Sixth International Workshop on Culturally-Aware Tutoring Systems (CATS2015)

held in conjunction with

Seventeenth International Conference on
Artificial Intelligence in Education (AIED 2015)

Monday, June 22, 2015
Madrid, Spain

Workshop Co-Chairs:

Ma Mercedes T. Rodrigo[1], Emmanuel G. Blanchard[2],
Amy Ogan[3], Isabela Gasparini[4]

[1]Ateneo de Manila University, Philippines
[2]IDÛ Interactive, Montreal, Canada
[3]Carnegie Mellon University. USA
[4]University of Santa Catarina State (UDESC), Brazil

http://cats-ws.org

Table of Contents

Preface i

Leveraging Comparisons between Cultural Frameworks: Preliminary Investigations of the MAUOC Ontological Ecology 1-10
Phaedra Mohammed and Emmanuel Blanchard

Exploring Power Distance, Classroom Activity, and the International Classroom Through Personal Informatics 11-20
David Gerritsen, John Zimmerman and Amy Ogan

Culture-Oriented Factors in the Implementation of Intelligent Tutoring Systems in Chile 21-30
Ignacio Casas, Patricia Fernandez, Marcia Barrera and Amy Ogan

More Fun in the Philippines? Factors Affecting Transfer of Western Field Methods to One Developing World Context 31-40
Juan Miguel Andres, Ma. Mercedes T. Rodrigo, Jessica O. Sugay, Michelle P. Banawan, Yancy Vance M. Paredes, Josephine S. Dela Cruz and Thelma V. Palaoag

Investigating the Impact of Designing and Implementing Culturally Aligned Technological Systems on Educators' Ideologies 41-46
Samantha Finkelstein

Preface

Culture has a profound effect on the way people interact with, react to, think and feel about knowledge, symbols, situations, etc. Yet it is underestimated in AIED research. Most of the currently influential learning systems have indeed been created by and for developed world contexts and with Western cultural perspectives in mind. However in recent years, more and more opportunities to design, develop, and deploy educational software for and in different contexts have emerged. This state of affairs naturally leads to broader questions. What features of culture are important to consider in the design process? Can software designed and developed in a specific cultural context transfer to other parts of the world and remain effective? The answers to these questions remain unclear although a growing body of research suggests that the use of AIED systems across cultural contexts results in variations of the knowledge acquisition process.

Over the last seven years, Culturally-Aware Tutoring Systems (CATS) workshops have been organized in conjunction with ITS2008, AIED2009, ITS2010, AIED2013, and ITS2014. The series is a venue for researchers to reflect on the universality of their work. CATS2015 thus proposes to discuss culture and AIED from five perspectives:

1. Developing both pedagogical strategies and system infrastructure mechanisms that incorporate cultural features to enculturate AIED systems;
2. Designing acquisition-oriented CATS, i.e. AIED systems to teach cultural knowledge and intercultural skills;
3. Designing adaptation-oriented CATS, i.e. AIED systems that can be personalized overtly or automatically based on users' cultural profiles;
4. Considering human features that are connected with the learning process, and that are culturally-sensitive, e.g. affect, behavior, cognition, or motivation; and
5. Considering cultural biases in the AIED research cycle.

In addition to describing the current state of the art in these domains, the workshop engages participants in working to expand the reach of AIED research to a greater global audience, including those disadvantaged due to a lack of resources or other obstacles.

Overseeing the quality of CATS2015 papers was a program committee of 37 members from Asia, Europe, North America, and South America. The program committee members were well-versed in AIED, culture, technology, and other relevant fields. The committee selected 4 full papers and 1 short paper for inclusion in this year's workshop.

We thank all the program committee members and authors for contributing their time and expertise to making CATS2015 possible. We also thank the Workshop Chairs and the Organizing Committee of AIED2015 for including CATS in this year's conference.

Ma Mercedes T. Rodrigo, Emmanuel G. Blanchard, Amy Ogan, & Isabela Gasparini
The CATS2015 Co-Chairs

Leveraging Comparisons between Cultural Frameworks: Preliminary Investigations of the MAUOC Ontological Ecology

Phaedra Mohammed[1], Emmanuel G. Blanchard[2]

[1]Department of Computing and Information Technology, The University of the West Indies, St. Augustine, Trinidad and Tobago
phaedra.mohammed@gmail.com

[2]IDÛ Interactive Inc.
Montreal, Canada
ebl@idu-interactive.com

Abstract. Many theoretical cultural frameworks have been proposed in the literature. For comparisons and critiques of these frameworks to make sense, community members have to assign similar-enough meanings to the terms that they use when interacting. This entails overcoming the challenge of dealing with the imprecise and interpretable definitions conveyed in frameworks due to the use of common language. The MAUOC Ontological Ecology (MOE) approach offers a strategy for dealing with this through reinterpretation of all cultural frameworks along a singular, common conceptual baseline. In this way, a far more cohesive, consistent, and controlled representation of cultural frameworks becomes available compared to just common language descriptions. The purpose of this paper is to clarify the MOE methodology, and report initial efforts into practically applying it to the Hofstede cultural framework.

Keywords: Culture, Heavyweight Ontology, Systematic Methodology, Hofstede Framework

1. Introduction

Culture is a key phenomenon in many academic disciplines such as psychology, anthropology, sociology, education, philosophy, and therefore has been studied from diverse perspectives. Consequently, many theoretical frameworks have been proposed, each with specific purposes as endorsed by different research communities. These frameworks are mostly described with common language terms which disguise the complexity and philosophical nuances within. For these reasons and others, frameworks are frequently prone to misinterpretation, and disagreements are common when conflicting claims are made regarding particular frameworks. A common source of dispute is the use of the same terminology across frameworks which may or may not refer to the same conceptualization, such as *Individualism* and *Collectivism* in the GLOBE and Hofstede frameworks [4].

As an emerging interdisciplinary field, research on Culturally-Aware Tutoring Systems (CATS) is driven by scholars with different profiles, both in terms of cultural backgrounds and expertise. This rich diversity places the CATS community in a unique position to properly tackle the techno-cultural objectives it has assigned to itself. However, the variety of existing cultural frameworks and the lack of time for many community members to deeply understand them creates challenges for cumulating research efforts and findings. Indeed, for comparisons and critiques to make sense, community members have to assign similar-enough meanings to the terms that they use when interacting. This is one way of overcoming the challenge of dealing with the imprecise and interpretable definitions conveyed in frameworks due to the use of common language.

The More Advanced Upper Ontology of Culture (MAUOC) aims to identify conceptual building blocks of the cultural domain, and it has several potential applications for CATS. The one that is considered in this paper is the possibility it offers for reinterpretation of all cultural frameworks along a singular, common conceptual baseline. In this way, a far more cohesive, consistent, and controlled representation of cultural frameworks becomes available compared to just common language descriptions. This would in turn promote objective comparisons between frameworks, and enhance interoperability between research efforts. Before this can be done, a structured, scientific methodology is necessary. One such strategy has been theorized and presented in [3]. It is referred to as the MAUOC Ontological Ecology (MOE) approach, and the purpose of this paper is to clarify this methodology, and report initial efforts into practically applying it to the challenges articulated earlier.

The remainder of the paper is organized as follows. Section 2 presents a justification for the choice of heavyweight ontology engineering as the basis for this research, and briefly describes the development processes behind MAUOC and the MOE approach which motivate the systematic methodology taken in the paper. Section 3 goes into the specifics of this methodology, briefly describes the Hofstede cultural framework, and gives insight regarding why this framework was chosen for analysis. The section then provides illustrative examples arising from the preliminary analysis of the Hofstede framework using the MOE approach, along with a brief discussion of each example. Section 4 discusses what is to be learnt from this preliminary investigation and identifies the limitations of the work so far. The paper concludes in Section 5 with future plans for the investigation.

2. Ontological Grounding of our Analytical Process

2.1 A Heavyweight Ontology Initiative

Heavyweight ontology engineering is strongly connected to the original philosophical meaning of 'ontology'. Whereas heavyweight and other (lightweight) ontologies look similar to non-specialists (simply put, they could be seen as a set of concepts/constructs interconnected with relations), the critical difference lies in the way heavyweight vs lightweight ontologies assign identities to these concepts/constructs and relations. Authors of lightweight ontologies commonly refer to a 'rule of thumbs' approach: they may look for, and accept a definition that makes sense to them in the context of the specific application(s) they have in mind, and according to their personal experience. This obviously limits its applicability while bringing risks of personal and socio-cultural biases. Heavyweight ontologies on the other hand must not target a specific application, but rather aim to capture the true essence of a domain or task (as in philosophy). A definition obtained following proper heavyweight ontological analyses can thus be reapplied in any situation related to the domain of interest.

Eventually, distinctions between heavyweight and lightweight ontologies are largely ignored by non-specialists. This is a major issue since these ontologies have very different properties. However, the purpose of this paper is not to reflect upon this point, and readers are invited to look at [8] for clarifications. Overall, if heavyweight ontologies are innately superior from a conceptual perspective, they have a major drawback: they are far more complex and consequently require more expertise and development time before being considered to be sufficiently stable for use. But for ontology specialists, these difficulties are overshadowed by the breadth of applicability and the subsequent interoperability that heavyweight ontologies allow once stable-enough. We therefore adopt a heavyweight ontological approach because capturing the philosophical essence of cultural frameworks requires careful, precise definitions that can bridge the operational data/solutions produced by different disciplines [3].

2.2 From MAUOC to MOE: Two Phases in Framework Reinterpretations

Initiated in 2008 [1], MAUOC is a heavyweight ontology initiative. Rather than describing MAUOC itself, which is prohibitive in this paper due to space constraints (see [3] for an overview), we will now make a brief presentation of MAUOC's development process. This is essential for understanding the remainder of the paper because it forms the basis for the systematic methodology described in the next section. The process has several objectives:

- Distinguishing 'natural concepts' (i.e. conceptual units which exist inherently in nature. See [8]) from 'constructs' (i.e. artificial conceptual units defined in the context of a framework to better carry out its message, connect with a user community, and/or facilitate its adoption and use) for the cultural domain,
- Providing precise definitions for natural concepts by figuring out their essential parts and properties. These features are 'essential' because the removal of one of them leads instances to be classifiable in more than one definition. In the same time, a proper definition has to respect Okham's razor principle, i.e. the simplest definition is always the best one.

The development process of MAUOC can thus be decomposed into five steps:

1. Acquiring a deep understanding of several cultural frameworks representing different schools of thought and disciplines
2. Identifying major framework terms as 'natural concept' candidates
3. Classifying the ideas behind these terms as trans-framework or framework-specific into a more restricted ensemble of 'natural concept' candidates while discarding those that are too specific or not innately cultural
4. Eliciting ontology-grade definitions for the remaining 'natural concept' candidates and their relations, and testing if the resulting ecology of concepts allows for expressing any cultural situations and issues that may arise
5. Iteratively repeating one or more of the previous steps if d) has failed, because this would mean that the current version of the ontology is incomplete, and/or includes inappropriately-defined elements.

In the course of its development, MAUOC has thus been revised many times before reaching the first version thought to be stable-enough [3]. Yet, one cannot be certain that the current version of MAUOC will not be challenged by cultural issues to be tested in the future. Developing MAUOC is both a top-bottom and bottom-up process that attempts to identify cultural building blocks by cross-analysing various frameworks. Now that a stable-enough version has been proposed, the MAUOC Ontological Ecology (MOE) aims to further this initiative by following a bottom-up approach where ontological translations of cultural frameworks will be designed and grounded on these building blocks. In other words, the goal of MOE is not to state what frameworks should or should not say, but rather to achieve clearer and more precise formulations of what they already intend to say.

Figure 1 presents a simplified view of MAUOC and MOE processes. Note that YAMATO is a top ontology, on which MAUOC is grounded (see [9]).

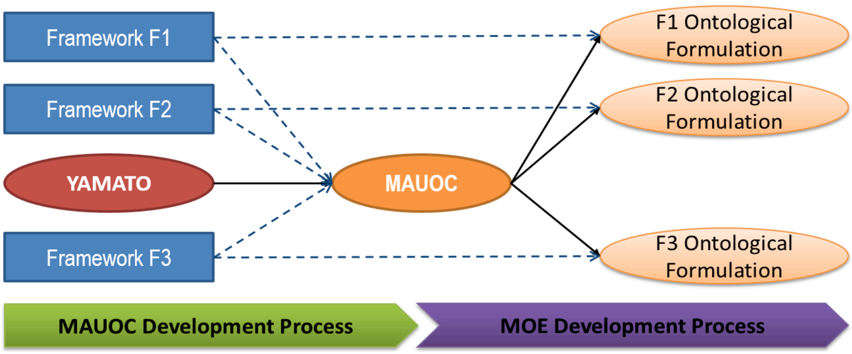

Figure 1. A Simplified View of the MAUOC and MOE Development Processes.

3. Applying the MOE Approach to Hofstede's Framework

3.1 A Systematic Methodology

The systematic methodology described in this section is framework-independent and therefore it can be applied to any cultural framework for which intercultural comparisons are desired using the MOE approach. It is important to note that this process first requires the perspective of external reviewers who have no connection to the particular framework being studied in order to guard against bias [2]. This is crucial since the analysis deals with matters of interpretation and comparison of meanings. At this early stage, only the two authors of the paper are solely involved in the process. Both authors are independent of the cultural framework to which the methodology is being applied and both have different cultural backgrounds which provide an additional layer for guarding against bias.

a) Identify major references for the cultural framework within the literature. Here, sources may include books, journal articles, or conference papers where the overarching quality is the frequency of reference.
b) Identify key terms and several corresponding quoted definitions within these references, by authors of the framework and/or the representative user com-

munity. Key terms, for our purposes, refer to words or phrases which define essential features or ideas that contribute towards the major theoretical underpinnings of a cultural framework.

c) Highlight any discrepancies, consistencies, and/or differences (if any) in the quoted definitions for the key terms. Two levels of analysis are performed in this step: Terminological analysis - which asks whether the definition is consistent over time from a grammatical and a lexical perspective, and Conceptual/ontological analysis - which asks whether the definition is precise enough. Consistency refers the number of changes in the grammatical and lexical structure across the quoted definitions, and it is used to assess whether those changes may alter the meaning in the definitions over time. Precision refers to the self-explanatory nature of expression used in the quoted definition, and the extent to which that expression is potentially subject to interpretations amongst readers.
d) Determine whether a coherent, durable definition can be extracted for each key term. In this step, a key term would still be expressed in common language, but it would now be ontology-ready. In other words, the term would have a logical and consistent structure that is made up of several other conceptualizations that fit together precisely.
e) Consult with experts of the cultural framework to assess the validity of the extracted definitions in keeping with the intended 'spirit' of the framework. If necessary, the definitions would be refined or modified to eventually come to a consensual definition that satisfies both the experts and reviewers while still remaining ontology-ready.
f) Interpret and convert the resulting common language, consensual definitions to MAUOC-grade formulations, using logical representations such as mathematical notations or those originating from HOZO.

Our approach currently focuses on achieving 'heavyweight ontology'-grade definitions for constructs articulated in various cultural frameworks, and as such it only partially reflects the vision stated in MOE. Subsequent and interleaved steps are thus required to clearly state relations and dependencies between these construct definitions in order to achieve true MAUOC-grounded ontologies.

3.2 Primer on Hofstede's Framework

The Hofstede cultural framework was chosen as the starting point in this research for several reasons. Firstly, it is the most popular one used in intercultural research as evidenced by the large body of work using the framework for theoretical and practical reference. Due to over 30 years of study, it is also one of the best documented and consequently one of the most attacked and critiqued of the available frameworks. This rich body of work and the clear evolution that naturally has taken place in the framework due to intense scrutiny, further provides a good distribution of terms upon which to test our methodology.

A brief description of the Hofstede framework is necessary at this point in order to give readers a sense of what the framework is about. The Hofstede framework takes an empirical, generalized approach towards studying cultural differences. It focuses on the identification of dimensions of national culture which were originally: Power Distance, Individualism, Masculinity, and Uncertainty Avoidance [5]. Since then, two more dimensions have been added to the framework: Long Term Orientation and Indulgence vs Restraint [7]. These dimensions are used to score and classify countries according to how members of those societies cope with problems and concerns that are basic to all human societies [7]. Using these scores and statistical relationships between the dimensions, the framework quantified the differences reported across 40 countries originally in 1980. The data set has since been extended to 107 countries [7]. Country clusters were used to account for cultural observations about behaviour which may apply at various levels (national, regional, individual). Table 1 shows definitions of the six Hofstede dimensions, as well as scores for three countries.

Table 1. Hofstede Dimensions and Country Scores for Three Sample Countries

Hofstede's Dimension	***Dimension Description***	***U.S.A.***	***Spain***	***Japan***
Power Distance	The degree to which the less powerful members of a society accept and expect power to be distributed unequally	40	57	54
Individualism	Preference for a loosely-knit social framework	91	51	46
Masculinity	Preference for achievement, material rewards, assertiveness over modesty, cooperation, caring	62	42	95
Uncertainty Avoidance	The degree to which members of a society feel uncomfortable with uncertainty and ambiguity	46	86	92
Long Term Orientation	The degree to which a society maintains links with its own past while dealing with challenges of the present and future	26	48	88
Indulgence vs Restraint	The degree to which a society allows relatively free gratification of basic and natural human drives over suppression and regulation with strict social norms	68	44	42

3.3 Illustrative Examples and Analyses

In applying the MOE systematic methodology to the Hofstede framework, three reference sources [5, 6, 7] were selected. These three refer to some of the most commonly cited sources of the framework, and together they cover over 30 years of the framework's evolution: the original source in 1980, the currently most cited source from 2001, and the most recent source in 2010. To illustrate part of the process, only 6 framework-specific terms were selected for analysis and presentation in this paper due to space constraints. The 6 key terms were chosen since they are core terms for the Hofstede framework (and most other frameworks), they test different situations in the methodology, and they are commonly used in the user community. These terms are considered according to their meaning in the scope of the Hofstede's framework. Hence there must be no confusion between some of these constructs (e.g. *value,* or *dimensions)* and heavyweight ontology concepts using the same labels (see [9]).

Table 2 below shows the directly quoted definitions (if present) extracted for each key term from each source. Summarized, unquoted descriptions are provided if there were no formal definitions found for a given key term. The sources [5, 6, 7] are referred to as 1), 2), and 3) respectively. It should be noted that only the first three steps of the systematic methodology were carried out on the Hofstede framework in this paper.

Table 2. Six Key Terms in Hofstede's Framework and their Representative Definitions in Reference Sources from 1980, 2001, and 2010.

Key Terms	***Key Term Definitions from Hofstede Sources***
Value	1) "*A value is a broad tendency to prefer certain states of affairs over others*." (1980, p.19) 2) "*A value is a broad tendency to prefer certain states of affairs over others*." (2001, p.9) 3) "*Values are broad tendencies to prefer certain states of affairs over others*." (2010, p.9)
Culture	1) "*The collective programming of the mind which distinguishes the member of one human group from another*." (1980, p.25) 2) "*The collective programming of the mind that distinguishes the members of one group or category of people from another*." (2001, p.9) 3) "*The collective programming of the mind that distinguishes the members of one group or category of people from others*." (2010, p.6)
Dimension	1) Empirically verifiable, independent phenomena (behaviours of individuals or situations, institutions, or organizations) on which cultures can be meaningfully ordered. (1980, p.36) 2) A dimension is described by two possible extremes which can be seen as ideal types. "*A dimension is rooted in a basic problem which all societies have to cope, but on which their answers vary*." (2001, p.28-29) 3) "*A dimension is an aspect of a culture that can be measured relative to other cultures*." A dimension groups together a number of phenomena in a society that were empirically found to occur in combination. (2010, p.31)
Individualism	*1)* "*... the relationship between the individual and the collectivity which prevails in a given society*." (1980) *2)* "*... the relationship between the individual and the collectivity that prevails in a given society*." (2001, p.209). "*Individualism stands for a society in which the ties between individuals are loose: Everyone is expected to look after her/his immediate family only*." (2001, p.225) *3)* "*Individualism pertains to the societies in which the ties*

	between individuals are loose: everyone is expected to look after him- or herself and his or her immediate family." (2010, p.92)
Collectivism	1) No formal definition in the 1980 source. 2) "*Collectivism stands for a society in which people from birth onwards are integrated into strong, cohesive in-groups, which throughout people's lifetime continue to protect them in exchange for unquestioning loyalty*." (2001, p.225). "*Collectivism is the degree to which individuals are supposed to remain integrated into groups usually around the family*." (2001, p. xx) 3) "*Collectivism pertains to societies in which people from birth onward are integrated into strong, cohesive in-groups, which throughout people's lifetime continue to protect them in exchange for unquestioning loyalty*." (2010, p.92)
IDV Dimension	1) "*It describes the relationship between the individual and the collectivity which prevails in a given society*." (1980) 2) "*It describes the relationship between the individual and the collectivity that prevails in a given society*." (2001, p.209) "*Individualism versus collectivism is related to the integration of individuals into primary groups*." (2001, p. 29). The IDV dimension is defined also by combining the Individualism and Collectivism definitions from 2) above.(2001, p.225) 3) The IDV Dimension is defined by combining the Individualism and Collectivism definitions from 3) above. (2010, p.92)

Value. Terminologically, the definition of value is cohesive from 1980 to 2010 with one grammatical change in 2010. The grammatical change, i.e. pluralisation, does not affect the meaning of the definition so it is cohesive from this perspective. However it is ontologically since inner terms leave room for interpretation (*state of affairs*, *broad tendency* – what do they refer to? Are these to be understood from a group, individual, or both levels?).

Culture. The definition is terminologically-inconsistent due to changes between 1980 and 2001 from *member* to *members*, and *one human group* to *one group or category of people*, and from *another* to *others* in 2010. In all of the definitions, comparisons are made between A and B, but the nature of A and B changes with each evolution of the definition. This has ontological implications for the cardinality of the comparisons namely a shift from a one-to-one comparison between two individuals in 1980 to a many-to-many comparison across individuals from two groups in 2001 to a broader comparison between not just two groups but amongst many groups in 2010. There are also imprecise inner terms: *collective programming of the mind* and *human group*.

Dimension. The first plain definition for dimension is found in the 2010 source. The term was used and described in 1980 and 2001 across a few pages, however neither source provides a precise definition; the salient parts are summarised in Table 2. Terminologically, there is no cohesion amongst the descriptions. Ontologically, the lack of more than one plain definition provides more room for interpretation. The 2001 quote is imprecise since inner terms (*rooted on*, *basic problem*) are subject to interpretation, whereas *society* is not clearly defined. The 2010 quote is also ontologically imprecise due to interpretable inner terms such as *aspect*, and *culture*. The measurable property of a dimension is however coherently and consistently articulated across all three sources.

Individualism. The quotes are terminologically cohesive for the first part between 1980 and 2001. The additional section added in 2001 is not cohesive with 1980, and not consistent with the 2010 due to two evolutions: society to societies and immediate family only to him or herself and his or her immediate family. Ontologically, there is a change in cardinality as in the culture definition, and the inner terms are imprecise in 1980 (*relationship*), and imprecise and subjective in both 2001 and 2010 (*ties, loose*).

Collectivism. Terminologically there is limited cohesion with no formal definition in 1980, and one evolution between the common quotes in 2001 and 2010: society changes to societies. Ontologically, the definitions in 2001 and 2010 are imprecise due to inner terms requiring further explanations (*strong, cohesive in-groups, society, protect* - from what, why, and by whom? -, *unquestioning loyalty* - allegiance to whom?, forced or voluntary? -).

IDV (Individualism-Collectivism) Dimension. The quotes from 1980 and the first part of 2001 are terminologically cohesive but ontologically imprecise due to inner terms requiring further definition (*relationship, collectivity)*. The quotes from the second part of 2001 and that of 2010 have the same outcome as the individualism and collectivism analyses above.

4. Discussion

The analysis in the previous section should not be construed as a criticism or praise of the Hofstede framework, nor should it be seen as an effort to create our own definitions for key terms. Rather, the intention is to raise awareness of the possible interpretations of the framework's core terms which can have wide-reaching implications for CATS research especially if misunderstanding and oversights are not cleared up. Contradictions from incorrect usage of framework term can lead to wrong conclusions in educational applications, and cascade dangerously in culturally-aware contexts. The goal is therefore to understand the cultural framework and confirm whether existing definitions are prone to significant misunderstandings.

At this point we cannot say that the MOE methodology is fully validated yet since the research is still in its early stages. More work is needed, and naturally there are limitations. Only three quotes were used for each term and we agree that more and deeper reflection is needed for each term in order to solidify the analysis. In addition, quotes were sourced from material written by authors of the framework only. User community quotes can help identify further misunderstandings as well as consensus from a broader perspective, and should be investigated as well. Finally, only the first three steps of the MOE systematic methodology were carried out on the Hofstede framework. Despite this, clear risks of misinterpretation were identified for key term definitions in the framework in these early, simple stages. As ontology-ready definitions are extracted and validated through consultation with experts of the cultural framework, the systematic process hopefully will reveal weaknesses in the MOE approach as well as provide additional validation of the soundness of existing concepts in MAUOC. For example, if a definition requires particular concepts that should have been defined in MAUOC, the missing concepts can be added to strengthen the ontology. If successful, this investigation will then create a baseline for analysing other existing cultural frameworks, and produce further validation of MAUOC as a deep ontological model of culture. Folk-based validation of definitions could also provide practical insight since ontologies, both lightweight and heavyweight, require a community of users. This type of validation however needs to be moderated since reliance on inexperienced users can lead to the design of a folksonomy. It is nonetheless still useful to be considered for future work.

5. Conclusion and Future Research

Derived from the MAUOC Ontological Ecology (MOE) approach, this paper presented a systematic methodology for overcoming the challenge of dealing with the imprecise and interpretable definitions conveyed in cultural frameworks due to the use of common language. Preliminary analysis of the Hofstede framework, using the MOE approach, indicates that the methodology is holding up. The next steps involve analysis of more Hofstede framework key terms, such as national culture, and country score for examples, and figuring out whether ontology-ready definitions are possible for the quoted definitions collected thus far in consultation with framework experts.

References

1. Blanchard, E.G., Mizoguchi R.: Designing culturally-aware tutoring systems: Toward an upper ontology of culture. In E. Blanchard, and D. Allard (eds.), Proc. Culturally Aware Tutoring Systems, 23-34. (2008)
2. Blanchard, E.G.: Is it adequate to model the socio-cultural dimension of e-learners by informing a fixed set of personal criteria? In Proc. 12th IEEE International Conference on Advanced Learning Technologies. 388-392. USA: IEEE Computer Society. (2012)
3. Blanchard, E.G., Mizoguchi R.: Designing Culturally-Aware Tutoring Systems with MAUOC, the More Advanced Upper Ontology of Culture. *Research and Practice in Technology Enhanced Learning* 9(1): 41-69. (2014)
4. Brewer, P., S. Venaik.: Individualism-Collectivism in Hofstede and GLOBE. *Journal of International Business Studies* 42: 436-445. (2011)
5. Hofstede, G.: Culture's Consequences: International Differences in Work-Related Values. Beverly Hills, CA: Sage. (1980)
6. Hofstede, G.: Cultures' Consequences: Comparing Values, Behaviours, Institutions, and Organizations Across Nations. Thousand Oaks, CA: Sage Publications Inc. (2001)
7. Hofstede, G., G-J. Hofstede, M. Minkov.: Cultures and Organizations: Software of the Mind: Intercultural Cooperations and Its Importance for Survival. NY: McGraw-Hill. (2010)
8. Mizoguchi, R.: Tutorial on ontological engineering - part 1: Introduction to ontological engineering. *New Generation Computing* 21(4): 365–384. (2003)
9. Mizoguchi, R.: YAMATO: Yet Another More Advanced Top-level Ontology, *Proceedings of the Sixth Australasian Ontology Workshop Adelaide (AOW2010)*, 1-16. (2010)

Exploring Power Distance, Classroom Activity, and the International Classroom Through Personal Informatics

David Gerritsen, John Zimmerman, Amy Ogan

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

Abstract. Research shows the benefits of active learning in American college classrooms. International graduate students in American universities may face difficulties in teaching students with different cultural dispositions. The current research uses power distance to explore cultural juxtapositions in classrooms and personal informatics design to propose an adaptive system for cultural acquisition. The work shows that even though instructors are aware of the distinctly Western value of speaking up in class, they do not employ it in their own classes. They show surprise at the amount of time they spend lecturing, but they express ambivalence about the importance of vocal contributions from the students. We describe a technical system design that supports the development of cultural fluency by providing ITAs with feedback such as visualizations of time spent lecturing and suggestions for strategy selection in culturally challenging scenarios. The system would reflect changes in classroom activity over time as a way for TAs to reflect on their own professional development.

Keywords: Power distance, international teaching assistants, classroom activity, personal informatics

1 Introduction

Research in the learning sciences has recently produced an explosion of experimental evidence that college students benefit from less lecture and more student activity. This evidence exists even for content-heavy science, technology, engineering, and mathematics (STEM) classes where instructors have traditionally emphasized the importance of covering and memorizing facts rather than exploring, curating, and constructing knowledge. Most of these studies have taken place in American classrooms and have not addressed questions of cultural dimensions of learning and teaching. Meanwhile, the number of international graduate students teaching introductory STEM classes in American universities continues to grow. These students tend not to have experienced the cultural shift toward active learning and its concomitant decrease in social distance to figures of authority that is familiar to most students from the U.S. This can lead to challenges for international graduate students in the U.S. when they are required to teach American students.

The CATS community has a history of developing systems to improve education and cultural awareness. We build on this line of research by focusing on new design methods that frame the instructor as both the learner and the agent of change in the classroom. Using methods from *Personal Informatics* (PI), we explore the state of international teaching assistants (ITAs) leading STEM classes in an American university, and propose a system that potentially simplifies the implementation of active learning in order to more fully engage students.

PI is an approach to behavior change and maintenance that gathers user data and generates digital artifacts for reflection, such as visualizations of change toward a behavioral goal. Very little research has looked at its value in education, and none has attempted to use it to better understanding culture. It incorporates methods of contextual design and development that may be valuable in improving educational outcomes while investigating culturally adaptive interactions.

To assess the feasibility of this line of research and development, we carried out several overlapping activities: classroom observation of ITAs in action in order to understand the context need for adaptive instruments, surveys and interviews in order to understand how ITAs might make sense of classroom behavior, and data visualization feedback for ITAs in order to understand and explore the potential interface for a PI system. Finally we constructed and evaluated a prototype classroom detection system to investigate if we could sense relevant behaviors.

We confirmed that ITAs' knew of the cultural value of classroom activity, yet their recitations were almost completely based on lecture, with little student participation. They were open to more classroom activity, but with some reservations. They shared an interest in monitoring their teaching behaviors and aligning their performance with expert models. Also, our technical system functioned with 85% accuracy. We propose that these findings support further investigation of PI methods for investigating and supporting the acquisition of cultural fluency in unfamiliar educational contexts.

2 Background

Several decades of research in U.S. higher education has produced a wealth of studies showing the benefits of active learning compared to passive lecture and fact memorization [1, 2, 3, 4]. These studies have investigated and advocated active learning tactics such as *think-pair-share* and *cooperative learning*, showing that students improve academically, socially, and psychologically [1, 4]. Like most education research, the studies tend not to include considerations of cultural dimensions of learning. Cultural dimensions of instructors and learners in American universities are poorly understood. Given the evidence that different cultures have different valuations of student activity in the classroom [5, 6], the call for increased student participation may create a tension when it fails to address how international instructors perceive and value active learning practices. This situation deserves attention as the number of international graduate students teaching STEM classes in the U.S. continues to grow [7].

One way to orient the conversation about cultural differences in praxis is to frame it in terms of power distance [8, 9]. Higher and lower national indices of power distance (PDI) attempt to describe the level of deference that individuals express toward members of higher and lower social status. Given the long history of measurable social distance between Asian students and American instructors [5, 6], power distance is a reasonable construct with which to study classroom practices. It seems to have a direct mapping to the differences students exhibit as a function of cultural orientation to learning [9]. A low PDI score of 40 in the U.S., compared to 77 and 80 in India and China [8], may partially explain these students' general tendencies to speak or remain silent when they attend American university classes, regardless of how well they know the material [6].

This distance is becoming increasingly important to address. International enrollment to American graduate schools has grown since 2005, with the most recent report showing a 17% jump in enrollment to engineering schools and a 40% increase in graduate students from India [7]. These students often fund their education by teaching small classes that act as a supplement to large introductory STEM courses. These small classes, normally called *recitations*, allow groups of undergraduates from a large class to review course material and interact more closely with each other and an expert instructor.

Although many states require ITAs to pass an oral proficiency exam before teaching, there is little support for developing cultural fluency (or even general teaching skills). In other domains, such as health and finance, PI has recently emerged as a technique for motivating changes in behavior [10–13] with only a small investment of time or conscious effort on the part of the user. It is a new class of socio-technical system based on self-monitoring through data visualization [14]. The process helps motivate people to make new decisions by increasing their awareness of behaviors that are normally obscure and hard to observe, such as encouraging more activity by showing people a record of how much (or how little) they move throughout the day. That awareness is a critical step in the process of making changes [12]. These systems have gained popularity due to advances in wearable technology and smartphones. Current PI systems can track a user's number of steps [10], hours and quality of sleep [15], levels of glucose in relation to food intake [16], consumption of non-renewable goods [17], and many more important activities that are hard to monitor without technological assistance.

Research investigating how people use and make sense of PI systems produced a five-stage model of behavior change that applies to a large number of general cases [14]. The model (Preparation, Collection, Integration, Reflection, and Action) describes the types of data users collect, the integration of data collection and reflection into a daily routine, and the transition from reflection to goal setting. The framework provides a list of barriers and design recommendations for each stage. Researchers have recently proposed that incorporating this framework into adaptive training systems may improve classroom interactions [18], but only one project has evaluated such an application. The Live Interest Meter is a PI system that tracks student engagement through a mobile app and provides data visualization to the instructor. It shows the potential to increase audience engagement and instructor responsiveness [19], but at the cost of increased cognitive demand by relying on live manual data input. Our system advances this work by investigating automatic detection of the presence of classroom features that may indicate enhanced learning, such as peer-to-peer interaction and student participation, both of which have been shown to correlate with students' critical thinking in American universities [20], and both of which would likely be difficult for cultural non-natives to enact in their classrooms [21]. Additional strategies for involving students include the use of student names, asking students to elaborate on ideas, and asking deep questions [22].

AIED work has addressed professional development for teachers by means of student tracking and data visualization [23, 24], but these systems have focused on online learning or blended classrooms, and did not offer instructors guidance on how to enact change in a live classroom. Other systems have attempted to visualize student participation (e.g., [25, 26]), but these have been deployed to support students' own self-reflection rather than to support the instructor, and only in online applications where participation can be tracked through clickstream data.

In our work, we advance the state of the art by focusing on the instructor as the primary agent of change. We focus on student participation in class as an achievable goal that is likely to provide academic benefits to students and cultural fluency for ITAs. The current stage of the work includes classroom observations and iterative phases of design for the adaptive system. Specifically, we wanted to answer the following research questions:

1. Do ITAs from a culture with a high PDI encourage active classrooms?
2. Are ITAs open to adapting their teaching style to an unfamiliar cultural context?
3. Are ITAs open to using PI to set and reflect on goals for their teaching?
4. Can we easily and inexpensively sense and create visualizations of classroom activity in terms of TA and student interactions?

3 Method

To answer the research questions, we recruited 5 ITAs, observed them teaching, issued surveys, conducted interviews, and showed them visualizations of their classroom data. We also developed a prototype technical system to detect instructor talk, student talk, and silence.

The TAs were all from India, male, and in their mid-twenties. India has a relatively high PDI (77) compared to the U.S. (40). Each TA had similar levels of teaching experience and content knowledge. None of them had received pedagogical training by the institution or the professor in charge of the course. We observed six to seven sessions of each TA's weekly course, a sophomore level computer science recitation, for a total of 32 sessions. We logged behaviors that would adduce attempts to engage active learning. We inferred activity from frequency and duration of student talk, as opposed to TA talk and silence. We logged the time and locus of all spoken contributions in order to extrapolate episodes of discussion vs. passive lecture.

We surveyed and interviewed the ITAs about their teaching experiences in and perspectives on American classrooms. The survey collected theoretical orientations toward cultural dimensions of learning via items such as demographics, definitions of terms (e.g., "classroom contribution"), and perceived locus of responsibility for learning (e.g., instructor, student, or a combination). We met with each TA three times during the semester (totaling 2.5 – 4 hours per TA) to discuss their survey responses, their perspectives on and motivations for teaching, and to explore their own teaching behaviors with data visualizations.

The data visualizations were initial sketches of what might exist in a PI system. These were meant as a probe for discussion that allowed TAs to reflect on the behaviors they most wanted to capture and view. This is a common technique in the design of new computing systems when there are no design patterns or social conventions to inform the design space [i.e., 27]. We gathered reactions to the visualizations, and redesigned them after each round of feedback. We also probed TAs on their willingness to try new teaching techniques, such as praising students, using students' names, encouraging elaboration, and asking difficult questions. To analyze the results we transcribed the interviews and iteratively searched for areas of strong agreement and disagreement amongst the participants' comments.

Finally, we developed an initial prototype system for a feasibility study, following a typical user-centered design process. We synthesized a set of system needs from the observations and interviews and proposed a minimal set of detection requirements. We developed a prototype system with two Microsoft Kinects and tested it with 20 students and a 60-minute lecture that included various kinds of classroom talk. We hand-coded the audio data with discrete categories of *instructor talk*, *no talk*, and *student talk*. Periods when students talked simultaneously were coded as *student talk*. We tested these categories against the Kinect's angle detection, confidence calculation, and audio amplitude, i.e., whether or not the device picked up sound and if so, where in the room it originated.

4 Findings

Exploring the presence of classroom activity, we observed that ITAs conducted nearly all recitation sections as lectures covering a subset of slides from the most recent primary course lecture. Instructor talk dominated the class, taking up 91.97% of class time (SD=3.6%). Student talk took up only 5.25% of class time on average (SD=2.3%), and the length of their contributions averaged 6.2 seconds (Median=3.4, SD=12.6). The most common prompt for student participation was to ask the class, "Do you have any questions?" The resulting patterns of speech were as follows:

1. TA-talk | silence | TA-talk
2. TA-talk | silence | Student-talk | TA-talk
3. TA-talk | silence | Student-talk | Student-talk

TAs were the first to speak after 85% of their pauses (SD=.088) (pattern 1). 13% of the time (SD=.088) students responded, followed by the TA again (pattern 2). These student contributions were typically brief. 2% of the time (SD .02) a different student contribution followed immediately from a prior student (pattern 3).

Student-student interactions were rare. From an active learning perspective, these interactions are useful as students build on each other's ideas. These conversations were typically animated discussions of the course content that took place in the few minutes before class began. TAs usually called a stop to such interactions in order to begin the lecture, and over the course of the semester most students stopped talking as soon as the TA entered the room. This matched an overall pattern of decreasing student talk (and attendance) for most classes over the semester.

ITAs did express that student participation was important to them, but they defined participation as *students asking or answering questions*. They used that information for diagnosis. TA-2: "*If you ... don't answer [a question asked by the instructor] there is no way for a teacher to know whether you are understanding what he is teaching or what is going on*." Nevertheless, the TAs made lecturing their priority, and student questions were a distraction from this goal. TA-5: "*Maybe I might want to involve their participation a bit more than what it is, but I also fear by doing so [that I won't] be able to complete the contents*."

To explore ITAs' positions on the cultural dimensions of the American classroom, we asked about their explanation for student silence (pattern 1). They speculated that students already understood the content, only had specific questions about their own work, feared appearing dumb, or that they would rather check with peers. When asked how one might increase participation, there were two types of response: ask students if they have questions (TA-1: "*Probably I should ask more times if they have questions*."), and push student to respond to recall questions (TA-3: "*I'll say ... at least take a guess ... I'm sure that one of them will say something*.").

Viewing visualizations of their teaching helped to assess the TAs' stance toward adopting new cultural strategies. At times these graphs triggered immediate motivation for change. When TA-1 saw he talked 99% of the time in the preceding class (Fig. 1), he shared that an interactive class was important to him and that he wanted to include the students more. Yet when he later viewed four weeks of data revealing that he never spoke less than 95% of the time (Fig. 2), he became frustrated with the students. "*I would prefer if the class had more [student participation]. I keep asking if there are any questions, but no one speaks so, I cannot help this one.*"

We probed TAs about their attitudes toward culturally specific strategies for teacher-student interaction. TAs generally agreed that lengthening the pause after asking students a question might be useful and expressed a familiarity with the idea. They showed interest in the tactic of pausing after a student stops talking, and were surprised that it might be valuable. When asked about asking students to elaborate, they expressed skepticism, sharing that students should only elaborate when the instructor does not understand them. We probed them on asking students deep questions from course content as opposed to simple recall questions. This met with mixed reactions. Most worried that asking hard questions would reduce the time needed to cover the material, and all were reluctant to slow down class. TA-5 described his technique of asking content questions in order to highlight important concepts, but only when the questions could be answered rapidly.

We raised the idea of calling on students by their name and of praising their contributions as approaches to create a supportive environment for student participation. Most TAs agreed that these ideas would help students feel valued and might improve their confidence in the learning process, but none of them were willing to employ these techniques. They worried that they might call a student by the wrong name and feel embarrassed, or that calling on a student directly might make them feel picked on. TA-2 shared that calling on specific students would point out that the student had not been speaking and that this might generate shame.

After looking at many visualizations of their classroom behaviors, including talk time, distribution of student participation, number of unique speakers per class, proportions of each event type per class, changes in rates across multiple classes, timelines of event types, and more, almost all TAs expressed an interest in eliciting more student talk, but each spoke about wanting explicit goals for different behaviors. How much is the *right* amount of student and TA talk? How long should the TA wait after asking a question? Are enough of the students participating? Most also asked how their individual data compared to the other TAs in the course. They were all open to the idea of using a PI system to empirically answer these kinds of questions.

Finally, as a first technical step towards a PI system, we built a prototype detector for speaker events meant to identify three states of classroom discourse that would indicate interesting patterns of events when viewed in sequence: (i) instructor speaking (in front of class), (ii) student speaking (from seats), and (iii) no one speaking for at least one second. Researchers have previously had success using microphone arrays for speaker localization [e.g. 28], a process that triangulates the angle of a noise source in relation to microphones placed in a line (the array). We chose to use the Microsoft Kinect, an inexpensive commodity device with a robust microphone array, a developers' kit, and a support community for software development.

In our 60-minute test of various kinds of classroom talk, we evaluated the accuracy of a single Kinect on one side of a classroom and the inclusion of a second Kinect at the front of the room facing the students. We used a *Nominal Logistic Fit for Categories* test (JMP V.10.0) with standard output from the device (angle detection and confidence), and were able to discriminate between students and the instructor with high accuracy (Table 1). We expanded the test to also detect silence by including average amplitude for each second of recorded audio as an input variable. This reduced accuracy overall, but much of that loss was amended by the inclusion of a second Kinect.

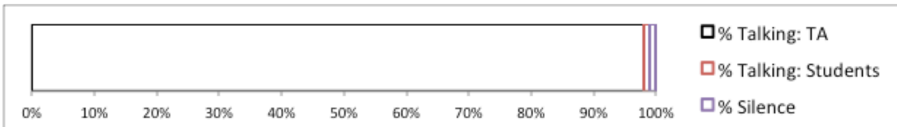

Fig. 1. TA-1's first day of recorded data.

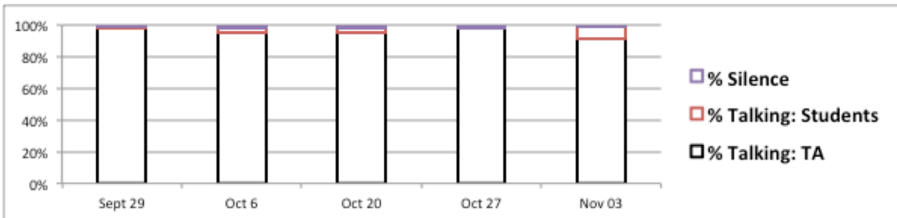

Fig. 2. Aggregate talk time for TA-1 across four classes.

Table 1. Accuracy of Kinects detecting instructor talk, student talk, and no talk

	1 Kinect	2 Kinects
Student/Instructor	94.78%	95.36%
Student/Instructor/Silent	77.70%	85.44%

5 Discussion and Conclusion

Our research explored classroom activity in a university STEM course taught by graduate students from a country with a PDI higher than the host country. We used design methods from PI to better understand the perspectives of ITAs who teach in an unfamiliar cultural context. This process led to the development of a prototype system for identifying levels of classroom activity based on speech events that could indicate higher order discourse phenomena. Our findings suggest that ITAs and their students may benefit from an adaptive feedback system built on measuring levels of classroom activity, and that international instructors would be open to using such a system.

ITAs were open to varying degrees of active learning techniques in their own classrooms. Some were easy for them to imagine using (e.g., pausing after students talk), and others were harder to accept (e.g., asking for elaboration). They showed reluctance to decrease the amount of time spent "covering" critical course material, yet they all valued when students got involved in the lecture. These tensions are clues that an adaptive system for cultural training may need do more than measure and report on behavior, but also provide scaffolding for implementing relatively low-cost active learning strategies, such as *think-pair-share*. The next step would be to assess the user's knowledge and stance toward different contextual behaviors and provide individualized instruction and adding more advanced scaffolding prompts as the TA becomes ready for them. Future research would need to navigate this complex space. To refine the detection system further and more easily differentiate between user states, it would be possible to include machine learning and more factors than we currently use, such as Kinect error rates, classroom details, pitch fluctuations and filters, and so on. With more tuning the system might identify individual speakers, leading to reflection opportunities based on individual student speaking patterns. Turn detection at this level could point out disproportionate properties of classroom talk, such as a group of dominant speakers.

There are aspects of the classroom that the proposed system would not be able to detect. ITAs were curious about whether they had lectured for "too long." They made reasonable requests, such as seeing when they had made a "good" explanation, or if students understood the material. A fully operational PI system would necessarily need supplemental human input to provide such feedback, which is already standard practice in current systems: much like annotating the quality of a recent jog when using a fitness-tracking app, our proposed system could request post-class assessments from students or the TA. Some TAs remarked that it would be a simple procedure to personally label the broad topic of the class, or the context of specific pauses throughout the lecture if they were able to review the data and access the audio. Although previous PI systems have not explored user input this deeply, such interactions would be possible to implement, and may be critical for system design.

Our study only observed one genre of recitation, but there are many others. It is critically important to assess how much the observed behaviors in this study were an artifact of culture, context, or simply being new to teaching. In our current work we are performing additional observations of a broad selection of classroom contexts taught by students from many different cultural backgrounds in order to assist in making these distinctions.

Research in professional development for teachers might note that our work did not address the quality of interactions, but only quantity and abstract patterns of discourse. As a first step, we argue that any increase in student talk would more closely align with the cultural context of the U.S. classroom, although in the future quality may prove to be a critical area of investigation. Currently, however, the space of cultural acquisition for graduate students and the professional development of novice instructors is under-investigated, and thus this early work makes a contribution.

The implications of this research are important in their potential to address the lack of research in supporting the cultural fluency of ITAs in a challenging new environment. Our work shows preliminary evidence that PI could be an approach to support reflection on classroom dynamics and an opportunity to adaptively expand an instructor's set of pedagogical tools. The impact of the work points to a better experience for international graduate students and potentially better learning for their students.

6 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.

7 References

1. Faust, J. L., Paulson, D. R.: Active learning in the college classroom. Journal on Excellence in College Teaching, 9(2), 3–24 (1998)
2. Michael, J.: Where's the evidence that active learning works? Advances in Physiology Education, 30(4), 159–67 (2006)
3. Lujan, H. L., DiCarlo, S. E.: Too much teaching, not enough learning: what is the solution? Advances in Physiology Education, 30(1), 17–22 (2006)
4. Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., Wenderoth, M.P.: Active learning increases student performance in science, engineering, and mathematics. Proc. Natl. Acad. Sci. U. S. A. 111, 8410–5 (2014)
5. Nisbett, R.: The Geography of Thought: How Asians and Westerners Think Differently... and Why. Simon and Schuster. (2010)
6. Liu, J.: Asian students' classroom communication patterns in US universities: An emic perspective. Greenwood Publishing Group. (2001)
7. Allum, J. R.: Findings from the 2013 CGS International Graduate Admissions Survey Phase III : Final Offers of Admission and Enrollment, 20036 (2013)

8. Hofstede, G., Hofstede, G. J., & Minkov, M.: Cultures and organizations: Software of the mind: 3rd edition. McGraw-Hill Professional. (2010)
9. Wang, M.: Designing online courses that effectively engage learners from diverse cultural backgrounds. British Journal of Educational Technology, 38(2), 294–311 (2007)
10. Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G., Strub, H.B.: Fish'n'Steps: Encouraging physical activity with an interactive computer game. UbiComp 2006. pp. 261–278. Springer-Verlag (2006)
11. Hsieh, G., Li, I., Dey, A., Forlizzi, J., Hudson, S.E.: Using visualizations to increase compliance in experience sampling. Proc. UbiComp '08. 164 (2008)
12. Li, I., Dey, A.K., Forlizzi, J.: Using context to reveal factors that affect physical activity. ACM Trans. Comput. Interact. 19, 1–21 (2012)
13. Levy, L., Jones, B., Robertson, S., Price, E.D.: Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. ACM Trans. Comput. Interact. 20, (2013)
14. Li, I., Dey, A., Forlizzi, J.: A stage-based model of personal informatics systems. CHI 2010. pp. 557–566. ACM, Atlanta, Georgia, USA (2010)
15. Kay, M., Choe, E.K., Shepherd, J., … Kientz, J.A.: Lullaby: A Capture & Access System for Understanding the Sleep Environment. UbiComp '12. ACM, Pittsburgh, USA (2012)
16. Mamykina, L., Mynatt, E.D., Kaufman, D.R.: Investigating health management practices of individuals with diabetes. Proc. CHI '06. 927 (2006)
17. He, H.A., Greenberg, S., Huang, E.M.: One size does not fit all: Applying the transtheoretical model to energy feedback technology design. CHI 2010. 927–936 (2010)
18. Muller, R., Long, Y., Koedinger, K.R.: Personal informatics for self-regulated learning. In: Personal Informatics in Practice: Improving Quality of Life Through Data — CHI 2012 Workshop. CHI, Austin, TX (2012)
19. Rivera-Pelayo, V., Munk, J., Zacharias, V., Braun, S.: Live interest meter: Learning from quantified feedback in mass lectures. Proc. 3rd Int. Conf. on Learn. Anal. and Knowledge. 23–27. ACM, Leuven, Belgium (2013)
20. Smith, D.G.: College classroom interactions and critical thinking. J. Educ. Psychol. 69, 180–190 (1977)
21. Gerritsen, D., Zimmerman, J., & Ogan, A.: A theoretical approach to the development of critical incidents for cultural training. 5th Int. Workshop on Culturally-Aware Tutoring Systems. 51–60 (2014)
22. Nunn, C. E.: Discussion in the college classroom: Triangulating observational and survey results. J. of High. Educ., 243–266 (1996)
23. Mazza, R., Milani, C.: Exploring Usage Analysis in Learning Systems: Gaining Insights From Visualisations. AIED'05 Work. Usage Anal. Learn. Syst. 65–72 (2005)
24. Scheuer, O., Zinn, C.: How did the e-learning session go? The Student Inspector. Proc. Conf. Artif. Intell. Educ. 487–494 (2007)
25. Janssen, J., Erkens, G., Kanselaar, G., Jaspers, J.: Visualization of participation: Does it contribute to successful computer-supported collaborative learning? Comput. Educ. 49, 1037–1065 (2007)
26. Smith, M.K., Vinson, E.L., Smith, J. a., Lewin, J.D., Stetzer, M.R.: A Campus-Wide Study of STEM Courses: New Perspectives on Teaching Practices and Perceptions. Cell Biol. Educ. 13, 624–635 (2014)
27. Davidoff, S., Lee, M., Dey, A., Zimmerman, J.: Rapidly Exploring Application Design Through Speed Dating. UbiComp 2007. 429–446 (2007)
28. Valin, J., Rouat, J., Dominic, L.: Robust Sound Source Localization Using a Microphone Array on a Mobile Robot. Proc. IEEE/RSJ (IROS); p. 1128–1233. (2003)

Culture-Oriented Factors in the Implementation of Intelligent Tutoring Systems in Chile

Ignacio Casas[1], Patricia Fernandez[1], Marcia Barrera[1], Amy Ogan[2]

[1]School of Engineering, Universidad Católica de Chile, Santiago, CHILE
[2]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA
{icasas, pifernac, mbarrerb}@uc.cl, aeo@andrew.cmu.edu

Abstract. With the aim of assessing the use of intelligent tutoring technology for math teaching in Chilean public schools, an experimental study was performed in the period 2013-2014. Although it was a successful experience in terms of number of participants and learning outcomes, it was not achieved without a number of difficulties which could be explained by focusing on the cultural challenges encountered in the endeavor. In this paper we explore the impact of cultural dimensions such as: organizational strategies and structure; organizational culture; pedagogical processes, human resources, and technology deployment. We characterize each one of these aspects by means of a qualitative study of the implementation process, involving tasks such as planning and technical support, class observations, interviews, and support to teachers in the classroom and lab. As a result, we propose a Diagnostic Chart which could help in the identification of pre-conditions to be solved at an earlier stage of the implementation phase.

Keywords: Intelligent tutoring experimentation; teaching strategies; country-specific developments; evaluation of CAL systems

1 Introduction

We describe a qualitative study focused on cultural issues encountered in the implementation of intelligent tutoring technology for Chilean public middle schools (5th to 8th grade in a K-12 system)[1]. The experimentation was carried out during two academic years (2013, 2014) and one of its objectives was to understand the challenges faced by teachers, students and authorities when engaged in the change of their teach-

[1] By implementation we refer to the complex endeavor of introducing new strategies and technology into the teaching-learning processes. This includes development and adaptation of software tools, planning, training, demos, on-line and field support.

ing-learning strategies by means of intelligent tutors[2]. The long range vision is to improve math learning in public education for underserved populations.

Based on the literature and the experimentations' findings, we have identified culture-oriented critical factors to be dealt with when implementing an intelligent tutoring system environment in the math class. From this characterization we construct a Diagnostic Chart which could help identifying pre-conditions to be solved at an earlier stage of the implementation process.

The implementation endeavor includes the development of a pedagogical framework that, considering scarce technological resources, takes advantage of personalized student-centered activities in the computer lab and collaborative-constructivist strategies in the classroom. Even though the ultimate goal has been to improve math learning among students, the core methodology has focused on the teachers: they provided training for teachers and implemented teaching support tools. In the training courses, the new technology-based strategies were socialized, situated and adapted to local contexts. We wanted to make sure teachers felt motivated and are willing participants-leaders of the required change process. After training, we provide constant support and follow-up of the implementation in the classroom and lab.

The focus is on the tools and support activities needed by teachers to adequately implement the new technology-enhanced teaching strategies. This involves substantial change in the teacher's attitude, motivations, activities, and plans. The teachers need training, time and support for studying and planning the new classroom-lab strategies. It involves major changes in planning, instructional design and the teaching processes itself; it is a complex task. We have identified that once the basic technology issues are resolved (computer labs with one functioning PC for each student, reliable local area networks, client software correctly installed, sufficient Internet access to the servers, and effective technical support), there are several cultural-organizational drawbacks that work against a successful implementation. Most teachers complain about the extra effort required for the process.

To understand the particularities associated with setting up a class on an intelligent tutoring environment, we first describe the technology and its strategies.

1.1 Cognitive Tutor Technology

Following the theoretical principles developed by Anderson [1], [2], a personalized digital learning system known as a Cognitive Tutor (CT) was built at Carnegie Mellon University and is maintained and operated by Carnegie Learning Inc.[3] In this software, each student has a personalized "problem-solving" space, with just-in-time feedback and detailed tracking of his or her progress [3]. 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

[2] We acknowledge the generous support of district-municipality authorities, school principals and teachers together with funding from the Inter-American Development Bank (grant ATN/KK-11117-RS) and CONICYT-Chile (project FONDEF-D10i1286).

[3] Cognitive tutoring technology is a trademark property of Carnegie Learning Inc.

help and solve problems. It provides personalized feedback and hints when errors are made in key points [4].

Cognitive tutors have shown considerable potential, and evidence in the literature indicates that they are effective in improving mathematics and science problem-solving skills [5], [6]. 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 [7].

1.2 Cognitive Tutor Strategies

The main objective of the CT software is to provide each student with a unique, enriched environment where he/she can interact with the system by solving specific problems. Multiple graphical representations can be explored by the student for creative thinking practice [8], [9], [10].

The software presents a problem and the student is requested to work towards the solution. Instead of jumping to the final answer, the software provides step-by-step scaffolding [11]. This divide-&-conquer strategy asks specific questions, from easier to more complex, so that 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. The next questions (posed by the software) guide the student towards the solution of the problem[4].

The student gets feedback (positive or negative points in a roster of skills to be achieved) whenever he/she answers questions within a problem. This immediate feedback is continuously represented via a “skill-o-meter” in the interface of the tutor [12]. Based on the “skill-o-meter” we have developed a web-based tool that provides teachers with a complete view of students’ 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 topics of struggle [13].

2 Experimental Study

The broad objective of the study is to understand how the culture-oriented challenges, that may be an obstacle for the implementation of an intelligent tutoring system in schools, can be characterized to detect deal-breaker barriers at an early stage of the implementation. We state that dealing with these obstacles is a condition sine qua non to successfully engage teachers, school authorities and students in an intelligent tutoring environment, hence the importance of achieving this goal.

The key questions are: Which are the culture-oriented challenges that can be identified during the experimentation? Which are the critical factors that can be deduced from the cultural challenges? Are there verifiable achievement indicators that can be

[4] There is extensive literature with thorough description of CT technology ([2], [4], [6], [7], [8]).

linked to those challenges? How can these indicators be arranged into an evaluation instrument to be used as a guideline for teachers and school authorities in the process of setting up an intelligent tutoring implementation?

2.1 Methodology

Building from experiences in USA, the Chilean initiative seeks an important innovation: the definition and application of new teaching strategies that, based on the CT technology, are adapted to the local educational context. This starts with the negotiation of change strategies with the district and school authorities. It follows with the involvement of teachers in training and instructional design blended-courses (90% of work is on-line) based on the CT. It culminates with the implementation of the technology-supported strategies in the math classroom.

At an early stage, we decided to work with public Chilean schools (totally or partially dependent of Municipalities) which enroll the largest percentages of vulnerable students and present the lowest learning results. These are the students with most diminished education opportunities explained by the lack of household economic resources. Once the schools were selected and authorities had committed their support, we provided training for teachers to engage them in the new strategies and technologies. Teacher involvement was the most critical issue in the implementation plan. The training goal was to achieve high motivation and strong commitment of the teachers towards the new technology-based strategies. However, a common denominator that plays against this goal is a dramatic lack of time for innovations on the part of the teachers. We also checked the technological infrastructure at the schools, providing support and solutions when needed[5].

In addition to the definition of the pedagogical strategies, we took an English version of the software content and, considering cultural and contextual differences, transformed it into a Spanish version. 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 culture. Finally, we have conducted activities to collect the data needed for constructing the Diagnostic Chart.

2.2 The Sample (2013-2014 Implementation)

In general, the selection of the participating districts was a difficult process. It is obvious that without full support and involvement of the district authorities, implementation was impractical. There were some initially invited districts that were necessary to discard due to their lack of real involvement. All schools within a district were invited to participate, but only a few of them decided to experiment with the CT technology.

During the implementation process, a number of treatment schools dropped out for different reasons: problems with infrastructure, lack of involvement in training, reluc-

[5] Even though the technological infrastructure of public schools in Chile is generally adequate, in some cases we needed to provide local servers and networks due to low connectivity.

tance toward teaching changes, and lack of support from school authorities. Due to the training process most participating teachers were enthusiastic and willing to adopt the new strategies and technology. Some teachers (about 20% of initial participants) didn't have enough time to complete the training. The later ones constituted drop-outs from the implementation and in some cases the school as a whole could not participate. Table 1 shows the total number of participants separated by geographic location (Villarrica is mainly a rural area.)

Table 1. Total number of participants by geographic location

	Schools	Teachers	Courses	Students
Santiago	17	36	76	2915
Villarrica	5	7	14	340
Others	4	6	8	95
TOTAL	26	49	98	3350

2.3 Culture-Oriented Challenges

Culture-oriented challenges continue to be a significant obstacle in the adoption of new technologies for the classroom and lab as means of improving teaching practices [14]. Based on the literature and best practices in industry [15], in our experimentation we have identified a number of these challenges, which rise up as significant barriers to be dealt with in the implementation of intelligent tutors[6]. We have grouped them in 5 categories or dimensions: (1) Pedagogical processes (teaching & learning); (2) Organizational strategies and structure; (3) Organizational culture (teacher's attitudes towards change and technology); (4) Human resources (teachers' skills and knowledge; student attitudes); (5) Technology acquisition and deployment.

A characterization of these dimensions can be obtained by a series of questions to be answered during the study (i.e., observations, interviews, empirical data analysis), as follows.

(1) **Pedagogical processes (teaching & learning):** Are the actual teaching processes adequate for improved learning? Are these processes student-centered or teacher-centered? Is the technology used to innovate (and improve) the teaching process or just to micro-improve a specific task (i.e., projectors for lectures, e-books for reading)?

(2) **Organizational Strategies and Structure:** Are the organization's structures and strategies adequate to motivate, lead and perform effective changes in the teaching processes? Is it feasible to implement changes in the classroom? Do authorities facilitate resources (equipment, time for training, planning, and implementation) to involved teachers?

[6] We focus here on "organizational" culture as opposed to "ethnical" culture. Notwithstanding, there are organizational issues that may be influenced by the local culture, such as dealing with scarce resources, poor planning and assessment, social unrest, vulnerable student communities, etc.

(3) **Organizational Culture:** Are teachers comfortable-satisfied with the actual pedagogical strategies? Are they committed to introduce changes for improvement? Using the CT technology, was it possible to change the classroom-lab processes? Were the resources assigned (by school authorities) adequate? Were there other critical factors? Do teachers perceive that the resources and support for innovation are adequate?

(4) **Human Resources:** Is the teacher's level of proficiency in the domain (math) adequate for teaching? Do teachers master the features present in the CT technology? Are the teachers confident on the contributions of technology for improved learning? Are they confident on the CT technology? What is the student's attitude towards learning, technology and math?

(5) **Technology Acquisition and Deployment:** Are there enough computers in the lab for a "one computer per student" strategy? Are there enough local area networks (e.g., Wi-Fi) to support the use of the new technology? Is there a sound Internet connection and Web services? Does the school have appropriate technical support?

3 Results and Discussion

Using assessment instruments such as interviews and surveys, during the experimentation we have identified specific factors for each dimension of culture-oriented challenges. These factors can be evaluated by means of achievement indicators. The set of dimensions, factors and achievement indicators provide a coherent characterization of culture-oriented challenges found in our study. What follows is a brief description of factors and indicators for each dimension.

3.1 Factors and Achievements for Culture-Oriented Dimensions

As shown in Table 2, within the "Pedagogical Processes" dimension we have identified two factors: teaching strategies and teaching tools.

Table 2. Factors and Indicators for Pedagogical Processes (Dimension 1)

Factor	Achievement Indicator
Teaching strategies	Facilitates a student-centered process v/s teacher-centered.
Teaching tools	Use of technology tools Use of other resources in the classroom (hands-on material, etc.)

The "Organizational Strategies and Structure" dimension addresses school's organizational structure and strategies for teaching-learning innovations. In this matter, school's authorities have the main saying; they should be motivators and promoters of transformations in the classroom. If authorities are open to changes, it is necessary to verify the feasibility of these transformations. Table 3 summarizes factors and achievement indicators for this dimension.

Table 3. Factors and Indicators for Organizational Strategies and Structures (Dim. 2)

Factor	Achievement Indicator
Authorities motivated towards changes	Interested in innovative pedagogical activities (with or without technology). Comfortable with current teaching strategies. Encourages teachers towards changes. Values the use of technology for teaching-learning. Positive evaluation of CT as a new learning strategies
Feasibility of implementation	Facilitates pedagogical innovations in the school. Facilitates the use of technology in the classroom.
Resources for teacher	Provides enough time for planning activities. Provides extra time for training activities. Provides enough time for implementation. Encourages school community involvement in innovation. Provides resources.

As part of the third dimension, organizational culture of a school, teachers are the most important agents of change and innovation in the classroom. Table 4 shows factors and achievement indicators for this dimension.

Table 4. Factors and Indicators for Organizational Culture (Dimension 3)

Factor	Achievement Indicator
Teacher's motivation towards change	Open to innovative pedagogical activities (with or without technology). Performs innovative pedagogical activities (with or without technology). Feels pleased about current teaching strategies. Encourages other teachers towards changes. Values the use of technology and CT for teaching.
Feasibility of implementation in the school	There is enough time for re-planning learning activities. There is enough time for attending training sessions. There is enough time to carry out the implementation. The school community is engage and supportive towards innovation. There are resources to carry out the innovation activities.
Training in new contents, methods and tools	Interest in training. Suggests training opportunities to his or her colleagues and school authorities. Participates in training sessions (school authorities initiative) Participates in training sessions (personal initiative)

Within the "Human Resources" dimension, we consider teachers and students as shown in Table 5.

Table 5. Factors and Achievement Indicators for Human Resources (Dim. 4)

Factor	Achievement Indicator
Teacher's tech skills	Mastering technology, at a user level: Internet, desktop tools.
Teacher's attitude towards technology	Introduction of technology into the annual or semester class planning Positive opinion towards the use of technology for teaching.
Teacher's self-perception towards math	Self-confidence on knowledge for domain area. Masters the learning objectives of the grade he/she teaches.

Factor	Achievement Indicator
Teacher's mastery level of CT software (technology and contents)	Check the lessons in "student" mode. Identifies fundamental strategies present in the CT software Understands CT methodology for problem solving and scaffolding
Teacher's confidence with technology based strategies	Self-confidence on his/her technology skills Comfort level regarding technology
Student's attitude towards technology	Interested in carrying out activities using technology Positive opinion towards the use of CT in the math classroom High level of comfort in using CT for math learning
Student's attitude towards math	Improved perception about math after using the CT technology

Factors and achievement indicators for the "Technology Acquisition and Deployment" dimension are shown in Table 6.

Table 6. Factors and Indicators for Technology Acquisition and Deployment (Dim. 5)

Factor	Achievement Indicator
Computers availability	Feasibility for adapting a one-computer-per-student strategy.
Internet connection and local networks	Sufficient Internet access and local area networks for full deployment of one-computer-per-student in a class.
Technical support	Permanent technical support staff for the lab. Lab administrator present during lab sessions.
Exclusive dedication of technical resources	Technical resources used exclusively for educational purposes (as opposed to administrative).

3.2 Diagnostic Chart

Following the dimensions, factors and indicators presented in the previous section, we have constructed a Diagnostic Chart of culture-oriented factors. With this tool we can pin-point those issues that seriously impact or endanger the feasibility of the implementation. Even though the chart is a result of our experimentation, it could be used in future studies to identify pre-conditions to be solved at an earlier stage of an intelligent tutoring endeavor.

Table 7. Diagnostic Chart Application: Critical Factors for Drop-Out Schools

School	Culture-Oriented Factors that Constrained the Implementation
School 1	**Dim 2:** Authorities (school principal and academic coordinator) were not motivated towards changing the actual teaching methodology. **Dim 4:** Lack of technological skills among teachers.
School 2	**Dim 2:** Authorities (school principal and academic coordinator) were not motivated towards changing the actual teaching methodology. **Dim 5:** No enough computers; lack of a reliable Internet connection; lack of technical support.
School 3	**Dim 2:** Authorities (school principal and academic coordinator) were not motivated towards changing the actual teaching methodology. **Dim 5:** Lack of a reliable Internet connection and local networks.
School 4	**Dim 5:** Lack of a reliable Internet connection and local networks.
School 5	**Dim 3, 4:** Teachers not open to change. Teachers do not value the use of CT technology.

We have used the diagnostic chart to assess the results of the experimentation with 26 schools in urban and rural areas. Out of 26 participating schools, 5 of them showed culture-oriented issues that endangered the implementation effort (resulting in drop-outs). These drop-outs and related inhibiting factors are shown in Table 7.

It could be inferred from Table 7 that the most frequent culture-oriented inhibitors (in our experimentation) are the ones related to "Technology Acquisition and Deployment" (Dim. 5), "Organizational Strategies and Structures" (Dim. 2) and "Human Resources" (Dim. 4).

4 Conclusions

Based on the analysis of culture-oriented factors encountered during our experimentation, we have constructed an instrument that helps identifying schools likely to drop out from an intelligent tutoring endeavor. Although the sample size is relatively small (5 out of 26 schools drop-out), observations in the field clearly highlight those factors which are critical in the implementation.

Cultural factors that had more impact on our experimentation (diminishing though the feasibility of implementation) are, in order of importance:

- Innovation is not facilitated by school authorities; no interest on innovative technologies.
- Lack of adequate Internet connection and local area networks.
- Lack of positive attitude towards changes (authorities and teachers).
- Teacher's claim that there are not enough resources to implement.

It can be noticed that there were no cultural issues related to students. According to our surveys and interviews, all drop-outs were due to problems with infrastructure, reluctance toward teaching changes, and lack of support from school authorities. Despite the sense that change was difficult for the teachers and administration, the fact that 100% of non-drop-out teachers and authorities want to continue using the CT technology in the future is an encouraging result that shows motivation and willingness to change once the value of the new technology is established.

5 References

1. Anderson, J.R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science 26,* 85-112.
2. Koedinger, K., Anderson, J.R., Hadley, W., & Mark, M.A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8,* 30-43.
3. Ritter, S. (2011). *The research behind the Carnegie Learning math series* [Online]. Available: http://www.carnegielearning.com.
4. Koedinger, K., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review 19,* 239-264.

5. Koedinger, K., & Corbett A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In *The Cambridge Handbook of the Learning Sciences* (pp. 61-78). Cambridge University Press.
6. Ritter, S., Anderson, J.R., Koedinger, K., & Corbett, A. T. (2007). The cognitive tutor: applied research in mathematics education. *Psychonomics Bulletin & Review 14(2)*, 249-255.
7. Arroyo, I., Woolf, B., Royer, J.M., Tai, M., & English, S. (2010). Improving math learning through intelligent tutoring and basic skills training. In Aleven, Kay and Mostow (Eds.), *Intelligent Tutoring Systems Part I* (pp. 423-432), *Lecture Notes In Computer Science 6094*.
8. Feenstra, L., Aleven, V., Rummel, N., & Taatgen, N. (2010). Multiple interactive representations for fraction learning. *Proceedings 10th Int. Conference on Intelligent Tutoring Systems (ITS), (*pp. 221-223).
9. Rau, M., Aleven, V., & Rummel, N. (2009). Intelligent tutoring systems with multiple representations and self-explanation prompts support learning of fractions. *Proceedings 14th Int. Conference on Artificial Intelligence in Education (AIED),* (pp. 441-448).
10. Wegerif, R., MacLaren, B.M., Chamrada, M., Scheuer, O., Mansour, N., Miksatko, J. & Williams, M. (2010). Exploring creative thinking in graphically mediated synchronous dialogues. *Computers & Education 54(3),* 613-621.
11. Luckin, R. (2008). The learner centric ecology of resources; a framework for using technology to scaffold learning. *Computers & Education 50,* 449-462.
12. Mitrovic, A., Ohlsson, S., Barrow, D.K. (2013). The effect of positive feedback in a constraint-based intelligent tutoring system. *Computers & Education 60,* 264-272.
13. Schnaubert, L., Andres, E., Narciss, S., Eichelmann, A., Goguadze, G., & Mellis, E. (2011). Student behavior in error-correction-tasks and its relation to perception of competence. In *Towards Ubiquitous Learning Proceedings of the 6th European Conference on Technology Enhanced Learning* (pp 370-383), Berlin: Springer.
14. Casas I., Imbrogno J., Ochoa S., Ogan A. "Cultural Factors in the Implementation and Use of an Intelligent Tutoring System in Latin America", Proceedings AACE E-Learn 2014 World Conference on E-Learning, New Orleans, USA, Oct. 2014.
15. Laube, D.R., Zammuto, R.F., Editors. "Business Driven Information Technology", Stanford University Press, 2003.

More Fun in the Philippines? Factors Affecting Transfer of Western Field Methods to One Developing World Context

Juan Miguel L. Andres[1], Ma. Mercedes T. Rodrigo[1], Jessica O. Sugay[1], Michelle P. Banawan[1,2], Yancy Vance M. Paredes[1,2], Josephine S. Dela Cruz[3], Thelma D. Palaoag[3]

[1] Ateneo de Manila University, Quezon City, Philippines
[2] Ateneo de Davao University, Davao City, Philippines
[3] University of the Cordilleras, Baguio City, Philippines
{mandres, mrodrigo, jsugay}@ateneo.edu
{mpbanawan, yvmparedes}@addu.edu.ph
{delacruzpen, tpalaoag}@gmail.com

Abstract. This paper presents some of the challenges encountered by a field research team when deploying an educational game for Physics. These included problems with site infrastructure and institutional support, logistical challenges, compliance with ethics requirements, launch delays, and student inattention or misunderstanding of directions. The paper shares these experiences with the wider community to help fellow researchers prepare, should they decide to conduct field studies in the Philippines.

Keywords: intelligent tutoring systems · research methods · field study· Physics Playground

1 Introduction

In 2012, two experienced human-computer interaction researchers said, "Fieldwork takes you to strange locations to meet new people. Despite the best-laid plans, surprises will happen and some amount of mayhem will ensue [5]." Nowhere is this more true than during attempts to transfer software or field methods from a developed country to the developing world. Because the software or field methods are usually designed in and for developed countries, the assumptions made during the design process and the circumstances surrounding deployment vary, sometimes extremely, from ground conditions in other countries. When describing the deployment of an American intelligent tutoring system in Brazil, Ogan and colleagues [4] found that most students had no computers in their homes, that teachers had little to no technology expertise and were not familiar with ways in which computers could be used for education. On a technical level, schools had a limited number of computers for student use and the ones that were available were often riddled with viruses. Other barriers discussed extensively in [2] include data costs, Internet reliability, the availability and reliability of electricity, and localization of content in terms of both culture and language.

Since 2006, the Ateneo Laboratory for the Learning Sciences (ALLS) has been conducting field studies in different schools all over the Philippines. In [6], key members of ALLS documented five of the challenges of transferring Western educational software and study methods to the Philippines. As in both [2] and [4], [6] observed that the overall level of technology adoption for education was generally low and that technology infrastructure was generally limited. [6] further added that school support, while essential, was not always easy to obtain. Students were culturally conditioned to be respectful of authority, therefore the presence of observers sometimes had an effect on behavior. Finally, typhoons are common occurrences in the Philippines. In one field experiment, they disrupted data gathering and introduced a possible confound: post-traumatic stress.

The goal of this paper is to present the challenges that confronted another ALLS research team during a more recent study. The goal of the paper is to describe additional considerations that researchers should take into account when planning field studies.

"It's More Fun in the Philippines" is the country's official tourism tagline, which presents how otherwise mundane activities such as commuting (as seen in Fig. 1) are more fun in the country by highlighting places, activities, and artifacts that are uniquely Filipino.

Fig. 1. Example poster of the "It's More Fun in the Philippines" tourism campaign.

2 Description of the Field Study

Data from 180 students was collected over three weeks from January to February 2015 in three schools (Sites A, B, and C) in different regions of the Philippines. The goals of the study were to assess the persistence and affect of students using an educational game for Physics, and to determine any differences among the different regional groups. The subsections that follow describe the methods and materials used to these ends.

2.1 Learning Environment

Data was gathered from students using Newton's Playground (now Physics Playground, PP). PP is a computer game for physics that was designed to help secondary school students understand qualitative physics. Qualitative physics is a nonverbal, conceptual understanding of how the physical world operates [7].

PP is a two-dimensional computer-based game that requires the player to guide a green ball to a red balloon. Two example levels are shown in Fig.1. PP has 74 levels that require the player to guide a green ball to a red balloon. The game presents these levels divided into eight different playgrounds. The player achieves this goal by drawing agents (ramps, pendulums, springboards, or levers) or by nudging the ball to the left or right by clicking on it. The moment the objects are drawn, they behave according to the law of gravity and Newton's 3 laws of motion [7].

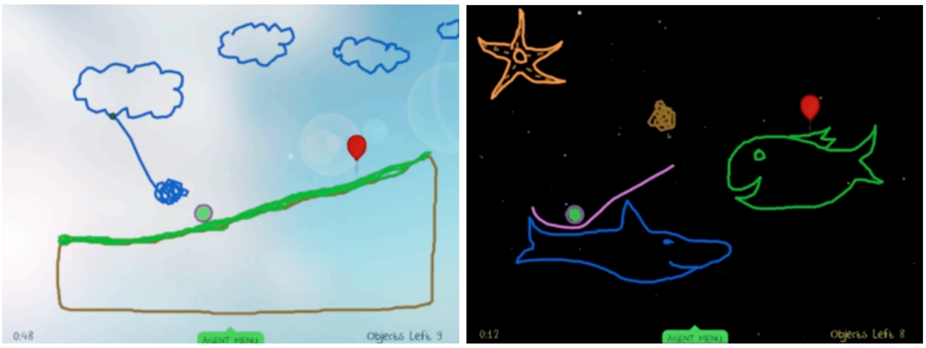

Fig. 2. Example PP levels.

A ramp is any line drawn that helps to guide a ball in motion. A ramp is useful when a ball must travel over a hole. A lever rotates around a fixed point, usually called a fulcrum or pivot point. Levers are useful when a player wants to move the ball vertically. A swinging pendulum directs an impulse tangent to its direction of motion. The pendulum is useful when the player wants to exert a horizontal force. A springboard stores elastic potential energy provided by a falling weight. Springboards are useful when the player wants to move the ball vertically. In Fig. 2, the level on the left requires a pendulum, and the level on the right requires a lever.

During gameplay, PP automatically generates log files. Each level a student plays creates a corresponding log file, which tracks every interaction the student has with the game in terms of particular counts and times for selected features of gameplay.

2.2 Participants

Data were gathered from 180 students in the Philippines, equally divided among three geographical locations in the country: 60 eighth grade students from Baguio City, 60 tenth grade students from Cebu City, and 60 eighth grade students from Davao City.

2.3 The Observation Protocol

The Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP) is a protocol for quantitative field observations of student affect and engagement-related behavior, described in detail in [3]. The affective states observed within Physics Playground in this study were engaged concentration, confusion, frustration, boredom, happiness, delight, and curiosity. The affective categories were drawn from [1].

Participants were divided equally among the two to three BROMP-certified observers present per session. Students were observed in 5 to 8 second intervals through each site's respective observation period, resulting in at least one observation per student per minute. If the student exhibited two or more distinct states during his or her respective observation period, the observers only coded the first state.

The observers recorded their observations using the Human Affect Recording Tool, or HART. HART is an Android application developed to guide researchers in conducting quantitative field observations according to BROMP, and facilitate synchronization of BROMP data with educational software log data.

2.4 Data Collection Methods

Before playing PP, students completed a demographics sheet and a 16-item multiple-choice pretest. Students then played the game for a certain period of time (i.e., 90 minutes in Site A, 75 minutes in Site B, and 30 minutes in Site C), during which the trained BROMP observers coded student affect and behavior on the HART application. After completing gameplay, participants completed a 16-item multiple-choice posttest. The pretest and posttest were designed to assess knowledge of physics concepts, and have been used in previous studies involving PP [7].

3 Challenges Encountered

Poverty is intrinsic to the Philippine situation, and as such, the adoption of information and communication technologies (ICTs) in the classrooms of the Philippines has been slow and marred by hindrances and limitations. Of the 46,000 public schools run by the country's Department of Education (DepEd), for example, about 8,000 have no power, and even more have no connectivity. There also exists a tremendous need for ICT integration in pre- and in-service teacher training in order to gain appreciation for the use of technology in the curriculum and in the classroom.

As in [6], infrastructure and institutional support remained challenging. This field study also introduced new challenges in terms of logistics, compliance with ethics requirements, launch delays, and student inattention or misunderstanding of directions.

3.1 Infrastructure

In preparation for data gathering, arrangements were made with on-site counterparts to have the software installed and tested prior to the arrival of the research team. PP requires several peripherals in order to launch smoothly. An error thrown by any of these necessary components can cause faulty data capture, which can result in having to throw out gathered data, or cause the game not to run at all. The three main components necessary for PP to run are 1) the software itself, 2) a steady Internet connection not blocked by a firewall or proxy, and 3) a webcam to record the participants' facial expressions.

A previous research project outside of this project's scope already required the team to install and debug the system in the past. Hence, the research team had solutions to problems encountered before. Site A, however, experienced problems with the installation of the software and hardware drivers, which required around three hours of debugging possible conflicts in the computer laboratory's system configurations, including webcam driver incompatibilities and the unstable Internet connection. PP had been running smoothly on one machine, but continued to encounter launch errors on every other machine in the computer laboratory. The team eventually found that the machines were configured to use a virtual environment, which was causing conflicts with the PP software installation and webcam drivers. Once the virtual environment was disabled, PP ran smoothly.

PP's Internet connection posed a technical challenge. The Internet connection was essential for the game's timing functionality to run smoothly. The timing functionality's main purpose is to synchronize all interaction events with Internet time, allowing for a unified set of timestamps for all the participants, as well as for the BROMP coders. Having to synchronize multiple data sources (including human-recorded data) into a single time-stream is a challenge all on its own; having to deal with time inconsistencies in the process makes the task much harder, and the resulting analyses less accurate.

This timing functionality on PP can be turned off optionally (though it is not advised), requiring the research team to take note of session start times manually. Computer labs are usually protected by firewalls and proxies, and as such, the research team had made it a point to request for a firewall exception and for proxies to be disabled a week before data gathering. The research team had to disable the timing functionality of the software in Site B because the administration would not allow addition of a firewall exception for the timeserver. Another solution to this issue could be the use of a local time server.

Another critical issue of PP is that, in order to ensure that the interaction logs and video files are properly saved to secondary storage, the software must exit cleanly. On several occasions, the research team observed that the software did not exit properly. This was consistently experienced in Site B, wherein the software had to be forcefully terminated before log files could be retrieved from temporary folders. Conversely, the problem was only encountered on two occasions in Site A, and never in Site C.

3.2 Institutional Support

Institutional support, in this case, refers to the willingness of the institution to participate in the study and their readiness to make adjustments to accommodate the arrangements required to properly conduct the study. These adjustments include, but are not limited to, scheduling of the experiment and access to the computer laboratories and the students.

The research team received some resistance from the school administration in Site B. Consent forms had been distributed to participants a week prior to data gathering, but had not been collected at the time of the research team's arrival. This caused concerns about research methods and scheduling, which ultimately led to the delay in system configuration and installation. School officials did not allow the local ground team to begin software and hardware installation until two days before the beginning of data gathering. Fortunately, installation and launching in Site B ran smoothly, and data gathering was able to proceed as scheduled.

The study was designed to be conducted over a period of three hours, allotting 30 minutes each for both the pretest and posttest, as well as delays in arrival and about 90 to 120 minutes of interaction with the software. Site B allotted only two hours for each session, including buffers for delay in arrival, introductions, and the administration of the pretest and the posttest. As a result, students were only able to interact with the software for 70-75 minutes per session.

Site C posed the most limitations in the schedule for data gathering. Instead of the prescribed three-hour period, each session was only allotted about 90 minutes, including the delayed arrival of the participants and the administration of the pretest and the posttest. To maximize the allotted time, PP was launched on each system before the participants arrived, which minimized the problems usually encountered when launching the software. As a result, students only interacted with the software for 30-45 minutes.

The final component of the study's design was the administration of a delayed posttest. Local teams in each site were instructed to administer a posttest exactly one week after a participant's interaction with the software. Due to the limited time, restricted by the school's schedule of activities as they were already on their final weeks of the semester, the delayed posttest was not administered to participants in Site C.

3.3 Logistics

Two local high schools took part in the study in Site A. Students here needed to travel from their high school campuses to the site where the study was conducted. School A had asked the research team to arrange for transportation of their participating students one week ahead of data gathering: from their high school to the data gathering venue, and vice versa once the session was over. As a result, members of the team were able to commission transportation for the 30 students coming from School A.

Conversely, School B instructed their students to proceed to the venue on their own. Because students had to manage their own transportation and because their commute was not properly managed, more than half of the time allotted (i.e., about an hour and a half) for the data gathering session was spent waiting for the participants to arrive. The delay caused the research team to shorten the interaction time with the software. For the succeeding groups of students from School B, the research team hired a shuttle service to transport the students to the venue in order to ensure timely arrival.

3.4 Compliance with Ethics Requirements

In line with university's guidelines on ethical research, the team was required to prepare and collect informed consent forms from each participant and his/her parents. While the study's data collection methods were non-invasive, the requirement applied to this study because interacting with the software required capturing the participant's face on video throughout the session.

Although arrangements were made with the partner schools in advance, only School A in Site A was able to distribute and collect the consent forms prior to the scheduled data gathering sessions. In effect, counterparts in Site A collected the consent forms from School B after the study was conducted, then sent the forms to the research team via courier. Similarly, counterparts in Site B also collected the consent forms one week after the study was conducted, and scanned copies were electronically sent to the research team.

Site C, being the last leg in the data gathering push, presented the most difficulty as their school year was already coming to a close. A week after data gathering had concluded, the research team's main counterpart in Site C said that, with the limited time and schedule constraints, it was going to be impossible to distribute and collect the consent forms. The team reached out instead to another member of the local team in Site C, and only after explaining the gravity of the situation and offering to compensate whoever can get it done was the request obliged. Consent forms were distributed, collected, and mailed back to the research team via courier within a week after contracting help.

3.5 Launch Delays

When launching PP, a number of technical problems sometimes occur. Most frequently, if the Internet is unstable when the game is launched, an error message will pop up saying that the game was unable to connect to the timeserver. Launching the game again usually resolves this issue. If the problem persists, however, the team had to resort to disabling the timing functionality of that specific machine.

Another frequent error that occurs has to do with the webcam malfunctioning. Previous experience with the webcam and its connection to PP has shown that when other applications on the machine are using the webcam, it was likely to malfunction when PP was launched. As a result, the research team usually quit all webcam-related software before launching PP. Webcam-related errors popped up on several occasions in Site A and Site B. Quitting and launching the game again usually resolves the problem as well.

Also, in order to better manage webcam software, the research team had its own set of webcams, which they install onsite immediately before data gathering. In Site C, however, because the school's officials wanted all students in each of the three participating classes to take part in the study, the research team had to use the built-in webcams of the site's machines. These built-in webcams had built-in webcam software that would pop up every time PP was launched. Because data gathering in Site C was already very limited time-wise, the research team resolved to launching the game before students arrived in order to address all launch delays before the session began.

3.6. Inattention to Directions

Not listening, reading, understanding, and paying attention to instructions also contributed to delays in gameplay. Because the timeserver synchronizes all student interactions in its logs with Internet time, it is important that all participants in each session begin at the same time. Once PP is launched, participants are asked to input a username (which is provided to them upon arrival), and to press OK. Participants will then be presented another screen to read, shown in Fig. 3, telling participants to wait for the moderator's go signal before pressing OK again. Clicking this OK button launches the game and begins the logging sequence.

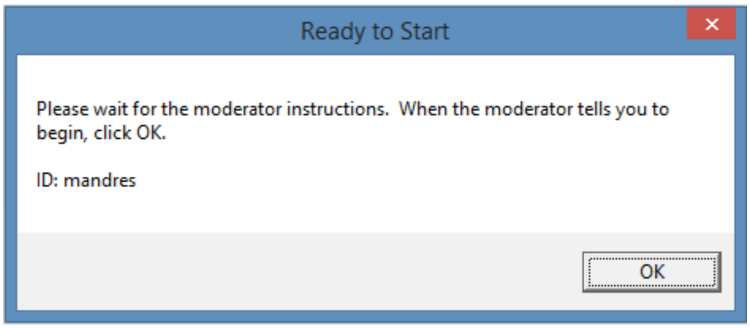

Fig. 3. Instruction screen telling participants to wait.

Participants are given the instructions to wait both verbally through the moderating member of the research team, and in writing through the pop-up screen in Fig. 3. Despite these, however, members of the research team have had to quit a game that was launched prematurely about two times every session. Once everyone is back on this screen and waiting for the go signal, participants are instructed to press OK, after which they are presented with a tutorial on how to play the game.

This tutorial ends with a string of text, instructing the students to "hit ESCAPE and select 'Quit'," as shown in Fig. 4. The research team noticed that almost half the participant population in each session gets stuck on this screen, possibly waiting for an "ESCAPE" button to come up on screen, as opposed to tapping the Escape button on the keyboard, which in turn brings the menu up, and allows the participants to click "Quit", which then brings them back to the game's main screen where they can choose what level they want to play.

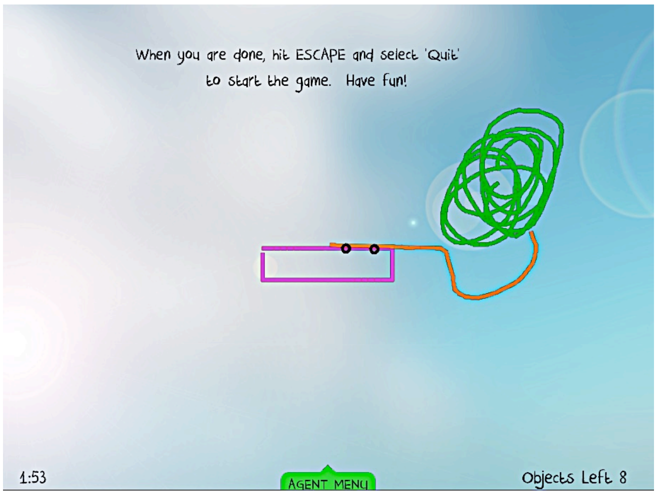

Fig. 4. Hit ESCAPE instruction.

4 Discussion and Conclusions

In an extensive literature review, [2] regards the Philippines as a significant producer of intelligent tutoring systems research outside of high-income nations. This finding implies an openness to new technology as well as commitment of Filipino researchers to collaborate with their counterparts abroad and to shepherd the deployment and study of technology use to improve educational institutions. However, many factors on the ground prevent adoption of these technologies. This paper describes some of the challenges that a Philippine team had to overcome to gather data from three local sites.

Infrastructure and institutional support were major roadblocks in the research method's smooth implementation. The learning environment used had three main components: the software itself, a stable Internet connection not blocked by a firewall or proxy, and a webcam. Any error produced by any of these three components results in faulty log capture, which eventually leads to data being thrown out. Having to ensure that each component runs without error in three separate data gathering sites in a country where education is only beginning to embrace the use of ICTs was the study's biggest hurdle to overcome. On top of this, resistance from and miscommunication with school administrators had caused the delay of both hardware/software setup and compliance with ethics requirements. The other challenges encountered during the study's execution were transportation arrangement, launch delays, and the students' inattention to directions.

All these challenges taken into consideration, there were some lessons learned in the process. In terms of dealing with institutional support and ethics compliance, starting the process early of arranging for data gathering schedules and the efficient distribution and collection of ethical consent forms. Avoiding the conduct of studies towards the end of the school year will give both the researchers and the partner institutions more time to fix issues that may have arisen during research execution. In terms of research execution itself, controlling transportation to and from the data gathering sites will ensure the participants' timely arrival, which is important especially when you are given only a certain number of hours for the session.

For educational technology adoption to widen, researchers must continue to plan for and address these challenges, and to share these experiences with the wider community to inform like-minded researchers about what to expect when conducting fieldwork in the Philippines.

Acknowledgements. We would like to thank the Bro. Robbie Paraan, S.J., Cecilie Villacrucis, and the officials at University of the Cordilleras, Bakakeng National High School, Ateneo de Davao University, Sacred Heart School – Ateneo de Cebu, Drs. Valerie Shute, Matthew Ventura, and Matthew Small of Florida State University for collaborating with us. This study was made possible through a grant from the Philippines' Department of Science and Technology Philippine Council for Industry, Energy and Emerging Technology Research and Development entitled "Stealth assessment of student conscientiousness, cognitive-affective states, and learning using Newton's Playground."

5 References

1. D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. In Proceedings of the Workshop on Affective Interactions: The computer in the affective loop workshop, International conference on intelligent user interfaces (pp. 7- 13). New York: Association for Computing Machinery.
2. Nye, B. D. (2014). Intelligent tutoring systems by and for the developing world: a review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education*, 1-27.
3. Ocumpaugh, J., Baker, R.S.J.d., Rodrigo, M.M.T. (2012) Baker-Rodrigo Observation Method Protocol (BROMP) 1.0. Training Manual version 1.0. Technical Report. New York, NY: EdLab. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.
4. Ogan, A., Walker, E., Baker, R.S.J.d., de Carvalho, A., Laurentino, T., Rebolledo-Mendez, G., Castro, M.J. (2012) Collaboration in Cognitive Tutor Use in Latin America: Field Study and Design Recommendations. *Proceedings of ACM SIGCHI: Computer-Human Interaction*, 1381-1390.
5. Portigal, S., & Norvaisas, J. (2012). Never eat anything raw: fieldwork lessons from the pros. *interactions*, *19*(4), 10-12.
6. Rodrigo, M. M. T., Sugay, J. O., Agapito, J., & Reyes, S. (2014). Challenges to Transferring Western Field Research Materials and Methods to a Developing World Context. *Research & Practice in Technology Enhanced Learning*, *9*(1).
7. Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and Learning of Qualitative Physics in Newton's Playground. The Journal of Educational Research, 106(6), 423-430.

Investigating the Impact of Designing and Implementing Culturally Aligned Technological Systems on Educators' Ideologies

Samantha Finkelstein

Human Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA
{slfink}@cs.cmu.edu

Abstract. Culturally sensitive educational technologies may be able to help improve under-represented students' learning and engagement when they are deployed in the classroom. However, there may be challenges integrating these systems into the classroom when the cultural components they incorporate are heavily stigmatized in contemporary society. In this on-going work, we are using an action research approach to investigate how involving teachers in the design of these technologies may not only affect the effectiveness of these interventions on students, but also teachers' own ideologies surrounding the targeted stigmatized cultural components.

Keywords: dialect, classrooms, teachers, culture, AAVE, action research

1 Introduction

The pervasive achievement gap between Euro-American and African American students is perpetuated by challenging and inter-related factors, including access to resources, socio-economic status, and racism (and vestiges of racism) in contemporary society [1]. One common manifestation of these vestiges of racism is a *deficit perspective* within the classroom, where the school system views certain aspects of a student's cultural background as a *challenge to overcome* rather than an *asset to leverage* [2]. For example, many African American students come into school as speakers of a non-standard dialect of English called African American Vernacular English (AAVE), which is rarely represented, or even accepted, within the classroom. Despite that AAVE has great cultural importance for its speakers and linguists regard AAVE as valid and grammatically consistent, it is common practice for educators to criticize or even shame students for their use of this dialect [3], such as by saying that they are speaking incorrectly, or even that they sound like they belong on the streets. However, some evidence suggests that when non-Standard English speakers are allowed to use their primary dialect within the classroom or when this dialect is represented in learning materials, students may improve on their task performance, academic engagement, self-efficacy, and even their use of Standard English [4, 5, 6]. While this evidence is promising, standard teacher training programs rarely incorporate enough background in language variation to prepare teachers for methods of incorporating students' dialect diversity into the classroom. For this reason, some researchers have proposed that culturally adaptive educational technologies may be a productive way for students to gain access to learning materials that may best support their learning [7, 8, 9].

Despite the potential promise of these systems, a notable challenge in the design of culturally adaptive classroom technologies is ensuring that they work with, and not against, the teacher. There is substantial evidence that teachers may be hesitant to incorporate classroom interventions that expose their students to stigmatized cultural behaviors such as non-standard dialect use. This is often due to lack of appropriate teacher training about cultural variation, misconceptions about the role of non-standard dialect use in their students' lives, and concern that they might accidentally cause offense and put their job at risk. As interventions are less likely to be successful if teachers do not believe that the systems are helping them meet their own goals [10], this may make even the most well-designed educational technologies unusable in real classroom settings. In this work, we are investigating how an *action research* (AR) approach may be used to both design technologies that best meet teachers' needs, while also helping them develop more progressive and positive ideologies about cultural variation. By action research, we refer to the cyclical process of researchers working alongside community partners (in this case, educators) to create knowledge by *learning through action* – taking steps, reflecting on the outcomes, and iterating together [11]. In AR, the researcher works alongside the community partners to open up productive lines of communication and facilitate activities expected to create change, rather than as a distanced observer of subjects. This method will allow us to work alongside educators to quickly iterate on different ways of incorporating a technology that can use AAVE into the classroom. This will help us understand what social and scientific impacts these interventions may have on the classroom culture, as well as investigate how this collaborative design process itself impacts teachers' ideologies about their students.

2 Previous Work on Culturally Aligned Technologies

Over the past two decades, there have been a small but notable number of educational technologies that have considered how to align to students' underrepresented cultural backgrounds. These projects demonstrate some of the potential scope for the impact culturally-aligned technologies may be able to have on students. For example, Pinkard's work on literacy learning for young African American students resulted in two systems, Rappin' Reader and Say Say Oh Playmate, which leveraged students' culturally-based knowledge of rhythm patterns and clap sequences to acquire early literacy components through writing rap lyrics [7]. Rap lyrics were also applied in Gilbert's African American Distributed Multiple Learning Styles System (AADMLSS) program, which is an intelligent tutoring system that additionally uses gaming components to allow students to practice math word problems where explanations are provided via rap lyrics that use AAVE features [8]. Other educational technologies have began exploring the potential impact of dialect congruence on students' performance in other non-standard dialects, such as Mohammad's Trinbago Adventures for Caribbean students, where students are allowed to customize the amount of dialect features they hear (and other cultural references) within the system [9]. Each of these systems has demonstrated success with the underrepresented population they had targeted, including both academic performance and student engagement. However, the teachers' response to these systems, and the potential impact that the deployment of these systems in the classroom had on the teachers over time, was either not performed or not reported.

There have also been a small number of investigations that examine the impact of simply manipulating only the dialect used in a system. For example, in our own previous work, we have found that when AAVE-speaking 3rd grade students were exposed to a system that provided them with identical science examples in either Standard English or AAVE, students demonstrated an average of two standard deviations improvement on the quality of their own science reasoning when they heard the example in AAVE [12]. However, in follow-up interviews with teachers, we found that they would be very uncomfortable with deploying such a system to their students in the future, regardless of the potential learning benefits. The impact of a German non-standard dialect was also investigated with German adults using a virtual agent who either spoke in Standard or Non-Standard German, finding that participants aligned their own dialect to match that of the agent, but that the Non-Standard agent was viewed as more likable [13]. In our current work, we are performing a similar analysis, and investigating how 3rd grade AAVE-speaking students' language use, self-efficacy, language ideologies, and science achievement is impacted by a virtual agent who either exclusively speaks Standard English or code-switches between Standard English and AAVE based on context over the course of six weeks. Previous work with this virtual agent, Alex, found that even during one session with the character, students switched between dialect features based on context along with the agent – even though they did not perform this type of code-switching with their teachers [14].

3 Educational Interventions to Impact Teacher Ideologies

Our previous research (in preparation) has found that teachers would be very hesitant to expose their students to AAVE via an educational technology, regardless of the potential learning benefits to students. This is consistent with what other researchers have found about integrating non-technical curricula into the classroom. However, research suggests that if teachers feel that an educational technology is working to support their overall goals, it is possible that teachers may experience a *pedagogical evolution* [10], whereby the technologies in their classrooms may support and structure class activities that the educator previously did not think possible. The challenge, then, is identifying methods for integrating these technological systems into a classroom in a way that is able to work *with,* rather than *against,* educators.

To address this problem, some designers of non-virtual curricula have found it effective to host professional development workshops with teachers to help teach them about linguistic variation [4, 15]. When paired with this knowledge, teachers become able to not just host the intervention within their classroom (such as is often the case with technologies), but also become active facilitators of the learning activities with their students. In fact, there is additionally evidence that when teachers have the opportunity to teach a pre-packaged learning activity involving linguistic variation to their students themselves, they develop a stronger positive change in their own ideologies compared to teachers who only attend professional development workshops [4]. These findings support the potential positive impact of action research on influencing teachers' ideologies, as action research involves many of these components, such as professional development discussions facilitated by researchers, reflection with other peer educators, and implementation of curricula within the classroom.

4 Investigating the impact of culturally aligned systems

The goal of this work is to employ AR approaches with urban elementary school teachers to promote a positive change in the often-negative classroom culture surrounding students from linguistically-diverse backgrounds. To do this, our approach will involve a combination of professional development workshops surrounding language variation, group reflection discussions about what learning goals they feel are important for their students to know regarding language variation, and hands-on activities to develop classroom activities to meet some of those identified learning goals. The classroom activities will involve the use of Alex, a virtual peer character capable of communicating to students about different science activities and some other social topics (e.g., video games) in either Standard English or AAVE (described above). Because one of the noted reasons that many teachers avoid talking about AAVE with students is many do not identify as speakers of this dialect, a system that is able to demonstrate dialect differences as a peer to the students may be a productive platform for helping to introduce this discussion. We additionally argue that providing educators with an existing technology that can be deployed differently in the context of different classroom activities may allow us to more efficiently iterate new ideas into the classroom.

In this planned work, we will work with approximately ten educators between two and four times a month for a full semester to facilitate and participate in these discussions and lesson plan design sessions. We will aim for teachers to deploy a new classroom activity surrounding the virtual character in the classroom approximately twice a month throughout the semester. We expect a large variation in the sorts of activities teachers design, for example, ranging from using the technology as part of a guided class discussion and worksheet, to a hands-on group activity where students are asked to make the character speak differently in different situations. The researchers and each of the teachers will observe how the students interact with the class activity, and bring their observations to the group discussion the following week. This discussion will spark teachers' iterations on their next class activity.

We will perform pre- and post-intervention measures including meta-linguistic awareness, language ideology, and dialect use for both teachers and students. These quantitative measures will be paired with qualitative measures of how different activities promoted different sorts of student interactions and responses and the types of interactions students and teachers shared throughout the lesson. We are currently performing a pilot analysis of this process with three elementary school teachers at a local, urban 100% African American charter school to help prepare us for the upcoming semester-long study. Through this pilot and the full-length study, we aim to gain a better understanding of how culturally-aligned educational technologies, and the collaborative process of designing them with teachers, may impact the classroom culture in ways that support positive social change.

Acknowledgments. Many smiles to the ArticuLab, HCII, and the Graduate Training Grant # R305B090023 from the US Department of Education (IES).

5 References

1. Ogbu, J. U. (2003). Black American students in an affluent suburb: A study of academic disengagement. Routledge.
2. Atkinson, J. L. Are We Creating the Achievement Gap? Examining How Deficit Mentalities Influence Indigenous Science Curriculum Choices. Cultural Studies and Environmentalism, 439-446. (2010)
3. Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.
4. Sweetland, J. (2006). Teaching writing in the African American classroom: A sociolinguistic approach. Stanford University.
5. Webb, L., & Webb, P. (2008). Introducing discussion into multilingual mathematics classrooms: An issue of code switching?. Pythagoras: Teaching and learning mathematics in multilingual classrooms: Special Issue 67, 26-32.
6. Wheeler, R. (2006). "What do we do about student grammar—all those missing -ed's and -s's?" Using comparison and contrast to teach Standard English in dialectally diverse classrooms. English Teaching: Practice and Critique, 5(1), 16–33.
7. Pinkard, N. (2001). Rappin'Reader and Say Say Oh Playmate: Using children's childhood songs as literacy scaffolds in computer-based learning environments.*Journal of Educational Computing Research*, *25*(1), 17-34.
8. Gilbert, J. E., Arbuthnot, K., Hood, S., Grant, M. M., West, M. L., McMillian, Y., ... & Eugene, W. (2008). Teaching Algebra Using Culturally Relevant Virtual Instructors. *IJVR*, *7*(1), 21-30.
9. Mohammed, P., & Mohan, P. (2011). Integrating culture into digital learning environments: studies using cultural educational games. *The Caribbean Teaching Scholar*, *1*(1).
10. Pajares, M. (1992) Teachers' Beliefs and Educational Research: cleaning up a messy construct, Review of Educational Research, 62, pp. 307-332.
11. Hayes, G. R. (2011). The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *18*(3), 15.
12. Finkelstein, S., Yarzebinski, E., Vaughn, C., Ogan, A., & Cassell, J. (2013, January). The effects of culturally congruent educational technologies on student achievement. In *Artificial Intelligence in Education* (pp. 493-502). Springer Berlin Heidelberg.
13. Kühne, V., Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2013, January). Using Linguistic Alignment to Enhance Learning Experience with Pedagogical Agents: The Special Case of Dialect. In *Intelligent Virtual Agents* (pp. 149-158). Springer Berlin Heidelberg.
14. Rader, E., Echelbarger, M., & Cassell, J. (2011, May). Brick by brick: iterating interventions to bridge the achievement gap with virtual peers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2971-2974). ACM.

15. Reaser, J., & Adger, C. (2007). Developing language awareness materials for nonlinguists: Lessons learned from the Do you speak American? curriculum development project. Language and Linguistics Compass, 1(3), 155-167.
16. Yocum, K. (1996) Teacher-centered Staff Development for Integrating Technology into Classrooms, Technology Horizons in Education, 24(4), pp. 88-91