USING VIRTUAL HUMANS TO SIMULATE LIMITED ENGLISH PROFICIENT HUMANS

By

DOAA T. EL SHEIKH

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2012

I dedicate this to my parents, my husband and my kids, Jumanah and Ali.

ACKNOWLEDGMENTS

This thesis would not have been possible without the guidance and the help of several mentors and colleagues who contributed and extended their valuable assistance in the preparation and completion of this work.

First and foremost, my utmost gratitude to my adviser Dr. Benjamin C. Lok, whose advice and encouragement were truly invaluable.

Dr. Ryan Moran, who dedicated his time and effort to make this study successful and helped with all the needed medical information gathering.

Dr. Erik Black, for his invaluable feedback and his generous funding for the study.

I would like to thank Dr. Jorg Peters and Dr. Erik Black for accepting to serve in my committee and providing their time and expertise to review my thesis.

I would like to thank my colleagues at the Virtual Experiences Research Group (VERG) for their continuous assistance and feedback.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

USING VIRTUAL HUMANS TO SIMULATE LIMITED ENGLISH PROFICIENT HUMANS

By

Doaa T. El Sheikh

December 2012

Chair: Benjamin Lok
Major: Computer Engineering

Virtual human technology has been recently exploited in medical training. In this thesis, we propose using virtual humans to simulate limited English-proficient humans to assess the quality of health care provided to pediatric patients whose parents have limited English proficiency.

Through a user study, virtual humans were used to represent parents with different levels of English proficiency in a virtual pediatric interview setting. In the user study, medical students and residents interacted with virtual parents of different levels of English proficiency.

The performance of medical students and residents when interviewing an English-proficient (EP) virtual mother and a limited English-proficient (LEP) mother was analyzed to assess the quality of the provided health care.

The results lead to some interesting conclusions. Firstly, virtual humans can effectively simulate different English proficiency levels. Secondly, the performance of the study participants regarding eliciting important information from patients was significantly higher with English-proficient patients. This conclusion agrees with the nationally observed fact that limited English-proficient patients get lower quality of health care than English-proficient patients mainly due to patient-provider language discordance. Thirdly, study participants learned how to recast their questions to adapt to the English proficiency level of the virtual mothers.

CHAPTER 1
VIRTUAL HUMANS AND LIMITED ENGLISH PROFICIENT PATIENTS

1.1 Virtual Humans

Virtual humans are conversational embodied agents that look and act like humans [16]. Virtual humans have been built to be used in a variety of applications. Virtual humans have been used in different types of interpersonal skills training such as in the military [9, 20], negotiation [8, 34], medical education [19, 24], public speaking [28] and team training [30].

Virtual humans have been used as virtual patients. Virtual patients are virtual humans that play the role of standardized patients. Standardized patients are actors that play the role of patients for training medical students. Virtual humans have benefits over standardized patients as virtual patients can simulate conditions that standardized patients are unable to simulate. For example, virtual patients can simulate a patient with cranial nerve palsy which is difficult for standardized patients to simulate [3].

Virtual humans can simulate humans with different races to elicit racial bias [31], simulate humans with different types of audience behaviors to elicit anxiety [28] and simulate humans with different levels of English proficiency as we will propose in this thesis.

1.2 Limited English Proficient Patients

In 2010, 25.2 million individuals in the U.S., amounting roughly to nine percent of the country's population, were considered to be of limited English proficiency [26]. Limited English-proficient (LEP) patients are less likely than English-proficient (EP) patients to obtain sufficient information from health care providers and to be encouraged to participate in medical decision making [13]. As a result, LEP patients typically receive lower quality of health care despite their level of income or access to health care insurance [7].

Assessment of the quality of health care provided to patients is usually conducted through surveys filled by patients in clinics or by phone [10, 29]. As a result, assessment of the quality of health care provided to patients depends only on the patients' opinion and ignores the physicians' side and hence the assessment may not be accurate.

The level of English proficiency in a patient is usually assessed using the patient's answer to the question "How well do you speak English?". If the patient answers with "very well" or "well", the patient is considered to be an EP. In case the patient answers with "not well" or "not at all", the patient is considered to be an LEP [1].

1.3 The Proposed Work

This thesis proposes the use of limited English proficient (LEP) virtual humans to address the impact of language barriers on the performance of health care providers. To address the impact of language barriers on the performance of health care providers, LEP virtual humans were used to assess and improve the quality of health care provided to LEP patients through assessing and improving the communication skills of physicians.

Four pediatric cases were created to help evaluate and improve the communication skills of both medical students and pediatric residents. The four pediatric cases represent the combinations of two medical cases and two levels of parental English proficiency (Figure 1-1). In each of the four cases, two virtual humans were used: one to simulate the child patient and another took the role of his mother.

In a user study, thirty one participants composed of medical students and pediatric residents from the University of Florida medical school were recruited. Participant interacted with two virtual pediatric patients of different medical cases (Heart and Lung cases) and with different parental English proficiency levels (EP or LEP).

The results showed that there was a significant difference between the performance of participants regarding the English proficiency level of the two interviewed mothers. The results showed that participants were able to elicit more information from the EP mothers than the LEP mothers.

	Heart case	Lung case
English Proficient (EP) mother	Linda and Bobby 	Trisha and Alex 
Limited English Proficient (LEP) Mother	Maria and Carlos 	Marta and Alejandro 

Figure 1-1. The four medical pediatric cases

Moreover, we explain how we used virtual humans to improve communication skills of the participants for interviewing LEP patients. Participants avoided using medical terminology and used simple sentences and common words to elicit information from the LEP parents. Participants rephrased their questions whenever the parent did not understand.

CHAPTER 2
PREVIOUS WORK

In the following two sections, we present some of the previous approaches that relate to our ideas of

- Using virtual humans
- Evaluating and improving the quality of health care for LEP patients.

2.1 Virtual Humans

2.1.1 Virtual Humans in The Medical Field

Johnsen et al. [19] compared the performance of medical students when interviewing a standardized patient and when interviewing a virtual patient to check for correlation. A correlation was found between the performance of students in both interviews. This correlation indicates that virtual-human interviews can be as effective as real-human interviews and hence can be used in education to improve the interpersonal skills of medical students.

Parsons et al. [27] created a system with virtual humans suffering from a mental disorder. In that system, health care providers were trained on interviewing a patient with a mental disorder using virtual humans. The health care providers who used the system established rapport with the virtual humans. The providers had useful discussions with the virtual humans about their traumatic experiences.

Rossen et al. [31] used virtual humans with different skin-tones to investigate skin-tone bias. Virtual humans were interviewed by Caucasian medical students. Each medical student interviewed either a black or a white virtual human. Virtual humans showed their ability to elicit behavior that is consistent with real-world skin-tone biases. Virtual humans were shown to be a good supplement for anti-bias training modules.

2.1.2 Virtual Humans in Military and Social and Cultural Training

In the Mission Rehearsal Exercise (MRE) project, USC's ICT group used virtual humans to improve the decision-making skills of unit leaders in the U.S. Army [20].

In the MRE project, participants were trained on making correct decisions in some situations by giving them the opportunity to interact with a larger than life-size virtual humans. The MRE project helps participant to become well prepared for similar situations in real life.

The Virtual Environment Cultural Training for Operational Readiness (VECTOR) project uses virtual humans for cultural training in the U.S. military [9]. The VECTOR project helps forces to develop specific skills needed for interacting with members of a culture of interest and hence helps these forces to survive in a foreign environment.

Bickmore el al. [6] used an animated character as a relational agent. The relational agent was designed to take the role of an exercise adviser that interact with elderly users daily for two months. The relational agent goal was to motivate users to exercise through walking. The relational agent system proved to be effective for health education applications for elderly users with low computer literacy.

Babu et al. [4] used virtual humans in a study to investigate their potential to teach people Indian social conversational traditions. Virtual humans taught participants the social traditions significantly better than the literature-based methods.

Kang el al. [21] conducted a study on real humans to elicit the different types of human nonverbal cues that are associated with different intimacy levels of verbal self-disclosure. The elicited nonverbal cues associated with verbal self-disclosure were then used to design a virtual counselor. The virtual counselor with the nonverbal cues was used in an online survey in order to evaluate the effect of those nonverbal cues. The virtual counselor with appropriate nonverbal cues for intimate self-disclosure proved to be liked by their human clients. The nonverbal cues indeed helped in creating a better rapport towards those counselors.

Pertaub et al. [28] assessed the anxiety responses of people giving presentations to virtual humans. The virtual humans had three different types of audience behaviors:

neutral, positive and negative. The virtual humans were shown to be capable of provoking appropriate arousal in people with public speaking anxiety.

2.2 LEP Latinos and Disparities in Health Care Quality

The unfavorable impact of the limited English proficiency is particularly more prominent in Latino groups.

Decamp [10] used the 2007 National Survey of Children Health (NSCH) as the data source for a study about medical home disparities for Latino children. Latino children with a Spanish-language parental interview were less likely to get primary-care access and good quality of health care than Latino children with an English-language parental interview. Moreover, both groups of Latino children were less likely than White children to get primary-care access and good quality of health care.

Pippins et al. [29] showed that insured LEP Latinos were more likely than insured EP Latinos to report not having a regular source of care, lacking continuity, long waits and difficulty getting information by phone.

2.3 Improving Health Care Quality for LEP Patients

Increasing access to language-concordant physicians can improve the experience of LEP patients [35]. LEP patients were shown in previous studies to be more likely to report better health care when having language-concordant physicians [14, 25, 33]. However, restricting LEP patients to language-concordant physicians may lead to lower quality of health care. The lower quality of health care is due to the possibility that language-concordant physicians are less qualified and have less access to important medical resources [5].

Professional interpreters can increase the quality of health care delivered to LEP patients [15, 17, 22]. However, interpreters may be underused even when available because physicians prefer to use their own limited second-language skills or family members as interpreters [11, 12]. Moreover, even with the use of interpreters, the majority of clinicians reported difficulties dealing with LEP patients [23]. In addition, LEP

patients may be less satisfied even with the use of interpreters compared to patients that are able to communicate directly with their physicians [25].

CHAPTER 3
THE VIRTUAL PEDIATRIC SIMULATOR

3.1 System Overview

The virtual pediatric simulator is composed of two main components (Figure 3-1):

- The script database that controls the virtual human responses to the user's questions.
- The interpersonal simulator (IPS) [18] that renders the virtual environment, connects to the script database and receives the user's questions to the virtual human through a graphical user interface.

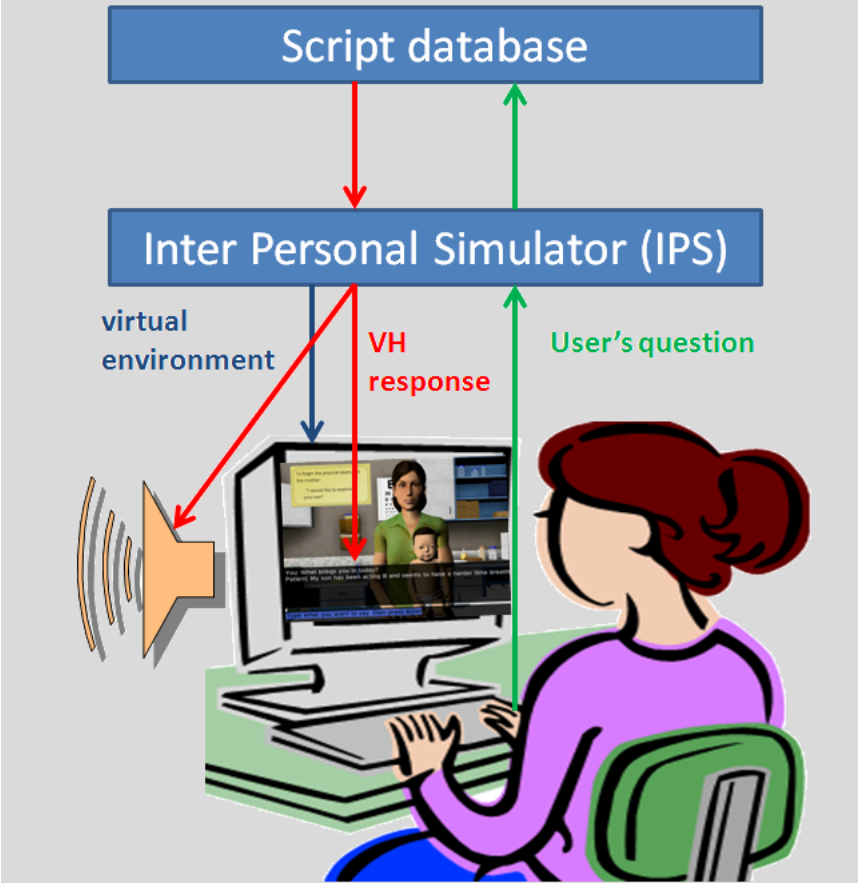

Figure 3-1. System overview

3.2 Script Database

3.2.1 VPF

Pediatric interviewing scripts for the virtual pediatric simulator were created using Virtual People Factory (VPF) [32]. VPF is a web-based application that allows developers to create virtual human scripts. A virtual human script is a corpus of all possible questions that a user can ask to the virtual human and their responses. Each response is connected to a group of questions (Figure 3-2). A response is triggered when one of its connected questions (or a similar one) is asked by the user. The triggered response is displayed on the screen in the user interface page of VPF and may also be heard if there is a voice recorded for that response.

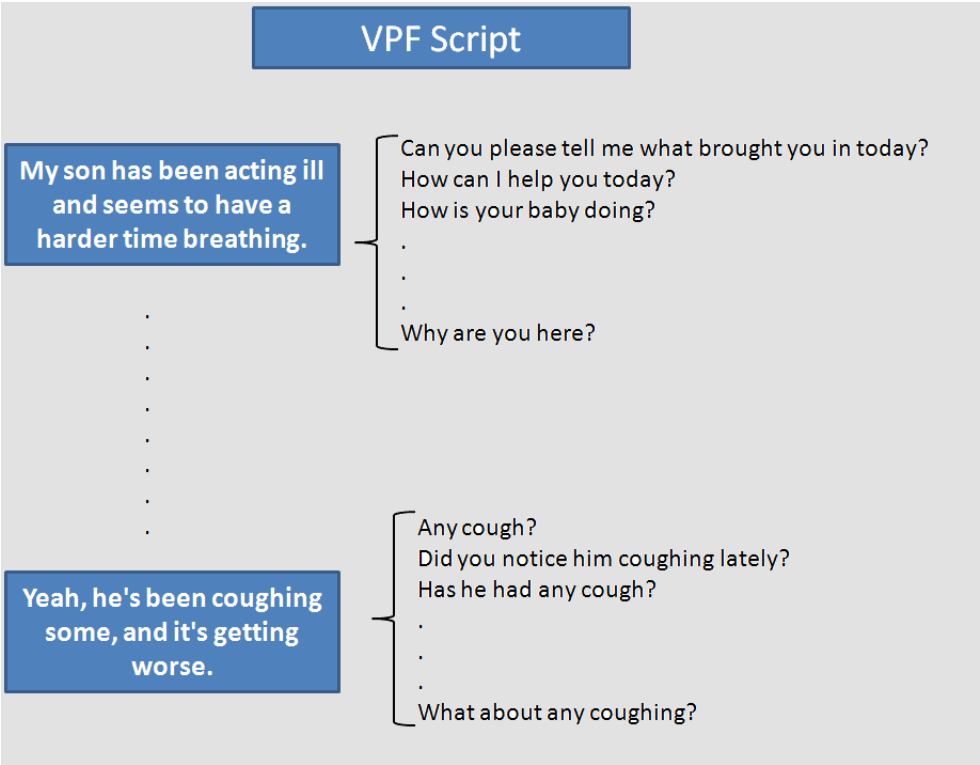

Figure 3-2. Questions and responses in a VPF script

3.2.2 Creating the Heart Case Script

The first step in creating the heart script was to choose a specific heart case. The heart case (as well as other cases) was chosen by Dr. Ryan Moran, an assistant professor in the Division of Pediatric Critical Care at the University of Florida. Dr. Moran wrote a detailed description about the medical case. The medical case was about a 12-month-old boy named Bobby. Bobby is suffering from Myocarditis which is inflammation of the heart muscle. Bobby is present in the clinic with his mother Linda. Bobby's case will be referred to as the Heart Case.

Following the case description, 35 responses and 52 questions were added to VPF to create a script for Linda. This means that Linda was able to give 35 different responses that were triggered by the 52 added questions.

The second step was to make some initial tests on the created script by asking questions to Linda. Whenever Linda was not able to respond to a question, the question was added to a suitable response in the script database. If there was no suitable response in the script database, the question would be added and connected to a new suitable response. After the initial tests, the created script grew to include 75 responses to 435 different questions and Linda's responses were recorded by an EP female.

The last step was to test the script by domain experts and general people. As a result of this step, new questions and responses were added to Linda's script. Linda's script reached a stable state where there was almost no new added responses with new testing. The testing was paused at this point. Currently, Linda's script has 216 responses to different 1193 questions.

3.2.3 Creating the Lung Case Script

A second medical case was chosen to be of the same level of difficulty of the Heart Case. The second medical case is about an 11-month-old boy named Alex. Alex has swallowed a foreign body (a big piece of carrot). Alex has some common symptoms with

Bobby such as hard breathing. Alex is present in the clinic with his mother Trisha. Alex's case will be referred to as the Lung Case.

The Heart and the Lung Cases were discussed between Dr. Moran and some colleagues in order to test if the two cases are of the same difficulty. The Heart and the Lung Cases were classified to be of the same difficulty.

The script for the Heart Case was reused to create a new script for Alex's mother, Trisha. The new script was modified to match the Lung Case and Trisha's voice was recorded by an EP female (not the same female that recorded Linda's voice in the Heart Case). The new script was created faster and with less testing due to the advantage of reusing the script for the Heart Case. Currently, Trisha's script has 222 responses to different 1198 questions.

The scripts created for the two medical cases has mostly the same questions but with different responses depending on each case. Also depending on each case, there are some different questions. For example in the Lung Case script, there is a question about the color of the produced mucous. This question is not in the Heart Case script because the heart patient does not have mucous so it does not make sense to ask about the color of the produced mucous.

3.2.4 Creating the LEP Scripts

The Heart Case description was translated into Spanish and was given to a Spanish-proficient female who is an LEP person. The LEP female studied the translated case description and then was asked all the questions in the Heart Case script in English. The LEP female responses were recorded and used as the responses for a virtual LEP mother (Maria) in a new script. The new script is about Carlos who has the same exact symptoms and conditions as Bobby (the heart patient) but with different parents.

The Lung Case description was translated into Spanish and was given to a different LEP Spanish-proficient female. The LEP female studied the translated case description

and then was asked all the questions in the Lung Case script in English. The LEP female responses were recorded and used as the responses for a virtual LEP mother (Marta) in a new script. The new script is about Alejandro who has the same exact symptoms and conditions as Alex (the lung patient) but with different parents.

Those two created LEP scripts have the same questions as their EP versions but with different responses due to the difference in the English proficiency levels of the LEP and the EP mothers.

In some situations, if the interviewer asks the mother in the EP script a couple of different questions that have the same meaning, he/she will get the same exact response. For example, the mother's response to the questions "Any ear discharge?" and "Any fluid coming from his ear?" will be the same: "No ear discharge". However, the response of the mother in the LEP script will be "Discharge? I don't know what that means" and "No, nothing is coming from his ear," respectively.

The variation in responses for the LEP scripts teaches health care providers that they need to choose their words carefully. Health care providers should always avoid using hard medical terminology with LEP patients.

3.3 Interpersonal Simulator and the Virtual Environment

Interpersonal Simulator (IPS) is a software system that enables users to practice communication skills with virtual humans [18]. IPS was used to render four virtual environments. Each virtual environment is connected to a different script database on VPF to create two different levels of parental English proficiency for two different medical cases which are the Heart and the Lung Cases. Each virtual environment represents a virtual exam room with virtual objects and two virtual humans that represent the patient and the mother.

3.3.1 The Exam Room and the Virtual Humans

In the virtual exam room, there exists an animated virtual mother. The virtual mother can respond verbally to the user's questions. The virtual mother accompanies

the patient who is represented by an animated virtual baby. The virtual baby suffers from a problem that causes him to have a fast and heavy breathing. The breathing was created in Maya and simulated by IPS.

The mothers in the Heart and the Lung Cases were represented by two different looking virtual females that can be perceived as either Caucasian or Latino to avoid race bias. The mothers were shown to ten different persons who agreed that the two mothers can be considered either Caucasian or Latino.

One virtual baby was used to play the role of the patient in the Heart and the Lung Cases. In both the Heart and the Lung Cases, the virtual baby has a matching eye ,hair and skin colors with his virtual mother. Within each of the Heart and the Lung Cases, the same virtual female was used to represent both the LEP and the EP virtual mothers (Figure 3-3 and Figure 3-4).

Figure 3-3. The heart patient and his mother.

Figure 3-4. The lung patient and his mother.

3.3.2 Interviewing the Virtual Mother

By having the four scripts, users can interview the virtual mothers through the graphical user interface of IPS and elicit information by asking relevant questions. These questions should be related to the health status of the patient such as “Has he been coughing?” and “Does he have any wheezing?”. These questions also include family history and some general information such as “Any medical problems in the family?” and “Do you have any other children?”.

When a user types a question to the mother that triggers a response in the database, the mother responds verbally. In addition to the verbal response of the mother, a text of her response is displayed on the screen (Figure 3-5).

In case the user’s question triggers different responses, a speech clarification is displayed. The speech clarification displays a maximum of three different possible questions that the virtual human thought the user might have asked. The speech

Figure 3-5. This picture shows the virtual mother's response to the user's question.

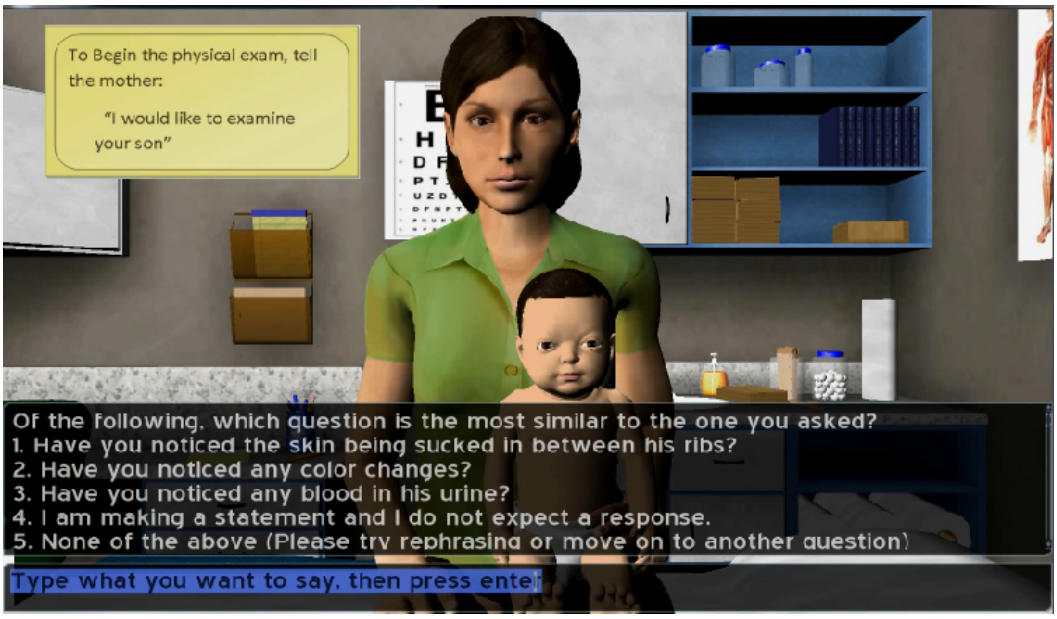

Figure 3-6. This picture shows the displayed speech clarification in case the user's question is "have you noticed any change in skin color?".

clarification also gives the user the option to choose that he/she did not mean any of the displayed questions or to choose that he/she was typing a statement (Figure 3-6).

When the user gets a response that has an important piece of information, a discovery is logged in a file. For example, if the mother's response is "He seems really congested in his chest," then the logged discovery should be "Breathing is characterized with chest congestion".

3.3.3 The Virtual Physical Exam

In addition to interviewing the virtual mother, the user can perform fifteen different physical exams on the virtual baby (Figure 3-7).

Figure 3-7. Fifteen different physical exams can be performed on the virtual patient.

The physical exams are divided into interactive and non-interactive exams:

- One interactive exam which is the Auscultate Heart and Lungs Exam. In this exam, the user can use the mouse to move a virtual stethoscope and listen to the baby's heart and lungs by lift-clicking (Figure 3-8). Depending on what location the user

listens to, he/she will detect different findings. The location of the stethoscope is logged periodically many times per second to help in evaluating the user's performance.

- Fourteen non-interactive exams which are the General Exam, the Head Exam, the Ears Exam, the Eyes Exam, the Nose Exam, the Throat Exam, the Neck Exam, the Abdomen Exam, the Genitourinary Exam, the Back Exam, the Anal Exam, the Extremities Exam, the Skin Exam and the Neuro Exam. All these fourteen non-interactive exams give information about what the user would find in case he/she performed the exam on the virtual baby. Among those exams, the Ears and the Throat exams have pictures along with the information (Figure 3-9).

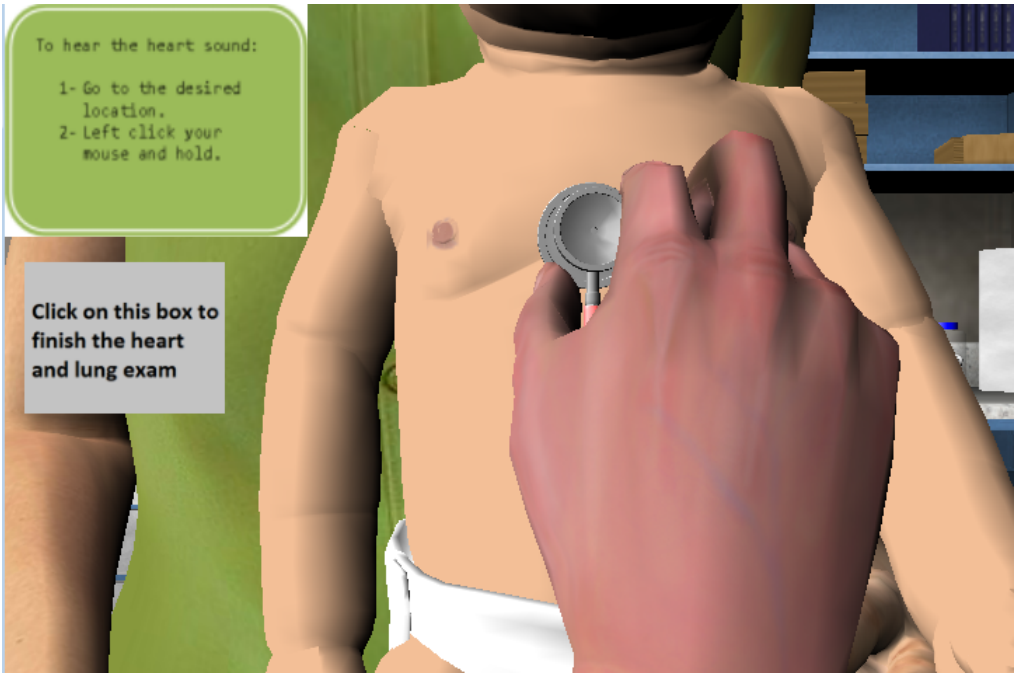

Figure 3-8. A user performing the Heart and Lungs Exam on the virtual patient

3.4 The Used Equipment

To interact with the virtual humans, commodity over-the-shelf equipment were used:

- A Mac Mini computer to run IPS and render the virtual environment and the virtual humans.
- A 15-inch standard monitor for displaying the virtual environment and the virtual humans.

Figure 3-9. A user performing the Ears Exam on the virtual patient

- Speakers for the voice of the virtual mothers and the heart and lungs sounds.
- A keyboard for typing questions to the virtual mothers.
- A mouse for performing the Heart and Lungs Exam on the virtual patient.

CHAPTER 4
LANGUAGE BARRIERS STUDY

4.1 Study Design

A mixed design study was conducted on a group of thirty one participants (eleven males and twenty females) to evaluate the communication skills of participants when interviewing EP and LEP patients and to check if participants developed better communication skills with the LEP patients over time.

Two participants were third-year pediatric residents, ten were second-year pediatric residents, three were first-year pediatric residents and sixteen were third-year medical students. All participants were recruited from University of Florida medical school. Each participant received a compensation of a $10 gift card for participating in the study. Participants were randomly placed into two different groups. Participants in each group interacted with one heart case and one lung case and the order of the cases was chosen randomly.

In the first group, one of the cases was chosen randomly to have an LEP mother and the other an EP mother whereas in the second group both cases were chosen to be of LEP mothers.

In each case, participants go through an interview session followed by a physical exam session. After finishing the physical exam , participants are allowed to go back to the interview and ask more questions to the virtual mother.

4.2 Hypotheses

- **Primary Hypothesis 1:** Participants will gather more critical findings (discoveries) from the virtual EP mothers than the virtual LEP mothers.
- **Primary Hypothesis 2:** Participants will improve their communication skills for dealing with the LEP mothers after using the system.
- **Primary Hypothesis 3:** Participants will identify more techniques for dealing with the LEP mothers after using the system.

4.3 Secondary Expected Findings

- Participants will be more likely to come to the correct diagnosis after interacting with the virtual EP mothers more than they would after interacting with the LEP mothers.
- Participants will ask more questions to the LEP mothers than the EP mothers.
- There will be a difference in the participants' responses to the survey questions between the EP and the LEP mothers.

4.4 Study Procedure

The study was conducted in a conference room at Shands hospital. Each participant went through the following steps:

- Was randomly assigned to a group and given a consent form to read and sign.
- Watched an interactive tutorial with a virtual doctor explaining how to use the system and how to perform the physical exam. The participant had the opportunity to practice interacting with the virtual doctor and perform two exams on a virtual baby.
- Filled a pre-survey.
- Interviewed the virtual mother in the first case to elicit information.
- Filled in an interim survey about the first case.
- Performed the physical exam on the virtual baby in the first case.
- Filled in another interim survey about the first case.
- Interviewed the virtual mother in the second case to elicit information.
- Filled in an interim survey about the second case.
- Performed the physical exam on the virtual baby in the second case.
- Filled in another interim survey about the second case.
- Filled in a post survey.

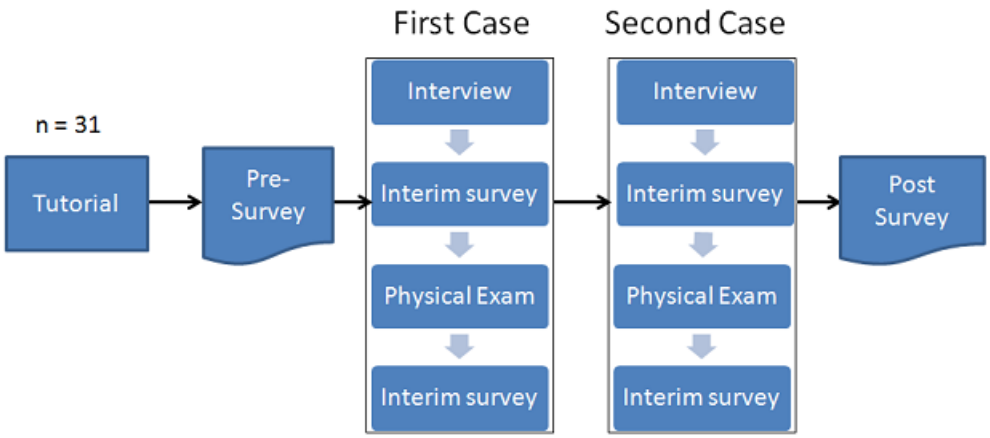

Figure 4-1. Study procedure block diagram

4.5 Metrics

4.5.1 Percentage of Identified Discoveries (Important Findings)

The percentage of identified discoveries is used for testing Primary Hypothesis 1 and Primary Hypothesis 2.

Discoveries are the useful information that a participant elicits from the virtual mother's responses. Discoveries were divided by a medical expert (Dr. Moran) into three classes were each class was analyzed separately:

- **Class A:** The important information that participants have to elicit from the mothers and which helps in arriving at the correct diagnosis.
- **Class B:** The information that participants may or may not elicit from the mothers and which helps in arriving at the correct diagnosis.
- **Class C:** The information that participants may or may not elicit from the mothers and which does not help in arriving at the correct diagnosis but it is elicited in each pediatric interview.

4.5.2 Technique Adaptation to Rephrasing Questions for LEP Cases

Participant's usage of rephrasing is used for testing Primary Hypothesis 3.

- **Anticipated Technique Score:** In the pre-survey, each participant was asked to provide the techniques he/she anticipated to use to elicit the information he/she needs from an LEP patient. If the participant's answer includes rephrasing or something similar he/she gets a point.
- **Used Technique Score:** In the post survey, each participant was asked to provide the techniques he/she indeed used to elicit the information he/she needed from the LEP virtual mother. If the participant's answer includes rephrasing or something similar he/she gets a point.
- **Technique Adaptation Score:** The difference between the used technique score and the anticipated technique score gives the technique adaptation score. If the technique adaptation score is positive, then the participant learned from the system how to rephrase questions for LEP patients. Otherwise, no rephrasing technique was learned from the system.

4.5.3 Differential Diagnosis Score

Differential diagnosis score is used for testing the secondary expected findings.

A differential diagnosis is a systematic diagnostic method used to identify the presence of an entity where multiple alternatives are possible [2].

In the interim survey after the interview session, each participant had the opportunity to identify five different differential diagnoses based on the information the participant was able to elicit. The differential diagnoses are ordered by the most likely first.

Depending on the rank of the correct diagnosis, each participant obtains a score from 0 to 5. If the correct diagnosis is the participant's first differential diagnosis, the participant obtains a score of 5. If the correct diagnosis is the participant's fifth differential diagnosis, the participant obtains a score of 1. If the correct diagnosis is not any of the participant's differential diagnosis, the participant obtains a score of 0.

4.5.4 Number of Questions Asked by Participants

Number of questions asked by participants to the mothers is used for testing the secondary expected findings.

It is important to compare between the number of questions asked to the EP and LEP mothers. The LEP mothers need to be asked more questions than the EP mothers to elicit important information due to the limited ability of the LEP mothers to understand

all of the participant's questions. Some of the participant's questions would need to be rephrased or broken into more elementary questions.

4.5.5 Survey Responses

For each case, participants had to fill in surveys that have questions about that specific case. Here are the survey questions that are of concern for the conducted study:

- I have the necessary skills and knowledge to elicit a focused history of this patient (Q1).
- I knew what pertinent information was needed to make the diagnosis for this patient (Q2).
- It was easy to get the information I wanted from the mother (Q3).
- The mother gave clear answers to all of my questions (Q4).
- I didn't need to rephrase or clarify my questions to the mother (Q5).
- I feel confident about the diagnosis based on the history alone (Q6).

After finishing the two cases and the surveys, participants had to fill in a post survey about each of the two cases they interviewed.

Here are the survey questions that are of concern for the conducted study:

- Eliciting the history was easy (Q7).
- The mother always understood my questions (Q8).
- The mother was always able to answer my questions (Q9).
- I never had to rephrase my questions in order to get information from the mother (Q10).

Participants had to respond to those questions once for the first patient and once for the second patient.

All of the previous questions have a 5-point Likert scale as follows:

- Strongly Disagree: 1
- Disagree: 2

- Neither Agree nor Disagree: 3
- Agree: 4
- Strongly Agree: 5

Participants responses to the survey questions test if there was a difference between the EP and the LEP mothers regarding the clarity of their answers and their English proficiency.

CHAPTER 5
STUDY RESULTS AND DISCUSSION

5.1 Primary Hypothesis 1 and Percentage of Identified Discoveries

We used independent factorial ANOVA with three independent factors: English proficiency (EP, LEP), case order (first, second) and case type (Heart, Lung) for analyzing the study results for the first group. The first group contains the participants that interviewed both EP and LEP mothers (23 participants).

5.1.1 Class A Discoveries

5.1.1.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.066, p = 0.798$). As a result, the three-way interaction effect was excluded from the model.

5.1.1.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 0.005, p = 0.946$) or between case order and case type ($F_{1,39} = 0.01, p = 0.921$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case order ($F_{1,41} = 10.116, p = 0.003$) which needed a further investigation. Further investigation was carried through pairwise comparisons between English proficiency and case order and the following was noticed (Figure 5-1):

- When the EP mothers were interviewed after the LEP mothers, the percentage of class A discoveries that participants identified from the EP mothers was significantly higher than the percentage of class A discoveries that participants identified from the EP mothers when interviewed before the LEP mothers ($F_{1,41} = 4.634, p = 0.037$). This indicates that there was a learning effect. When the EP mothers were interviewed after the LEP mothers, participants learned from the LEP mothers in the first interview and hence the participants were able to elicit more information than they did when the EP mothers were their first experience.

- When the LEP mothers were interviewed before the EP mothers, the percentage of class A discoveries that participants identified from the LEP mothers was significantly higher than the percentage of class A discoveries that participants identified from the LEP mothers when interviewed after the EP mothers ($F_{1,41} = 5.489, p = 0.024$). When the LEP mothers were interviewed before the EP mothers, it was possible that participants had low expectations for the system so they did their best to elicit information from the LEP mothers. On the other hand, when the EP mothers were interviewed first, participants had high expectations for the next interview. The high expectations for the next interview caused the participants to expect from the LEP mothers what they experienced from the EP mothers which was wrong because eliciting information from the LEP mothers was harder. Hence participants elicited significantly less information from the LEP mothers than when interviewed first.

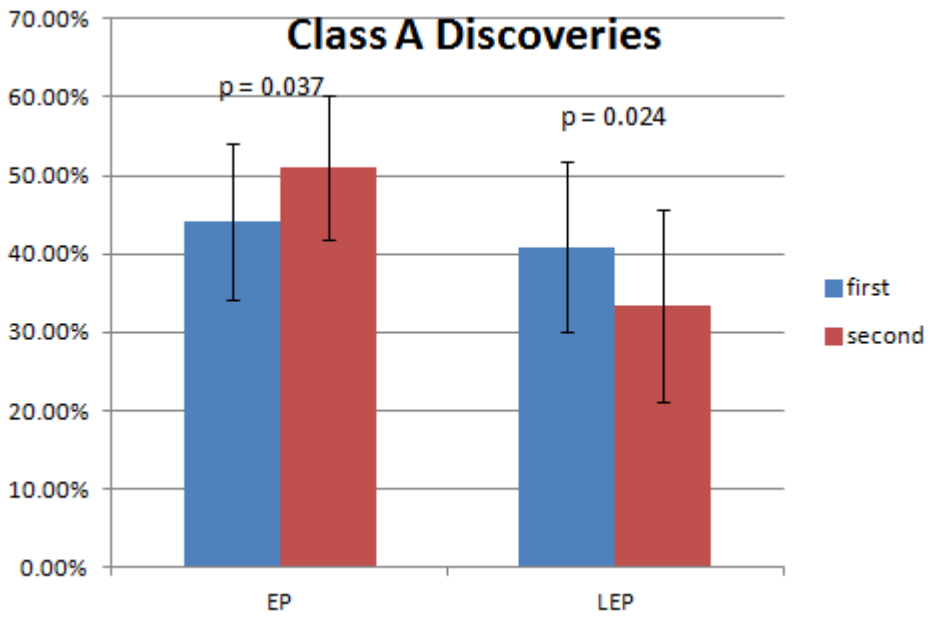

Figure 5-1. The interaction effect between English proficiency and case order for class A discoveries.

5.1.1.3 Main effects

Due to the significant interaction effect between English proficiency and case order, the data was divided into two groups to get rid of the order effect. One group has the data for participants that interviewed an EP mother followed by an LEP mother. The

other group has the data for participants that interviewed an LEP mother followed by an EP mother.

English proficiency: For the participants who interviewed an EP mother followed by an LEP mother, the percentage of class A discoveries that participants identified from the EP mothers was significantly larger than the percentage participants identified from the LEP mothers ($F_{1,19} = 9.110, p = 0.007$). For the participants who interviewed an LEP mother followed by an EP mother, the percentage of class A discoveries that participants identified from the EP mothers was significantly larger than the percentage participants identified from the LEP mothers ($F_{1,21} = 11.447, p = 0.003$) (Figure 5-2).

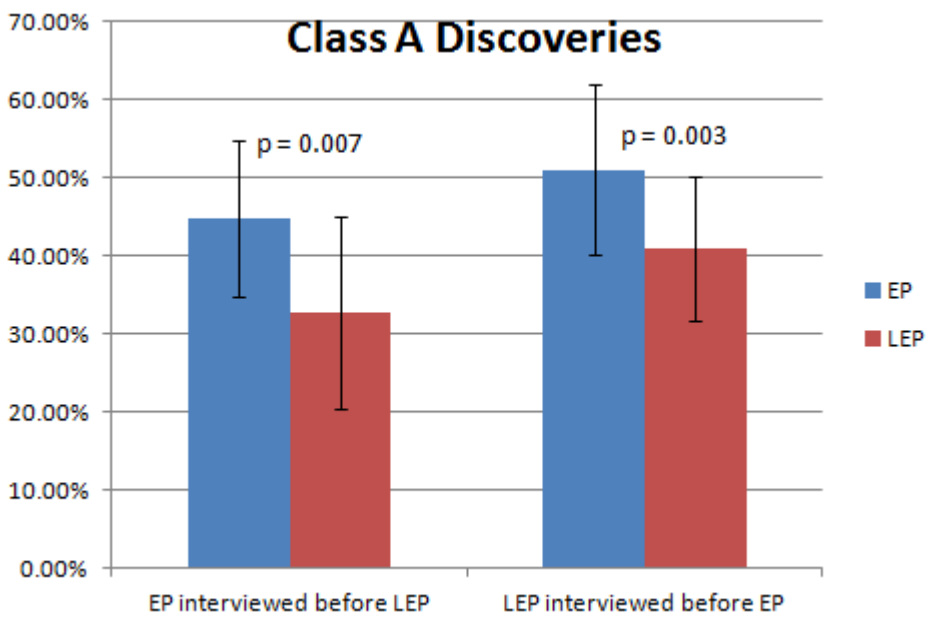

Figure 5-2. Results for percentage of identified class A discoveries from the EP and the LEP mothers.

For the two groups, the results agree with Primary Hypothesis 1. It was expected that interviewing the EP mothers would be easier than interviewing the LEP mothers. Interviewing the LEP mothers was hard due to the language barriers that prevented the

LEP mothers from understanding some of the participants' questions and hence not being able to answer those questions.

When the LEP mother understands a question, she may give an answer that has less information than the EP mother would give. For example the EP mother in the Heart Case would respond to the question "Can you describe his cough?" with "His cough sounds wet and deep, he does not really cough anything up though". The LEP mother in the same Heart Case would respond to the same question with "Well, for me cough is cough". The response of the LEP mother was the actual response we got from the LEP female that played the role of the mother in that case.

Due to the lack of enough information in some of the answers of the LEP mother, it makes sense that participants would elicit less information from the LEP mothers than they would elicit from the EP mothers and hence identify a lower percentage of discoveries.

Case type: For the participants who interviewed an EP mother followed by an LEP mother, the percentage of class A discoveries that participants identified from the mothers in the Heart Case was significantly larger than the percentage participants identified from the mothers in the Lung Case ($F_{1,19} = 17.818, p < 0.001$). For the participants who interviewed an LEP mother followed by an EP mother, the percentage of class A discoveries that participants identified from the mothers in the Heart Case was significantly larger than the percentage participants identified from the mothers in the Lung Case ($F_{1,21} = 21.018, p < 0.001$) (Figure 5-3).

For the two groups, The percentage of class A discoveries that participants identified from the Heart Case was significantly larger than the percentage participants identified from the Lung Case which indicates that the Lung Case was of a higher difficulty regarding eliciting important information. However, there was no interaction effect between the English proficiency and case type because within each case, the

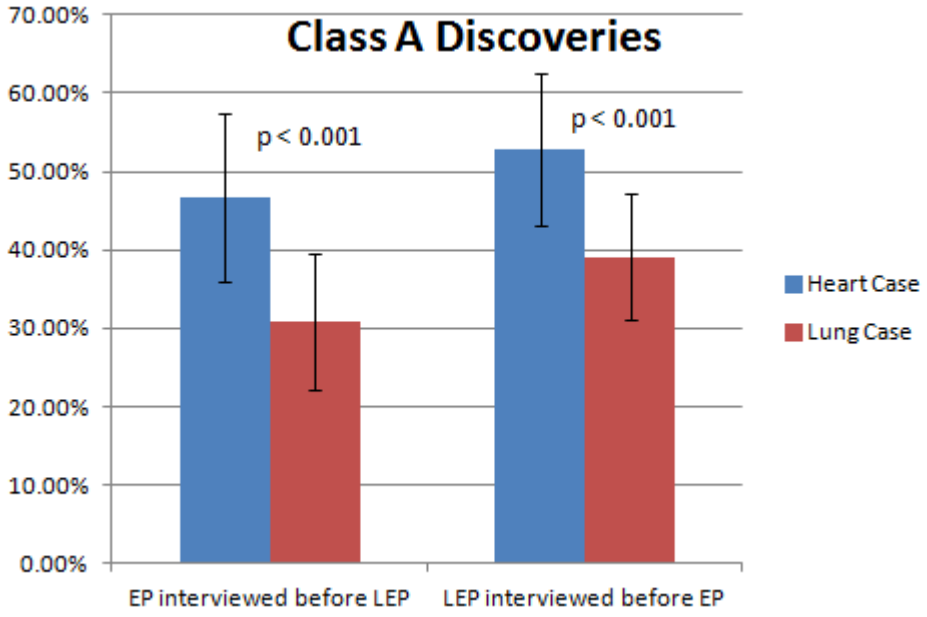

Figure 5-3. Results for percentage of identified class A discoveries from the mothers in the Heart and the Lang Cases.

number of class A discoveries that the participants identified from the EP mothers was larger than the number that the participants identified from the LEP mothers.

5.1.2 Class B Discoveries

5.1.2.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.019, p = 0.892$). As a result, the three-way interaction effect was excluded from the model.

5.1.2.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 2.123, p = 0.153$) or between case order and case type ($F_{1,39} = 0.940, p = 0.338$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case order ($F_{1,41} = 4.983$, $p = 0.031$) which needed a further investigation. Further investigation was carried through pairwise comparisons between English proficiency and case order and the following was noticed (Figure 5-4):

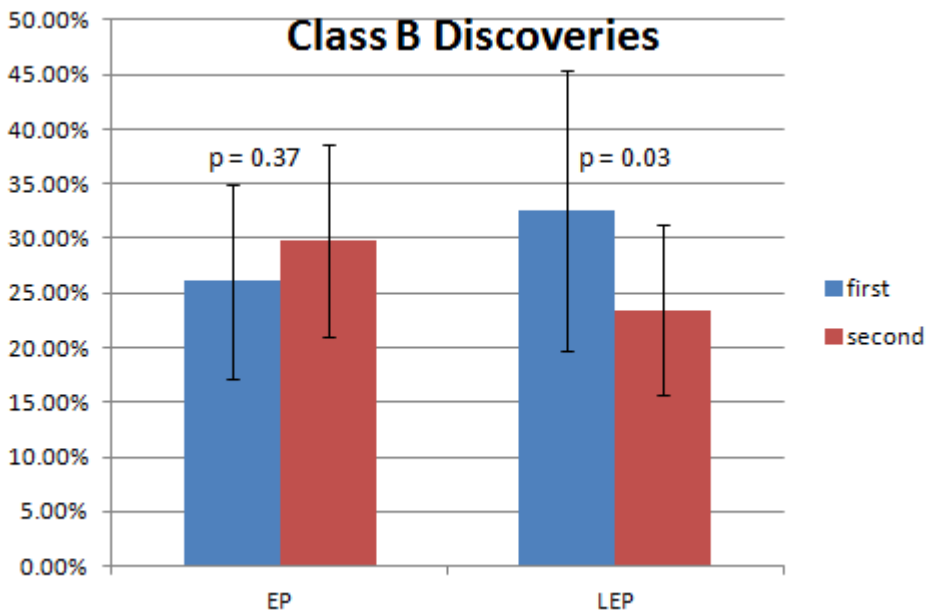

Figure 5-4. The interaction effect between English proficiency and case order for class B discoveries.

- There was no significant difference between the percentage of class B discoveries that participants identified from the EP mothers when the EP mothers were interviewed after the LEP mothers and the percentage of class B discoveries that participants identified from the EP mothers when interviewed before the LEP mothers ($F_{1,41} = 0.822$, $p = 0.37$).
- When the LEP mothers were interviewed before the EP mothers, the percentage of class B discoveries that participants identified from the LEP mothers was significantly higher than the percentage of class B discoveries that participants identified from the LEP mothers when interviewed after the EP mothers ($F_{1,41} = 5.058$, $p = 0.03$). This supports our argument about class A discoveries.

5.1.2.3 Main effects

Due to the significant interaction effect between English proficiency and case order, the data was divided into two groups to get rid of the order effect. One group has the data for participants that interviewed an EP mother followed by an LEP mother. The other group has the data for participants that interviewed an LEP mother followed by an EP mother.

English proficiency: For the participants who interviewed an EP mother followed by an LEP mother, there was no significant difference between the percentage of class B discoveries that participants identified from the EP mothers and the percentage participants identified from the LEP mothers ($F_{1,19} = 0.544, p = 0.470$). For the participants who interviewed an LEP mother followed by an EP mother, there was no significant difference between the percentage of class B discoveries that participants identified from the EP mothers and the percentage participants identified from the LEP mothers ($F_{1,21} = 0.384, p = 0.542$) (Figure 5-5).

Case type: For the participants who interviewed an EP mother followed by an LEP mother, there was no significant difference between the percentage of class B discoveries that participants identified from the mothers in the Heart Case and the percentage participants identified from the mothers in the Lung Case ($F_{1,19} = 0.544, p = 0.470$). For the participants who interviewed an LEP mother followed by an EP mother, there was no significant difference between the percentage of class B discoveries that participants identified from the mothers in the Heart Case and the percentage participants identified from the mothers in the Lung Case ($F_{1,21} = 0.384, p = 0.542$) (Figure 5-6).

In general, among all participants, the average percentage of identified class B discoveries was 28.8%. This percentage is low compared to the average percentage of identified class A discoveries which was 40.6%.

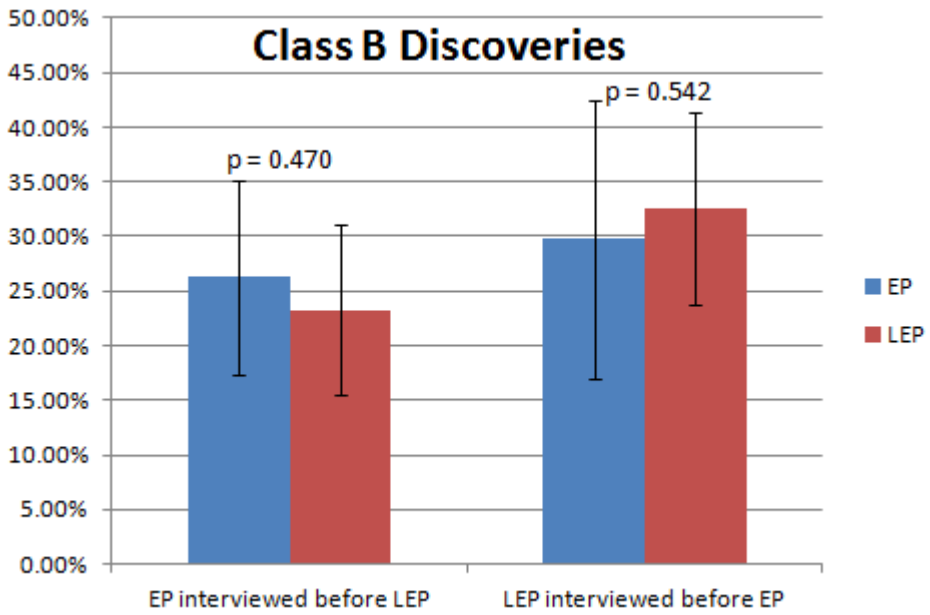

Figure 5-5. Results for percentage of identified class B discoveries from the EP and the LEP mothers.

The low average is possibly because class B discoveries are of a lower importance than class A discoveries and hence participants did not focus on that kind of discoveries. The low average makes it harder to see a significant difference.

5.1.3 Class C Discoveries

5.1.3.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.691, p = 0.411$). As a result, the three-way interaction effect was excluded from the model.

5.1.3.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 2.342, p = 0.134$) or between case order and case type ($F_{1,39} = 0.044, p = 0.836$). There was also no significant two-way interaction effect

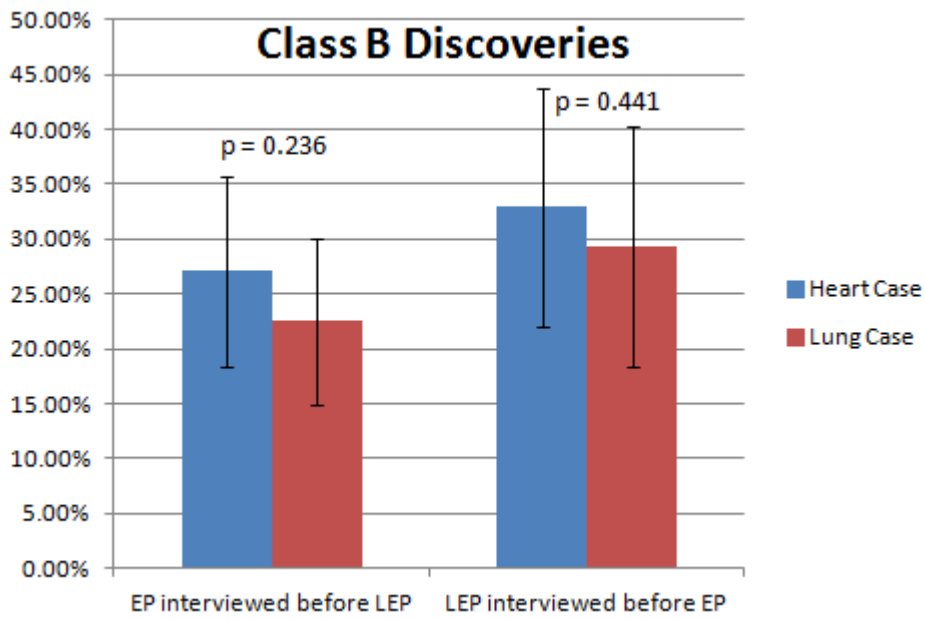

Figure 5-6. Results for percentage of identified class B discoveries from the mothers in the Heart and the Lang Cases.

between English proficiency and case order ($F_{1,39} = 1.398$, $p = 0.244$). As a results all two-way interaction effects were excluded from the model.

5.1.3.3 Main effects

There was no significant difference between the percentage of class C discoveries for the EP and the LEP mothers ($F_{1,42} = 0.879$, $p = 0.354$) or the mothers in the Heart Case and the mother in the Lung Case ($F_{1,42} = 0.879$, $p = 0.354$). There was also no significant difference between the percentage of class C discoveries for the first and the second interviewed cases ($F_{1,42} = 0.01$, $p = 0.922$) (Figure 5-7).

In general, among all participants, the average percentage of identified class C discoveries was 6.9%. This number is very low compared to the average percentage of identified class A discoveries which was 40.6%.

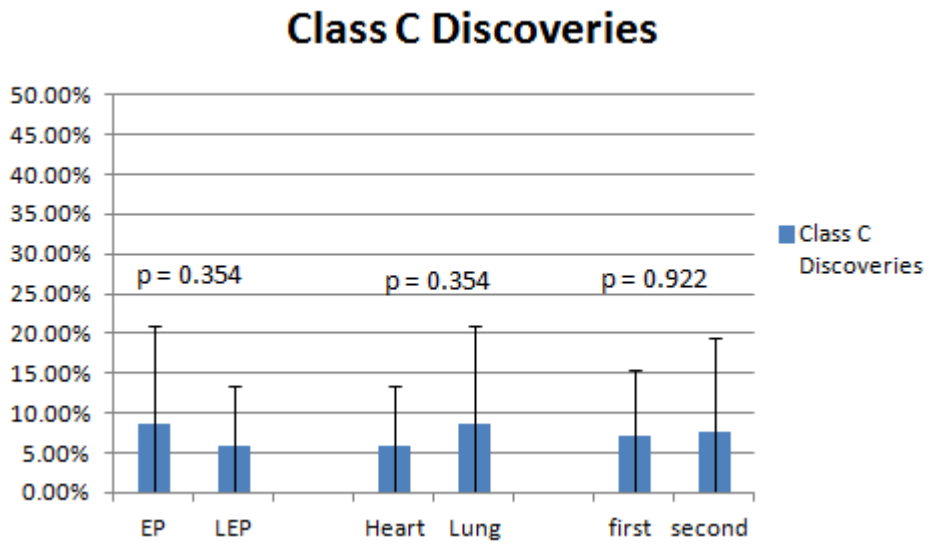

Figure 5-7. Results for percentage of identified class C discoveries from the EP and the LEP mothers, from the mothers in the Heart and the Lang Cases and from the first and the second interviewed mothers.

The low average is possibly because class C discoveries are of a much lower importance than class A discoveries and hence participants did not focus on that kind of discoveries. The low average makes it harder to see a significant difference.

5.2 Primary Hypothesis 2 and Percentage of Identified Discoveries

We used independent factorial ANOVA with two independent factors: case order (first, second) and case type (Heart, Lung) for analyzing the study results for the second group. The second group contains the participants that interviewed two LEP mothers (8 participants).

The results for number of class A, B and C discoveries regarding case order (first and second) are summarized in Table 5-1.

Table 5-1. Percentage of identified discoveries from the first and the second interviewed LEP mothers for each class of discoveries

	First LEP case mean (st. dev)	Second LEP case mean (st. dev)	P value
Percent of identified class A discoveries	33.6% (6.7)	36.6(10)	0.248
Percent of identified class B discoveries	33.76% (10.3)	27.67 (10.7)	0.283
Percent of identified class C discoveries	3.1% (8.8)	7.8 (9.3)	0.319

5.2.1 Class A Discoveries

5.2.1.1 Two-way interaction effect

There was no significant two-way interaction effect between case type and case order ($F_{1,12} = 1.525, p = 0.241$). As a result, the two-way interaction effect was excluded from the model.

5.2.1.2 Main effects

Case order: There was no significant difference between the percentage of class A discoveries identified from the first interviewed LEP mothers and the second interviewed LEP mothers ($F_{1,13} = 1.466, p = 0.248$).

This result rejects Primary Hypothesis 2 about participants improving their communication skills for dealing with the LEP mothers after using the system. One factor that may have affected the result is that there were only 8 participants (out of 31) that interviewed two LEP mothers one after another. Also, participants might have learned from the system in both interviews and hence no significant difference existed between the percentage of identified discoveries in the two interviews.

Case type: The percentage of class A discoveries that participants identified from the mothers in the Heart Case was significantly larger than the percentage participants identified from the mothers in the Lung Case ($F_{1,13} = 27.784, p < 0.001$) (Figure 5-8). This agrees with our argument about the Lung Case being more difficult than the Heart Case.

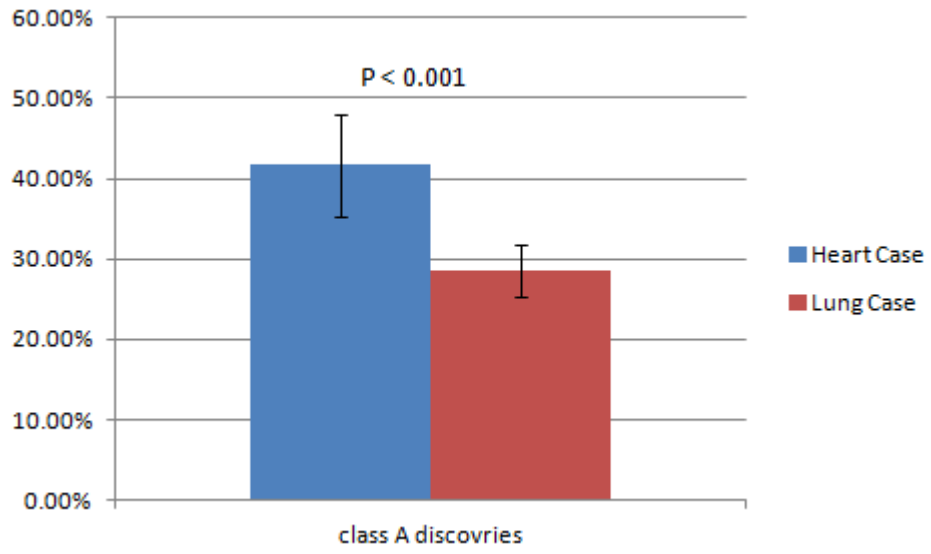

Figure 5-8. Results for percentage of identified class A discoveries from the mothers in the Heart and the Lang Cases (for participants in the second group).

5.2.2 Class B Discoveries

There was no significant difference between the percentage of class B discoveries that participants identified from the LEP mother in the first case and the LEP mother in the second case ($F_{1,13} = 1.253$, $p = 0.283$).

5.2.3 Class C Discoveries

There was no significant difference between the percentage of class C discoveries that participants identified from the LEP mother in the first case and the LEP mother in the second case ($F_{1,13} = 1.073$, $p = 0.319$).

5.3 Primary Hypothesis 3 and Technique Adaptation to Rephrasing Questions

Twelve participants out of the thirty one (38.71 %) mentioned that they used rephrasing to elicit information from the LEP mothers (Technique Adaptation Score = 1). Table 5-2 shows the response of two of these participants about the techniques they anticipated to use to elicit information from an LEP patient before and after their

Table 5-2. Technique adaptation to rephrasing questions for two participants.

	Before the interview with an LEP mother	After the interview with an LEP mother
Participant 1	"*Use other family members who can interpret. Call an interpreter service. Use my hands to gesture and try and describe my questions.*"	"*I tried rephrasing questions with some success. I also used more common words, such as "poop" instead of "stool" when the mother did not understand.*"
Participant 2	"*I try to use simple words. But if their levels of understanding of English is too low, I would get an interpreter or use language line. I haven't been in that situation yet.*"	"*Using more simple words. For example, she didn't understand the word fussy. So I worked around that.*"

actual interview with the LEP mothers. The rest of the participants might have used rephrasing without realizing and hence those participants did not mention that. In general, a stronger measure is needed to be used to support Primary Hypothesis 3.

5.4 Secondary Findings

We used independent factorial ANOVA with three independent factors: English proficiency (EP, LEP), case order (first, second) and case type (Heart, Lung) for analyzing the study results for the first group. The first group contains the participants that interviewed both EP and LEP mothers (23 participants).

5.4.1 Differential Diagnosis Score

5.4.1.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 1.865, p = 0.18$). As a result, the three-way interaction effect was excluded from the model.

5.4.1.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 0.534, p = 0.469$) or between English proficiency and case order ($F_{1,39} = 0.083, p = 0.775$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between case type and case order ($F_{1,41} = 5.193, p = 0.028$). Further investigation was carried through pairwise comparisons between case type and case order and the following was noticed (Figure 5-9):

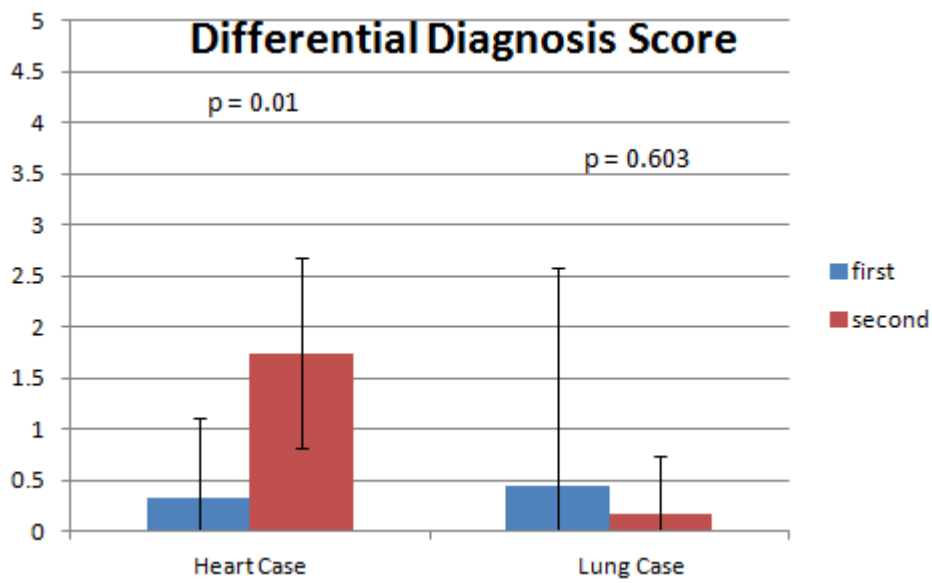

Figure 5-9. The interaction effect between case type and case order for the differential diagnosis score.

- For the Heart Case, the differential diagnosis score that the participants obtained from their second interview was significantly larger than the score that the participants obtained from their first interview ($F_{1,41} = 7.275, p = 0.01$). This result possibly indicates that the participants learned from the Lung Case in the first interview and hence obtained a higher score for their differential diagnosis regarding the Heart Case in the second interview.

- For the Lung Case, there was no significant difference between the differential diagnosis score that participants obtained from their first and second interviews ($F_{1,41} = 0.274, p = 0.603$). It is possible that the Heart Case in the first interview biased the differential diagnosis for the Lung case (which proved to be harder than the Heart Case) in the second interview.

5.4.1.3 Main effects

Due to the significant interaction effect between case type and case order, the data was divided into two groups to get rid of the order effect. One group has the data for participants that interviewed an EP mother followed by an LEP mother. The other group has the data for participants that interviewed an LEP mother followed by an EP mother.

English proficiency: For the participants who interviewed an EP mother followed by an LEP mother, there was no significant difference between the differential diagnosis score that the participants obtained after interviewing the EP mothers and the score that the participants obtained after interviewing the LEP mothers ($F_{1,19} = 2.792, p = 0.111$). For the participants who interviewed an LEP mother followed by an EP mother, there was no significant difference between the differential diagnosis score that the participants obtained after interviewing the EP mothers and the score that the participants obtained after interviewing the LEP mothers ($F_{1,21} = 0.214, p = 0.649$) (Figure 5-10).

However, this result does not mean that language barriers have no effect on physician's differential diagnosis because there was a factor that affected the results. The factor that affected the results was the difficulty of the Heart and Lung Cases. The chosen Heart and Lung Cases are rarely presented in clinical training. Presenting these cases through the virtual humans is still good for training purposes.

In general, the differential diagnosis scores that the participants got for both the EP and the LEP mothers were very low as most of the participants scored zero (Figure 5-11). This supports our argument that the result does not mean that language barriers have no effect on differential diagnosis. The results merely indicates that the cases we presented to the participants were not capable of showing a significant difference.

Case type: For the participants who interviewed an EP mother followed by an LEP mother, the differential diagnosis score that the participants obtained after

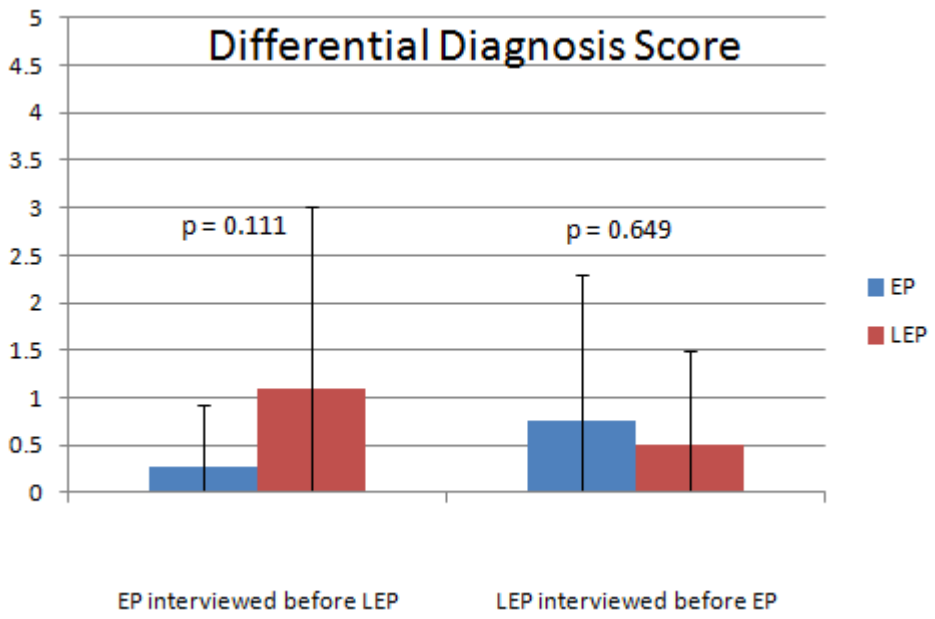

Figure 5-10. Results of differential diagnosis score for English proficiency (EP, LEP).

interviewing the mothers in the Heart Case was significantly larger than the score that the participants obtained after interviewing the mothers in the Lung Case ($F_{1,19} = 5.142, p = 0.035$). For the participants who interviewed an LEP mother followed by an EP mother, there was no significant difference between the differential diagnosis score that the participants obtained after interviewing the mothers in the Heart Case and the score that the participants obtained after interviewing the mothers in the Lung Case ($F_{1,21} = 0.214, p = 0.649$) (Figure 5-12). This possibly indicates that identifying the diagnosis for the Lung Case was harder than identifying the diagnosis for the Heart Case but there is no obvious reason, why was that significant for the first group only. This might be due to some differences between the participants in the two groups.

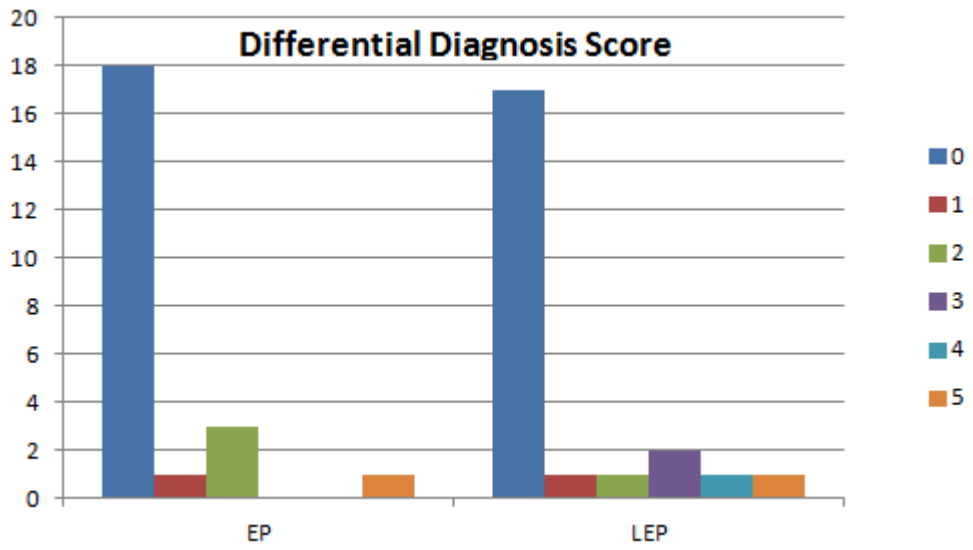

Figure 5-11. Histogram for differential diagnosis scores

5.4.2 Number of Questions

5.4.2.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.586, p = 0.449$). As a result, the three-way interaction effect was excluded from the model.

5.4.2.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 3.746, p = 0.06$) or between case order and case type ($F_{1,39} = 0.122, p = 0.729$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case order ($F_{1,41} = 4.227, p = 0.046$).

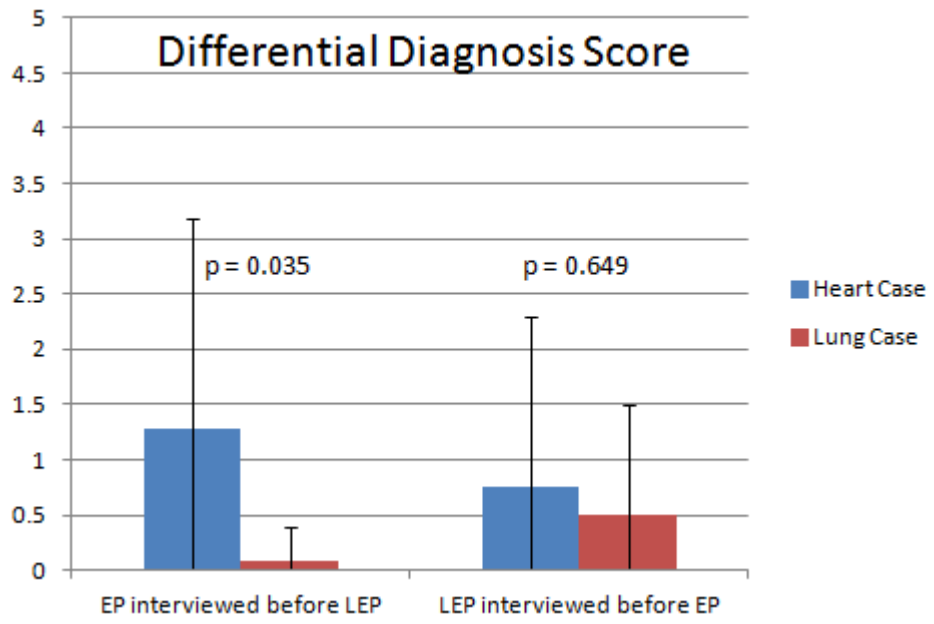

Figure 5-12. Results of differential diagnosis score for case type (Heart, Lung).

5.4.2.3 Main effects

Due to the significant interaction effect between English proficiency and case order, the data was divided into two groups to get rid of the order effect. One group has the data for participants that interviewed an EP mother followed by an LEP mother. The other group has the data for participants that interviewed an LEP mother followed by an EP mother.

English proficiency: For the participants who interviewed an EP mother followed by an LEP mother, there was no significant difference between the number of questions that participants asked to the EP mothers and the number of questions that participants asked to the LEP mothers ($F_{1,19} = 0.333, p = 0.571$). For the participants who interviewed an LEP mother followed by an EP mother, there was no significant difference between the number of questions that participants asked to the

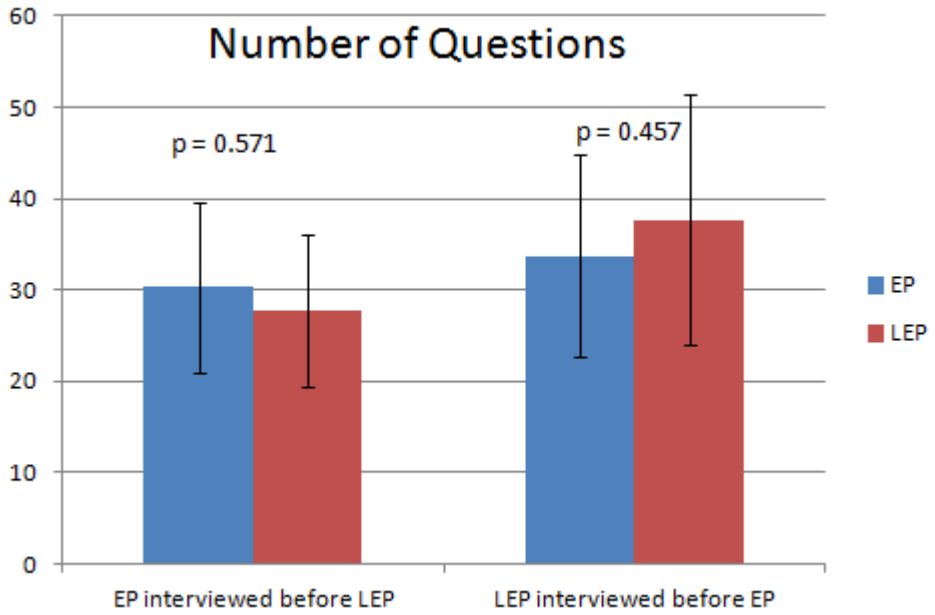

Figure 5-13. Results for numbers of questions asked to the EP and the LEP mothers.

EP mothers and the number of questions that participants asked to the LEP mothers ($F_{1,21} = 0.574$, $p = 0.457$) (Figure 5-13).

Case type: For the participants who interviewed an EP mother followed by an LEP mother, there was no significant difference between the number of questions that participants asked to the mothers in the Heart Case and the number of questions that participants asked to the mothers in the Lung Case ($F_{1,19} = 0.539$, $p = 0.472$). For the participants who interviewed an LEP mother followed by an EP mother, there was no significant difference between the number of questions that participants asked to the mothers in the Heart Case and the number of questions that participants asked to the mothers in the Lung Case ($F_{1,21} = 0.162$, $p = 0.691$) (Figure 5-14).

The insignificant results are possibly due to the fact that all questions asked by the participant were taken into account including repeated questions. Repeated questions

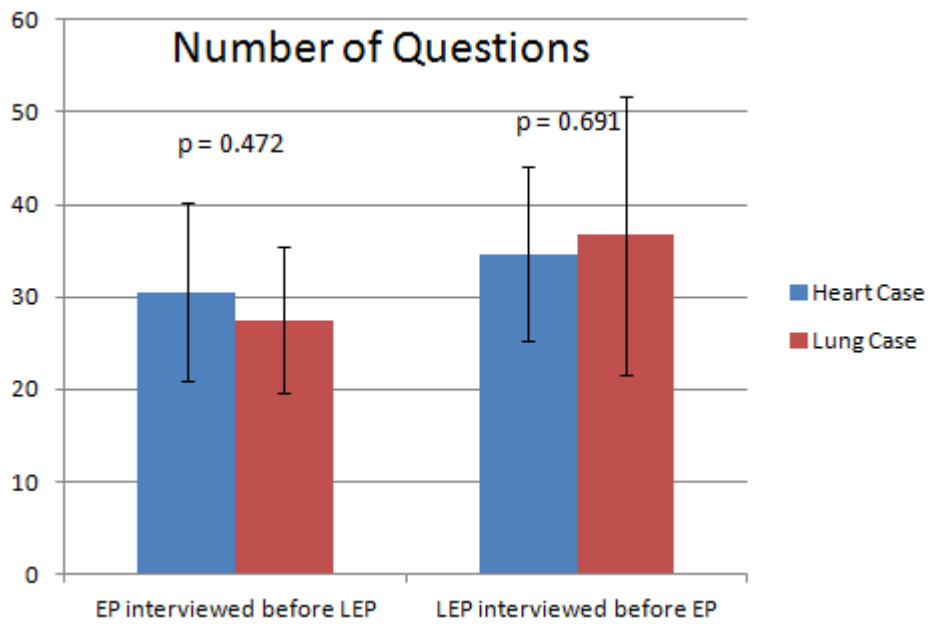

Figure 5-14. Results for numbers of questions asked to the mothers in the Heart and the Lung Cases.

were sometimes due to the limitation of the system to recognize those questions but some participants insisted on asking the same question more than once in different ways. It would be better to break down the asked questions into categories to capture significant variations.

5.4.3 I Have the Necessary Skills and Knowledge to Elicit a Focused History of This Patient (Q1)

5.4.3.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.186, p = 0.669$). As a result, the three-way interaction effect was excluded from the model.

5.4.3.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 0.248, p = 0.621$) or between case order and case type ($F_{1,39} = 1.255, p = 0.269$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case order ($F_{1,41} = 7.354, p = 0.01$).

5.4.3.3 Main effects

Due to the significant interaction effect between English proficiency and case order, the data was divided into two groups to get rid of the order effect. One group has the data for participants that interviewed an EP mother followed by an LEP mother. The other group has the data for participants that interviewed an LEP mother followed by an EP mother.

English proficiency: For the participants who interviewed an EP mother followed by an LEP mother, there was no significant difference between the level of interviewing confidence of participants who interviewed the EP mothers and for those participants who interviewed the LEP mothers ($F_{1,19} = 0.122, p = 0.731$). For the participants who interviewed an LEP mother followed by an EP mother, there was no significant difference between the level of interviewing confidence of participants who interviewed the EP mothers and for those participants who interviewed the LEP mothers ($F_{1,21} = 0.574, p = 0.457$) (Figure 5-15).

It is obvious from the survey results that the participants in general were confidant about about their skills and knowledge for eliciting a focused history for their patients regardless of the level of English proficiency of the interviewed mother.

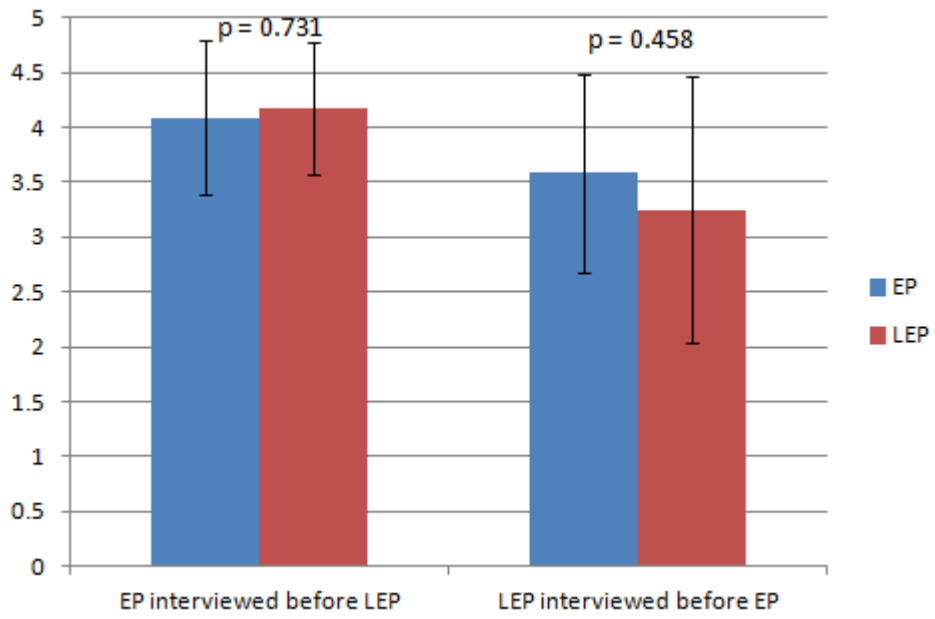

Figure 5-15. Results for Q1 (I have the necessary skills and knowledge to elicit a focused history of this patient).

5.4.4 I Knew What Pertinent Information Was Needed to Make the Diagnosis for This Patient (Q2)

5.4.4.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.155, p = 0.696$). As a result, the three-way interaction effect was excluded from the model.

5.4.4.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 0.033, p = 0.857$) or between case order and case type ($F_{1,39} = 2.649, p = 0.112$). There was also no significant two-way interaction effect between English proficiency and case order ($F_{1,39} = 2.649, p = 0.112$). As a results all two-way interaction effects were excluded from the model.

5.4.4.3 Main effects

English proficiency: There was no significant difference between the knowledge of information needed for diagnosis for participants that interviewed the EP mothers and for those participants who interviewed the LEP mothers ($F_{1,38} = 0.112, p = 0.740$) (Figure 5-16).

It is obvious from the survey results that the participants in general were confidant about the information needed to make the diagnosis for their patients regardless of the level of English proficiency of the interviewed mother.

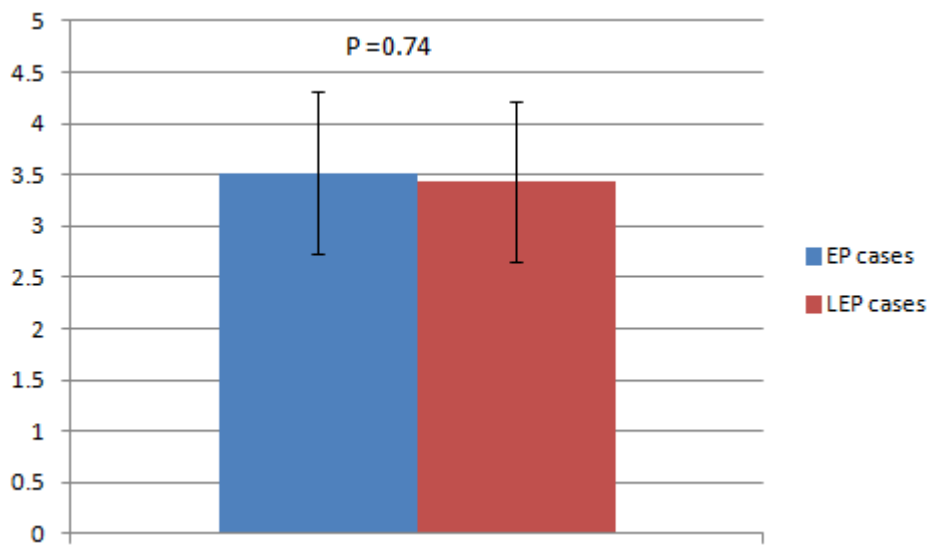

Figure 5-16. Results for Q2 (I knew what pertinent information was needed to make the diagnosis for this patient).

5.4.5 It Was Easy to Get the Information I Wanted From the Mother (Q3)

5.4.5.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.494, p = 0.486$). As a result, the three-way interaction effect was excluded from the model.

5.4.5.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case order ($F_{1,39} = 0.031, p = 0.86$) or between case order and case type ($F_{1,39} = 0.014, p = 0.907$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case type ($F_{1,41} = 6.769, p = 0.013$).

5.4.5.3 Main effects

Due to the significant interaction effect between English proficiency and case type, the data was divided into two groups to get rid of the case type effect. One group has the data for participants that interviewed an EP mother for the Heart Case and an LEP mother for the Lung Case. The other group has the data for participants that interviewed an EP mother for the Lung Case and an LEP mother for the Heart Case.

English proficiency: participants thought that there was no significant difference between eliciting information from the EP mothers in the Heart Case and eliciting information from the LEP mothers in the Lung Case ($F_{1,21} = 1.679, p = 0.209$). participants thought that eliciting information from the EP mothers in the Lung Case was significantly easier than eliciting information from the LEP mothers in the Heart Case($F_{1,21} = 7.174, p = 0.015$) (Figure 5-17).

These results might have been be due to an unintentional difference between the two EP mothers which might have caused eliciting information from the EP mother in the Heart Case to be almost as difficult as eliciting information from the LEP mother in the Lung Case. Or it might be due to some random choices for the participants.

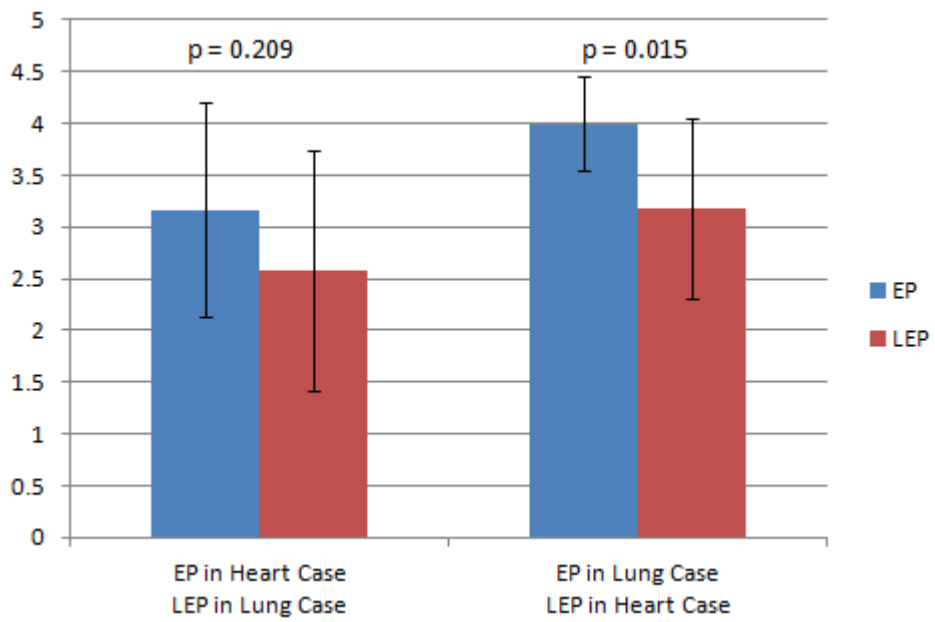

Figure 5-17. Results for Q3 (It was easy to get the information I wanted from the mother).

5.4.6 The Mother Gave Clear Answers to All of My Questions (Q4)

5.4.6.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.001, p = 0.970$). As a result, the three-way interaction effect was excluded from the model.

5.4.6.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case order ($F_{1,39} = 0.525, p = 0.473$) or between case order and case type ($F_{1,39} = 0.641, p = 0.428$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case type ($F_{1,41} = 5.194, p = 0.028$).

5.4.6.3 Main effects

Due to the significant interaction effect between English proficiency and case type, the data was divided into two groups to get rid of the case type effect. One group has the data for participants that interviewed an EP mother for the Heart Case and an LEP mother for the Lung Case. The other group has the data for participants that interviewed an EP mother for the Lung Case and an LEP mother for the Heart Case.

English proficiency: participants thought that the answers of the EP mothers in the Heart Case were significantly clearer than the answers of the LEP mothers in the Lung Case ($F_{1,21} = 15.464$, $p = 0.001$). Participants thought that there was no significant difference between the clarity of the answers of the EP mothers in the Lung Case and the clarity of the answers of the LEP mothers in the Heart Case ($F_{1,19} = 0.958$, $p = 0.34$) (Figure 5-18).

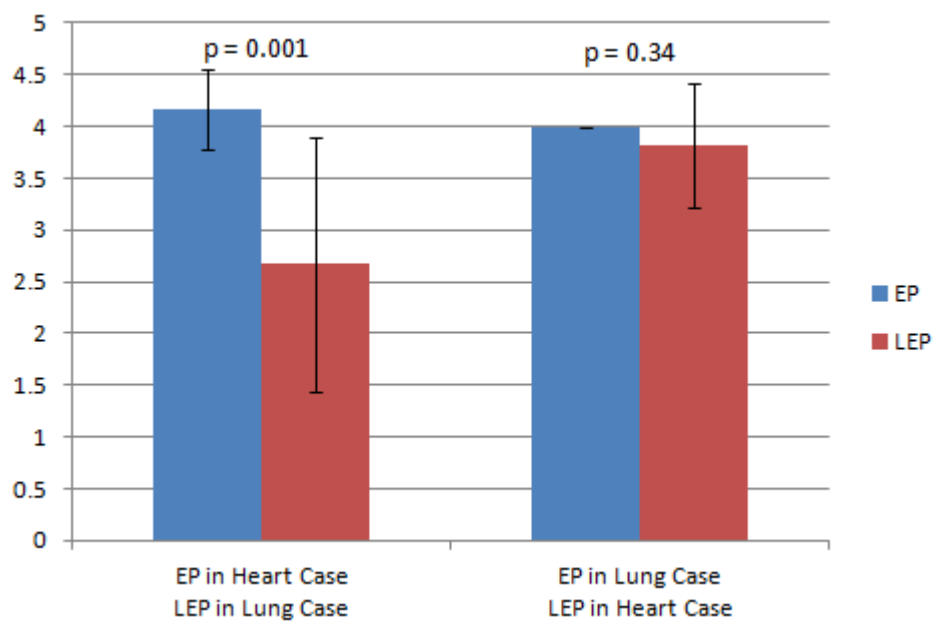

Figure 5-18. Results for Q4 (The mother gave clear answers to all of my questions).

These results might have been be due to an unintentional difference between the two LEP mothers which might have caused the answers of the LEP mother in the Heart Case to be almost as clear as the answers of the EP mother in the Lung Case. Or it might be due to some random choices for the participants.

5.4.7 I Did Not Need to Rephrase or Clarify My Questions to the Mother (Q5)

5.4.7.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.24, p = 0.627$). As a result, the three-way interaction effect was excluded from the model.

5.4.7.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case order ($F_{1,39} = 1.277, p = 0.265$) or between case order and case type ($F_{1,39} = 0.005, p = 0.944$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case type ($F_{1,41} = 4.520, p = 0.04$).

5.4.7.3 Main effects

Due to the significant interaction effect between English proficiency and case type, the data was divided into two groups to get rid of the case type effect. One group has the data for participants that interviewed an EP mother for the Heart Case and an LEP mother for the Lung Case. The other group has the data for participants that interviewed an EP mother for the Lung Case and an LEP mother for the Heart Case.

English proficiency: Participants thought that there was no significant difference between the need of rephrasing for the EP mothers in the Heart Case and the need of rephrasing for the LEP mothers in the Lung Case ($F_{1,21} = 0.645, p = 0.431$). Participants thought that there was no significant difference between the need of

rephrasing for the EP mothers in the Lung Case and the need of rephrasing for the LEP mothers in the Heart Case ($F_{1,19} = 0.262, p = 0.614$) (Figure 5-19).

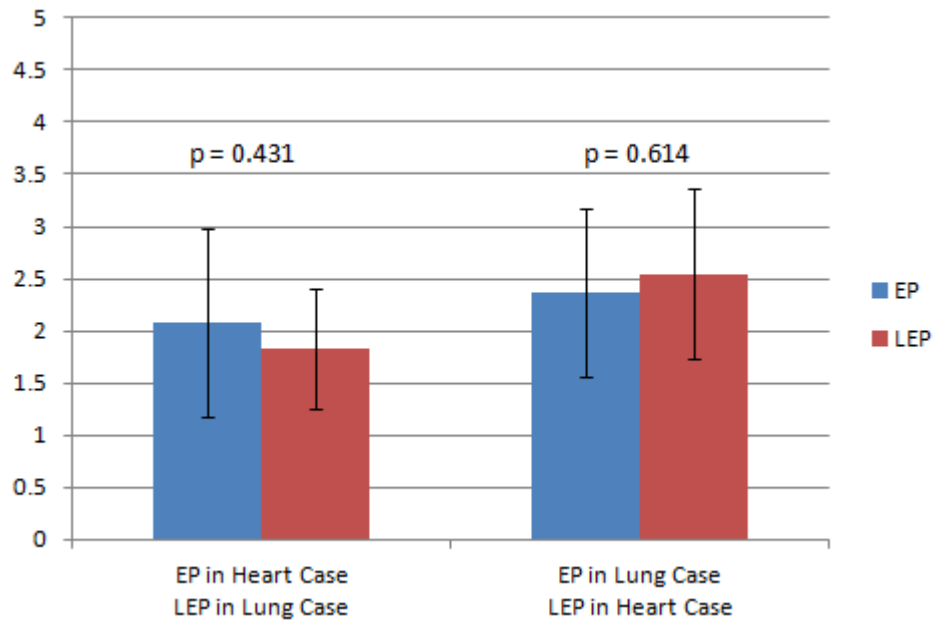

Figure 5-19. Results for Q5 (I did not need to rephrase or clarify my questions to the mother).

The insignificant difference might have been be due to the need for rephrasing in general for both EP and LEP mothers because of the limitations in the system database.

5.4.8 I Feel Confident About the Diagnosis Based on the History Alone (Q6)

5.4.8.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.907, p = 0.347$). As a result, the three-way interaction effect was excluded from the model.

5.4.8.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case order ($F_{1,39} = 0.417, p = 0.522$) or between case order and case type

($F_{1,39} = 2.271, p = 0.14$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case type ($F_{1,41} = 7.935, p = 0.007$).

5.4.8.3 Main effects

Due to the significant interaction effect between English proficiency and case type, the data was divided into two groups to get rid of the case type effect. One group has the data for participants that interviewed an EP mother for the Heart Case and an LEP mother for the Lung Case. The other group has the data for participants that interviewed an EP mother for the Lung Case and an LEP mother for the Heart Case.

English proficiency: There was no significant difference between the confidence of diagnosis for participants that interviewed the EP mothers in the Heart Case and the confidence of diagnosis for participants that interviewed the LEP mothers in the Lung Case ($F_{1,21} = 0.074, p = 0.789$). There was no significant difference between the confidence of diagnosis for participants that interviewed the EP mothers in the Lung Case and the confidence of diagnosis for participants that interviewed the LEP mothers in the Heart Case ($F_{1,19} = 0.252, p = 0.622$) (Figure 5-20).

These results shows that in general the confidence about the diagnosis was low regardless the level of English proficiency of the mothers. These results are consistent with the low diagnosis score that the participants obtained in general.

5.4.9 Eliciting the History Was Easy (Q7)

5.4.9.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.403, p = 0.529$). As a result, the three-way interaction effect was excluded from the model.

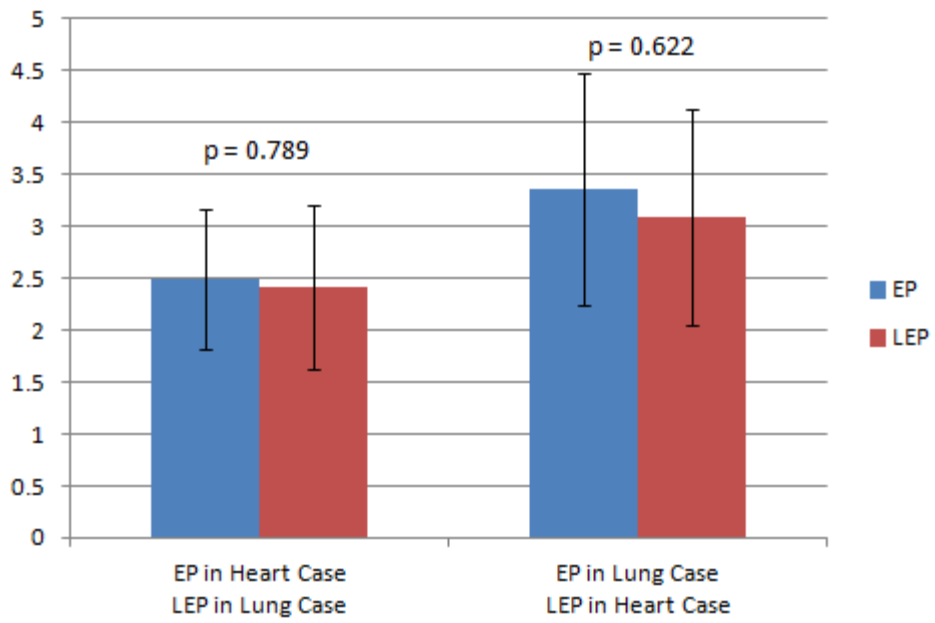

Figure 5-20. Results for Q6 (I feel confident about the diagnosis based on the history alone).

5.4.9.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 1.805, p = 0.187$) or between case order and case type ($F_{1,39} = 0.495, p = 0.486$). As a results those two-way interaction effects were excluded from the model.

There was a significant two-way interaction effect between English proficiency and case order ($F_{1,41} = 8.114, p = 0.007$).

5.4.9.3 Main effects

Due to the significant interaction effect between English proficiency and case order, the data was divided into two groups to get rid of the order effect. One group has the data for participants that interviewed an EP mother followed by an LEP mother. The

other group has the data for participants that interviewed an LEP mother followed by an EP mother.

English proficiency: For the participants who interviewed an EP mother followed by an LEP mother, there was no significant difference between eliciting information from the EP mothers and eliciting information from the LEP mothers ($F_{1,19} = 0.922, p = 0.349$). For the participants who interviewed an LEP mother followed by an EP mother, eliciting information from the EP mothers was no significantly easier than eliciting information from the LEP mothers ($F_{1,21} = 4.565, p = 0.045$) (Figure 5-21).

There is no logical explanation, why the significance exists only in the second group.

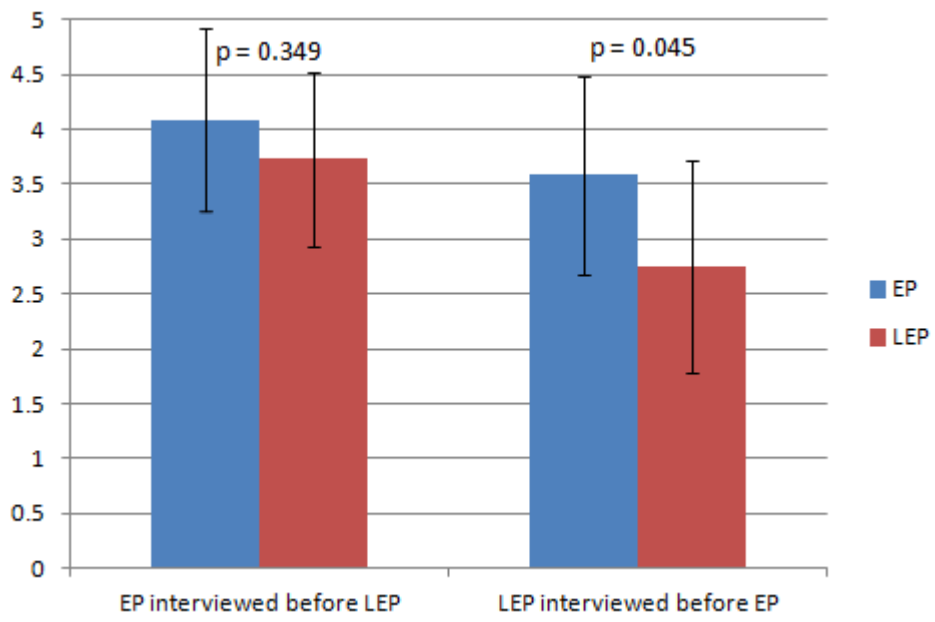

Figure 5-21. Results for Q7 (Eliciting the history was easy).

5.4.10 The Mother Always Understood My Questions (Q8)

5.4.10.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.406, p = 0.528$). As a result, the three-way interaction effect was excluded from the model.

5.4.10.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case order ($F_{1,39} = 0.0, p = 1.0$) or between case order and case type ($F_{1,39} = 0.085, p = 0.772$). Also, there was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 0.085, p = 0.772$). As a results all two-way interaction effects were excluded from the model.

5.4.10.3 Main effects

English proficiency: Participants thought that the ability of understanding the participants' questions for the EP mothers was significantly higher than the ability of understanding the participants' questions for the LEP mothers ($F_{1,42} = 4.366, p = 0.043$) (Figure 5-22).

This result supports the capability of the virtual humans to simulate humans with different levels of English proficiency as in general EP humans are able to understand questions asked in English significantly better than LEP humans.

5.4.11 The Mother Was Always Able to Answer My Questions (Q9)

5.4.11.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.216, p = 0.645$). As a result, the three-way interaction effect was excluded from the model.

5.4.11.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case order ($F_{1,39} = 0.932, p = 0.34$) or between case order and case type

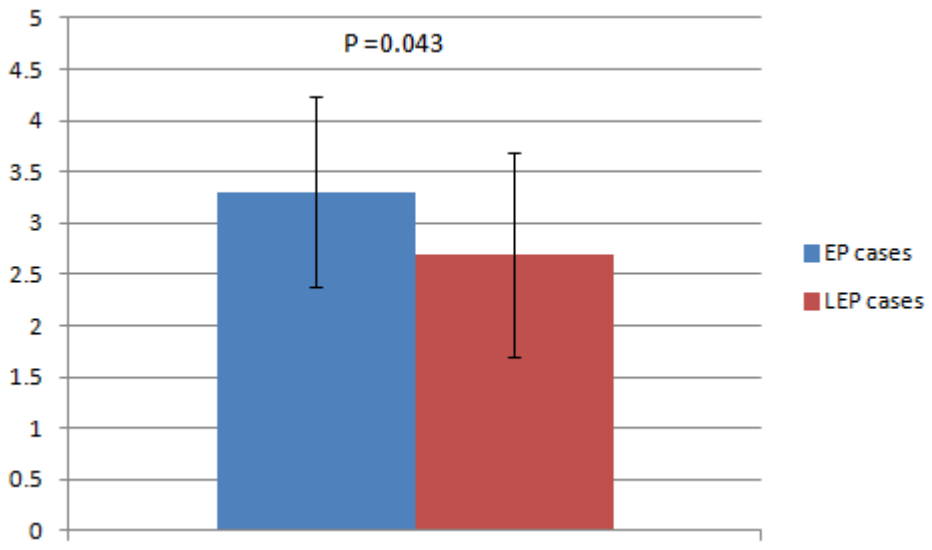

Figure 5-22. Results for Q8 (The mother always understood my questions).

($F_{1,39} = 0.476, p = 0.494$). Also, there was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 1.541, p = 0.222$). As a results all two-way interaction effects were excluded from the model.

5.4.11.3 Main effects

English proficiency: Participants thought that the ability of answering the participants' questions was significantly higher for the EP mothers than for the LEP mothers ($F_{1,42} = 5.706, p = 0.021$) (Figure 5-23).

This result supports the capability of the virtual humans to simulate humans with different levels of English proficiency as in general EP humans are able to answer questions that are asked in English significantly better than LEP humans.

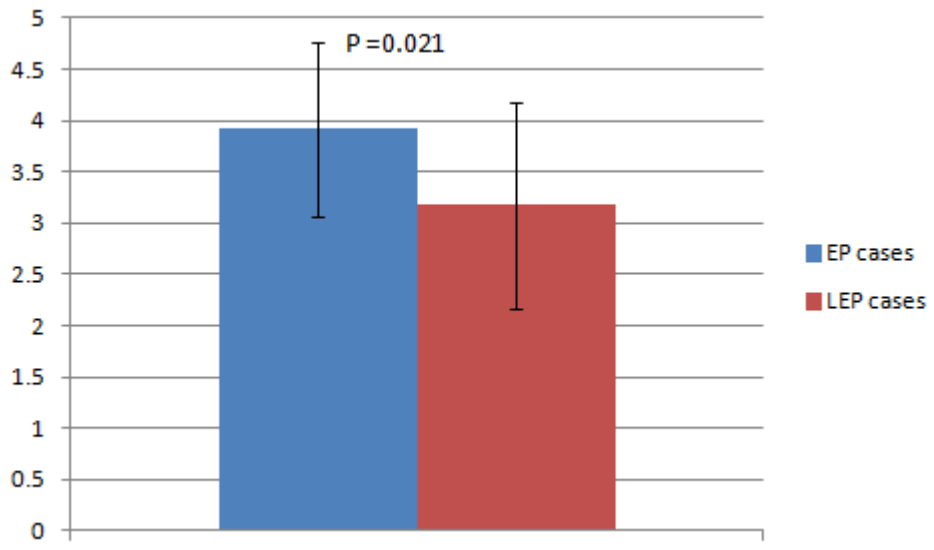

Figure 5-23. Results for Q9 (The mother was always able to answer my questions).

5.4.12 I Never Had to Rephrase My Questions in Order to Get Information From the Mother (Q10)

5.4.12.1 Three-way interaction effect

There was no significant three-way interaction effect between English proficiency, case type and case order ($F_{1,38} = 0.205, p = 0.654$). As a result, the three-way interaction effect was excluded from the model.

5.4.12.2 Two-way interaction effect

There was no significant two-way interaction effect between English proficiency and case order ($F_{1,39} = 0.292, p = 0.592$) or between case order and case type ($F_{1,39} = 1.880, p = 0.178$). Also, there was no significant two-way interaction effect between English proficiency and case type ($F_{1,39} = 0.913, p = 0.345$). As a results all two-way interaction effects were excluded from the model.

5.4.12.3 Main effects

English proficiency: Participants thought that there was no significant difference between the need of rephrasing questions for the EP mothers and the need of rephrasing questions for the LEP mothers ($F_{1,42} = 0.662$, $p = 0.421$) (Figure 5-24).

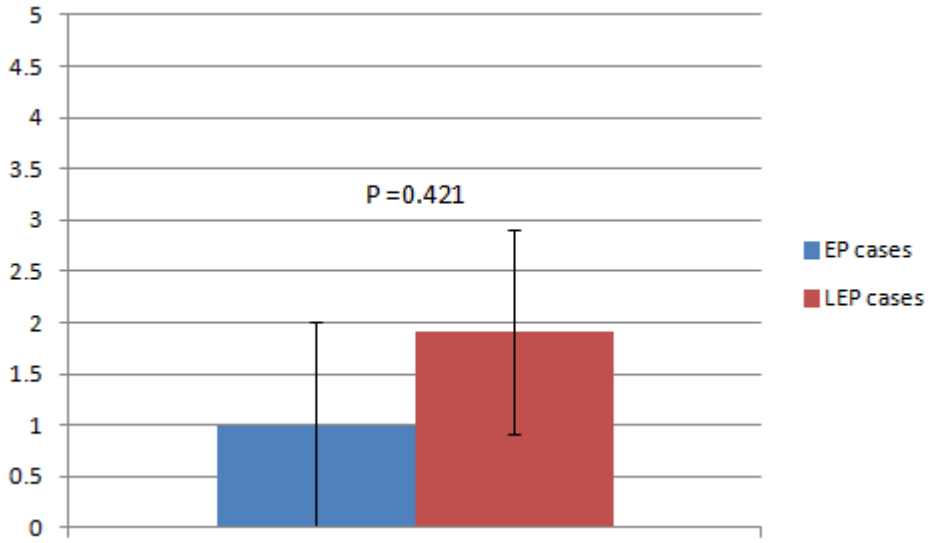

Figure 5-24. Results for Q10 (I never had to rephrase my questions in order to get information from the mother).

The insignificant difference might have been be due to the need for rephrasing in general for both EP and LEP mothers because of the limitations in the system database.

CHAPTER 6
LIMITATION AND FUTURE WORK

6.1 Limitations

6.1.1 Limited Database

The system has 220 responses to 1200 different questions in the EP scripts and 260 responses to 1200 different questions in the LEP scripts. This limited number of questions in the system database may have had a negative effect on the performance of the participants. This limited number of questions may have caused the participants to rephrase their questions to the mothers. Consequently, the participants were not able to distinguish between rephrasing for the system to recognize the question and rephrasing for the LEP mothers to understand the question.

6.1.2 Difficulty of Medical Cases

Two medical cases were used in the study: Heart Case and Lung Case. Both medical cases were of high difficulty for participants to come to the correct diagnosis. It would be better to have more common medical cases.

6.1.3 Limited Number of Participants

The number of participants was reasonably enough to draw conclusions about the effects of English proficiency on medical interviewing outcomes. However, some of the statistical conclusions might be strengthened by a larger sample of participants.

6.2 Future Work

6.2.1 Improving the Database

The number of questions in the script database can be increased by adding all the unmatched questions of the participants to the database. By adding all those questions to the database, the mothers will be able to respond to a larger number of questions which will improve the experience of the user with the system.

6.2.2 Having the System Online

Transferring the system into a new platform so as to have it available online can increase the number of participants and hence improve the results.

6.2.3 Giving Immediate Feedback

The system improves the capabilities of health care providers in eliciting information from an LEP mother by giving the providers responses from the virtual LEP mother that reflect her understanding of their questions. For example, an answer like “I don’t know what is discharge” tells providers that the mother could not understand their question.

In order to elicit the information providers want, providers need to rephrase their questions and to use simpler words. Currently the system gives the providers the impression that they need to rephrase their questions but it does not tell providers explicitly what is wrong with their questions.

For future work, we are planning to make the system more educational when dealing with LEP cases by giving health care providers immediate feedback about what was wrong with their questions and what they need to ask instead. For the previous example, it would be nice that the system could tell providers that “discharge” is hard for an LEP parent to understand. Moreover, it would also tell health care providers that they can replace the word discharge by fluids for instance.

6.2.4 Possible Studies

6.2.4.1 Order effect on LEP patients

When the EP mothers were interviewed after the LEP mothers, the percentage of class A discoveries that participants identified from the EP mothers was significantly higher than the percentage of class A discoveries that participants identified from the EP mothers when interviewed before the LEP mothers. This result makes sense because of the learning effect. However, when the LEP mothers were interviewed before the EP mothers, the percentage of class A discoveries that participants identified from the LEP mothers was significantly higher than the percentage of class A discoveries that

participants identified from the LEP mothers when interviewed after the EP mothers. Further studies are needs to be conducted to understand the previous result. Would participants always perform worse with an LEP parent when interviewed secondly or was it because the LEP mothers were interviewed after EP mothers?

The results from that future study can be very useful for medical training. Medical training may need to always start on LEP patients which will also benefit EP patients. If medical training is initialized on EP patients, that might hinder the performance with LEP patients later.

6.2.4.2 Improving the quality of health care for LEP patients

Further studies are needed to be conducted to investigate the possibility that our system has the ability to improve the quality of health care that LEP patients get by possibly following up with our thirty one participants and assessing their current performance with LEP patients.

Also new studies can be conducted to check for improvement in participants' performance when they use our system over time especially after having the system deployed online.

CHAPTER 7
CONCLUSION

In this thesis, we introduced a novel virtual environment for assessing the quality of health care that LEP patients get in comparison with the quality of health care that EP patients get.

We demonstrated through a user study that the educational benefits that medical students and residents got from interacting with virtual mothers were as good as those gained from dealing with real-life cases. As well, the medical staff learned how to rephrase their questions to overcome misunderstanding and confusion as they dealt with LEP mothers.

The results drawn from the proposed system are highly encouraging and show a great potential for using virtual humans for simulating LEP humans to assess the quality of health care in LEP groups. The usefulness of the proposed system can be further improved by enlarging the associated medical script database and deploying the system on the web.

Using virtual humans for simulating LEP humans may not only be used in the medical field. Using virtual humans for simulating LEP humans may be used in the military or any other field.

Further studies are needed with different medical cases that are more common than the Heart and Lung cases that we used in our study. Also further studies are needed to check for order effect as it was obvious from the results that participants performed significantly worse with the LEP mothers when they were interviewed after the EP mothers.

REFERENCES

[1] Guidance to federal financial assistance recipients regarding title vi prohibition against national origin discrimination affecting limited english proficient persons.

[2] *Ferri's differential diagnosis* :, 2nd ed. ed. Elsevier/Mosby,, Philadelphia, PA :, c2011.

[3] A, K., JC, C., K, J., and B, L. Simulation of a virtual patient with cranial nerve injury augments physician-learner concern for patient safety. *Journal of Bio-Algorithms and Med-Systems 6*, 11 (2010), 25–34.

[4] Babu, S., Suma, E. A., Barnes, T., and Hodges, L. F. Can immersive virtual humans teach social conversational protocols? In *VR*, W. R. Sherman, M. Lin, and A. Steed, Eds., IEEE Computer Society (2007), 215–218.

[5] Bach, P. B., Pham, H. H., Schrag, D., Tate, R. C., and Hargraves, J. L. Primary care physicians who treat blacks and whites. *New England Journal of Medicine 351*, 6 (August 2004), 575–584. Access to full text is subject to the publisher's access restrictions.

[6] Bickmore, T. W., Caruso, L., Clough-Gorr, K., and Heeren, T. 'it's just like you talk to a friend' relational agents for older adults. *Interact. Comput. 17*, 6 (Dec. 2005), 711–735.

[7] Brian D. Smedley, A. Y. S., Alan R. Nelson, Editors, C. o. U., Racial, E., and in Health Care, E. D. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care (full printed version)*. The National Academies Press, 2003.

[8] de Melo, C. M., Carnevale, P., and Gratch, J. The effect of expression of anger and happiness in computer agents on negotiations with humans. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (Taipei, Taiwan, May 2011).

[9] Deaton, E., Barba, C., Santarelli, T., Rosenzweig, L., Souders, V., McCollum, C., Seip, J., Knerr, W., and Singer, J. Virtual environment cultural training for operational readiness (vector). *Virtual Real. 8*, 3 (May 2005), 156–167.

[10] Decamp, L., Choi, H., and Davis, M. Medical home disparities for latino children by parental language of interview. *J Health Care Poor Underserved 22*, 4 (2011), 1151–66.

[11] Diamond, L. C., and Reuland, D. S. Describing physician language fluency: deconstructing medical spanish. *JAMA 301*, 4 (2009), 426–8.

[12] Diamond, L. C., Schenker, Y., Curry, L., Bradley, E. H., and Fernandez, A. Getting by: underuse of interpreters by resident physicians. *J Gen Intern Med 24*, 2 (2009), 256–62.

[13] Ferguson, W. J., and Candib, L. M. Culture, language, and the doctor-patient relationship. *Fam Med 34*, 5 (2002), 353–61.

[14] Fernandez, A., Schillinger, D., Warton, E. M., Adler, N., Moffet, H. H., Schenker, Y., Salgado, M. V., Ahmed, A., and Karter, A. J. Language barriers, physician-patient language concordance, and glycemic control among insured latinos with diabetes: The diabetes study of northern california (distance). *J Gen Intern Med 26*, 2 (2011), 170–6.

[15] Gany, F., Leng, J., Shapiro, E., Abramson, D., Motola, I., Shield, D. C., and Changrani, J. Patient satisfaction with different interpreting methods: a randomized controlled trial. *J Gen Intern Med 22 Suppl 2* (2007), 312–8.

[16] Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., and Badler, N. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems 17*, 4 (July 2002), 54–63.

[17] Jacobs, E. A., Lauderdale, D. S., Meltzer, D., Shorey, J. M., Levinson, W., and Thisted, R. A. Impact of interpreter services on delivery of health care to limited-english-proficient patients. *J Gen Intern Med 16*, 7 (2001), 468–74.

[18] Johnsen, K., Dickerson, R., Raij, A., Lok, B., Jackson, J., Shin, M., Amy, J. H., and Lind, D. S. Experiences in using immersive virtual characters to educate medical communication skills. In *In Proc. IEEE Virtual Reality*, IEEE Press (2005).

[19] Johnsen, K., Raij, A., Stevens, A., Lind, D. S., and Lok, B. The validity of a virtual human experience for interpersonal skills education. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, ACM (New York, NY, USA, 2007), 1049–1058.

[20] Jr., R. W. H., Gratch, J., Marsella, S., Rickel, J., Swartout, W. R., and Traum, D. R. Virtual humans in the mission rehearsal exercise system. *KI 17*, 4 (2003), 5–.

[21] Kang, S.-H., Gratch, J., Sidner, C., Artstein, R., Huang, L., and Morency, L.-P. Towards building a virtual counselor: Modeling nonverbal behavior during intimate self-disclosure. In *Eleventh International Conference on Autonomous Agents and Multiagent Systems* (Valencia, Spain, June 2012).

[22] Karliner, L. S., Jacobs, E. A., Chen, A. H., and Mutha, S. Do professional interpreters improve clinical care for patients with limited english proficiency? a systematic review of the literature. *Health Serv Res 42*, 2 (2007), 727–54.

[23] Karliner, L. S., Perez-Stable, E. J., and Gildengorin, G. The language divide. the importance of training in the use of interpreters for outpatient practice. *J Gen Intern Med 19*, 2 (2004), 175–83.

[24] Kotranza, A., Johnsen, K., Cendan, J., Miller, B., Lind, D. S., and Lok, B. Virtual multi-tools for hand and tool-based interaction with life-size virtual human agents. In

Proceedings of the 2009 IEEE Symposium on 3D User Interfaces, 3DUI '09, IEEE Computer Society (Washington, DC, USA, 2009), 23–30.

[25] Ngo-Metzger, Q., Sorkin, D. H., Phillips, R. S., Greenfield, S., Massagli, M. P., Clarridge, B., and Kaplan, S. H. Providing high-quality care for limited english proficient patients: the importance of language concordance and interpreter use. *J Gen Intern Med 22 Suppl 2* (2007), 324–30.

[26] Pandya, Chhandasi, Batalova, J., and McHugh, M. Limited english proficient individuals in the united states: Number, share, growth and linguistic diversity., 2011.

[27] Parsons, T. D., and Rizzo, A. Virtual human patients for training of clinical interview and communication skills. In *Proceedings of the 2008 International Conference on Disability, Virtual Reality and Associated Technology* (Maia, Portugal, Sept. 2008).

[28] Pertaub, D.-P., Slater, M., and Barker, C. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence 11*, 1 (2002), 68–78.

[29] Pippins, J. R., Alegria, M., and Haas, J. S. Association between language proficiency and the quality of primary care among a national sample of insured latinos. *Medical care 45* (2007 2007), 1020 – 1025.

[30] Rickel, J., and Johnson, W. L. Virtual humans for team training in virtual reality, 1999.

[31] Rossen, B., Johnsen, K., Deladisma, A., Lind, S., and Lok, B. Virtual humans elicit skin-tone bias consistent with real-world skin-tone biases. In *IVA*, H. Prendinger, J. C. Lester, and M. Ishizuka, Eds., vol. 5208 of *Lecture Notes in Computer Science*, Springer (2008), 237–244.

[32] Rossen, B., Lind, S., and Lok, B. Human-centered distributed conversational modeling: Efficient modeling of robust virtual human conversations.

[33] Tienda, M., and Faith Mitchell, Editors, C. o. T. O. C. D. N. R. C. *Hispanics and the Future of America*. The National Academies Press, 2006.

[34] Traum, D., Marsella, S. C., Gratch, J., Lee, J., and Hartholt, A. Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *Proceedings of the 8th international conference on Intelligent Virtual Agents*, IVA '08, Springer-Verlag (Berlin, Heidelberg, 2008), 117–130.

[35] Wilson, E., Chen, A., Grumbach, K., Wang, F., and Fernandez, A. Effects of limited english proficiency and physician language on health care comprehension. *J Gen Intern Med 20*, 9 (2005), 800–6.

BIOGRAPHICAL SKETCH

Doaa El Sheikh graduated with a BS degree in electrical engineering in 2004 from Cairo University, Egypt. She received her MS degree in computer science from the University of Florida in fall 2012.

She joined the Virtual Environment Research Group (VERG) in 2010 where she has been investigating the potential of virtual humans in overcoming language barriers between physicians and patients. She developed a system that allows medical students to practice performing a physical exam on virtual pediatric patients whose parents exhibit different levels of English proficiency.

Her research interests are in human computer interaction and virtual environments.