SIDNIE: Scaffolded Interviews Developed by Nurses in Education

Lauren Cairco Dukes, Toni Bloodworth Pence, Larry F. Hodges
Clemson University School of Computing
McAdams Hall, Clemson, SC 29634
{LCairco, tbloodw,LFH}@clemson.edu

Nancy Meehan, Arlene Johnson
Clemson University School of Nursing
Edwards Hall, Clemson, SC 29634
{nmeehan, aejohns}@clemson.edu

ABSTRACT
One of the most common clinical education methods for teaching patient interaction skills to nursing students is role-playing established scenarios with their classmates. Unfortunately, this is far from simulating real world experiences that they will soon face, and does not provide the immediate, impartial feedback necessary for interviewing skills development. We present a system for Scaffolded Interviews Developed by Nurses In Education (SIDNIE) that supports baccalaureate nursing education by providing multiple guided interview practice sessions with virtual characters. Our scenario depicts a mother who has brought in her five year old child to the clinic. In this paper we describe our system and report on a preliminary usability evaluation conducted with nursing students.

Author Keywords
Virtual characters; virtual patient; pediatric nursing; virtual environment; scaffolded learning; interview simulation

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION

Nursing students have limited opportunities for interaction with real patients, especially with pediatric patients, and do not often receive immediate and impartial feedback on their performance during their patient interaction. In order to provide alternative educational opportunities, we have created a system for Scaffolded Interviews Developed by Nurses In Education (SIDNIE), which allows nursing students to interact with virtual characters acting as patients to practice their interviewing skills, while receiving guidance and feedback from a virtual nurse educator. Experiential learning through simulation may help students develop the skills necessary for clinical practice and help develop the self-efficacy and critical thinking skills they need to provide the safest care possible. Also, this system has the potential to provide consistent experiences for all students with feedback and repetitive practice.

IUI'13, March 19–22, 2013, Santa Monica, CA, USA.

There is currently a shortage of registered nurses and nurse educators, resulting in the need for new educational pedagogies for clinical nurse education [23]. The need for registered nurses could increase as much as 500,000 by 2025. In 2007 and 2008, 84 percent of nursing schools were actively recruiting nursing faculty [23]. A shortage in nursing faculty often results in larger class numbers. Our system provides a tool to aid nursing faculty in providing scored practice opportunities for student nurses. Since children are a vulnerable population, nurses must develop appropriate pediatric skills necessary to provide safe care to infants, children, and teens. Because nursing students have limited access to pediatric patients and may not have the opportunity to practice with children, we focus on a mother and child scenario.

RELATED WORK

Simulation Learning and Patient Safety
The Institute of Medicine has recommended the use of simulation training as a method to improve health care delivery [25], and patient safety has been identified as one of the six competencies necessary to improve health care education [18]. Simulation training is an effective strategy to help promote safe clinical practices [17] and impacts the development of self-efficacy and judgment skills for nurses, that are essential to provide the safest and most effective care possible [15]. Simulation learning that mimics real world scenarios is beneficial to nursing students and will provide standardized experiences in which students can practice problem solving techniques and clinical decision making abilities [15] Five advantages to using simulation in nursing education have been identified: 1) providing opportunity for interactive learning without risk to patients, 2) boosting students' self confidence and reducing anxiety in the practice setting, 3) allowing nursing students to practice clinical decision making, and critical thinking in a controlled environment, 4) allowing skills and procedures to be repeated until proficiency is reached, and 5) providing

immediate feedback [13]. We designed SIDNIE specifically to provide repetitive practice as well as the benefits of immediate feedback.

Simulation Learning

One of the current simulation exercises in the School of Nursing involves students acting out written scenarios, where one student acts as the patient while the other acts as the nurse. SIDNIE can replace the student patient with a virtual patient, which is an artificially intelligent human representation that behaves similarly to a human patient under the same circumstance. Researchers have used virtual patients to teach communications skills to medical students, and students have rated the virtual patient experience as being as effective as a standardized patient (a paid actor) [15]. Medical students have also used virtual patients to help practice patient interviewing skills with a high level of immersion. Results indicate that using life-size virtual characters with speech recognition is useful in their education [20]. Adult virtual patients are fairly common, but virtual pediatric patients are rare. The use of virtual pediatric patients was first addressed in [13], where they were used for training and assessment for medical students. The scenario in this system included a mother and daughter, but the technology for creating virtual human behaviors had not focused on children's behaviors, therefore the realism needed for this project was not available [13]. Even with the lack of realism, the results of their study were positive in that many of the participants stated that they gained valuable experience.

Communicating with Children and Parents

In addition to the challenges presented in displaying appropriate animations for children, the dynamics related to the nurse-family-child relationship are extensive due to the many factors that enter into this relationship, including ethnicity, age, culture, and illness. During an assessment, the nurse must obtain information (verbal and nonverbal) from the parent(s) and child, and observe any interactions between them [15]. Studies by pediatric experts have shown that the nurse-family-child relationship is heavily dependent upon effective communication, which is a skill that is developed through interaction with different kinds of pediatric patients and families [10]. Student nurses must be aware of the interactions that may negatively or positively affect their communication skills, therefore affecting the relationship. This study also showed that a positive nurse-family-child relationship will promote the health of the child [10].

Scaffolding in Learning

Since SIDNIE is designed for nursing student education, we want to provide an efficient pathway for students to be able to achieve some level of competence in interviewing techniques before they reach the clinical setting. We previously developed a prototype system to support freeform interviewing with a pediatric patient using speech recognition. However, during a usability study, we found that even experienced nurses had difficulty using the system, since there was no guidance for which questions our system could answer, and because of problems with visual and behavioral fidelity. Additionally, to give a user feedback on his or her performance required the attention of a nurse, since there was no automated scoring system. To better support learning and reduce difficulty, we decided to build a scaffolded system for interview practice.

Instructional scaffolding is a teaching and learning strategy that provides temporary support to a learner while he or she achieves competency in a task that he or she would not be able to accomplish unaided [24]. Traditionally, scaffolding occurs in an apprentice-learner context, with a subject expert aiding a less experienced learner; however, any method of support provided during learning is considered a form of scaffolding [16]. Holton and Clarke have classified scaffolding techniques based on agency (who provides the scaffolding) and domain (whether the scaffolding helps teach a concept, or helps teach a heuristic method) [16]. There are three possible scaffolding agencies: expert, when a domain expert aids a learner; reciprocal, when peers working together aid each other through their individual strengths; and self, when an individual can apply his or her related knowledge to a new problem. Conceptual scaffolding provides problem-specific aid, while heuristic scaffolding provides principles that are useful for solving generalized problems. Any agency can provide either of the domains. Since the final goal of scaffolding is learner independence, self-scaffolding is especially desirable [16].

Educators have successfully applied scaffolding to many domains, including mathematics [16], medicine [19], and science [22]. Several researchers have applied expert scaffolding to pharmacy students learning to conduct patient interviews. Planas et al. developed a scaffolded communications skills development system that students used as a part of their course on clinical communications [21]. This system provided scaffolding through several means. Students worked in groups, providing reciprocal scaffolding, and received feedback from faculty members and standardized patients, providing expert scaffolding. Students also were required to evaluate their own performance, encouraging self-scaffolding. As the course progressed, support in the form of peer and instructor guidance lessened and students became more competent. Comparing the assessments of a standardized patient interview before the course and one after the course, students performed significantly better ($p < 0.01$) as rated by faculty, peers, standardized patients, and students themselves [21]. Stupans et al. developed several tools for scaffolding patient counseling skills, including rubrics for student evaluation, a flexible protocol for counseling sessions, tools to aid in self-critique and peer-critique of interviews, roleplaying, and use of expert scaffolding, especially early in the experiential learning process. Although they have not yet tested its effect on performance, students generally responded positively to the idea of scaffolded opportunities for learning [26].

Crossing into the electronic domain, Simmersion has developed several systems designed to teach interaction for specific scenarios, providing aid through a "virtual coach" who gives feedback as users ask questions [6]. Simmersion systems show a videotaped human who acts as the interview partner in a large portion of the screen. The user selects a question from a list of questions to ask the interview partner, then the partner responds through a videotaped clip. A virtual coach provides nonverbal feedback on the question selection through motions such as applauding or shaking her head, and provides text feedback if clicked. One example product is a system that gives people with autism practice in social conversation. With eleven learning objectives, a complex conversational model determines how the interview partner responds to user questions. The system keeps a points-based score for user choices and allows the user to review the interview script. A user study on this system showed that the users enjoyed their interaction with the system but would have appreciated a system with more conversational choices, and that the mean scores improved from the first use of the system to the last use of the system [27]. The Simmersion systems provide expert, conceptual scaffolding, since the virtual agent gives choice-by-choice feedback specific to the current conversation. Due to the complexity of the learning objectives and the videotaped interactions, these systems are difficult to extend. Additionally, there is no leveled scaffolding—the same feedback is provided to the user by the virtual coach regardless of the user's experience level.

Extending the success of these systems and the apprenticeship model already present in nursing education, we designed our system to provide leveled heuristic scaffolding through expert agency while supporting maximum extensibility. Our virtual agent serves as the expert scaffolder, teaching heuristic knowledge by providing information on criteria that pediatric patient interviews should meet and broad categories that an interview should cover. The virtual agent provides scaffolding by giving instruction on the criteria and categories, and by giving feedback on user selections. We encourage progressive learner independence by providing several levels of scaffolding, ranging from heavy guidance to minimal feedback. Additionally, we preserve extensibility through our database-driven design, criteria-based scoring, and virtual patients, whose actions and speech can be easily modified.

KEY CONTRIBUTIONS

There are no reports in the literature of applying scaffolded learning and standardized feedback to training in interview skills for any e-learning scenario, much less in the pediatric nursing domain. There are existing e-learning systems for teaching communication skills that offer question-by-question feedback or overall scenario scoring. However, no system found in the literature offers scaffolded support for learners by progressively decreasing guidance through the interview task, or standardized scoring based on specific criteria. Our work is novel in that it offers a framework for scaffolding and feedback for any interaction task, given a set of criteria and scored question-answer pairs. It is also a contribution to the nursing community, since the current training techniques of peer role play and standardized patient (actor) interviews are typically unscored and students receive feedback based only on a peer or instructor's personal perception of the interaction.

SYSTEM DESCRIPTION

SIDNIE is designed to teach nursing students pediatric patient interview techniques by providing interview practice with guidance and feedback from a virtual agent named Sidnie. Sidnie guides the user through several scaffolded practice opportunities and provides feedback on user choices. The user conducts an interview with the virtual patients by selecting questions from a preset list of questions developed by experienced nurses, and the virtual patients respond appropriately. The user can also view interview playbacks and score them as if they were an instructor, encouraging knowledge application to new situations. The application runs on a standard desktop computer and a single monitor, and the user interacts with the system using the mouse.

Scenario Development

We worked with experienced nursing faculty to develop a patient-nurse interaction scenario for a five year old child with an earache. The nurses generated a set of possible questions and scored each question according to two criteria important for pediatric interviews: age appropriate, and unbiased. Age appropriate in this context primarily means that questions addressed to the child used language children would understand, and that necessarily complex questions were addressed to the parent. For example, "Hello Sarah, I am the RN who will be your nurse" is not age appropriate, since a child would not know that RN stands for "Registered Nurse." Unbiased questions do not make implicit assumptions or lead the patient to respond in a particular way. For example, "Sarah, your tummy hasn't been hurting, has it?" leads the child to respond negatively, and is therefore biased. Additionally, the nurses categorized the questions into broad categories and subcategories, and provided us with the subset of categories that are essential for a nurse to cover in every patient interview. The subset of categories and subcategories that are essential are: Introduction/Introducing Yourself, Introduction/Confirming Patient Identity, Chief Complaint, History of Present Illness, Related Systems, and Past Medical History. This detailed structure enables us to provide feedback to users on their question selections and interview thoroughness. This meets our goal of providing heuristic scaffolding, since these criteria and categories can be applied to any pediatric patient interview.

Sidnie

A male nurse named Sidnie serves as our virtual nurse educator during the interview simulation. Sidnie's role is to

Figure 1. SIDNIE permits a user to interview the virtual patients by selecting question categories and subcategories using a tabbed interface at the bottom of the screen. The user clicks a question, then the virtual patients respond with speech and animations.

provide instruction on how to use the other interaction modules, to give feedback to the user on his or her performance, and to demonstrate patient interviews. Sidnie is represented by a static image of a male nurse in the lower right hand corner of the screen. We chose not to animate Sidnie or support additional interaction with him because we felt that an additional animated character would distract from the patients' behaviors and animations. Sidnie has a quote bubble above his head that is used to display instructions or give feedback.

Tutorial

When each session begins, Sidnie introduces himself through text and guides the user through a tutorial that outlines how to use the level using screenshots with accompanying text. The user must click "Next" to proceed through each slide until the tutorial is finished. Sidnie can give additional text or image instructions according to what is present in the database.

Electronic Health Record

After the tutorial, our system displays a simple electronic health record (EHR) for the patient. The EHR features information on the patient's current statistics including a photo of the patient, the patient's birthday, weight, temperature, immunization record, medications, allergies, parents' names and contact information, and reason for the visit. The user can access this information using buttons to select categories of information. Once the user has reviewed the medical record, he or she clicks a button to proceed with the scenario.

Pediatric Patient Interview

During the scenario, the majority of the screen displays the virtual patients in their environment. For our current simulation, a mother and her five year old daughter sit in a pediatric patient room. In an idle state, the virtual characters sit quietly. When responding to questions, the patients perform appropriate speech and animations. Depending on the chosen scaffolding level, the user may either conduct a patient interview with guidance or may observe an interview and score it for the chosen criteria.

Conducting an Interview

Our database stores a set of questions that the user may ask, and the corresponding character responses. For ease of use, we display the questions according to their categories and subcategories in the question selection interface. The question selection interface is displayed at the bottom of the screen when the user is permitted to conduct an interview with patients. On the left side of the screen, colorful tabs permit the user to select top level categories, which are dynamically loaded in from the database. Users may select subcategories using a set of tabs along the top of the list of questions. Under the tabs, we display the list of questions belonging to the selected category and subcategory. We randomize the order the questions are displayed with each run of the application so that repetitive use is still slightly

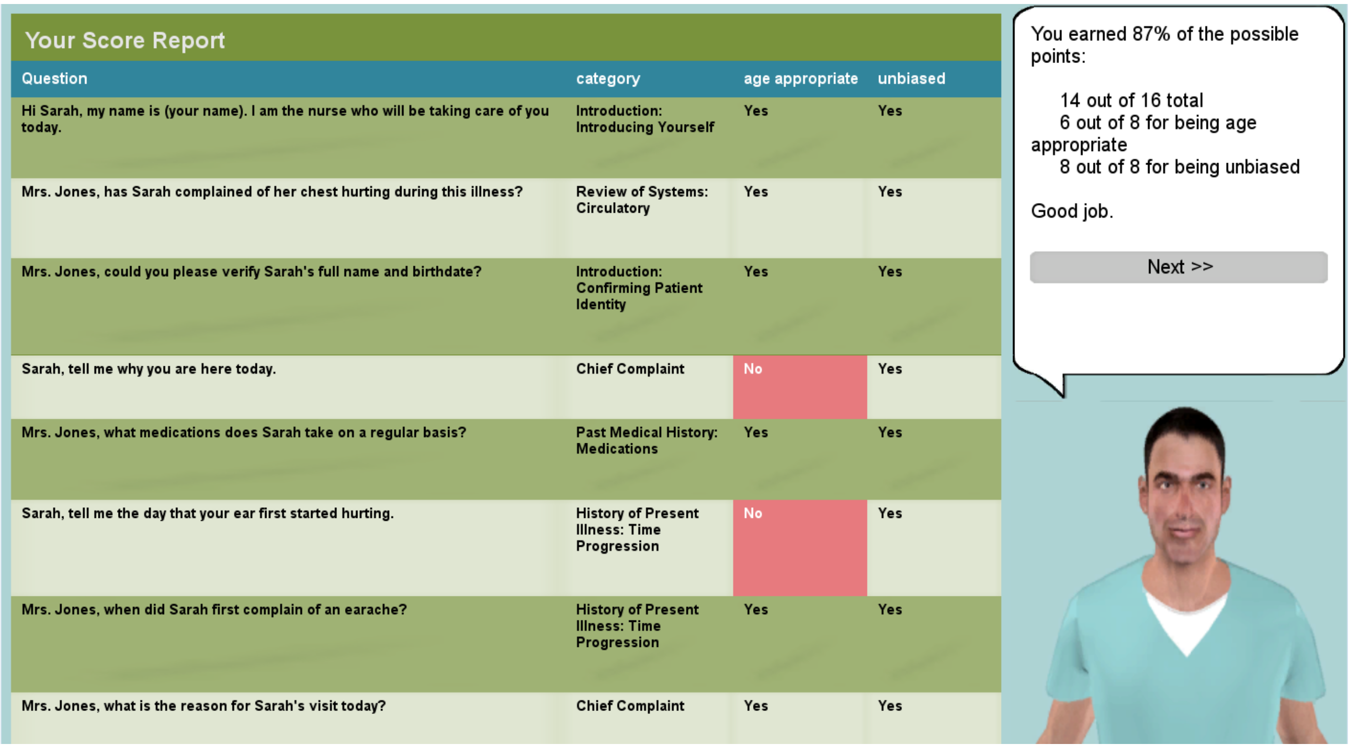

Figure 2. Sidnie gives feedback in chart and percentage form after a scaffolding level is completed.

varied. To select a question, the user simply clicks on the displayed question. The question is highlighted, then the characters respond appropriately with speech and/or animation. After the characters respond, the user may proceed to select questions until the session is complete.

Grading an Interview

Sidnie may also prompt the user to grade another interviewer's questions to encourage the user to apply his or her knowledge to a new problem. In this module, the application retrieves a prerecorded sequence of questions from the database. Sidnie "reads" a stored question using text to speech, then the virtual patient and/or parent responds with the corresponding answer stored in the database. After the characterss finish responding, in his quote bubble, Sidnie prompts the user to grade the question on the stored criteria, and to categorize the question into the correct category & subcategory. We provide a checkbox for each criteria, where the user should check the checkbox if it meets the criteria, and then a scroll down list, where the user must select a category before proceeding. After submitting the score, the interview playback proceeds.

Feedback Mechanisms

To promote user learning, we provide two types of feedback for the user depending on the selected scaffolding level. Immediate feedback provides selection-by-selection scoring so that the user can correct his or her mistakes immediately during the interview, and is used in the lower scaffolding levels. Summative feedback is provided at the end of each scaffolding level and provides the user with an overall score and overview of his or her performance for the entire scenario.

Immediate Feedback

When the user is conducting an interview, Sidnie may give immediate feedback on user question selection in his quote bubble. When the user selects a question, we query the database to see whether the question meets the stored criteria. We use the query results to generate a string that Sidnie may immediately display in his quote bubble to indicate how the user performed. For example, for a question the user selected that meets both criteria, Sidnie may say, "Good job! The question you asked was unbiased and age-appropriate."

When the user is scoring an interview, Sidnie also may provide immediate feedback on the submitted score. When the user submits a score, we query the database to find the actual question's scoring and category, then compare them to the user selected characteristics and build a string to provide feedback. For example, the correct category was chosen but both criteria were scored incorrectly, Sidnie may say, "You selected the appropriate category for this question: 'Introduction: Confirming Patient Identity.' You scored the question incorrectly as: unbiased when it was not unbiased, not age appropriate when it was age appropriate."

Summative Feedback

At the end of each session, the user may receive summative feedback in chart form as well as an overall percentile score for his or her performance. Each row in the chart represents a question the user asked, along with its scoring data.

If the user conducted the patient interview, we highlight the cells for the scored criteria based on whether the selected questions met the criteria. Pink cells show the user that the question did not meet the criteria, while green cells show that the question met the criteria. Additionally, we display a pink row with the category name for each essential category that the user neglected. If the user scored a patient interview, we highlight the cells based on whether they scored the question correctly. Pink cells show the user that they made a mistake in scoring or categorizing the question.

In the summative feedback screen, Sidnie displays percentage and proportion scores in his quote bubble. If the user conducted an interview, we calculate the percentage score and proportion score by counting how many questions met the criteria and how many of the essential categories the user covered. We generate speech for Sidnie based on those scores. For example, for a user who asked seven questions, and covered each essential category, but did not meet the scoring criteria for every question, Sidnie may say, "You scored 12 out of 20 points, for a score of 60%. You scored 6 out of 6 points for category coverage, 5 out of 7 points for being unbiased, and 1 out of 7 points for being age appropriate." We calculate scores similarly for users who score an interview, except we score them for choosing the correct category and for correctly grading the question for the criteria instead of choosing questions that met criteria and covered essential categories.

SCAFFOLDING LEVELS

Our goal in this system is to provide leveled scaffolding so that an inexperienced student may progressively gain enough confidence, knowledge, and simulated experience to ease the transition into actual patient interviews. Currently we have implemented four scaffolding levels for our simulation designed to progressively remove scaffolding until the student can perform an interview without aid.

We based our scaffolding levels on scaffolding used for in-person pharmacology interviews[21,26]; however, our approach is innovateive in that we use one set of criteria to score each question, so that students learn overall concepts to be applied for each question in addition to the steps needed for a complete interview.

Level 1: Baseline Measurement

To enable measurement of user progress, we first require the user to conduct a patient interview using our system without any training or guidance. During the tutorial phase, Sidnie explains how to select a question using our system. The user reviews the medical record and proceeds with the interview, selecting as many questions as he or she desires. Sidnie does not give immediate feedback on the question choices. When the user is finished with the interview, the session concludes without any summative feedback.

Level 2: Guided Interview with Immediate Feedback

The second level begins the scaffolded learning by guiding the user through a short pediatric patient interview. During the tutorial phase, Sidnie explains how to select a question using our system, and then describes the criteria that the questions should meet and the categories that an interview should cover. The user initially reviews the medical record then continues to the guided interview. The system preselects the first essential question category in the question selection interface and disables the user from changing the category. Then Sidnie instructs the user to select a question. When the user selects a question, the patients respond, then Sidnie gives immediate feedback on whether the question met the criteria. If the question did not meet each of the criteria, Sidnie informs the user of which criteria were not met and instructs the user to try again. The user must continue to select questions until he or she chooses a question that meets both criteria. Then, the system advances to the next essential category. The process continues until each essential category is covered, then shows summative feedback to the user.

Level 3: Grading an Interview

The third level requires the user to score a prerecorded interview by watching and listening to each question and response, then identifying the category. The tutorial phase explains the interview playback process and the grader interface. The user then views Sarah's electronic medical record and proceeds to the interview. As described in the previous section on grading an interview, the user grades the interview question by question, and Sidnie gives immediate feedback on his or her grading, but does not require the user to correctly grade the question before moving to the next question. Once the user finishes scoring the interview, he or she receives summative feedback on the scoring.

Level 4: Evaluation

Finally, the fourth level evaluates the user's learning by providing unguided interaction similar to Level 1. The tutorial phase again explains the system and the scoring criteria, and then the user proceeds through the interview by selecting as many questions in as many categories as they wish. The user clicks a button when he or she is finished. Lastly, they receive summative feedback on their interview based on their category coverage and whether their questions met the criteria.

IMPLEMENTATION DETAILS

Our system is implemented primarily in Unity 3D [8] as a set of related modules implemented in Boo or C#. We created the virtual environment using Blender [1] and imported it into Unity. We rigged and animated the characters using Poser Pro [5] and Blender [1]. Using Unity's plugin interface, we built a C++ plugin as an interface to SAPI 5.3 [4]. We used IVONA voices [2] for our virtual characters: Sarah's voice is Ivy, Mrs. Jones' voice is Kendra, and Sidnie's voice is Joey. In addition to the animations triggered by responding to user questions, our characters exhibit baseline animations such as breathing and blinking.

Although our virtual characters and environment are static at this point, the remainder of our program is database-driven, providing flexibility to change scenarios, scoring criteria, instructions, and required categories. We store our scenario data in a SQLite 3 database [7] and access it through Unity's plugin interface. Additionally, each question selection and scoring action is timestamped and recorded in the database, keyed on the user ID and the scaffolding level for easy data processing.

Character Response Queueing

When a user clicks on a question, we retrieve the answer the characters should give from the database. We store the answer as a sequence of animations and text to speech, with each record in the response representing one animation or line of speech. Each action is accompanied by an identifier that associates it with one of the virtual characters, and a delay that denotes when in time the action should occur. Within our Unity application, we keep a queue of these actions with their delays. While the queue still contains items, we use a coroutine to keep track of elapsed time and check whether it is time to perform an action. When it is time to perform an action, we call the appropriate function and remove the action from the queue.

Animating a character is a simple Unity function call to trigger a preset animation. When a character's response requires speech, the process is more complex. Because timings in the database may be inexact, and creating many SAPI instances causes performance problems, we also keep a text-to-speech queue that holds any text to speech to be rendered. When a new question is selected, we count the number of speech items contained in the response, then send that number ahead to the plugin so that it can indicate to Unity that it is busy until there are no more items to be spoken. Using a coroutine, while there is speech in the queue, our application checks to determine whether the text-to-speech engine is currently rendering speech. When the engine is not busy, it sends the next item in the queue, then waits for it to be non-busy again. To keep the count of items to speak up to date, within the plugin, we handle the SAPI event for finished speech and use it to decrement the counter.

We send the string to SAPI via the C++ plugin, along with parameters to select an appropriate voice based on the character's age and gender. This increases our system's extensibility since it would cause the most appropriate voices to be selected on a system that does not have our chosen voices installed. Since Unity's default behavior is to call a plugin as a part of its own thread, and SAPI voice loading events are blocking function calls, within our plugin we launch a thread to render the text to speech so that the Unity application may continue to render the virtual characters with appropriate animations and lipsyncing. Within the launched thread, the plugin loads the appropriate SAPI voice and renders the string through text-to-speech. Inside the plugin we store the current viseme for each character (mouth shape) by handling SAPI viseme events as the text to speech is rendered. In the Unity application's update function, we query the plugin to gather each character's current mouth shape. If the mouth shape has changed since the previous update, we trigger an animation to change the character's mouth shape to match the speech SAPI is rendering, using a modification of the Metamorph script[3] to apply shapekey morphs to our characters' faces.

In order to encourage the user to be attentive to the characters' responses, we disable the question selection interface while the character is performing animation or speaking. We do this by checking to determine whether the response queue is empty and whether the text to speech plugin has more speech to render. While the queue is nonempty or text to speech has more lines to render, we overlay the question selection interface with a translucent screen and place text over the interface and in Sidnie's text bubble to indicate that the user should wait for the characters to finish their response. After the character has finished her response, the question selection interface is enabled and Sidnie provides appropriate instructions for the scaffolding level.

EXPERIMENTAL PROCEDURE

We conducted an experiment to gather initial impressions on SIDNIE's usability and to measure learning outcomes for a student completing all three scaffolding levels. Each of the scaffolding levels used the same scenario: a mother with her five year old daughter, who has an earache.

Participants completed a demographics questionnaire and read the informed consent form the morning before the experiment. When the participants came to the experiment site, we greeted them then gathered their demographic questionnaires, which included questions about previous computer usage, experience in healthcare, and exposure to virtual environments. The participant then signed the informed consent. We seated the participant at a desktop computer and began the SIDNIE system on scaffolding level 1. We asked the participant to view the beginning tutorial screens, which instructed them on how to use the system but included no information on the required question categories or the scoring criteria, then to conduct an interview with the patients to the best of their ability, and to let us know when they had finished.

After the participant completed the initial interview, we began the SIDNIE system on level 2. We told the participants that SIDNIE would guide them through their interview this time, and that they would receive feedback on their choices. The participants then proceeded through the guided interview and notified the experimenter upon completion. The participant continued through levels 2 and 3 similarly, with the experimenter briefly commenting on each level then starting the application.

After completing every level, the participant filled out the System Usability Scale (SUS) [12] and was given the

	Age appropriate		Unbiased		Category Coverage	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Level 1	75.12	12.04	70.65	10.76	91.67	19.34
Level 2	94.22	10.14	93.46	11.18	n/a	
Level 3	82.14	8.75	80.10	13.19	n/a	
Level 4	93.52	9.72	90.26	8.39	100.0	0.00

Table 1. Percentage means and standard deviation for each scaffolding level.

opportunity to provide any written feedback. Finally, the participant completed an interview with the experimenter, where we asked questions about his or her opinion of the system.

Our primary hypothesis was that participant performance in terms of essential category coverage and choosing questions that met the criteria of age appropriateness and unbiasedness would increase between the level one baseline measurement and the level four evaluation measurement. We also expected participant performance to increase between each scaffolding level, and expected participants to report our system to be usable.

RESULTS

Fifteen undergraduate students seeking their Bachelor of Science in Nursing (BSN) at Clemson University participated in our experiment as an optional extra credit assignment for their Nursing Care of Children class. All of the participants already have received a bachelors degree in another field. Fourteen of the fifteen students are a part of the accelerated second degree program, while the remaining participant is a traditional BSN student.

Twelve of the participants were female, and three of the participants were male. Participants' ages ranged from 23 to 48 (mean=30.13, sd=7.28). Participants reported a high level of computer use. Participants reported some level of healthcare experience, experience with children, and experience in a healthcare setting, and somewhat lower levels of experience with children in healthcare, virtual human exposure, and virtual reality exposure. See Table 1 for detailed demographic information. When asked about their previous experience with virtual reality, virtual humans, or avatars, participants cited their experience in the nursing simulation lab, which has physical dummies that exhibit instructor-controlled symptoms. Most of the students also participated in an experiment in the previous semester that required them to take vital signs for a virtual human in a virtual hospital displayed on a desktop computer and record the results in an electronic health record.

Technical difficulties caused five sessions to terminate prematurely. We gave these participants the option to try the session again, or to continue on with the next level. Three participants chose to redo the session, while two participants continued onto the next level. We were able to recover sufficient data for all but one participant, who chose not to complete level three. We excluded her data from the quantitative analysis but kept her usability data. We recovered the three participants' data who made a second attempt at level completion by splicing together their data from the two session attempts by keeping all choices made in the first attempt before the program crashed and discarding all choices in the second attempt up until they reached the same point in the level as they were in the first session. For the final participant who chose not to retry a level, we considered her data in terms of proportions only, to give a fair basis for comparison to others who completed the level.

Our primary interest in this evaluation was to determine whether user scores improved between the level one baseline measurement and the level four evaluation measurement. Levels one and four were identical except for the instructions during the tutorial phase and the summative feedback at the end of the interview in level four. We calculated the percentage of questions the user asked that were age appropriate, the percentage of questions the user asked that were unbiased, and the percentage of the essential categories that the user covered. Using a two-tailed t-test for paired samples, we found significant differences in each of the scores for the criteria, indicating that participants' performance improved as a result of the use of our system ($p < 0.01$ for each criteria). We did not find a significant difference in the percentage of categories covered. See Table 1 for detailed statistics.

Progress through Scaffolding

To gauge which of our scaffolded levels were effective for learning, we analyzed whether participants' scores improved over each level. Levels two and three are significantly different from levels one and four, so it is somewhat difficult to make a direct comparison in user performance. We chose to compare the levels based on the two criteria: age appropriateness, and bias. Comparing category coverage was not possible since level two forced all essential categories to be covered, and level three played back a predetermined interview that covered all essential categories. We calculated the criteria percentages for levels one, two, and four for each of these criteria by finding the proportion of questions the user asked that met the criteria. For level three, since users were scoring an interview instead of conducting an interview, we calculated the proportion of questions correctly scored in regards to that criteria.

Using a two-tailed t-test for paired samples, we found significant differences in scores between each of the consecutive scaffolding levels (every $p < 0.01$), however, the mean scores were not always increasing, as we would have expected. The scores for level three were significantly lower than the scores for level two and level four (see Table 1). For nonconsecutive levels, there was no significant difference between levels two and four, but a significant difference in both criteria between levels one and three

	Mean	SD
Healthcare experience	3.80	1.37
Experience with children	4.40	1.96
Experience in a healthcare setting	3.60	1.77
Experience with children in healthcare	2.80	1.78
Computer use	6.47	0.74
Virtual human exposure	2.87	1.51
Virtual reality exposure	2.20	1.26

Table 2. Demographic information for the study participants (1= no experience, 7= very experienced)

(both p’s < 0.03), with higher scores in level three.

Usability and Qualitative Feedback

Overall, participants seemed to find the system easy to use. Our system received an average SUS score of 80.33% (sd=12.20%). The highest scoring prompts were, “I didn’t need to learn a lot of things before I could get going with this system” and “I didn’t think there was too much inconsistency in the system”, both with means of 5.6 out of 6 and the lowest standard deviation of any question as well (sd =0.63). The lowest scoring prompt was “I felt very confident using the system” with a mean score of 3.93 out of 6 (sd=1.75). See Table 2 for other detailed results.

When asked, every participant said that they thought they and their peers would use our system if it were made available to them. Comparing this technique of interview practice to other techniques (such as roleplay, standardized patients, or reading case studies), all participants felt they would have performed the same or worse using another practice technique. Reasons cited for improved performance using our system included less performance anxiety, being able to select from provided choices instead of coming up with questions on their own, being able to get immediate feedback, and being able to take our system more seriously than roleplay with their peers.

Three of the fifteen participants said that the child’s symptoms were not realistic for a five year old with an earache. Two participants were not sure since they did not have much experience with children. The remaining ten participants said the child’s symptoms were realistic, citing examples such as the data in the electronic medical record, the child pulling her ear, not responding to questions that weren’t age appropriate, and responding nonverbally. Participants suggested that the patients should have appropriate facial expressions, and that Sarah could be more whiny, fidgety, or irritable. Three participants thought that the mother’s voice sounded “mean”, but one participant commented that a disagreeable parent is realistic in a clinical setting.

When asked what they liked about the system, nine of fifteen participants mentioned some aspect of the program that related to the scaffolding, including guidance and feedback, time to think about choices, and being able to try multiple questions in an interview with no risk of offending a patient. The most common aspects that participants disliked were the mother’s voice and the scoring interface.

DISCUSSION

Our primary hypothesis was confirmed—participants’ performance did improve from the baseline to the evaluation. However, the scores did not monotonically increase over each scaffolding level as we expected, and determining which scaffolding levels are effective for learning is difficult. There was a significant difference in performance between levels one and two, where the only training received was the tutorial preceding level two, and the immediate feedback for each question selected in level two. At level two, it seems that the scores hit a “ceiling”, with participants scoring upwards of 93% for each criteria. In fact, nine out of the fifteen participants showed perfect scores in the second level. These scores do not leave much room for improvement in the similar task for level four, and in fact, we did not find significant differences in the performance between level two and level four.

The “ceiling effect” in the scores for levels one, two, and four may be explained by considering literature in the learning domain such as the Revised Bloom’s Taxonomy [14], [9]. Our goal is to encourage the highest levels of learning, which in the new taxonomy are titled “creating” and “evaluating”. In the first and fourth scaffolding levels, we aimed to permit users to “create” an interview scenario with limited choices. However, because there were only a few choices in each category and subcategory, and a participant was potentially exposed to each of those questions and their corresponding responses and scores multiple times, by the time the user reached the fourth level, the task became a low-level learning exercise of “remembering” or “understanding”, or at best, “applying” the meaning of the criteria and categories to select appropriate questions.

In retrospect, given our limited question base, a more appropriate baseline and evaluation may have been to require the user to score an interview before and after some training. The scoring task reaches the “evaluating” level of the learning taxonomy, and the lower percentage scores in the third level of scaffolding can possibly be attributed to the higher level of learning required.

The lower scores for the third level are also likely the result of usability problems. Several participants were confused by the scoring interface. Since “unbiased” is a negative word and was paired with a checkbox, several participants asked questions like, “Do I check this box if it is biased, or unbiased?” In future iterations of the design, we will change to a radio button interface where the user can select “Yes” or “No” to clarify the scoring. Additionally, several users mentioned that they did not realize until far into the session that they could scroll down to select categories besides

	Mean	SD
I think that I would like to use this system frequently	4.2	0.94
The system was not unnecessarily complex	5.13	1.25
I thought the system was easy to use	4.13	1.85
I don't think I would need the support of a technical person to be able to use this system	5.13	0.92
I found the various functions in the system were well integrated	4.4	1.18
I didn't think there was too much inconsistency in the system	5.6	0.63
I would imagine that most people would learn to use this system very quickly	4.67	1.59
I didn't find the system very cumbersome to use	5.4	0.83
I felt very confident using the system	3.93	1.75
I didn't need to learn a lot of things before I could get going with this system	5.6	0.63

Table 3. Results of the System Usability Scale (SUS). Scores range from 0 (strongly disagree) to 6 (strongly agree). In this table we have modified the original questionnaire and scores so that higher scores are always better.

those in immediate view. Unity3D's default scrollbars are dissimilar to standard scroll controls and do not feature up and down arrows, so we will modify their appearance to match more traditional interfaces, or we will change the scrollable view to a drop-down box. Lastly, although our system currently supports a question belonging to only one category, several students felt that some of the questions could have fit in multiple categories. In the future, we will consider allowing questions to be part of multiple categories.

Although there is always room for improvement, our SUS scores as well as the feedback from our debriefing interview shows that our system has good usability, confirming our final hypothesis. The lowest scoring SUS question was about user confidence. Although we did not gather data on specific sources of anxiety, since each participant reported that they and their peers would use the system if it were offered to them, and that they would expect similar or worse performance using another interview technique, the low score for this question may reflect a user's lack of confidence in their ability to conduct pediatric interviews instead of their confidence in actually using the system.

FUTURE WORK

In addition to correcting usability problems, there are two primary areas we plan to extend our project. First, we will extend our project to provide a broader range of interview opportunities along several dimensions. For the existing scenario, we intend to add variety and attempt to reduce learning effects by adding more questions and answers within each category to the database, then showing randomly selected subsets of the questions each session. We will also add more scoring criteria to increase complexity. Additionally, we will develop more scenarios including different medical conditions as well as patients from various socioeconomic, ethnic, religious, and health backgrounds. During the debriefing interview, several students mentioned their interest in practicing interviews with people who were outside of the perceived norm, since those patients may require extra care during interaction.

Second, we would like to extend support for learning. From the scaffolded learning perspective, our current system covers several levels of the Revised Bloom's Taxonomy[9]. We promote "remembering" and "understanding" by requiring users to recall question categories and criteria in each scaffolding level, and enable "applying" that knowledge by requiring the user to select questions for their interviews. We also support "evaluating" by letting users score prerecorded interviews. However, for future work, we would like to provide scaffolding to cover the remaining levels of "analyzing" and "creating". We are currently working with natural language processing techniques to enable freeform patient interviews by matching typed or spoken user input to questions present in our database, enabling the user to truly "create" an interview while maintaining automatic scoring and feedback. We may also be able to support "analysis" by requiring users to describe or interpret why questions meet or fail to meet the given criteria. We may also consider incorporating the knowledge dimension of the Revised Bloom's Taxonomy into our scaffolding system by promoting factual, conceptual, procedural, and meta-cognitive knowledge to support holistic learning.

ACKNOWLEDGEMENTS

This research was supported, in part, by the NSF Graduate Research Fellowship Program (fellow numbers 2009080400 and 2011095211), an Interdisciplinary Research Innovations Grant from the College of Health Education and Human Development at Clemson University, and a grant from the Agency for Healthcare, Research, and Quality (RO3#HS020233-01). We would also like to thank Jeffrey Bertrand for the many hours he spent sharing his expertise in character animation.

REFERENCES

1. Blender. 2012, 1/19/2012. http://www.blender.org/

2. Anonymous. IVONA Text-to-Speech Voices | TTS Voices | Text To Voice 2012, 6/29/2012. http://www.ivona.com/en/

3. Anonymous. MetaMorph for Unity | Rezzable 2012, 6/29/2012 http://rezzable.com/blogs/foolish-frost/presenting-metamorph-toolkit-blender-and-unity

4. Anonymous. Microsoft Speech API (SAPI) 5.3 2012, 6/29/2012. http://msdn.microsoft.com/en-us/library/ms723627(v=vs.85).aspx

5. Anonymous Poser 3D Animation & Character Creation Software - Official Website 2012, 6/29/2012. http://poser.smithmicro.com/

6. Anonymous. SIMmersion.com - SIMmersion LLC, your training systems solution. 2012, 7/5/2012. http://www.simmersion.com/

7. Anonymous. SQLite Home Page 2012, 6/29/2012 ().

8. Anonymous. UNITY: Game Development Tool 2012, 1/19/2012. http://www.sqlite.org/

9. Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P., Raths, J. and Wittrock, M. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. (2000).

10. Ball, J. W., Bindler, R. C. and COWEN, K. J. Child health nursing. (2006).

11. Bloodworth, T., Cairco, L., McClendon, J., Hodges, L. F., Babu, S., Meehan, N. K., Johnson, A. and Ulinski, A. C. Initial Evaluation of a Virtual Pediatric Patient System. Carolinas Women in Computing, (2012).

12. Brooke, J. SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(1996), 194.

13. Durham, C. F. and Alden, K. R. Enhancing patient safety in nursing education through patient simulation. Patient safety and quality: An evidence-based handbook for nurses, 6, 3 (2008), 221-250.

14. Forehand, M. Bloom's taxonomy. Emerging perspectives on learning, teaching, and technology, (2005).

15. Hockenberry, M. J. and Wilson, D. Wong's nursing care of infants and children. Recherche, 67(2006), 02.

16. Holton, D. and Clarke, D. Scaffolding and metacognition. International Journal of Mathematical Education in Science and Technology, 37, 2 (2006), 127-143.

17. Hubal, R. C., Deterding, R. R., Frank, G. A., Schwetzke, H. F. and Kizakevich, P. Lessons learned in modeling virtual pediatric patients. Stud. Health Technol. Inform., (2003), 127-130.

18. Johnsen, K., Dickerson, R., Raij, A., Lok, B., Jackson, J., Shin, M., Hernandez, J., Stevens, A. and Lind, D. S. Experiences in using immersive virtual characters to educate medical communication skills. In Anonymous *Virtual Reality, 2005. Proceedings. VR 2005. IEEE.* . IEEE, 2005, 179-186.

19. Lua, J., LAJOIE, S. P. and WISEMAN, J. Scaffolding Collaboration in Simulated Medical Emergencies. In Anonymous *International Conference on Computers in Education.* (). Hong Kong: Asia-Pacific Society for Computers in Education., 2009, 285-292.

20. McKeon, L. M., Norris, T., Cardell, B. and Britt, T. Developing patient-centered care competencies among prelicensure nursing students using simulation. J. Nurs. Educ., 48, 12 (2009), 711-715.

21. Planas, L. G. and Nelson, L. E. A systems approach to scaffold communication skills development. Am. J. Pharm. Educ., 72, 2 (2008).

22. Quintana, C. and Fishman, B. Supporting science learning and teaching with software-based scaffolding. American Educational Research Association (AERA), (2006).

23. Rothgeb, M. K. Creating a nursing simulation laboratory: A literature review. J. Nurs. Educ., 47, 11 (2008), 489.

24. Sawyer, R. K. *The Cambridge handbook of the learning sciences.* Cambridge Univ Pr, , 2006.

25. Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raij, A., Harrison, C., DiPietro, M., Allen, B., Ferdig, R. and Foti, S. The use of virtual patients to teach medical students history taking and communication skills. Am. J. Surg., 191, 6 (2006), 806-811.

26. Stupans, I., Owen, S., Ryan, G., Woulfe, J. and McKauge, L. Scaffolding patient counselling skills in Australian University Pharmacy programs. Asia-Pacific Journal of Cooperative Education, 11, 2 (2011), 29-37.

27. Trepagnier, C. Y., Olsen, D. E., Boteler, L. and Bell, C. A. Virtual conversation partner for adults with autism. Cyberpsychology, Behavior, and Social Networking, 14, 1-2 (2011), 21-27.