewees were from, classes they had taught, and so on. We then asked open-ended questions inspired by Fiedler’s method [4], such as “What did you find surprising when you first came to live and study in the US?” We discussed their scores from the CDLF survey and asked how accurate the ratings appeared, making note of any discrepancies between their results and their perceptions. Specific to step four, we then presented the forward scenario simulations.

We found that each forward scenario uncovered very different reflections. One such, the late assignment request described in step two, caught our former novices by surprise. Although we conceived this vignette with the help of our experts, and knew it was something that does in fact happen, the ITAs had difficulty believing any student would request additional time to understand and work on an assignment. We recognized this as an example of either something our former novices had not yet encountered or were not able to identify as a cultural mismatch. They had no trouble providing various strategies to anoth-
er scenario that described student “interruptions” during class. They were all familiar with the level of interjection in American classrooms.

5. Remove initial incidents, replace with theory-based prompts. Continue interviewing.

Step five replaces the forward scenario simulations of step four with an introspection format [11]. This means that for the rest of the interviews we ask the interviewee to propose specific classes of problems that reflect the dimensions of the survey, as well as how they might solve those problems. We should note that if time were unlimited we would most likely keep the forward scenario section. Expert-driven scenarios help to drive solution generation, or to identify situations that may be unfamiliar to some former novices. At the same time, forward scenarios are likely to restrict the range of confounding incidents we might otherwise discover. The introspection format serves the greater purpose: finding a larger sample of theoretically grounded critical incidents.

The timing of the transition from step four to step five depends in part on how well the survey appears to capture the cultural values of the interviewees. In our experience, evidence that the CDLF survey could ground the interviews along theoretical lines became apparent around interview 10. For the remaining interviews (as of printing we had conducted 17), we spent more time detailing the CDLF and answering questions about it before then showing them a comparison between their scores and those of our sample target population. While we remained open to their thoughts on any of the eight dimensions, we intentionally probed for memories that would reflect their personal experiences with only the most interesting dimensions. For example, if the former novice was at least one standard deviation away from the average student in our sample on equality/authority, we asked them how that dimension had been relevant in any of their own teaching experiences.

This approach was particularly effective at providing a rich set of responses. Our data indicate that experiences exhibiting different surface features can in fact reflect a single dimension. For example, a mismatch in equality orientations can make it difficult for ITAs to learn how to ask a professor for help, but can also create confusion when American students occasionally send “rude” emails with “demanding” language. These common breakdowns indicate a potential need for cultural training on this dimension, and suggest a hypothesis that different scenarios may train the same underlying values.

One empirical question at this point is exactly how many interviews to perform, and the answer is likely to depend on the specific cultural theory and practical application of each study. In our case, we are still gathering data, but it is not necessary to complete all interviews before initiating the coding phase. Beginning early can even inform and improve the interview protocol.

The segmentation and coding process depends on the practical domains and cultural theories under investigation, but categories should at least identify the *topics of conversation*. These topics should then be grouped by *theoretical dimension*. This step is an iterative process that will eventually produce a set of specific narratives distributed amongst theoretically grounded cultural dimensions (as described in step seven).

For our study with ITAs, we chose a two-tiered segmentation scheme with three layers of codes. Our first tier was topical, maintaining a large amount of context, including the interviewer’s input when necessary. We labeled interactions with short summaries, such as *reticence in asking a professor for help* or *receiving a “rude” email from a student*, and then grouped each topic within the relevant cultural dimension of learning, such as equality/authority [7.15]. This step required many careful readings of and discussions about the data in order to achieve agreement between the researchers.

Our second level of segmentation was at the statement level. To these we applied three codes: <noticing>, <reflection>, and <output>. These codes reflect stages that have been identified as critical components to gaining cultural competence [10]. A more formal definition is beyond the scope of this paper, but the categories can be thought of generally as *recognition of events in the world, consideration of the potential meaning of events*, and *responsive behaviors*. An example from our data includes this comment by a Chinese ITA teaching an Introduction to Computer Systems recitation: “Maybe one or two good, meaningful questions [from the students] is good. I would prefer more structured schedule.” To this we applied the following codes: <Topic: Class Discussion>, <Dimension: Equality/Authority>, <Statement Type: Reflection>.

7. **Build final episodes.**

After the last iteration of segmenting, coding, and clustering the data, it is time to consolidate and summarize the results. This means searching for common themes amongst the clustered samples and building multiple representative narratives for each dimension. In Fiedler et al.’s original design of the culture assimilator [4], the researchers transformed the unique stories from each interview into *episodes* (i.e., a revised version of the original story with names obscured and irrelevant detail removed). They tested these episodes amongst some number of experts to determine which were critical incidents. With our approach, we have already determined the types of incidents that can confuse novices, and we create a final episode for each type.

We are still iterating through our own data coding, but can already see the cultural experiences that were most salient for our former novices. The majority of the responses to our reflection probes (45% of all topics), had to do with some aspect of open discussion in American classrooms: how it happens “here” more than in the home country, how it is often required for participation points, etc. On the surface, this may not seem like an equality issue, but these responses tended to focus not on the fact that discussion *happens*
more in the US, but that teachers in the US encourage and expect it. Most of our former novices had some version of the following response, “I sort of knew what to expect when I came here to study, but I was surprised by how much the teachers care what students think.” This is clearly an important issue for international students to accommodate, and even more relevant to ITAs who hope to succeed at teaching in American universities.

Using the example of discussion in the classroom, one incident could be framed:

*Katie is a Chinese graduate student who is teaching recitation for an introductory physics class. She feels that the material is clear and requires no discussion, so she is surprised at how often her students interject with comments—not just questions. How would you explain the students’ behavior?*

Because this event is identified as being part of the equality/authority dimension, we can create additional critical incidents from this same category by referring back to the clustered data and finding other examples that fit the dimension. Relevant themes that we found included Asking for help, Giving class presentations, Eating in the classroom, and Personal interactions between students and TA. Each of these themes can produce multiple scenarios representing the entire dimension, and we hypothesize that training novices on a subset of these incidents will lead to successful transfer of learning to unfamiliar scenarios that are also rooted in equality and authority.

4 Conclusion

In this paper we have described a new approach for generating critical incidents that align with both cultural theory and real-world experience. Involving experts early in the process, implementing a survey of target culture natives and former cultural novices, and focusing the interview on specific theoretical dimensions generated critical incidents that clustered around specific cultural values. We described the approach in steps, and we gave an example from our own work acquiring knowledge about the environments and problem states experienced by international teaching assistants in an American university.

As we continue to explore this approach, we intend to instruct our learners with one subset of scenarios, and assess them with another. We predict that training on scenarios that are built around specific cultural dimensions will increase transfer of knowledge and skills to new, unfamiliar scenarios that reflect the same constructs. Likewise, we will identify which scenarios are most likely to support transfer. The intended outcome would be learners who are ready to notice and reflect on new incidents in their real lives, and independently arrive at adequate strategies for avoiding conflict in the classroom.

In developing a new culture tutoring system, our next step will involve using these critical incidents to build models of the situations in which ITAs tend to perceive breakdowns. By using a general theory of culture and its influence on education in an American university, we have uncovered specific examples of incidents that represent general cate-
categories of environments and behaviors to expect in the classroom. Our models will include the context and content of the class, the range of breakdown rates, and the different behaviors that ITAs might enact to solve (or fail to solve) common problems [18]. The system will prompt novices for reports on their experience, determine their progress toward the learning goals, and make training recommendations involving reflective journaling and various game modules.

Beyond our specific case, we believe that researchers and educational designers could implement this approach even without building formal models. We have combined strategies from the tradition of culture assimilators with techniques familiar to knowledge engineers, which will provide coherent scenarios that trainers could use in the design of scenario-based games, simulations, or tutoring systems, and additionally to align instructional activities with learning goals and assessments.

Acknowledgements. This work is supported in part by Carnegie Mellon University’s Program in Interdisciplinary Education Research (PIER) funded by grant number R305B090023 from the US Department of Education.

References