FACILITATING THE AUTHORING OF MULTIMEDIA SOCIAL PROBLEM SOLVING SKILLS INSTRUCTIONAL MODULES

A Dissertation
Presented to
The Academic Faculty

by

Fatima A. Boujarwah

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Computer Science in the
School of Interactive Computing

Georgia Institute of Technology
May 2012

FACILITATING THE AUTHORING OF MULTIMEDIA SOCIAL PROBLEM SOLVING SKILLS INSTRUCTIONAL MODULES

Approved by:

Dr. Gregory D. Abowd, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Mark O. Riedl
School of Interactive Computing
Georgia Institute of Technology

Dr. Rosa I. Arriaga, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Brian Magerko
School of Interactive Computing
Georgia Institute of Technology

Dr. Andrea Tartaro
Computer Science Department
Furman University

Date Approved: March 30, 2012

ACKNOWLEDGEMENTS

I wish to thank my family; Nasser Al-Daihani, and Mercy, Abdulazeez, Nadia, and Saud Boujarwah, without whose support this work would not have been possible. I would also like to thank my academic family; my academic parents and advisors Rosa Arriaga and Gregory Abowd, and my academic siblings in the Ubicomp Lab for both challenging and supporting me throughout my four years at Georgia Tech. My committee; Mark Riedl, Brian Magerko, and Andrea Tartaro also deserves great recognition for their detailed feedback, and overall dedication to helping me do the best research possible.

I would also like to thank my collaborators, especially Jackie Isbell, Hwajung Hong, and Husayn Versee. Their contributions to the work were invaluable, and played a critical role in its evolution into the research presented here.

This work would also not have been possible without the support of Kuwait University who advised and financially supported me throughout my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

HFA	High-Functioning Autism
Refl-ex	Reflection and Experience
REACT	Refl-ex Authoring and Critiquing Tool
ASD	Autism Spectrum Disorders
CAI	Computer-Assisted Instruction
ITS	Intelligent Tutoring System
AMT	Amazon's Mechanical Turk
HIT	Human Intelligent Task

SUMMARY

Difficulties in social skills are generally considered defining characteristics of High-Functioning Autism (HFA). These difficulties interfere with the educational experiences and quality of life of individuals with HFA, and interventions must be highly individualized to be effective. I explore ways technologies may play a role in assisting individuals with the acquisition of social problem solving skills.

This thesis presents the design, development, and evaluation of two systems; Refl-ex, which is a collection of multimedia instructional modules designed to enable adolescents with HFA to practice social problem solving skills, and REACT, a system to facilitate the authoring of a wider variety of instructional modules. The authoring tool is designed to help parents, teachers, and other stakeholders to create Refl-ex-like instructional modules. The approach uses models of social knowledge created using crowdsourcing techniques to provide the authors with support throughout the authoring process.

A series of studies was conducted to inform the design of high-fidelity prototypes of each of the systems and to evaluate the prototypes. The contributions of this thesis are: 1) the creation of *obstacle-based branching*, an approach to developing interactive social skills instructional modules that has been evaluated by experts to be an improvement to current practices; 2) the development of an approach to building models of social knowledge that can be dynamically created and expanded using crowdsourcing; and 3) the development a system that gives parents and other caregivers the ability to easily create customized social skills instructional modules for their children and students.

CHAPTER 1

INTRODUCTION AND MOTIVATION

Autism is a phenomenon that has touched the lives of many families around the world, and encompasses individuals with a wide range of needs and abilities. The work in this thesis is targeted towards individuals with high-functioning autism (HFA), which despite being a subset of the general autism phenotype, remain a highly heterogeneous population [Baron-Cohen, 2008].

Difficulties in social skills are generally considered defining characteristics of HFA [Howlin, 2003]. Social skills can be defined as specific behaviors that result in positive social interactions and encompass both verbal and non-verbal behaviors necessary for effective interpersonal communication [Gresham, 1986]. Difficulties in social skills interfere with the educational experiences and quality of life of individuals with HFA, and interventions must be highly individualized to be effective. For this reason, this thesis explores ways in which technology may play a role in facilitating the creation of customized instructional modules that can assist individuals with the acquisition of social skills. Indications are that the target population responds well to computer-assisted instruction [Williams and Wright, 2002]. Furthermore, there was a general call for more technologies that specifically target social skills training [Putman and Chong, 2008], and several recent studies have explored the use of technology for this purpose [Laffey et. al., 2009, Hopkins et. al., 2011].

This thesis explores how technology can be used to help an individual practice social skills. In preliminary research, adolescents and young adults were targeted because they are underrepresented with respect to applicable therapies, they are more likely to have complex social skills needs, and research indicates that they should be targeted for social skills intervention [Rao et. al, 2008]. For example, an adolescent with HFA may want to go to a movie theatre without the assistance of a parent or guardian. Can a software module be developed to help that individual prepare for that social context? Furthermore, can a system be developed that helps parents and caregivers to author these modules themselves, for individuals in a broader age range and with various levels of functioning? Such a system would address one of the most challenging aspects of teaching students with autism: the need for individualized instruction for a highly heterogeneous population. These are questions that I would like to answer with my research.

Purpose of Research

The goal of this research is to develop a system that can help teachers and caregivers author instructional modules that individuals with HFA can use to practice their social problem solving skills. In the first stage of research (chapter 3), Refl-ex (short for Reflection and Experience) was built and tested [Boujarwah et. al 2010]. Refl-ex is designed to allow adolescents with autism to practice these skills by experiencing social situations and choosing appropriate responses to unexpected events. The introduction of an obstacle and possible solutions to the obstacle into the scenario creates a branching structure in the modules that is unique and different from current approaches. We call this structure *obstacle-based branching.* Refl-ex also assists and supports reflecting on the social experiences, providing the individual with an opportunity to process information that can be recalled for later use.

The exploratory study of the Refl-ex system yielded promising results with respect to the effectiveness of the instructional approaches used. All of the participants were able to successfully navigate the software, and the software appeared to provide sufficient scaffolding to support the participants' varying levels of language ability. In addition, the study logs and discussions with the participants showed that they did not all struggle with the same social situations. These findings support the importance of individualizing interventions for this population of students and of providing a variety of scenarios for them to practice. The problem is that the modules are very time-consuming to author.

The second system this thesis presents is called REACT (Refl-ex Authoring and Critiquing Tool), and is an authoring tool designed to help parents, teachers, and other caregivers to create Refl-ex-like instructional modules. As described, the modules present the individual with a social situation in which an unexpected obstacle arises, guiding them through a problem solving process to overcome the obstacle. Everyday life consists of many complex social situations within which a wide variety of obstacles may arise. In addition, there is no one correct way to overcome these obstacles. Therefore, authors may require support in order to generate the content of the modules. It can be seen that, though the instructional modules are useful, the large variety of potential content, and the time consuming manual process for creating the branching modules [Bruckman, 1990; Riedl and Young, 2005] is not practical for teachers or parents to adopt when designing for a specific individual. The goal of this work is to empower parents and educators to create these modules; therefore it is important to consider how to accelerate the authoring process.

The nature of social problem solving skills intrinsically requires human input, and will not likely be automatable in the near future. In this work, I present an approach to exploiting human computation to develop complex models of social knowledge that can be used to facilitate the authoring process. This problem is currently intractable using existing artificial intelligence techniques, but luckily it is highly conducive to a crowdsourcing approach.

The rich models of social knowledge developed using the crowdsourcing approach will be used to provide suggestions to the authors as they create customized instructional modules for a particular child. The suggestions will include possible next steps in the social situation, obstacles that may arise at each step, and solutions to these obstacles. In this way, the authoring of the interactive software is facilitated with the aid of models of social knowledge. For clarity, throughout this document, the user of Refl-ex will be referred to as the student, and the individual using REACT will be referred to as the author.

Thesis Statement, Research Questions and Contributions

I propose the following thesis statement:

> An authoring tool can be developed that uses crowdsourced models of social knowledge to help parents easily author individualized *obstacle-based branching* instructional modules, a structure experts evaluate to be an improvement to current approaches to social skills instruction for children with autism.

In particular, I address the following research questions:

- *RQ1a: How can software modules be developed to help a student with autism prepare for various social contexts?*

This question is addressed through the exploratory study that was conducted to evaluate the Refl-ex modules (Chapter 3). This study looked at how the students interacted with the software, and whether or not the design decisions made during software development were appropriate and provided effective scaffolding to enable the students to practice their social skills.

Contribution: The development of interactive software modules that adolescents with HFA can use to independently practice social skills.

- *RQ1b: What value do experts perceive in obstacle-based branching scenarios, and how do they compare to the current approach of using sequential stories to teach social skills?*

  For this question, a study was conducted that allowed experts to see social skills instructional material presented in three different ways, sequentially, sequentially with an obstacle, and in the *obstacle-based branching* format (Chapter 3). In this way they were able to compare across the presentations and evaluate them.

  Contribution: The creation of *obstacle-based branching*, an approach to developing interactive social skills instructional modules that experts confirm is an improvement to current practices.

- *RQ2: What is a mechanism for generating rich models of social knowledge that are consistent with the obstacle-based branching approach to problem solving and can be used to provide scaffolding for the authoring of social skills instructional modules?*

This question is addressed through the description and evaluation of an approach to using crowdsourcing to dynamically populate and expand a knowledge base of social scripts of everyday tasks (Chapter 4). By this what is meant is information like, for instance, what are the steps you take to go to lunch at a fast food restaurant? It is apparent that one person, or several people, can easily answer this, and many other such questions. Crowdsourcing enables the system to have access to these answers.

Contributions: The development of an approach to building models of social knowledge that can be dynamically created and expanded using crowdsourcing.

- RQ3: *How will parents use a tool that employs rich models of social knowledge to facilitate authoring, and will the tool enable them to produce good instructional modules?*

The goal of the REACT system is to enable authors to create customized social skills instructional modules. This question addresses the usability aspects of the system and the evaluation of the output of REACT. The usability was measured through a usability study with parents of children with autism (Chapter 6). To evaluate the output of the REACT system a study was conducted in which experts (i.e., specialists who teach social skills) were asked to evaluate the modules (Chapter 6). The phrase "good instructional modules" is operationalized here to mean both individualized and rated by experts as being suitable to teach appropriate social skills.

Contributions:

The development of an authoring tool that: 1) is easy to use; 2) empowers parents and other caregivers to easily create customized social skills instructional modules for their children and students; and 3) has been confirmed by experts to enable authors to create good individualized social skills instructional modules.

Thesis Overview

In this thesis, the design, development, and evaluation of two systems are presented. The first is Refl-ex, which is a collection of multimedia instructional modules designed to enable adolescents with autism to practice social skills, and the second is REACT, a system to facilitate the authoring of instructional modules for individuals in a broader age range and with varying levels of functioning.

In Chapter 2, background and related work in the areas of autism, social skills instruction, computer-assisted instruction, and crowdsourcing are presented. Chapter 3 presents Refl-ex, describing the technology development, and two studies that were conducted to evaluate it. In Chapter 4, the crowdsourcing approach that was used to build the authoring tool's knowledge base is presented along with a preliminary evaluation. Next, in Chapter 5, the design of the authoring tool, and a description of the technology development are presented. This is followed by a description of the two-study evaluation of REACT, and a presentation of the analysis of the study results. Finally this document concludes with a summary of the contributions of this work, and actionable steps for future work.

Table 1 summarizes the research questions that are answered with this thesis, and the studies that address them.

Table 1.1: Summary of research questions and studies.

#	Research Question	How it was Addressed
1a	*How can software modules be developed to help a student with autism prepare for various social contexts?*	Refl-ex exploratory study (Chapter 3)
1b	*What value do experts perceive in obstacle-based branching scenarios, and how do they compare to the current approach of using sequential stories to teach social skills?*	Branching Validation Study (Chapter 3)
2	*What is a mechanism for generating rich models of social knowledge that are consistent with the obstacle-based branching approach to problem solving and can be used to provide scaffolding for the authoring of social skills instructional modules?*	• Preliminary crowdsourcing study (Chapter 4) • Description and Mechanical Turk evaluation of the full approach (Chapter 4)
3	*How will parents use a tool that employs rich models of social knowledge to facilitate authoring, and will the tool enable them to produce good instructional modules?*	• Parent Study (Chapter 6) • Expert evaluation (Chapter 6)

CHAPTER 2

BACKGROUND AND RELATED WORK

In this chapter, I discuss some background and work related to the creation of Refl-ex and REACT. The research I conducted is such that there are two distinct classes of users, the student and the author, and the authoring tool's output is a secondary system (REACT helps authors create Refl-ex-like modules). For these reasons there are several distinct areas of related work that are relevant and must be presented. In particular, in addition to presenting background information on autism, I describe how my work fits into the areas of social skills instruction, computer-assisted instruction, and crowdsourcing. This section serves as an overview of related work in these areas.

Autism and Social Skills Instruction

Background

Kanner [Kanner, 1943] and Asperger [Asperger, 1944] are generally credited with first identifying and describing individuals with autism in the 1940's. Today we have improved our understanding and awareness of autism and recognize it as a spectrum, clinically referred to as autism spectrum disorders (ASD) [APA, 2000]. Though not a clinically differentiated subgroup, individuals who are diagnosed with ASD, but do not exhibit language impairments, are often referred to as having high-functioning autism (HFA). Impaired social functioning is the central feature of HFA. A lack of social competency can result in significant difficulties in daily living, academic achievement, and poor adult outcomes related to employment and social relationships [Howlin, 2003; Klin and Volkmar, 2003].

Researchers and educators have attempted to develop and implement interventions that lead to social competency. The results of one recent meta-analysis, however, suggest that current school-based interventions were minimally effective for children with autism [Bellini et. al., 2007]. In order to improve the status quo they recommend increasing the intensity or frequency of the intervention, and developing interventions that address the individual needs of the child. My research provides the means to implement these changes; the child can practice these skills as frequently and for as long as necessary via Refl-ex, and REACT will enable the parent or other caregiver to easily create modules that address the child's needs.

Current Approaches

Social skills training interventions are an important part of the education of children with HFA. Due to the lack of a recognized best practice, educators use a variety of techniques, often in combination, to teach these skills. Power Card [Gagnon, 2001] and Social Stories™ [Gray, 1995] are examples of non-technological interventions. In Social Stories™, which is the paradigm more commonly used, parents or teachers develop stories that are related to some event in the child's life. Each story is meant to help the child learn appropriate behavior for a particular situation. These stories are written in a manner that is instructive, however they do not demand active child involvement [Reynhout and Carter, 2009]. Refl-ex augments this approach by engaging the student in the evolution of the story and guiding them as they practice social problem solving skills.

Recently, Social Stories called Storymovies have been made commercially available that include video recorded vignettes of the behavior being taught [Storymovies, 2012]. These videos provide richer visual cues, and in the style of children's television shows, ask the child questions about what they have just seen on the video. However, since the video has no way to know what the child has answered, or even if they have answered, the video proceeds to provide the correct response in all cases. Refl-ex, defers from this approach in that it engages the child by asking a question, waiting for a response, and then providing feedback on the response the student has given.

Interventions that incorporate technology range from those in which technology plays a marginal role, acting as a complement to other activities, to those in which technology plays a more central role. One approach to social skills training that uses a combination of technological and non-technological practices is the Junior Detective Training Program [Beaumont and Sofronoff, 2008], which consists of group social skills training, parent training, teacher hand-outs, and a computer game. The social competence that was sought was operationally defined as engaging in reciprocal positive interactions with others and responding appropriately to others' behavior. This program was tested with 44 students between the ages of 7 and 11. Parent-reported social skills of those in the treatment group improved from the clinically significant range to within normal range.

The "I can Problem-Solve" program [Bernard-Opitz et. al., 2001] is a completely software-based intervention used with children between the ages of 5 and 9. During the training sessions, the trainer presented a problem situation and a solution via pictures and animations. Children were then asked to suggest new solutions, and were reinforced with a variety of sensory or natural conditions (e.g. lines and spirals or a child jumping on a trampoline). Children with ASD produced fewer solutions than neuro-typical children, but the number of solutions produced by children with ASD increased with repeated usage of the software. The findings from the evaluation of both these interventions reinforce the importance of developing such software systems, and show that they are an effective way to teach social skills. The student cannot use these systems independently, however. Refl-ex enables this independent practice.

Other experimental technological approaches to ASD intervention include virtual reality simulations, and virtual peers for language learning, an important aspect of social interaction. In several studies researchers use virtual animated characters to invite language learning [Tataro and Cassell, 2008; Bosseler and Massaro, 2003]. Tartaro and Cassell in particular cite the advantages of using a virtual human over actual human interactors: virtual humans have more patience and can be made to consistently use strategies to elicit responses from the student.

Researchers have also created virtual reality environments designed to familiarize individuals with ASD with social settings [Parsons et. al., 2004, Laffey et. al. 2009]. Parson's and her colleagues' work helps students learn to identify roles and procedures in a social environment. Similarly, Laffey and his coauthors have created a virtual environment, called iSocial, which enables social interaction, and helps students learn and rehearse the use of meta-cognitive strategies, self-monitoring and self-regulation. Refl-ex differs from these approaches by simulating the progression through a social situation in which the student must exhibit social problem solving skills in a simplified manner. As will be seen in Chapter 3, Refl-ex uses simple and straightforward imagery to ensure the student focuses on the right aspects of the situation, and is not distracted by the quality of the virtual reality.

Computer-Assisted Instruction

Computer-assisted instruction (CAI) generally refers to any use of software in the instruction of students. Research in this area has evolved significantly in the more than 50 years since it began, and has been shown to effectively augment traditional instruction and interventions [Suppes and Morningstar, 1969; Anderson et. al. 1995]. The following two sections will present current research in intelligent tutoring systems, which are a specialized form of CAI, and authoring tools to aid in the development of these systems.

Intelligent Tutoring

When CAI software incorporates artificial intelligence to model real human tutoring practices, it is referred to as an Intelligent Tutoring System (ITS). ITSs employ models of instructional content that specify what to teach, and teaching strategies that specify how to teach. Some of the first approaches to Intelligent Tutoring were model-tracing tutors, and Cognitive Tutors [Koedinger et al., 1997; Anderson et. al., 1995], which interpret and assess student behavior with reference to a cognitive model that can solve problems in the way that competent students can. Constraint-based tutors [Ohlsson, 1992], which interpret and assess student work with respect to a set of constraints, and example-tracing tutors [Aleven et. al. 2009], which interpret and assess student behavior with reference to generalized examples of problem-solving behavior, are also approaches currently employed in ITSs. As such Intelligent Tutoring is a subclass of CAI that is theoretically more suited for personalized interactions.

Authoring Tools

Refl-ex shares many similarities with ITSs. In particular, both Refl-ex and ITSs require labor-intensive authoring of new content and problems, and require knowledge from diverse teams of experts. To facilitate the development of these systems, a great deal of research has been done on ways to develop ITS Authoring Tools. Blessing and his co-authors, for instance, have developed and evaluated an authoring tool to aid in the development of model-tracing ITSs [Blessing et. al. 2006, Blessing et. al. 2008]. They employ a three-part system to help users develop the cognitive models, define the rules of the task to be taught, and create problem instances. Similarly, Aleven and his coauthors have created an authoring tool that allows individuals to author example-tracing tutors without having any programming expertise; they use a drag and drop user interface to create the student-facing portion and a straightforward interface for creating the generalized examples. Both of these systems attempt to enable users to produce expert quality ITSs. The aim of the work described in this thesis is to facilitate the authoring of social problem solving skills instructional modules. While my work is not strictly considered an ITS, it is greatly informed by research in ITSs and ITS authoring.

Authoring tools that aid users in the creation of instructional stories for children with autism have begun to appear, in the form of commercial tools, both for the classroom and for parents to use at home. Intellitools [Intellitools, 2012] is an example of a tool developed for use in the classroom that is currently being used to develop instructional stories. Though not designed for that purpose, anecdotal evidence indicates that teachers take advantage of the text and sound functionality, combined with the color graphics provided by the software to create stories for their students. Stories2Learn is one of many iPad apps that have recently been developed for use with children with autism [Stories2Learn, 2012]. It is designed for use both by teachers and parents, and facilitates the creation of personalized instructional stories in the standard sequential narrative approach. These systems simply enable the creation of the story, but do not guide the author with respect to the content. REACT will not only provide the author with suggestions for content, but will also allow them to create interactive multi-path scenarios.

Crowdsourcing

The vastness of the literature that exists on crowdsourcing, despite the newness of the notion, is testament to its potential. In the following sections I will present some background on crowdsourcing, present examples of systems that facilitate it, and give an overview of some of the ways it has been used to date.

What is Crowdsourcing?

The first known use of the term crowdsourcing was in 2006 in Jeff Howe's article in WIRED magazine titled "The Rise of Crowdsourcing" [Howe, 2006]. Howe defines crowdsourcing, from a business perspective, as "*the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call.*" He explains that the crucial prerequisite is the use of the open call format and the large network of potential laborers.

The idea of such an approach predates the article by several years. One example is the work done on human computation [von Ahn, 2005]. Von Ahn defines human computation as: *"harnessing human time and energy for addressing problems that computers cannot yet tackle on their own."* He describes a scenario were human brains are treated as processors in a distributed system, each performing a small part of a massive computation.

In addition to crowdsourcing, and human computation, the notion of asking many people to complete a task that could not as easily or cheaply be done by a single person or a computer is sometimes referred to as human-based computation, citizen science, or using the wisdom of the crowd, far-flung genius, distributed intelligence or collective intelligence to name a few. The basic idea is, however, that people are realizing that under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them [Surowieki, 2005]. In addition, asking many people to contribute small amounts of time can make it possible to do a great deal of work faster, and as well, if not better, than a small group of experts or a computer [von Ahn et. al, 2008].

Crowdsourcing Facilitating Systems

In recent years a number of systems that facilitate the use of crowdsourcing have emerged. Yahoo! Answers, for instance, enables individual users to pose questions to the crowd [Yahoo! Answers, 2012]. On the Yahoo! Answers website, users can ask and answer each other's questions, but the process is restricted in that a user can only ask a few questions before he or she must answer some questions. This ensures that users are contributing to the process.

InnoCentive, is another example, in this case targeted towards companies, that provides a marketplace for them to outsource research and development work to individuals [InnoCentive, 2012]. The system supports four different types of problems:

ideation, theoretical, reduction to practice, and electronic request for proposal. Due to the complex and potentially time consuming nature of the problems, companies using InnoCentive typically award between $10,000 to $100,000 USD for solutions.

One of the most well-known and widely used crowdsourcing systems is Amazon's Mechanical Turk (AMT) [Mechanical Turk, 2012]. Requests on AMT are referred to as Human Intelligent Tests (HITs), and are generally tasks that humans can perform more cost-effectively than computers. AMT offers extensive options for creating customized questionnaires, and the results are made available in standard formats so they can easily be processed. The creator of a HIT has the right to decide whether or not to pay the worker who accepts the HIT, regardless of the worker's performance completing the task, and payment generally ranges between $.01 and $10. At the time of writing this document more than a quarter million tasks were available, making the popularity of AMT apparent. For these reasons, AMT was the most suitable system for me to use to build the knowledge base for REACT.

Example Uses

Wikipedia, is arguably the most pervasive, well known, and widely used example of crowdsourcing. In the Wikipedia entry about Wikipedia, it is defined as "a free, web-based, collaborative, multilingual encyclopedia project supported by the non-profit Wikimedia Foundation" [Wikipedia, 2012]. Volunteers around the world have written all of the articles in Wikipedia. The diversity of the volunteer pool is exemplified by the fact that less than 25% of the articles on Wikipedia are written in English. In addition, the site's policies promote verifiability and neutrality in the content of the articles. This is especially important since almost anyone who accesses the site can edit an article. The paradigm employed by Wikipedia has been so successful, that researchers want to better understand the reasons why people choose to contribute to the site, or become "wikipedians," [Bryant et. al., 2005].

Researchers across disciplines have also begun to explore the potential of crowdsourcing to advance their research. Researchers have explored its use in philosophy, for instance, in assembling concept hierarchies [Eckert et. al., 2010]. These researchers found that AMT users gave responses in the same deviation range as experts completing similar tasks. This confirmed crowdsourcing's potential to provide a cost effective way in which to collect this data.

Workshops have also been held to explore the use of crowdsourcing. In 2010, for instance, the North American Chapter of the Association for Computational Linguistics - Human Language Technologies held a workshop dedicated to exploring the ways in which speech and language data could be created using AMT. One of the papers at the workshop presented a way in which AMT was used to evaluate commonsense knowledge so that it could be used by a reasoning system [Gordon et. al., 2010]. These researchers indicated that their experiments taught them that providing users with some background on the goals of the research lead to higher-quality responses. They speculate that this is because they found the task more interesting or worthwhile. In addition, an overview paper presented the results of 9 experiments that were conducted in a range of linguistic disciplines from semantics to psycholinguistics [Munro et. al. 2010]. The authors argued that, with crowdsourcing, linguists have a reliable new tool for experimentally investigating language processing and linguistic theory, because it enables systematic, large-scale judgment studies that are more affordable and convenient than expensive, time-consuming lab-based studies.

Crowdsourcing Games

Games have also been developed as a means to use human computation to tackle important and often difficult tasks. Researchers realize that gamers spend large amounts of time playing single and mulit-player games online, why not use those cycles to do something beneficial? Phetch is one example of such a game, and is designed to attach descriptive paragraphs to arbitrary images on the Web, thereby improving accessibility (making output from screen readers more rich and accurate for the visually impaired) [von Ahn et. al., 2006]. Phetch is designed as an online game played by 3 to 5 players, where one of the players is chosen at random as the "Describer" while the others are the "Seekers." Since only the Describer can see the image, her or she must help the Seekers find it by giving a textual description of it. The descriptions are given iteratively, in theory, improving with every iteration. A scoring scheme is used, but essentially the descriptions that lead to a Seeker finding the image are saved as good descriptions. The researcher's goal is to have people engage in the game not because they want to do a good deed but because they enjoy it.

Similarly, scientists have used games to help predict protein structures. Foldit, is a multiplayer online game that engages non-scientists in solving hard protein folding prediction problems [Cooper et. al., 2010]. Players interact with protein structures using direct manipulation tools and user-friendly versions of algorithms from an advanced structure prediction methodology, and compete and collaborate to optimize the computed energy. Researchers found that not only did players excel at solving challenging structure refinement problems, but that players working collaboratively develop a rich assortment of new strategies and algorithms that, unlike computational approaches, explore not only the conformational space but also the space of possible search strategies. Their experiences indicate that the integration of human visual problem-solving and strategy development capabilities with traditional computational algorithms is a powerful new approach to solving complex scientific problems.

The game most relevant to the work in this thesis is The Restaurant Game [Orkin and Roy, 2009]. Players of this multi-player online game are anonymously paired to play the roles of a customer and a waitress in a 3D virtual restaurant. The incentive to play the game is social interaction and contributing data for a new collaboratively authored game. The authors used the data captured from over 11,000 players of the game to create a Plan Network, a statistical model that encodes context-sensitive expected patterns of behavior and language, with dependencies on social roles and object affordances. In other words this model is a graph showing probabilistically how events follow each other; as illustrated by the interactions in the game, thereby showing all the ways in which a restaurant experience can unfold. This data was then used to create a Collective Artificial Intelligence system that generates behavior and dialogue in real-time. This work shows that despite giving little to no guidelines people generally follow social conventions for this sort of interaction. This work lends credence to my assertion that crowdsourcing can be used to build realistic and complex social scripts and guided my use of crowdsourcing to develop the knowledge base for the system.

Crowdsourcing in Intelligent Tutoring Systems

Crowdsourcing has also been used in several intelligent tutoring systems. Rosa and Eskenazi [2011], for instance, use crowdsourcing to build the training data set for their vocabulary tutoring system. Crowdsourcing has also been used by researchers in math tutoring systems to automatically rate user generated math solutions [Aleahmed et.al., 2010], and in creating a personal learning environment for mathematics [Corneli, 2010]. Additional work has been done on using the crowd to help create an intelligent authoring tool that provides intervention strategies in response to the detected mental state of the student [Banda and Robinson, 2011]. The researchers refer to this as Multimodal Affect Recognition and use the crowd to analyze the video and audio training data. This is a case that uses workers complete a task that humans are generally quite good at: recognizing emotion.

Other Example Uses

To date human computation techniques have been used in a variety of other ways that are related to the work presented here. Researchers have explored its use, for instance, in enabling higher-level interactions between computers and users [Singh et. al., 2002]. This was done by inviting the general public to visit a website and teach computers common sense (i.e. build a database of commonsense knowledge using simple English sentences). At the site people are presented with a series of statements and asked if they are true, false, or sort of true (e.g. "You are likely to find a weasel in the dessert." Or "You are likely to find dessert in the supermarket."). This work confirms that the crowd can be called up to provide commonsense knowledge, which is an important component of social knowledge and skills.

Other research has explored the notion of crowdsourcing general computation [Zhang et. al., 2011]. This process involves problem decomposition to harness the crowd to perform general problem solving tasks, including local search and divide-and-conquer, and ends with the intriguing notion of creating a crowd-based advice system. The approach used is iterative, and is facilitated by Turkit, a toolkit that enables requestors on AMT to create programs that are iteratively executed by human workers [Little et. al. 2009]. This work shows that there is great potential to use the crowd to help solve problems like that of representing the intricacies that make up our social world.

Friendsourcing, or soliciting one's friends to answer particular social or personal questions, was also recently introduced and defined [Bernstein et. al., 2010]. When specific information is desired about a person or situation, asking one's friends can be a very effective solution. The work in this paper differs in that, not only is diverse information desired, but also because it is undesirable for the data to converge to a single "best" response. The goal is instead to have many distinct and relevant responses in order to provide parents with suggestions as they author the instructional modules.

The extent to which crowdsourcing has been used makes it impossible to present all the examples of its use. Some additional examples include, using crowdsourcing for audio transcription [CastingWords, 2012], information retrieval [Alonso et. al, 2008], problem solving [Brabham, 2008], and to help digitize old texts [von Ahn et. al. 2008]. The latter, a system called reCAPTCHA, again uses otherwise wasted human cycles for beneficial purposes. Most of us have encountered a website where you are asked to decipher distorted letters before being allowed to log in. That system, CAPTCHA, which was also developed by von Ahn, is meant to prevent non-humans from gaining access to these sites [von Ahn et. al., 2003]. reCAPTCHA, demonstrates that old print material can be transcribed, word by word, by having people solve CAPTCHAs (where one of the words to be deciphered is a word that an optical character recognition system was unable to convert correctly) throughout the Web.

Lastly, in a recent review of crowdsourcing experiments and systems, Wightman presents a framework for describing the trends in crowdsourcing and advice for researchers and other individuals planning crowdsourcing work [Wightman, 2010]. In the framework the author described four classes of what he refers to as crowdsourced human-based computation (CHC): 1) non-competitive direct motivation tasks, 2) non-competitive indirect motivation tasks, 3) competitive indirect motivation tasks, and 4) competitive direct motivation tasks. The work I propose in this thesis falls into the third category of this framework for which the author urges researchers to consider: 1) opportunities to reduce large tasks to sequences of smaller tasks, 2) the costs of paying users, and 3) opportunities to get results without paying users. In the preliminary crowdsourcing study I conducted, all users were volunteers, which addresses the last consideration. The other two bits of advice were addressed in this thesis, and are discussed in more detail in chapter 4.

It is apparent from this presentation of related work that many researchers from a variety of research domains have confirmed the utility of crowdsourcing. Their experiences informed the work I did to build the knowledge base of the system. The advice presented as lessons learned helped me to avoid some of the pitfalls, and use crowdsourcing effectively to facilitate the authoring of the instructional modules.

CHAPTER 3

REFL-EX: COMPUTER ASSISTED INSTRUCTION OF SOCIAL PROBLEM SOLVING SKILLS FOR ADOLESCENTS WITH HIGH-FUNCTIONING AUTISM

In this chapter, I describe an approach to social skills training for adolescents and young adults with HFA [Boujarwah et. al 2010]. The prototype system allows the student to role-play through social scenarios – such as going to the movie theatre—while providing appropriate scaffolding to support the student's effective practice of social problem solving skills.

Technology Development

The Refl-ex prototype system was built by myself, and two other researchers[1], in Adobe Flash and Adobe Flex Builder, and is made up of three interactive scenarios; Going to a Movie, Going to a New Restaurant, and Unlocking the Door [Refl-ex, 2012]. The system is inspired by Social Stories^{TM} [Gray, 1995], and is designed to present the student with real life social situation in which an unexpected obstacle arises. Social Stories^{TM} presents the student with a relevant story, but is not interactive; students are simply required to passively view the story. I augment this approach in several ways.

The three scenarios that were created are of differing levels of difficulty and familiarity to the students. The Unlocking the Door Scenario, for instance, presents a social situation that takes place at school. This scenario was considered to be the least

---

[1] Jackie Isbell and Hwajung Hong

complex due to the nature of the social problem the student is faced with, and their familiarity with the setting. The Going to a New Restaurant Scenario presents a scenario that is set in the home environment and requires more planning than the first scenario. Lastly the Going to a Movie Scenario, is considered the most complex as it takes place in the community and involves both complex interpersonal interactions and a dynamic environment. These scenarios were chosen because of their relevance to the student, both in terms of content and setting, and because they allowed for varying levels of difficulty.

Each scenario has two main components, an experience section and a reflection section. In the experience section the student is guided as they navigate a situation and overcome an unexpected social obstacle. This process creates a structure I call *obstacle-based branching,* which will be described in more detail in the next section. In the reflection section, the student revisits the decisions they made during the experience portion, and is able to reflect upon how they successfully found a solution to the social problem they encountered. Throughout each scenario, an invisible recording system logs data related to the choices the student makes as they interact with the system and the time it takes them to make those choices. This data gathering feature allows us to understand how a student progresses and differences between students. Following is a description of the various key design criteria incorporated into the system.

The Experience Section

In the experience section the social scenario is presented through text, audio narration, and picture book style images that correspond to the specifics of the situation (Figure 3.1). This is similar to the approach used in Social Stories™ which uses cartoon figures and text-bubbles to represent speech. One advantage of this visual presentation is that it allows for the inclusion of the meaningful information, and the exclusion of distracting material. In addition, like social stories, the system presents the scenarios in the first person perspective. This helps the student to identify with the character, and immerse himself or herself in the storyline. Individuals with autism prefer environments with high degrees of certainty, repetition, and predictability. Dealing with others in social settings introduces a high degree of uncertainty for those with autism. It is common for individuals with HFA to rehearse for situations before hand. Unfortunately, rehearsal is not always effective as those with autism might learn cues specific to only one environment or one person [Heflin and Alberto, 2001]. Our visual design choices are meant to help avoid the learning of incorrect cues by limiting the information in each picture.

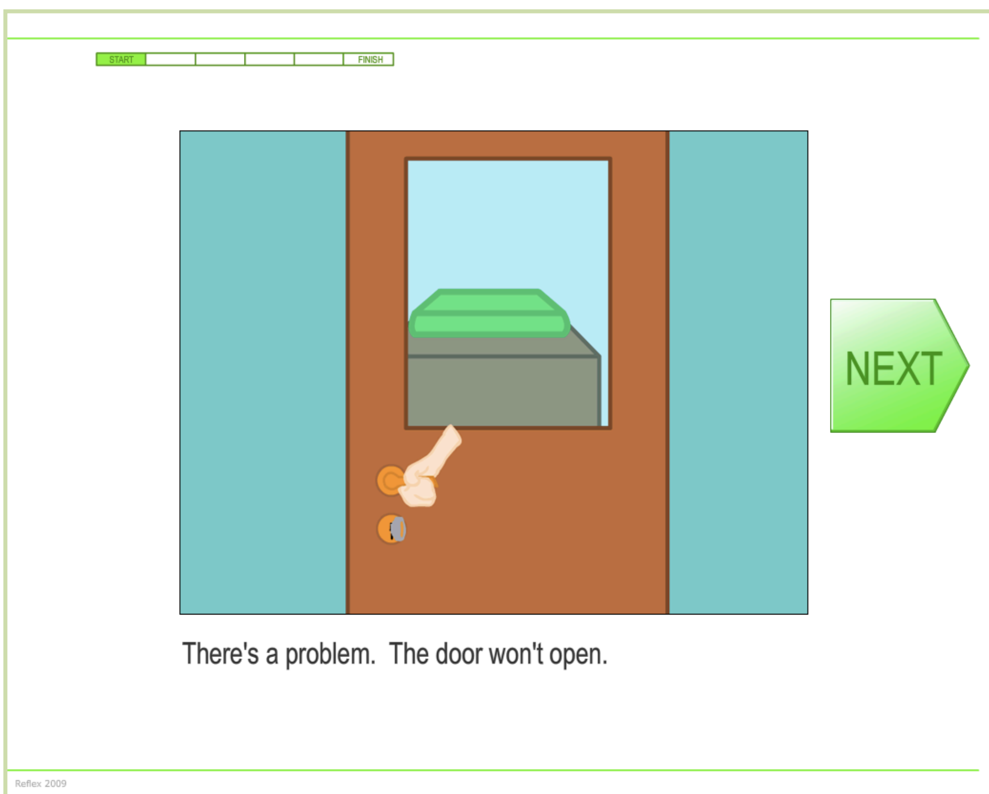


Figure 3.1. Introduction of the obstacle in the Unlocking the Door Scenario.

In order to ensure that the student acknowledges the narration, throughout the experience section all the buttons are disabled until the audio narration has ended. In this way, the software prevents the student from clicking through the frames without perceiving the information that has been presented.

Creating an *Obstacle-Based Branching* Story

In order to make the system interactive, we approached each situation as a narrative with multiple paths, where a narrative is defined as a series of actions that describe how a situation unfolds. All the paths together resemble a branching story, which is a graph structure such that each node represents a segment of narrative and a choice point. The canonical branching story systems are Choose-Your-Own-Adventure novels. However, recent research has explored computational approaches to branching stories [Riedl et. al., 2008]. Refl-ex can be considered to be a branching narrative where each possible narrative is based on productive, unproductive, and counter-productive possible executions of social skills in response to obstacles that arise in specific contexts. The *obstacle-based branching* story we use in the experience portion consists of three major decision points in which the student has to make a series of choices to overcome the obstacle, proceed to the next stage, and ultimately successfully navigate the social situation.

At each branching point in the story the student is presented with a series of choices. The system plays all the choices individually then the student is taken to a screen where they are prompted to make a decision (Figure 3.2). The student makes their choice by clicking on the image or button on the screen. This action takes them to a page where they can review the option before confirming their choice.

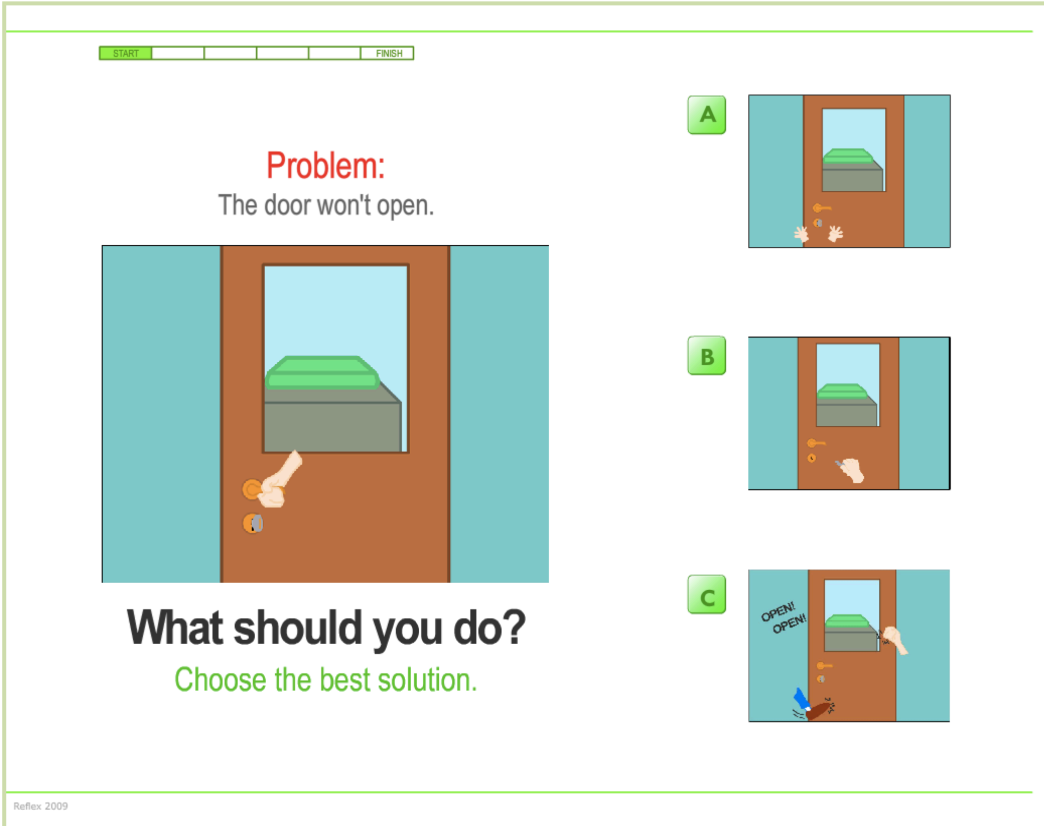


Figure 3.2. Decision point in the Unlocking the Door Scenario.

<u>Errorless Learning</u>

We follow an approach of errorless learning, which means that the student is not allowed to fail. When the student makes a choice that is considered unproductive or counter-productive the system explains the possible consequences of that action without using negative language, and directs the student to rethink their choice (Figure 3.3). When the student returns to the decision point the undesirable choice that they have already explored is grayed out to prevent it from being chosen again. In this way the system provides immediate feedback and helps the student to correct their error. Errorless learning is often used with individuals with HFA to avoid the possibility that they acquire incorrect skills; individuals with HFA are extremely prone to repetition so it is essential to avoid reinforcing anything other than the desirable execution [Heflin and Alberto, 2001].

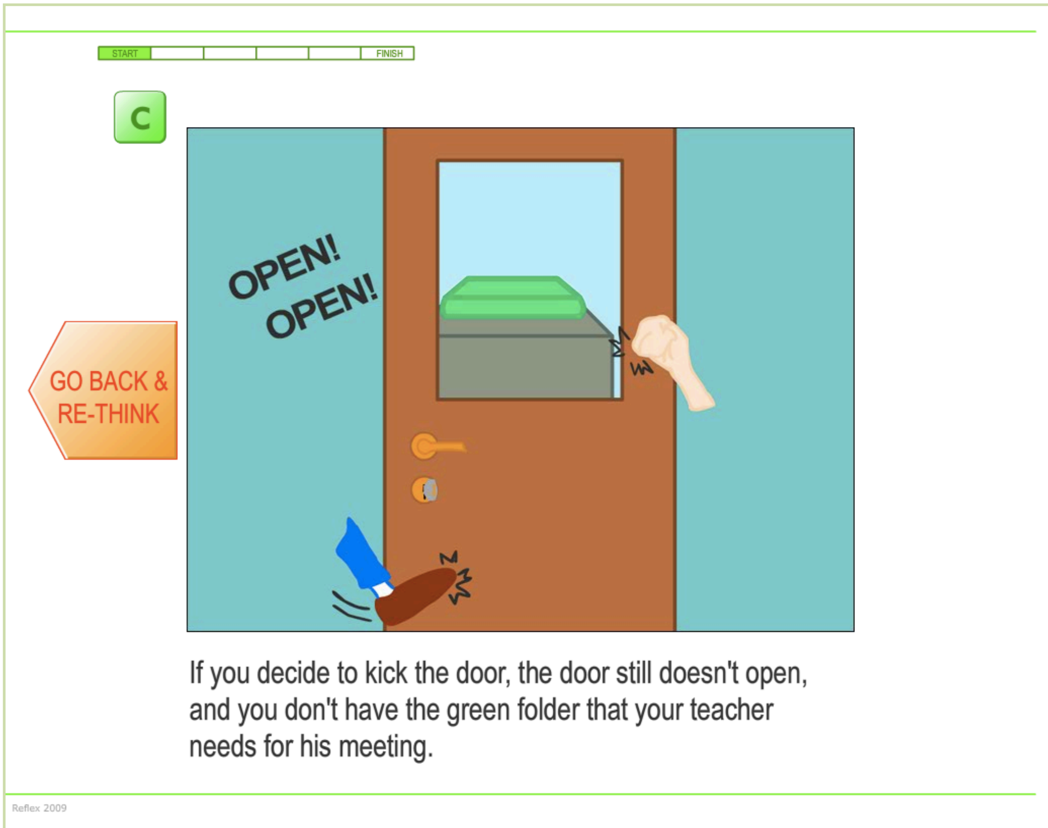


Figure 3.3. Explanation of the consequences of an undesirable choice.

The Reflection Section

Once the student has successfully navigated the experience portion of the scenario, they are asked to reflect on their decisions by recreating the social story. The student is presented with a puzzle piece for each decision they made, and is asked to recreate the story by sequencing the puzzle pieces correctly on a timeline (Figure 3.4). The pictures used in this section are the same as those the student saw in the experience portion. For this portion the student is given both text and audio instructions.

In the reflection portion the student is again given immediate feedback. If the student drags a puzzle piece to the wrong location on the timeline, they are prompted to "Try again" by a message that appears on the screen. When the student places a puzzle piece in the correct location the action is reinforced with a "Good job!" message. Once the student has successfully recreated the social story, the story is played back to them. Each picture appears on the screen with the corresponding text and audio narration. By prompting the student to revisit the story they created, the system reinforces the social problem solving skills that were used to solve the problem. In this way, the technology assists and supports the student as they reflect on the decisions they made and provides the student with an opportunity to process information that can be recalled for later use.

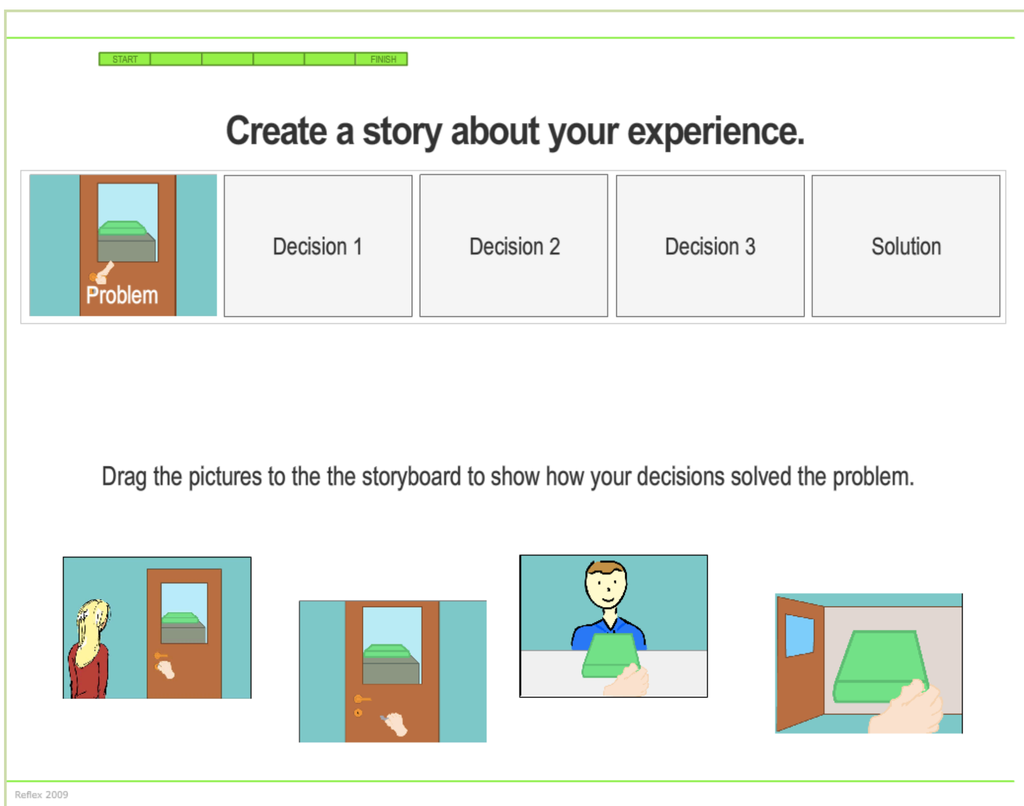


Figure 3.4. Reflection section as presented by the system.

The Exploratory Study

In an effort to validate the *obstacle-based branching* instructional approach, we conducted two studies, the first of which was an exploratory study with eight individuals with HFA. These individuals were students between the age of 13 and 19 years who were attending a special needs school in Atlanta. Half of the participants were 18 years old or older, but were still in school due to their disability. We used the Test of Problem Solving for Children and Adolescents (TOPS2-A), which has an interview format with questions involving social situations. Because this test has been shown to correlate to actual problem-solving ability [Griswold et al., 2002], it was used to assess the participants before they interacted with the system. None of the questions on the TOPS2-A are directly related to the scenarios in the system. We also asked the participants to complete the Social Problem Solving Inventory (SPSI-R) [D'Zurilla et. al., 2002], which has a multiple-choice format, and is designed to determine an individual's perceptions of their own problem solving skills. Lastly, we asked the participants to complete each of the three scenarios in our prototype system. The methodological procedure was as follows.

Pre-tests

Prior to beginning testing, all parents were given a description of the software and asked to give their consent to allow their child to participate in the study. Once parent consent had been obtained the participant's assent was collected. On the first visit, we administered the standardized tests. All of the testing took place in the school in room usually used as a community room (Figure 3.5a).

TOPS-2A

The TOPS2-A test was administered individually to the participants. A researcher read a passage aloud while the participant read the same passage silently if they chose. The researcher then asked questions about the passage and instructed the participants to answer verbally. Myself and another researcher were present during the testing and both scored the participant's responses. Immediately after each test was completed the two of us discussed the scores to ensure inter-rater reliability and accuracy in the scoring. Since the TOPS-2A has only been validated for use with individuals age 12 to 17 and 11 months, the test was only administered to the four participants that were in this age range.

SPSI

The following week, before using the software, the participants were asked to complete the SPSI. A researcher read the questions aloud and asked the participant to respond verbally. Participants were also allowed to complete the SPSI on their own.

Software Testing

On the second week the hardware was set up in the same community room. Three computers were set up such that the participants had their backs to each other and would not distract each other (Figure 3.5a). This allowed the participants to complete the scenarios in parallel, thereby reducing the disturbance to their normal routine. The participants completed the scenarios in three groups, two groups of 3, and a group of 2. Each participant was given a new headset that was labeled, so that the participant could use it on all three days of the software testing. Since the software only required mouse interaction, participants were not provided with a keyboard (Figure 3.5b).

Figure 3.5a. Setup of the community room in which the study was conducted.

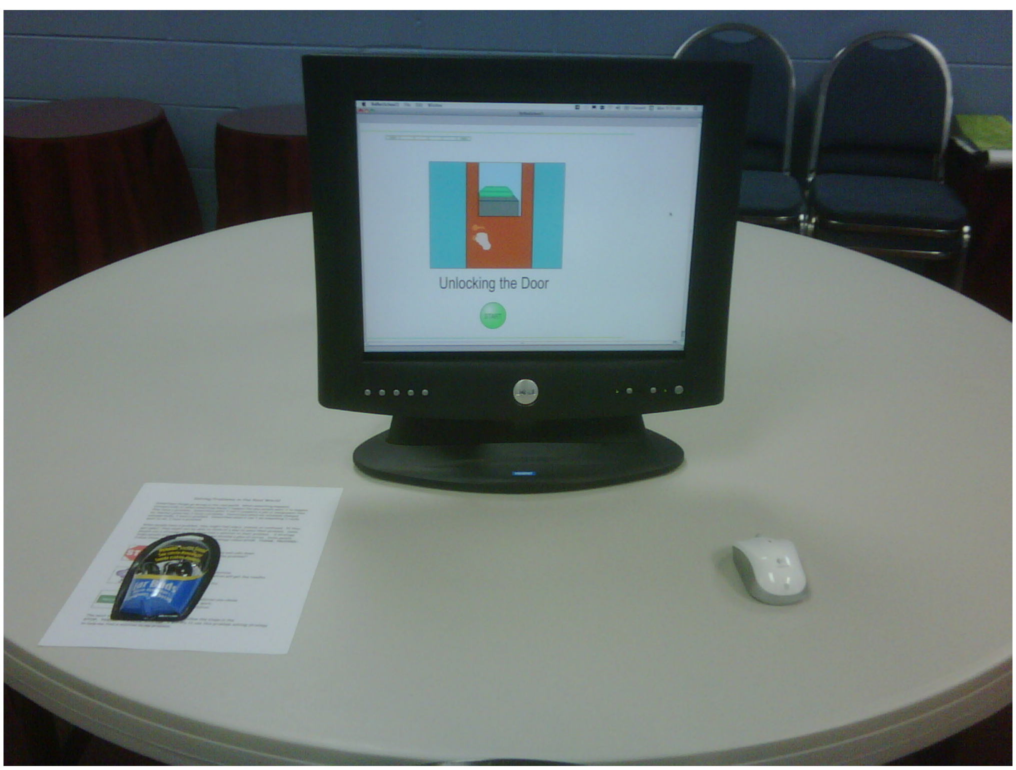
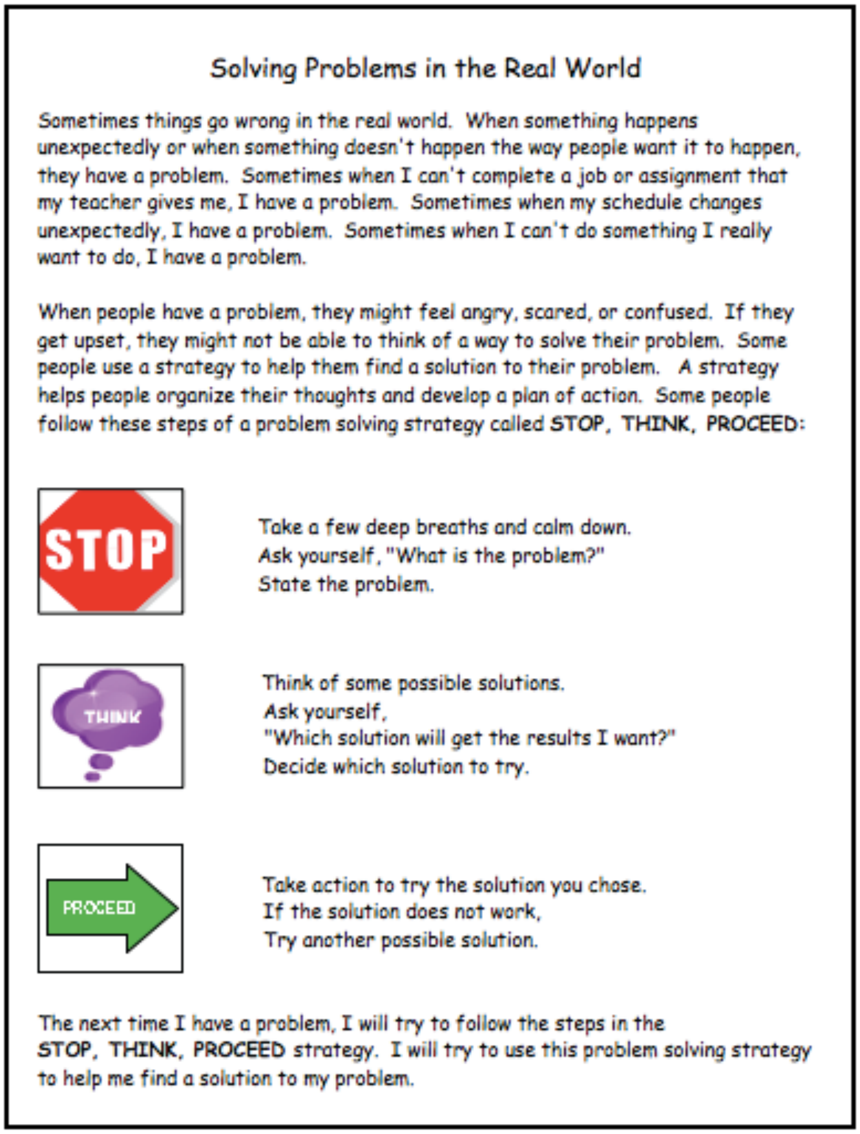


Figure 3.5b. Setup of the experimental equipment for each participant.



Solving Problems in the Real World

Sometimes things go wrong in the real world. When something happens unexpectedly or when something doesn't happen the way people want it to happen, they have a problem. Sometimes when I can't complete a job or assignment that my teacher gives me, I have a problem. Sometimes when my schedule changes unexpectedly, I have a problem. Sometimes when I can't do something I really want to do, I have a problem.

When people have a problem, they might feel angry, scared, or confused. If they get upset, they might not be able to think of a way to solve their problem. Some people use a strategy to help them find a solution to their problem. A strategy helps people organize their thoughts and develop a plan of action. Some people follow these steps of a problem solving strategy called **STOP, THINK, PROCEED**:

STOP

Take a few deep breaths and calm down.
Ask yourself, "What is the problem?"
State the problem.

THINK

Think of some possible solutions.
Ask yourself,
"Which solution will get the results I want?"
Decide which solution to try.

PROCEED

Take action to try the solution you chose.
If the solution does not work,
Try another possible solution.

The next time I have a problem, I will try to follow the steps in the **STOP, THINK, PROCEED** strategy. I will try to use this problem solving strategy to help me find a solution to my problem.

Figure 3.6. Problem solving process as presented to the participants.

Each day before they began to interact with the software we read the participants a short passage that introduced a problem solving process (Figure 3.6). This passage explicitly presented the problem solving process that the participant is lead through in the software. Following the Social Stories', format this passage was presented on a sheet of paper with brightly colored icons.

Each participant then completed one scenario per day for three consecutive days. We presented the scenarios in order of difficulty; asking the participants to complete the Unlocking the Door Scenario the first day, the Going to a New Restaurant Scenario the second day, and the Going to a Movie Scenario the third day. In order to reward the participants for their efforts, they were allowed to play a computer game after completing each day's scenario.

Results

The following is a detailed discussion of the log data collected, and the results from the standardized testing.

Standardized Testing

Only two of the four participants administered the TOPS-2A were able to successfully complete it, P3 and P6. P6 gave responses that resulted in a standard score in the low average to average range for his age of 15 years and 9 months. His test score resulted in an age equivalency of 12 years and 6 months. P3 gave responses that resulted in a standard score in the below average to low range for his age of 13 years and 5 months, and an age equivalency below the minimum measured by the test, which is 11 years and 2 months.

P4 attempted the test, but was unable to complete it. He became agitated and began repeating parts of the question as his response. For this reason, his testing was stopped and scoring his test was not possible. We were also unable to score P2's test. He had very limited language, and it was not possible to complete his test because we could not be sure that he understood the passage or the questions.

We only attempted to administer the SPSI to the two adolescent participants who successfully completed the TOPS2-A (P3 and P6). P6 was able to complete the SPSI independently, scoring 124, which is in the above average range. P6 responses classified him as having a positive problem orientation, with an avoidant style of problem solving. P3, showed signs of stress early in the survey and we did not ask him to complete it.

Only 2 of the participants over 18 were able to complete the SPSI. P8 was able to complete the inventory independently, scoring 89, which is in the low average range. P7 was also able to complete the SPSI, however, he asked that a researcher read it to him, indicating that "reading is part of [his] disability." P7 scored 90 on the SPSI, which is in the average range. Both these participants were classified as having a negative problem orientation, and an avoidant style of problem solving. Of the two remaining participants over 18, only P5 attempted the test, but began showing signs of distress almost immediately so we stopped the test. We did not feel that P1 had the verbal language ability to be able to complete the inventory, so we did not attempt it.

Unlocking the Door Scenario

All the participants completed all three scenarios successfully during the software testing. The logs collected by the software provided us with a detailed description of the participants' interaction with the system; recording mouse clicks and timings. We began by analyzing the data for the Unlocking the Door Scenario. As can be seen in Figure 3.7, seven out of the eight participants chose a path considered to be complex, in that the participant was required to navigate three decision points in order to successfully solve the social problem. In addition, all seven of these participants chose the same complex path (2C, getting help from a teacher).

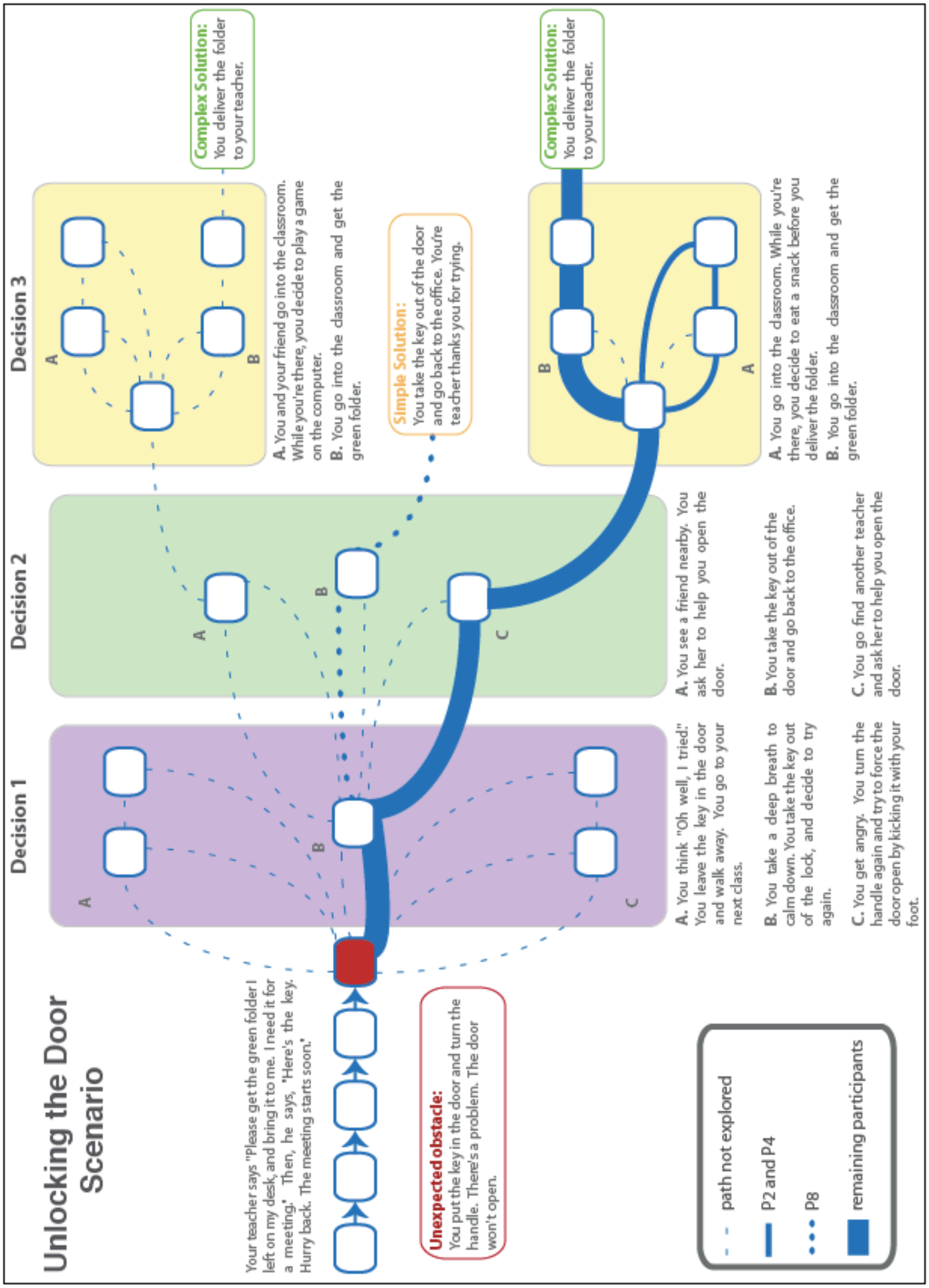


Figure 3.7. Paths chosen by the participants in the Unlocking the Door Scenario.

There was a great deal of variation, however, in the time it took the participants to complete the scenario. Participant P5, for instance, completed the complex path in the shortest time, taking only 156 seconds, and P6 took the longest time, taking 252 seconds (Figure 3.8).

The logs also allowed for the analysis of data regarding the number of times a participant clicked a particular button. This enabled us to see which participants did not wait until the audio narration had completed before attempting to proceed.

As can be seen in Table 3.1, this data also varied greatly across participants. In this case a low number indicates that the participant more often waited to hear the complete narration, and a zero in decision 3 indicates that the student chose the simple solution, and therefore did not have to complete the third stage of the scenario.

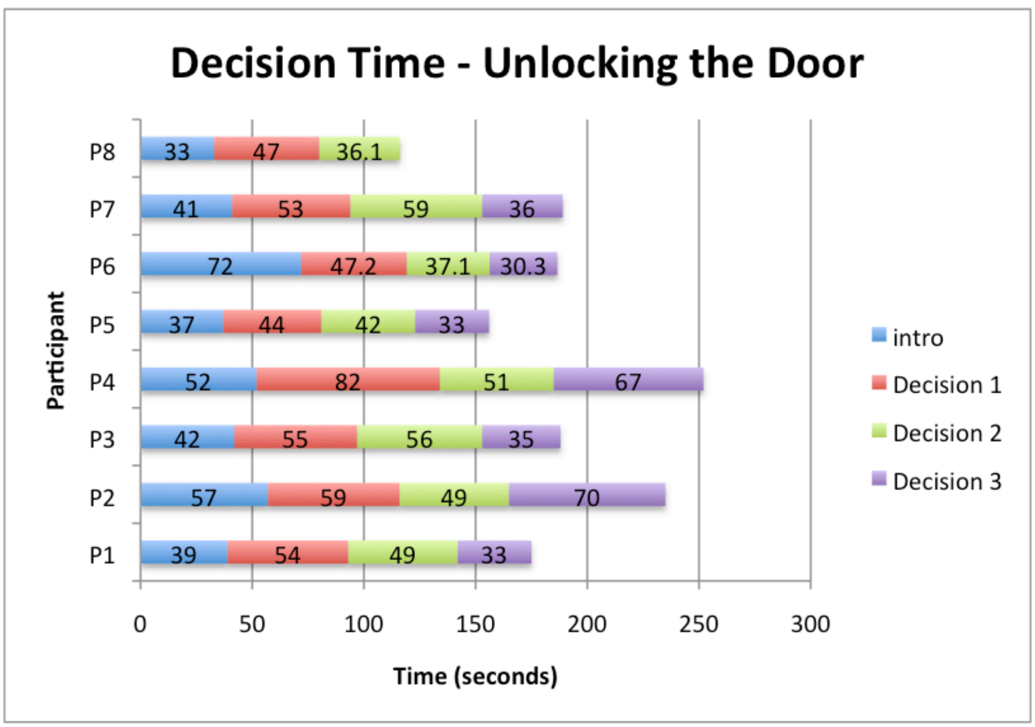


Figure 3.8. Time to complete each stage in the Unlocking the Door Scenario.

Table 3.1: Number of mouse clicks logged in each stage of the Unlocking the Door Scenario

	Intro	Decision 1	Decision 2	Decision 3	Total
P1	5	3	3	4	15
P2	5	3	3	7	18
P3	5	3	3	5	16
P4	5	5	4	7	21
P5	5	6	7	6	24
P6	5	12	15	19	51
P7	5	4	3	3	15
P8	8	6	7	0	21

Log data was also collected from the Reflection portion of the software. This data showed long pauses in some of the participants' transitions from the Experience to the Reflection portion. Following the pause, however, very few errors were made. Six out of the eight participants (all except P2 and P7) made one or fewer errors before successfully populating the timeline.

Post-Hoc Analyses

Following the conclusion of the study, the other researcher that was present during the standardized testing and throughout the software testing and I ranked the participants with respect to their language abilities and visible characteristics of autism. The other researcher is a PhD student in Education with more than ten years of experience teaching children with special needs. Despite the differences in our backgrounds, we both ranked the participants in the exact same order. These rankings can be seen in Table 3.2, where one indicates the participant that was observed to have least noticeable characteristics of autism and the most language and four indicates the participant perceived to have the most characteristics of autism and the least language.

Table 3.2: Subjective post-hoc ranking of participants.

Minor	Rank	Over 18	Rank
P6	**1**	**P7**	**1**
P3	**2**	**P8**	**2**
P2	**3**	**P5**	**3**
P4	**4**	**P1**	**4**

The participants were divided into those that were over 18 years of age and those that were minors. This allowed for the timing data and the click data to be correlated to the ranking. For the minors in the study, we found that in all three decisions the number of clicks was negatively correlated with the participants subjective ranking. What this mean is, the participants that had more language tended to be more impatient to proceed and so clicked more. We also found that the timing data was very strongly positively correlated with the ranking in the minor participants, with the total time having a 0.95 correlation with the ranking. This means that the minor participants with more characteristics of autism and less language tended to take longer to complete the scenario. These correlations were not as clear in the data from the participants over 18, however.

Discussion

Only 3 of the 8 participants (1 minor, 2 over 18) were able to complete the standardized testing component of this study. The goal of using standardized testing measures was to have an objective measure of the students' problem solving ability before beginning our intervention. Given that our sample size was small we cannot make claims about the general feasibility of using these two measures with a population with HFA, however, it does indicate that an alternative manner of assessing problem solving skills in this population is needed. One consideration is that both of the current assessments required that the participants propose solutions, in other words they needed to recall information. Research indicates that for individuals with HFA "recalling" information taxes their executive function and thus may be more difficult than other manners of responding [Griswold et al., 2002]. Thus, researchers should consider adapting the problem-solving test so that it is more conducive to this population, and enables them to exhibit their knowledge by other means, for instance by recognizing appropriate options (multiple-choice). This manner of response is less taxing cognitively and may result in a higher rate of successful completion. This would enable us and other researchers to have a better means of assessing social skills acquisition.

Despite the difficulties we encountered with the standardized testing, our experience with individuals with HFA made us confident that our participants would interact more favorably with the technology. Our expectations were confirmed during the study, as all the participants were able to successfully complete all three scenarios. This supports the idea that the software was able to provide the scaffolding necessary to enable even the participants that struggled with the testing to be successful during the software intervention.

In the Unlocking the Door Scenario, participants had the choice of either giving up and returning the key, or asking a friend or teacher for help. The latter is more complex because it requires interpersonal interaction, and the navigation of a third decision point. Seven out of eight participants chose the same complex solution, which involved the student asking a teacher to help them open the door (Figure 3.7). Our study was conducted at a school in which the students interact one-on-one with a counselor for most of the day. For this reason, we believe that this choice is a reflection of the environment in which these participants find themselves.

The results also indicated that longer pauses were logged at the transition between the experience and the reflection sections. We believe that this is the result of the fact that the actions expected from the student in this section were notably different from those in the experience portion. All the participants were able to complete the reflection portion successfully indicating that some time to transition between the tasks was all that was required.

In addition, our post-hoc subjective ranking allowed us to analyze the log data further (Table 3.2). We found that the minors that were ranked higher were more impatient and did not wait to hear the audio. We believe that this is likely because these participants were able to read the text quickly, and therefore were ready to proceed before the audio had completed.

The correlations between subjective participant ranking and log data for the participants over 18 were not as telling. This can be explained in several ways. In conversation with P1 and his counselor, for instance, we learned that, despite exhibiting limited verbal language, P1 was a proficient reader and Scrabble™ player. These skills enabled him to benefit from the text provided in the software and navigate the scenarios successfully. Similarly P7 exhibited particular interests in technology and suggested that we modify the system to allow for those that can, to proceed faster. Knowing that the buttons would not respond until the sound had completed, he chose to wait before attempting to proceed.

Contributions

The work done in this portion of this thesis suggests that participants with HFA responded favorably to a prototype software system designed to help them independently practice social problem solving skills. This is in line with other studies that have found that individuals with HFA do well with computer-assisted interventions [Beaumont and Sofronoff, 2008, Bernard-Opitz et. al., 2001, Rao et. al, 2008, Tartaro and Cassell, 2008].

As expected the log data confirmed the heterogeneity of this population, and provided us with insights into the process that the students used to solve a given scenario. This work also shows that a computer based intervention can be adapted to be used in a school setting and that students could, in fact, have a period where they are able to engage in an intervention then go back to their normal routine.

This study addressed research question one which was: *How can software modules be developed to help a student with autism prepare for various social contexts? Contribution:* The development of interactive software modules that adolescents with HFA can use to independently practice social skills.

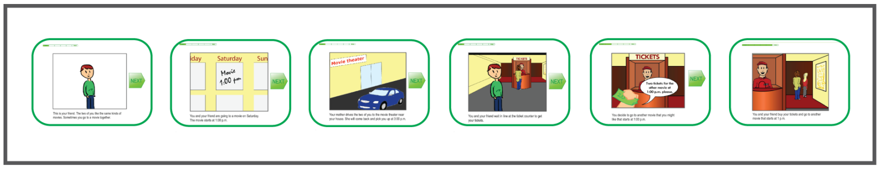
Now that it had been confirmed that students with autism responded favorably to the software, the next step was to systematically obtain expert evaluation of the *obstacle-based branching* approach. Following is a description of the study that was conducted to this end.

The Branching Validation Study

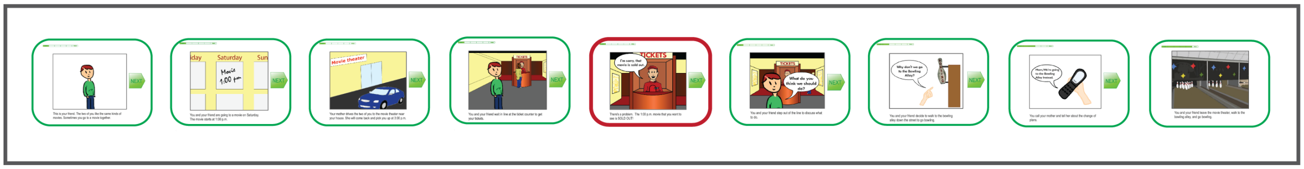
Before pursuing the challenging goal of creating an authoring tool to facilitate the creation of individualized Refl-ex modules, it was crucial to systematically validate the *obstacle-based branching* approach used in the modules. For this reason, I conducted a study in which experts were shown three versions of the instructional software:

1. *A sequential story in the standard sequential format (Figure 3.9a).* In this format the student is presented with a story that exhibits appropriate behavior for a particular social scenario, like going to a movie.
2. *A sequential story in which an unexpected obstacle arises (Figure 3.9b).* In this format an obstacle arises, but the story continues and the student is lead through one possible solution to overcome the obstacle (e.g. the movie we want to see is sold out, so we go to the bowling alley instead).
3. *A story in the obstacle-based branching format (Figure 3.9c).* In this format, an unexpected obstacle arises, and the student must navigate a series of branching/decision points in order to find a way to overcome the obstacle. In addition, the student is presented with explanations of the consequences of inappropriate behavior. This format was described in more detail in Chapter 3.

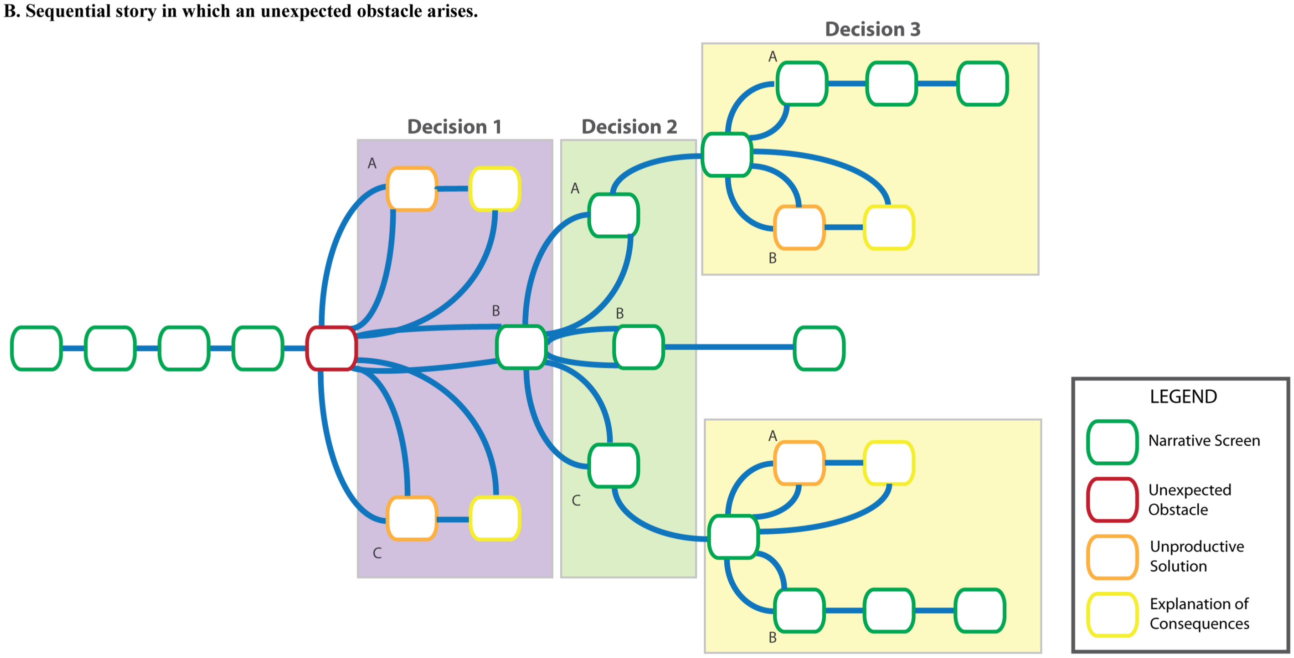
The three versions of the instructional modules were created for all three scenarios; the Going to a Movie Scenario, the Going To A New Restaurant Scenario, and the Unlocking the Door Scenario.



A. Sequential story in the standard format.



B. Sequential story in which an unexpected obstacle arises.



C. Story in the Refl-ex format.

Figure 3.9. a-c (a) Sequential story in the standard format. (b) Sequential story in which an unexpected obstacle arises. (c) Story in the *obstacle-based branching* format.

Participants and Recruitment

The participants in this study were individuals who have experience working with students with autism and other developmental disabilities. In particular, I recruited participants who have experience in social skills instruction. Given the importance of social skills in the education of children with autism, many special education teachers fit these criteria. The participants were recruited by word of mouth. Since I have shared my research with many in the autism community in Atlanta, I made it a point to recruit participants who had not interacted with Refl-ex in the past.

Participant Demographics and Qualifications

Sixteen experts participated in the study. All the experts were white females, with four indicating that they were between the ages of 25 and 34, four between the ages of 35 and 44, six between the ages of 45 and 54, and two between the ages of 55 and 64. The majority of the participants had master's degrees in special education or a related field (13 of 16), two had PhD's in clinical psychology, and one had a bachelor's degree in special education. Twelve of the 16 experts were teachers of students with autism, and the four remaining participants were therapists. They worked with students at all age levels with the most working with high school students (7/16), five worked with elementary school students, two with middle school students, and two indicated that they worked with individuals with autism throughout their lifespans. Most importantly, all the participants indicated that they provide social skills instruction.

Procedure

Each interaction with a participant was conducted one-on-one, as an informal questionnaire driven semi-structured interview. This allowed for rich qualitative data to be collected verbally, and quantitative data to be collected in writing.

In order to obtain feedback systematically I began with a questionnaire requesting demographic information, and in particular: degrees achieved, certifications and other qualifications, the extent of their experience working with children with autism, their experience teaching social skills and their current social skills instructional practices (Appendix A).

Participants were then given the second questionnaire, and presented with the 3 versions of the software each depicting a different scenario. The order of presentation was randomized so all participants did not see them in the same order, or depicting the same scenario. For example, participant one saw the unlocking the door scenario in the *obstacle-based branching* format first (Figure 3.9c), followed by the movie scenario in the standard sequential format (Figure 3.9a), then the restaurant scenario in the sequential with obstacle format (Figure 3.9b). The questionnaire the participants were given was used along the way to obtain individual evaluations of each version. At the end of the questionnaire participants were asked to provide overall evaluations of the three versions. These questions facilitated the discussion with the participant regarding the strengths and weaknesses of each of the three formats. The interviews were audio recorded to facilitate the analysis of the participants' responses.

The Evaluation Questions

The questions asked in the questionnaire given to the participants after their interaction with the software were informed by a literature review of studies in which approaches to addressing social skills have been evaluated. For instance, Reynhout and Carter [2009] conducted a study to evaluate the perceived efficacy of social stories. The questions used in the study included questions about:

1. The types of behaviors educators would use social stories to address, for instance: social interaction, conversation, to introduce changes/new routines, to reduce inappropriate behaviors.
2. The settings in which educators would use stories, for instance: classroom, playground, at home.

In addition to asking these questions, I asked participants what they liked and disliked about each format, how useful they thought each format was, and how appropriate they thought each format was for their students. I also asked participants about their desire/motivation to customize the stories. Finally I asked four straightforward questions that provided assessments of: 1) the perceived utility of introducing an obstacle; 2) the perceived utility of incorporating interaction (i.e. giving the student options for solutions); 3) which format they liked best; 4) which format they liked least. The complete questionnaire is available in Appendix B.

Results

<u>Perceptions and Usage of Social Stories</u>

As can be seen from the questionnaires, participants were asked several questions about their perceptions of Social Stories, their use of the approach, and how useful they think customizing them is to the student. The data from this portion of the study can be seen in Table 3.3 and 3.4.

Table 3.3: Social Stories usage data.

Question	**Answers**			
Do you use Social Stories?	Yes 9	No 4	Sometimes 3	
How long does it take you to create a customized story?	50 minutes (average across the 16 responses)			
What types of behaviors do you use Social Stories to address?	Social Interaction Skills 13	Conversation Skills 9	Address/introduce changes/new routines 13	Address/reduce inappropriate behaviors 14
What settings would you use Social Stories in?	Classroom 14	Playground 5	At home 10	

Table 3.3 shows the participants responses to the questions related to their usage of Social Stories. It can be seen that 12 of the 16 participants use Social Stories to some extent, and that on average the participants indicated that it takes them 50 minutes to create a custom story. Further, the participants use the stories to address a variety of different behaviors. In addition to the ones given as options (see Table 3.3), the participants indicated that they also use Social Stories to address: behavior in public; hygiene and self-help; directions; peer interaction; bullying; the hidden curriculum[2]; and expected behavior in specific settings. Social stories are also used in a variety of settings. In addition to those offered as options the participants added: in the community, therapy, social life (prom), and karate. These findings confirm the fact that the potential content for social skills instructional materials is vast and varied.

It is apparent from the data below (Table 3.4) that all of the participants would like to use Social Stories more often. Despite the fact that four of the participants indicated that they did not use Social Stories (Table 3.3), the approach was rated as being useful (5.1/6). Most notably, the participants indicated that creating custom stories makes a significant difference to the student (5.7/6). This data greatly motivates the need for a tool like REACT to help parents and teachers create customized socials skills instructional modules.

---

[2] A **hidden curriculum** is a side effect of an education, "[lessons] which are learned but not openly intended" such as the transmission of norms, values, and beliefs conveyed in the classroom and the social environment [Wikipedia, 2012]

Table 3.4: Social Stories assessment.

Question	Answers	
How useful do you find Social Stories?	5.1 (6 = very useful)	
Would you want to create Social Stories more often?	Yes	No
	16	0
Do you think it makes a difference for a story to be customized?	5.9 (6 = a significant difference)	

<u>Evaluation of the *Obstacle-Based Branching Approach*</u>

The final type of data that the study produced was data assessing the usefulness and appropriateness of each format of the software (Figure 3.9), overall ratings for the best and worst formats, and individual assessments of the usefulness of introducing an obstacle and of introducing interactions (i.e. options for the students to choose from). This data can be seen in Tables 3.5 and 3.6.

Table 3.5: Assessment of the three software formats.

Format	**Useful** (6 = very useful)	**Appropriate** (6 = very appropriate)
Sequential Story	3.8	4.2
Sequential Story with Obstacle	4.5	4.8
Refl-ex Format	5.1	5.2

In the questionnaires participants were asked to rate the usefulness and the appropriateness of each of the formats to their students on a 6 point Likert Scale (1= "not useful at all"/ "not appropriate at all" to 6 = "very useful"/ "very appropriate"). As indicated by the data, the *obstacle-based branching* format rated highest in both measures, and the sequential story rated lowest. Further, when asked which format they thought was the best, and which format they thought was the worst, 14 of the 16 participants indicated that the *obstacle-based branching* format was the best, and 14 of the 16 participants indicated that the Sequential format (which is most like the approach currently being used) was the worst. This data provides expert confirmation to the assertion that the modifications we made to the current approach improved it.

Table 3.6: Participants overall ratings of the three software formats.

Format	**Best** (votes)	**Worst** (votes)
Sequential Story	0	14
Sequential Story with Obstacle	2	2
Refl-ex Format	14	0

One of the most interesting findings in the study was that at least four of the 16 participants indicated that each of the formats could be useful either for a particular student (based on their level of development or functioning) or for a particular situation. These participants indicated that they would like to be able to create all three formats. This finding implies that an authoring tool should allow for the creation of all three formats.

To more directly get at assessments of the specific design decisions, participants were also asked how useful they thought the introduction of the obstacle was, and how useful they thought that incorporating interaction was. The responses to these questions were overwhelmingly positive with the introduction to the obstacle receiving an average score of 5.6 and the interaction receiving an average score of 5.8 (6 = "very useful" in both questions).

Contributions

Altogether, the findings of this study not only addresses research question 1b, by providing expert confirmation that branching and interaction improve the current approach, but the responses related to the time it takes to create a customized story, and the vastness of the potential content strongly support the need for an authoring tool.

Contribution: The creation of *obstacle-based branching*, an approach to developing interactive social skills instructional modules that experts confirm is an improvement to current practices.

Conclusions and Contributions

The population of young adults with high functioning autism (HFA) is growing. Many of these individuals can function effectively and autonomously, but need assistance to handle the complexities of society. The work done in this portion of this thesis confirms that the *obstacle-based branching* approach used in the Refl-ex modules is evaluated by experts to be an improvement to the current approaches to teaching social skills. Further, the findings of the exploratory study suggest that participants with HFA responded favorably to the prototype software that was designed to help them independently practice their social problem solving skills.

In summary, the work presented in this chapter addresses research questions 1a and 1b which were:

1.a. How can software modules be developed to help an adolescent or young adult with HFA to prepare for various social contexts?

1.b. What value do experts perceive in obstacle-based branching scenarios, and how do they compare to the current approach of using sequential stories to teach social skills?

Contributions:

- The development of interactive software modules that adolescents with HFA can use to independently practice social skills.
- The creation of *obstacle-based branching*, an approach to developing interactive social skills instructional modules that experts confirm is an improvement to current practices.

CHAPTER 4

BUILDING A KNOWLEDGE BASE TO SUPPORT THE AUTHORING OF SOCIAL SKILLS INSTRUCTIONAL MODULES

How do neurotypical individuals organize their knowledge and experiences in such a way that allows them to know what appropriate behavior is in a particular situation? For instance, how do you know that you are supposed to pay for your food before you sit down at a fast food restaurant, but not until after you have eaten at other restaurants? Schank and Abelson present the notion that we develop scripts, or standard event sequences, which enable us to subconsciously know what to expect [Schank and Abelson, 1977]. These scripts are not rigid, and instead contain multiple contingencies. They explain that people develop these scripts early in life, based on their experiences. Research has shown that these scripts develop early in childhood, and that children with autism generally generate fewer well-organized scripts [Trillingsgaard, 1999; Volden and Johnston, 1999].

Is it possible to ask people to describe their scripts? Further, can those scripts, once collected, be turned into a model that can be used to provide authors with suggestions for content during their authoring process? These are questions I attempted to answer with the work described in this chapter.

The Preliminary Study

I developed and conducted a study to explore the potential of eliciting these scripts using crowdsourcing techniques, like those discussed in Chapter 2. I asked participants to describe the steps they take to complete everyday tasks, namely, going to a restaurant and going to a movie. I also asked participants what could go wrong at each step. The goal was to use this data to create a model of these everyday tasks in a manner inspired by Orkin and Roy's [2009] Collective Artificial Intelligence. The data collected would be analyzed to develop a model similar to a Plan Network that shows probabilistically how events follow each other and all the ways in which a restaurant experience can unfold. I planned to use this model to enable the system to provide the author with suggestions for subsequent steps. The design and results of this study were presented in a poster presented at the Human Computation Workshop [Boujarwah et. al., 2011].

Participants and Recruitment

To minimize costs (I just bought a few chocolates) my participants were students at Georgia Tech that were over 18 years of age. I recruited these participants by word of mouth, namely distributing flyers in two classes, an undergraduate psychology course, and a joint undergraduate/graduate computer science course that is cross-listed with industrial design. In addition, myself and another student spent 2 hours at the Student Center stopping passersby and asking them to complete the study. We provided potential participants with background on the study, explaining that we were working on developing software to help children with autism practice social skills. We found that this made people much more likely to stop, and they appeared to put more effort into their responses.

Procedure

I used a series of dynamically generated online questionnaires to collect the data [Cognitive Models Study, 2010]. After indicating their consent to be in the study (by clicking to proceed to the next page), participants were asked to provide some minimal demographic information, namely; age, gender, ethnicity, highest level of education, and whether or not they have a learning disability (Figure 4.1).

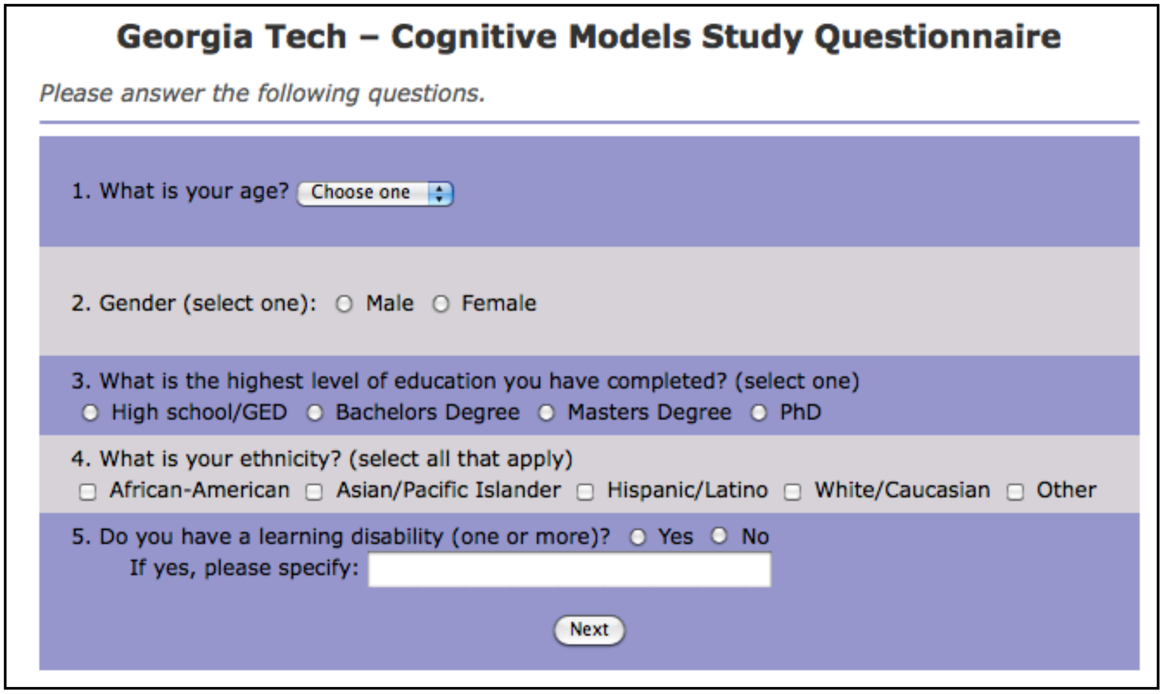


Figure 4.1. Demographic data collected in the Cognitive Models Study.

Next the participants were asked to describe the steps they use to navigate a particular situation. In order to motivate a certain level of granularity in their responses, examples of potential steps were given. In addition, the form began with one step (Figure 4.2), and participants could add and remove steps as needed (with at least 1 step and at most 12 steps) (Figure 4.3). The structure of the study was such that all participants were asked to describe their scripts for going to a restaurant and going to a movie, the order in which they were asked to provide these was balanced to counteract the effects of learning and of participants getting tired or bored. Also, after describing the steps they would use for going to a restaurant, participants were asked to indicate which restaurant they were thinking of (Figure 4.4).

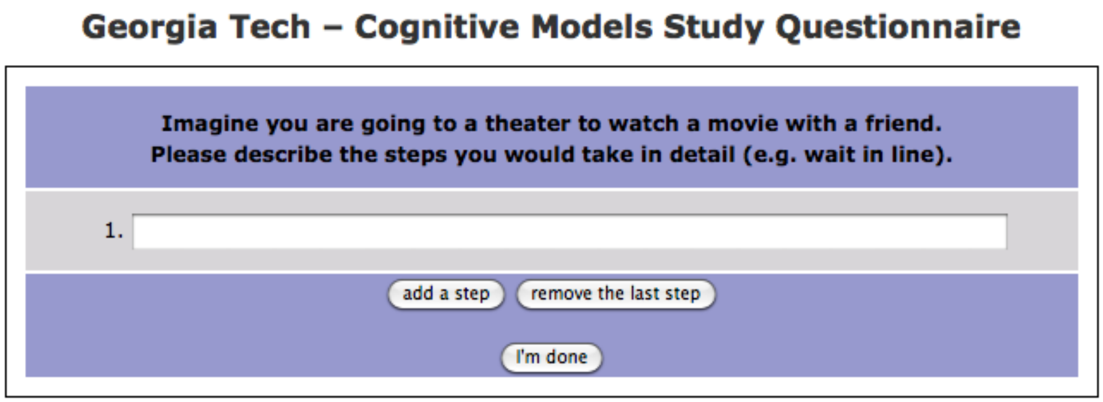


Figure 4.2. Webpage requesting participants' steps for going to a movie.

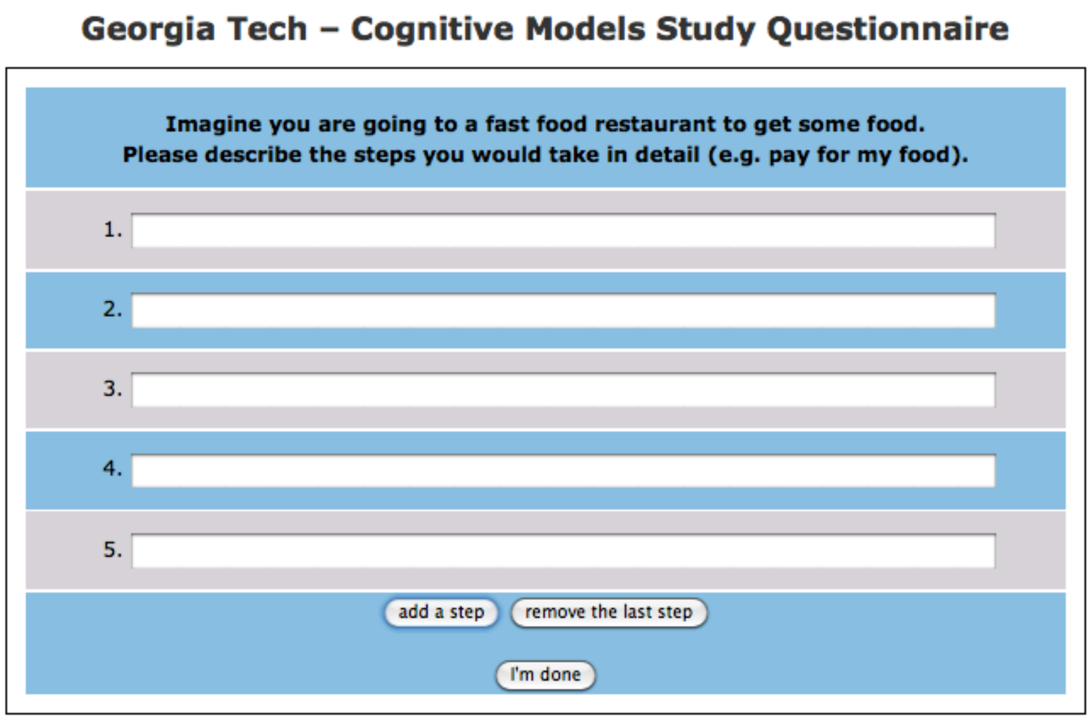


Figure 4.3. Webpage requesting participants' steps for going to a restaurant (with 4 steps added).

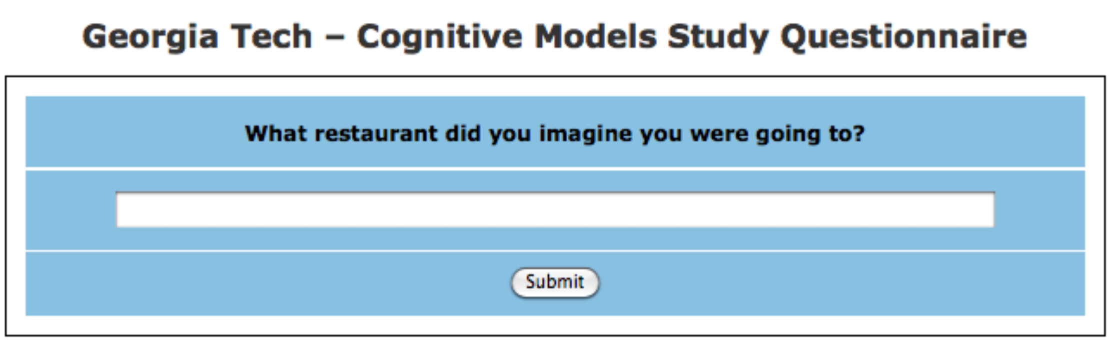


Figure 4.4. Webpage asking participants what restaurant they were imagining.

After providing all the steps in their script for one task, pages were generated dynamically that showed participants their steps, three at a time, and asked them to give an example of something that could go wrong at each step (Figure 4.5). The same procedure was then used for the second task.

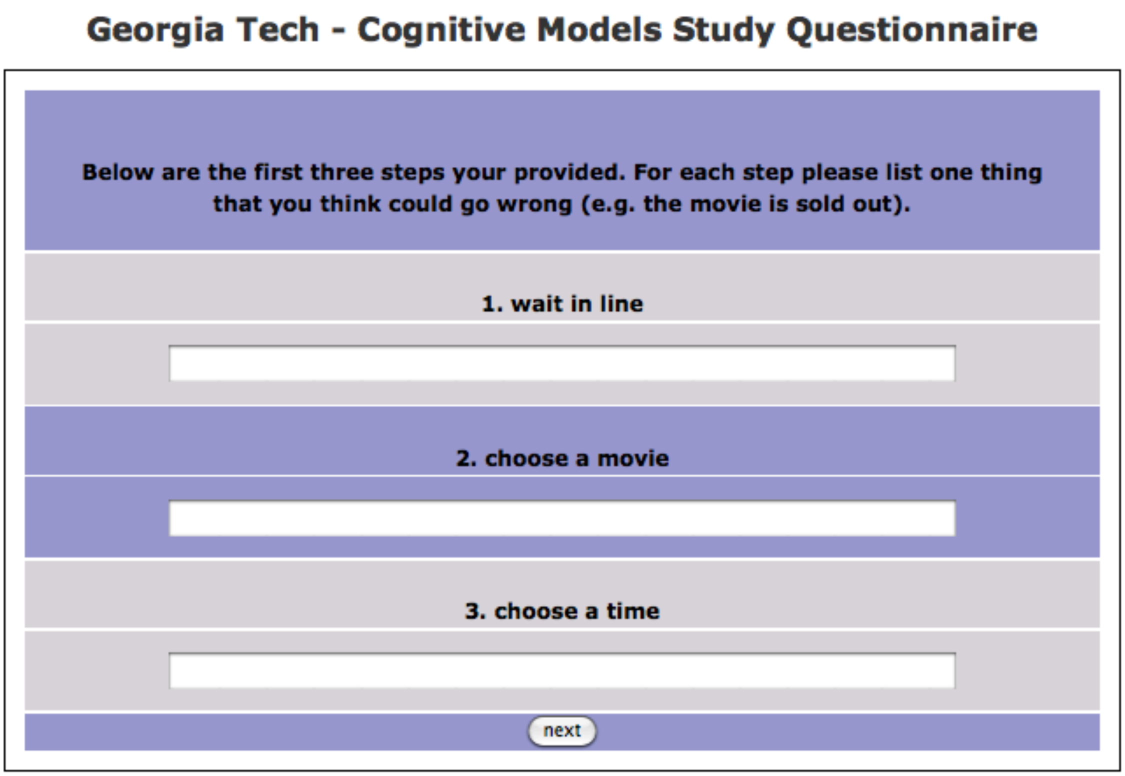


Figure 4.5. Webpage asking participants what could go wrong at each step.

Results

Once enough data had been collected to work with, I began to analyze it with the goal of using it to create a model. At the time of data analysis, 38 participants had completed the study. A summary of the demographics of the participants is presented in Table 4.1. The data indicates that participants were somewhat diverse with most between the ages of 20 and 24, a little more than one-fourth were female, and approximately half were graduate students. The data also shows that the algorithm used to balance the order in which participants were asked for their steps was not perfect (23 saw the restaurant first, and 14 saw the movie first). I believe that having a larger number of participants would have made these number more balanced. Also, two participants did not complete the second script, once when the movie was second, and once when the restaurant was second. Due to the fact that the data from the two tasks was not related, I could still use the data they had provided.

Table 4.1. Demographic information from the preliminary crowdsourcing study (38 participants).

Age		**Gender**		**Education**	
<20	3	Female	10	High School	23
20-24	30	Male	28	Bachelors	13
>24	3			Masters	2
Ethnicity		**Learning Disability**		**Order of Responding**	
African American	5	Yes	1	Restaurant first	23
Asian	8	No	(ADHD)	Movie first	14
Latino(a)	2		4		
White	24				
Other	2				

I chose to analyze the restaurant data, since more participants saw it first, and I planned to focus on this scenario when creating the authoring tool. Thirty-seven participants provided the steps they use when going to a fast food restaurant. Despite only starting with one step on the screen, on average eight steps were provided. The smallest number of steps provided was three, and seven participants provided 12 steps. Three of those seven participants' last steps appeared to indicate that they were not done, and would have liked to continue.

After doing a high-level analysis of the data, I proceeded to manually create a model, like the one that was previously described, representing the probability with which steps follow each other (Figure 4.6). First, I noticed that two participants had described the steps for going through the drive-thru of a fast food restaurant. The steps of this task are different enough from entering the restaurant that I decided to exclude this data. I then proceeded to classify the steps. I began by identifying steps that used the same word, (e.g. “line,” “menu,” “order,” and “condiment”). I then combined the steps that represented the same steps. This process created most of the classes, the remaining steps were clustered appropriately such that all steps representing the same action were in the same class, and all the steps had been classified. I used the classified steps to create a graph with weighted directed edges. Each node was a class of step, and an edge E from node N1 to N2 with weight .33 indicated a 33% probability that N2 followed N1 in the scripts. Edges with probability zero were excluded to prevent confusion. Therefore, the sum of all outgoing edges from a node is one. Also, I marked nodes that appeared in more that 75% of the participants scripts in red.

Discussion and Conclusions

From this graph of event sequences in the restaurant script, even though I only had 35 scripts to work with, a rich description of possible restaurant event sequences was created. The information in this graph would enable the system to suggest possible next steps for the author. For example, if the author just wrote a screen in which the student has looked at the menu, the most probable next step would be to “decide what to eat” (45%), and other suggestions could be ordered by probability; “order food” (.32), “greet cashier” (.09), etc. In addition, this structure could be used to help the system keep track of the nodes the author has visited, or the steps that have already been completed, and the ones that remain to be done.

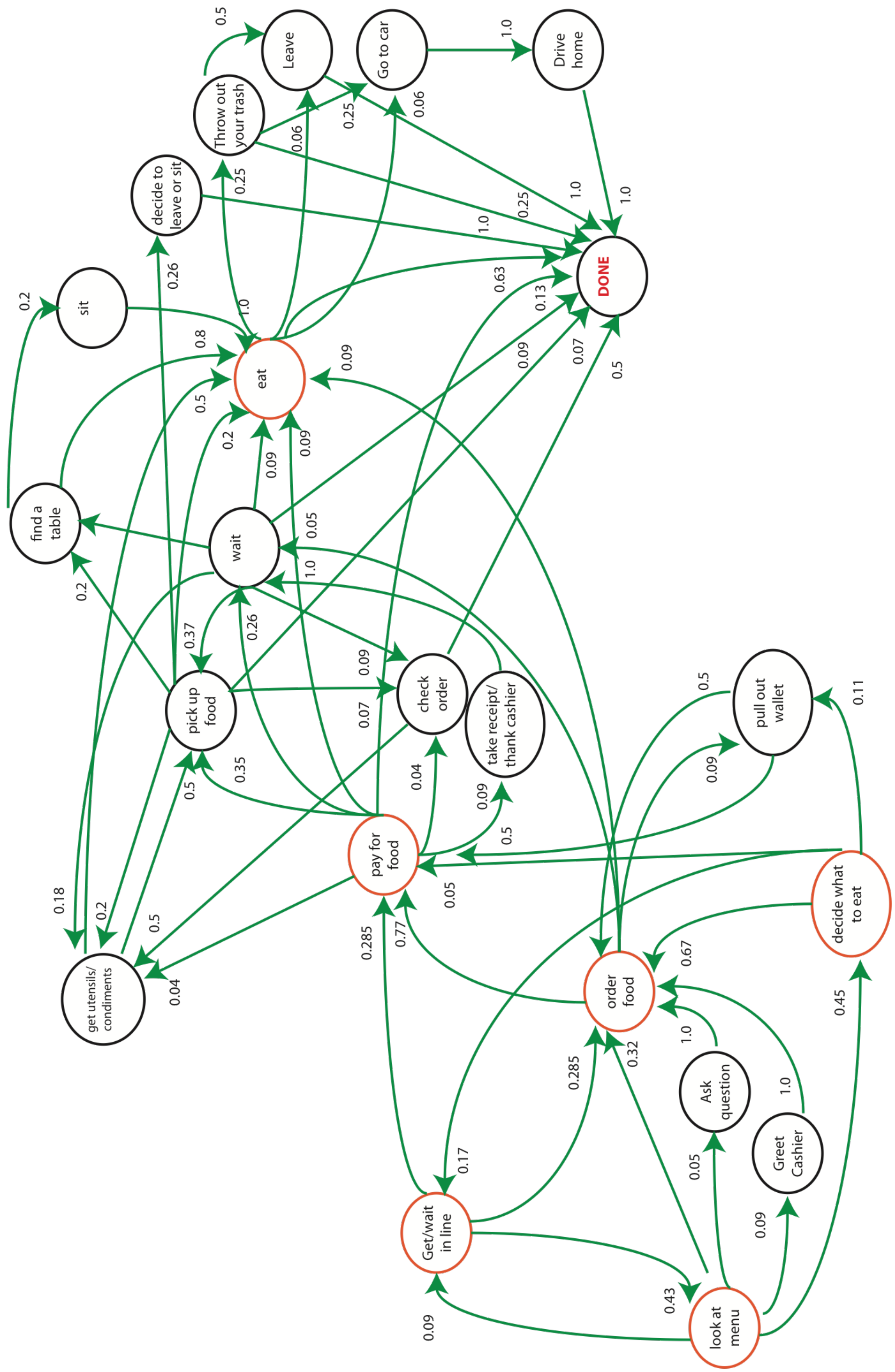


Figure 4.6. Model representing the probability with which steps follow each other (35 participants).

In addition to confirming that rich data could be collected using crowdsourcing techniques, and converted into a model manually, this study taught me a great deal about how to refine this approach in a way that would allow me to successfully collect data on Amazon's Mechanical Turk (AMT). In particular:

1. Provide participants with background on the purpose of the research, or how their responses will be used.
2. Be as specific as possible to constrain the range of responses. For example, instead of saying "Imagine you are going to a fast food restaurant…" say "Imagine you are eating lunch at a fast food restaurant" and provide an explicit starting and ending point for the workers' responses.
3. Ask for less from each participant/respondent. Instead of asking the same person for the entire script, and things that could go wrong, only ask for the script or possible obstacles.

The lessons learned from the preliminary analysis of the data from this study were similar to those presented by researchers in recent crowdsourcing literature (chapter 2), and were used to improve the approach used for crowdsourcing the knowledge base for the authoring tool (described in the next section). In addition, the graph that emerged from the data confirmed that it is possible to get good structured information when eliciting scripts from the crowd.

How to Crowdsource Social Scripts

The preliminary study provided confirmation that it was possible to elicit scripts from a limited crowd; Georgia Tech students. The next step was to truly crowdsource the scripts using a crowdsourcing facilitating system, like Amazon's Mechanical Turk

(AMT) [MECHANICAL TURK, 2012]. To attempt this, I have developed an approach to creating social script models using an iterative crowdsourcing approach [Boujarwah et. al., 2012]. These models are meant to be used to provide the author with suggestions during the authoring process for steps to include in the module, obstacles to introduce, and solutions to offer the individual. For this reason, three types of data are needed—steps, obstacles, and solutions. The complete models show how events follow each other and many of the ways in which an everyday experience can unfold, with deviations from the path in the form of obstacles followed by solutions. After completing the preliminary study, it became clear that the probabilities with which steps follow each other were not helpful in this context. Instead of offering the author the most common next step, it was more useful to provide all the possible next steps in a random order, as unexpected events is what parents are trying to prepare their child for.

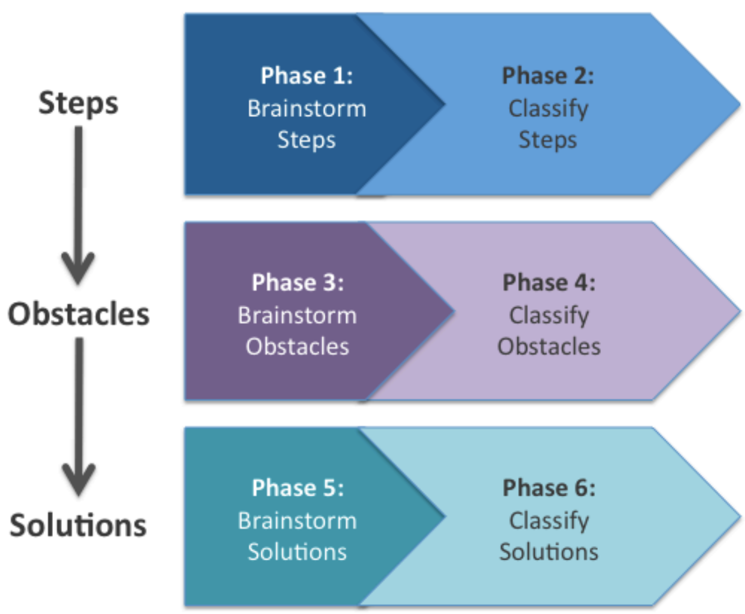
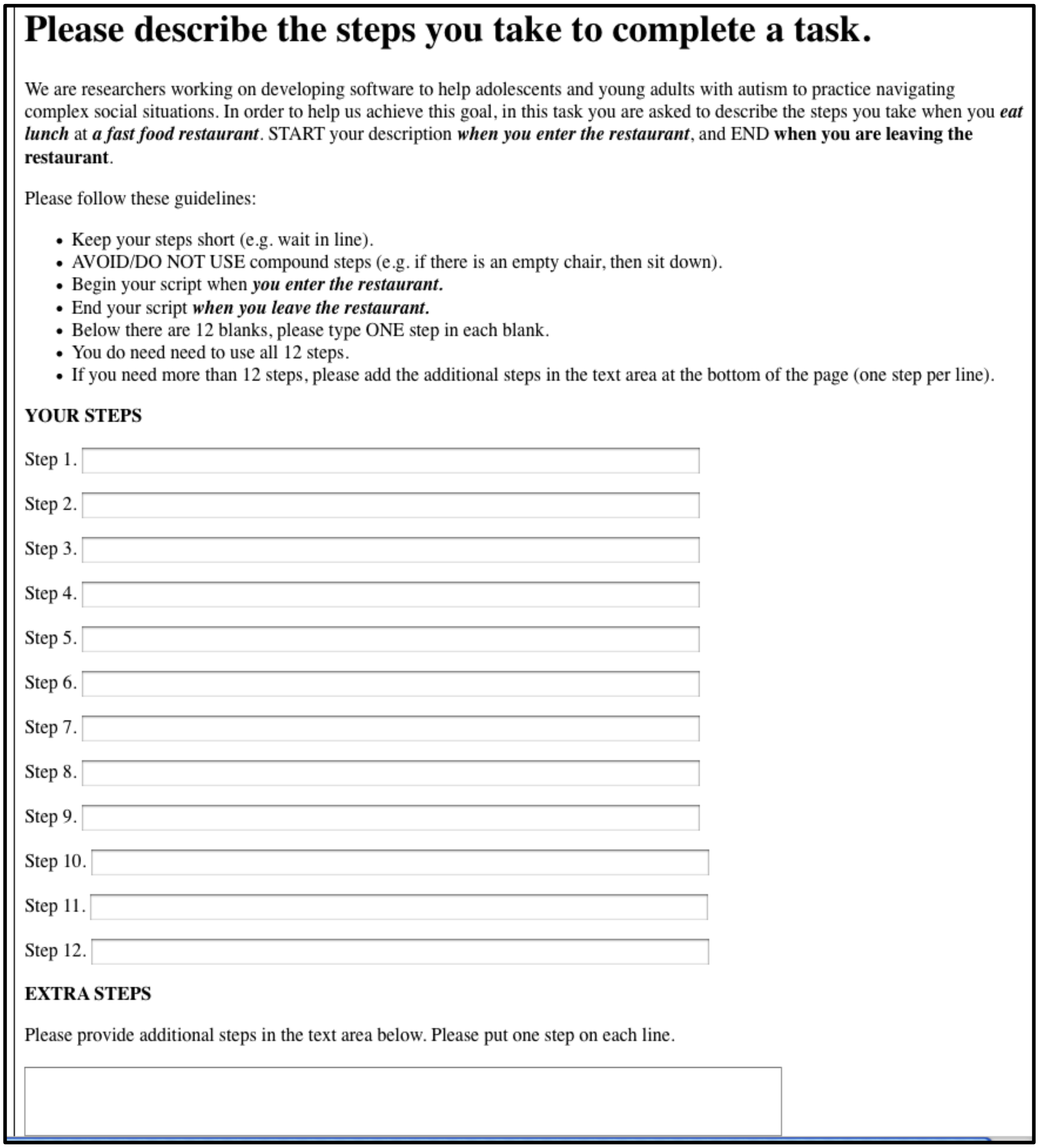


Figure 4.7. 6-Phase process for creating the models.

To build the models I conduct two phases of data gathering for each of the 3 types of data; a brainstorming phase and a classification phase. This results in a 6-phase process (Figure 4.7). First, an assortment of input data must be specified, namely; the location and task or activity the script is to be created about, where and when the script should start and where and when the script should end. The idea is that, in the future, the person seeking to author a module will specify this data. To illustrate, the HIT examples below will be shown with the task *"eat lunch,"* the location *"a fast food restaurant,"* the starting point *"when you enter the restaurant,"* and the ending point *"when you exit the restaurant."*

The process proceeds as follows:

Phase 1: *Brainstorm Steps*—A HIT (AMT Human Intelligence Task) is created using a template like that shown in Figures 4.8 and 4.9. As shown, each of the HITs were preceded by a presentation of the reason for the data collection, in order to motivate workers to give better responses (Figure 4.8). For each HIT, workers were asked to provide demographic information (age, gender, location, whether or not they have a disability) and any comments they might have (Figure 4.9). I felt it was important to collect this information because AMT HITs are accessible by people around the world, but what is considered socially appropriate behavior is not the same around the world. Knowing more about the worker's social context gives us greater insight into their responses.



Please describe the steps you take to complete a task.

We are researchers working on developing software to help adolescents and young adults with autism to practice navigating complex social situations. In order to help us achieve this goal, in this task you are asked to describe the steps you take when you ***eat lunch*** at ***a fast food restaurant***. START your description ***when you enter the restaurant***, and END **when you are leaving the restaurant**.

Please follow these guidelines:

- Keep your steps short (e.g. wait in line).
- AVOID/DO NOT USE compound steps (e.g. if there is an empty chair, then sit down).
- Begin your script when ***you enter the restaurant.***
- End your script ***when you leave the restaurant.***
- Below there are 12 blanks, please type ONE step in each blank.
- You do need need to use all 12 steps.
- If you need more than 12 steps, please add the additional steps in the text area at the bottom of the page (one step per line).

YOUR STEPS

Step 1.

Step 2.

Step 3.

Step 4.

Step 5.

Step 6.

Step 7.

Step 8.

Step 9.

Step 10.

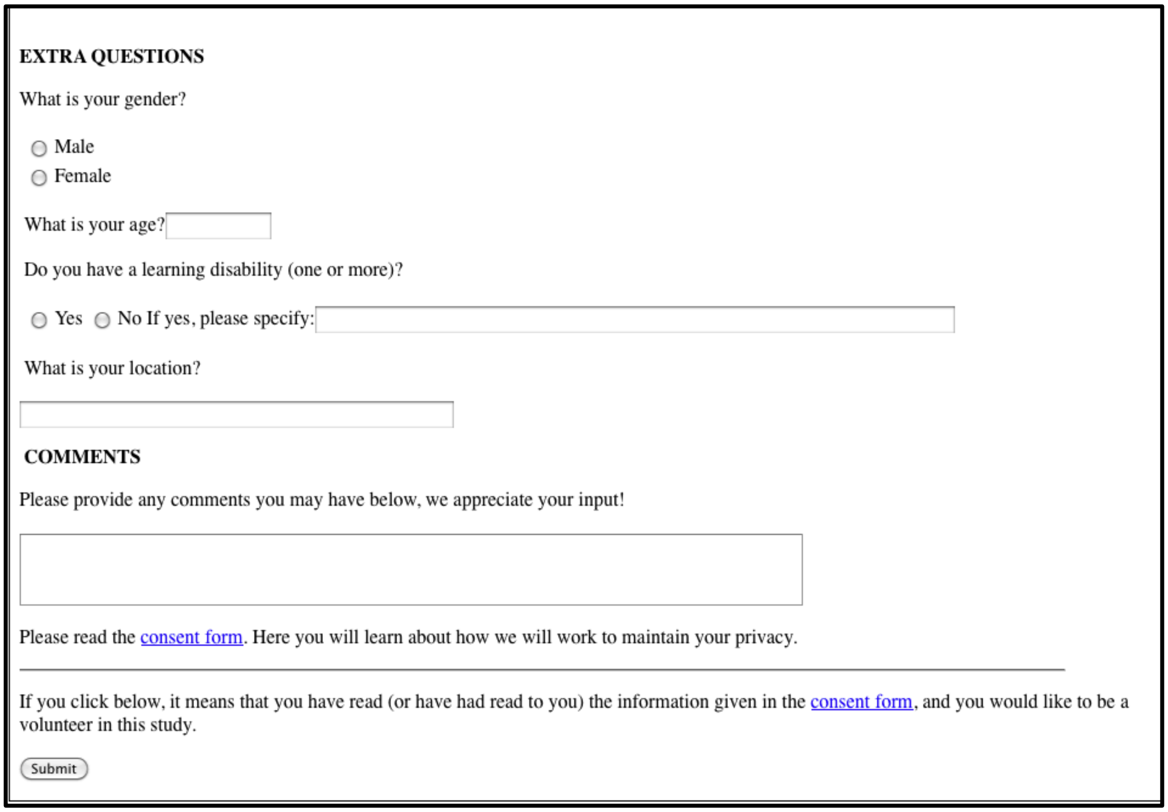
Step 11.

Step 12.

EXTRA STEPS

Please provide additional steps in the text area below. Please put one step on each line.

Figure 4.8. First half of an example of a phase 1 HIT.



EXTRA QUESTIONS

What is your gender?

- Male
- Female

What is your age?

Do you have a learning disability (one or more)?

Yes No If yes, please specify:

What is your location?

COMMENTS

Please provide any comments you may have below, we appreciate your input!

Please read the consent form. Here you will learn about how we will work to maintain your privacy.

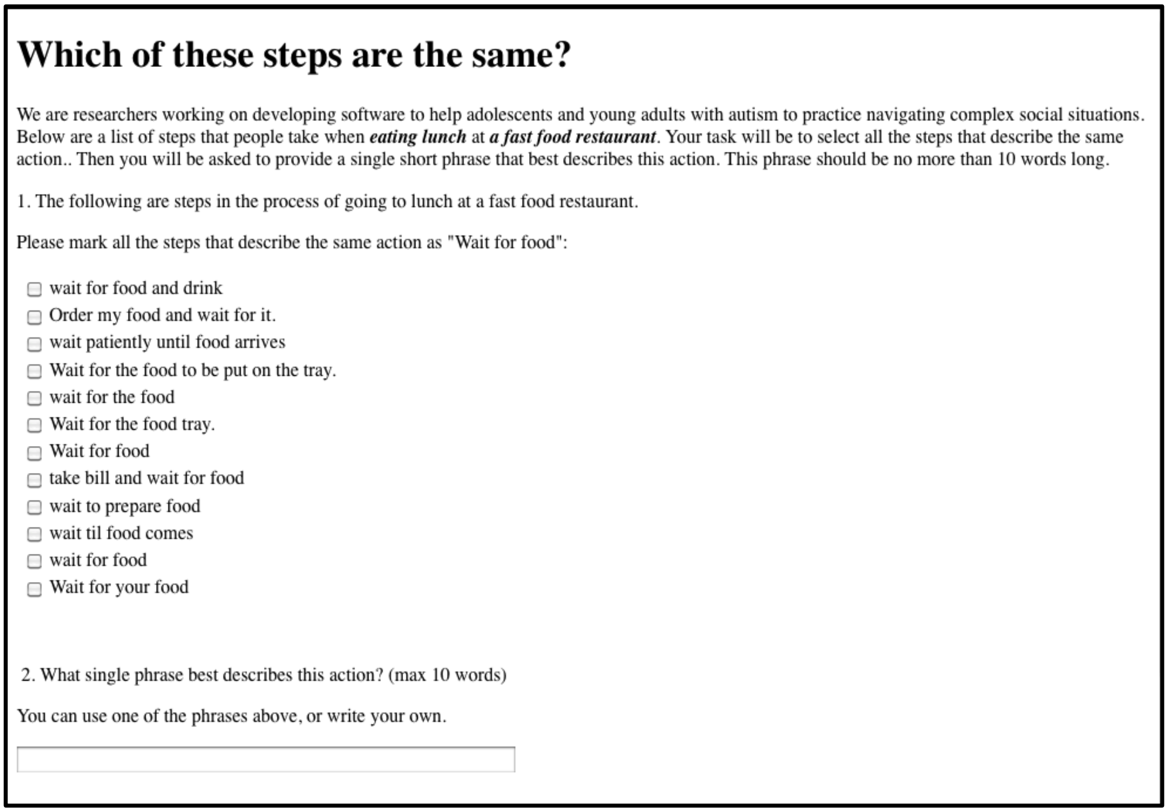
If you click below, it means that you have read (or have had read to you) the information given in the consent form, and you would like to be a volunteer in this study.

Submit

Figure 4.9. Second half of an example of a phase 1 HIT.

Phase 2: *Classify Steps*—There are many ways to say the same thing. In this phase the steps provided in the previous phase were first processed using natural language processing techniques (described in the next section) to identify steps that are potentially similar. Groups of two or more steps marked as similar were used to create a HIT that asks "Which of these steps are the same?" (Figure 4.10).

Again the task and location are those specified in the input data. The sample step is randomly selected from the group of preprocessed steps so the workers could have something to compare against. In this way, the similar steps were marked, and a classifying phrase or label was collected. The second half of this HIT, and all the HITs in the remainder of this presentation were identical to the one used in phase 1 (Figure 4.9).



Which of these steps are the same?

We are researchers working on developing software to help adolescents and young adults with autism to practice navigating complex social situations. Below are a list of steps that people take when ***eating lunch*** at ***a fast food restaurant***. Your task will be to select all the steps that describe the same action.. Then you will be asked to provide a single short phrase that best describes this action. This phrase should be no more than 10 words long.

1. The following are steps in the process of going to lunch at a fast food restaurant.

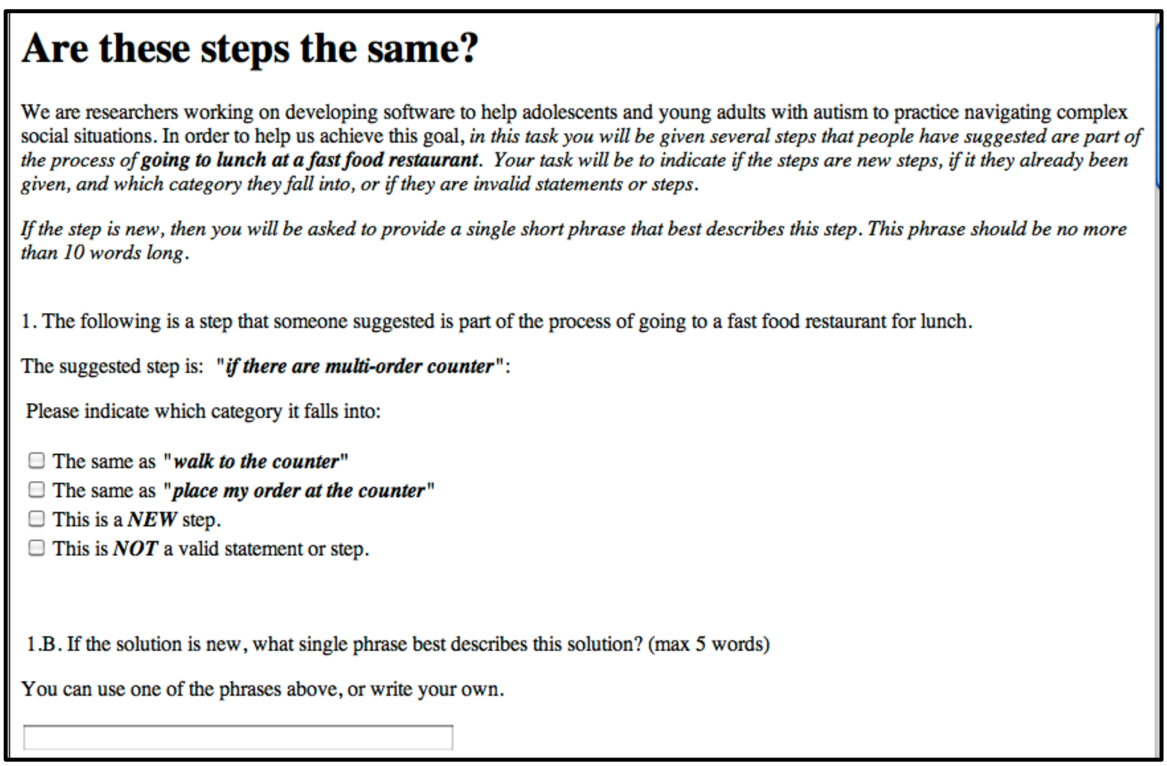
Please mark all the steps that describe the same action as "Wait for food":

- ☐ wait for food and drink
- ☐ Order my food and wait for it.
- ☐ wait patiently until food arrives
- ☐ Wait for the food to be put on the tray.
- ☐ wait for the food
- ☐ Wait for the food tray.
- ☐ Wait for food
- ☐ take bill and wait for food
- ☐ wait to prepare food
- ☐ wait til food comes
- ☐ wait for food
- ☐ Wait for your food

2. What single phrase best describes this action? (max 10 words)

You can use one of the phrases above, or write your own.

Figure 4.10. First half of an example of a phase 2 HIT.



Are these steps the same?

We are researchers working on developing software to help adolescents and young adults with autism to practice navigating complex social situations. In order to help us achieve this goal, *in this task you will be given several steps that people have suggested are part of the process of **going to lunch at a fast food restaurant**. Your task will be to indicate if the steps are new steps, if it they already been given, and which category they fall into, or if they are invalid statements or steps.*

If the step is new, then you will be asked to provide a single short phrase that best describes this step. This phrase should be no more than 10 words long.

1. The following is a step that someone suggested is part of the process of going to a fast food restaurant for lunch.

The suggested step is: "***if there are multi-order counter***":

Please indicate which category it falls into:

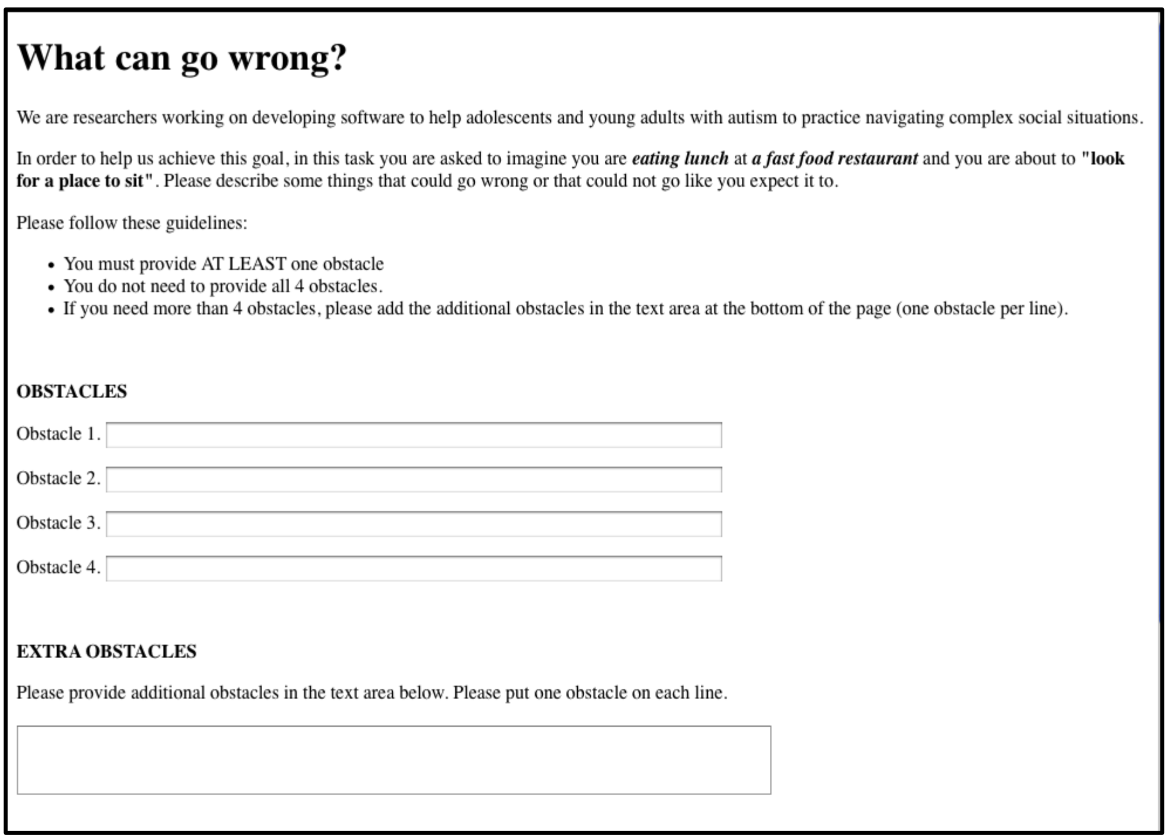
- ☐ The same as "***walk to the counter***"
- ☐ The same as "***place my order at the counter***"
- ☐ This is a ***NEW*** step.
- ☐ This is ***NOT*** a valid statement or step.

1.B. If the solution is new, what single phrase best describes this solution? (max 5 words)

You can use one of the phrases above, or write your own.

Figure 4.11. Example of one question in the second type of phase 2 HIT.

The groups that emerged that had a single step during the preprocessing, and those steps that were not marked in the classification HIT above, had to be classified separately. To achieve this classification a second template was used that asked whether a step in questions is similar to an existing category, a new solution or an invalid statement or solution. HIT workers were presented with up to 5 steps to classify at a time (one question shown in Figure 4.11). Each of the five steps had its own set of options. The categories that are provided (the first two options in the example) are chosen from among the already generated classes using similar natural language processing techniques as those used in the initial data processing (the process used to identify similar steps). In addition to removing redundancy in the data, these two HITs also served the purpose of handling the rare cases where the data provided by the worker was unusable.



What can go wrong?

We are researchers working on developing software to help adolescents and young adults with autism to practice navigating complex social situations.

In order to help us achieve this goal, in this task you are asked to imagine you are ***eating lunch*** at ***a fast food restaurant*** and you are about to **"look for a place to sit"**. Please describe some things that could go wrong or that could not go like you expect it to.

Please follow these guidelines:

- You must provide AT LEAST one obstacle
- You do not need to provide all 4 obstacles.
- If you need more than 4 obstacles, please add the additional obstacles in the text area at the bottom of the page (one obstacle per line).

OBSTACLES

Obstacle 1.

Obstacle 2.

Obstacle 3.

Obstacle 4.

EXTRA OBSTACLES

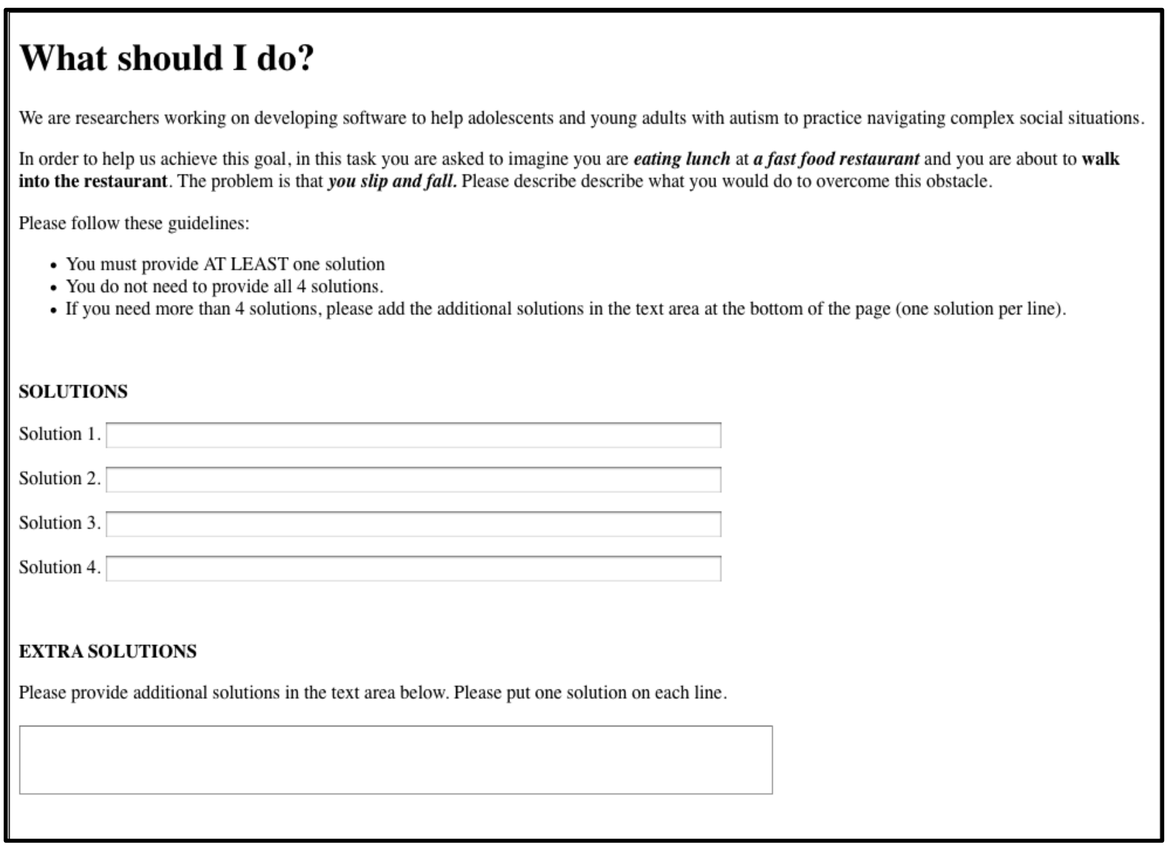
Please provide additional obstacles in the text area below. Please put one obstacle on each line.

Figure 4.12. First half of an example of a phase 3 HIT.

Phase 3: *Brainstorm Obstacles*—To gather the obstacle data, HITs were created for each class of step that emerged from phase 1 that asked participants to suggest 4 obstacles that could arise. The HIT looked like the one shown in Figure 4.12.

Phase 4: *Classify Obstacles*—HITs were created to enable the classification of the obstacle data. These HITs were very similar to those shown in phase 2 with the appropriate changes made to reflect the type of data being classified.

Phase 5: *Brainstorm Solutions*—To collect the solution data, HITs were created for each class of obstacle that emerged from phase 4. The HIT template was as follows:



What should I do?

We are researchers working on developing software to help adolescents and young adults with autism to practice navigating complex social situations.

In order to help us achieve this goal, in this task you are asked to imagine you are ***eating lunch*** at ***a fast food restaurant*** and you are about to **walk into the restaurant**. The problem is that ***you slip and fall.*** Please describe describe what you would do to overcome this obstacle.

Please follow these guidelines:

- You must provide AT LEAST one solution
- You do not need to provide all 4 solutions.
- If you need more than 4 solutions, please add the additional solutions in the text area at the bottom of the page (one solution per line).

SOLUTIONS

Solution 1.

Solution 2.

Solution 3.

Solution 4.

EXTRA SOLUTIONS

Please provide additional solutions in the text area below. Please put one solution on each line.

Figure 4.13. First half of an example of a phase 5 HIT.

Phase 6: *Classify Solutions*—HITs were created to enable the classification of the solution data. These HITs were again very similar to those shown in phase 2 with the appropriate changes made to reflect the fact that the data contains solutions to a potential obstacle.

Data Management

My approach required some backend processing to be effective. Currently, the transitions between phases and the iterations are being handled semi-automatically with the help of a backend data management and processing application that was created. In the future, it is my goal to fully automate the process using a toolkit like Turkit [Little et. al. 2009].

The three types of data that are needed are sequentially dependent on one another; steps determine potential obstacles, which determine potential solutions. Therefore, the steps needed to be collected first. As shown in the example of Phase 1 of data collection, the steps were collected as complete scripts from each worker. In order to maintain the integrity of the data after it is classified, it was critical to maintain an identifier that would allow for each step to be mapped back to the script that it came from after it has been classified. This identifier ensured that the ordering of the steps was maintained.

To this end, a Java application was written that read in the data downloaded from AMT, converted each worker's response into a Script object that is made up of Step objects, and assigned it the ID of the worker that provided the script, as an identifier (Figure 4.14). These scripts act as the starting points for the model. Once the steps are classified, each is assigned the class label that was determined by the crowd. The labels combined with the ordering that is preserved from the raw script data allows for the steps to be transformed into a usable model that shows many of the ways the particular scenario can unfold (Figure 4.15).

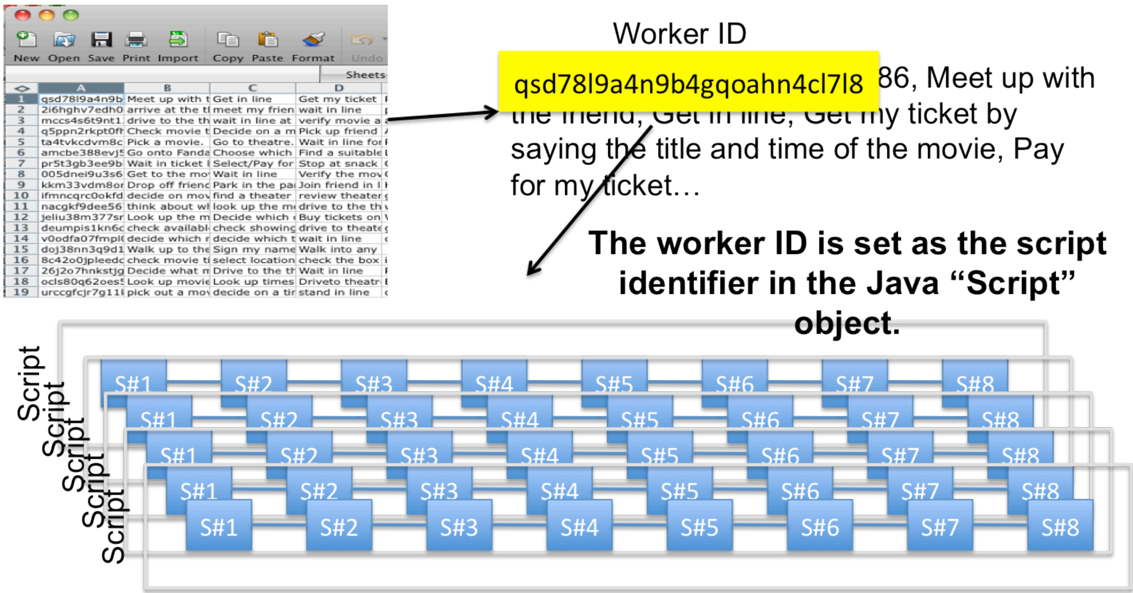


Figure 4.14. Approach for managing the step data.

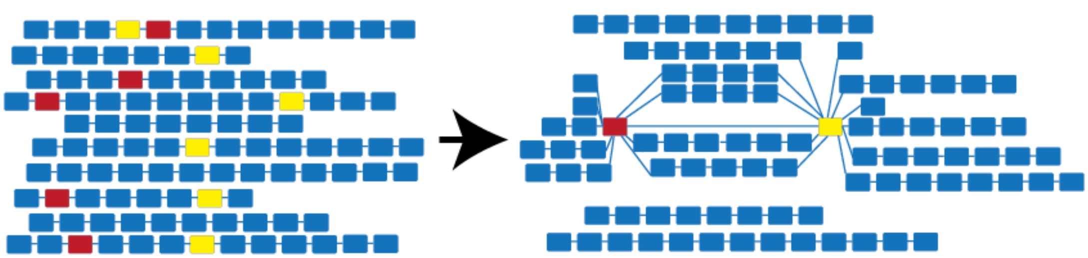


Figure 4.15. Data after two steps have been classified.

Once the steps are collected and classified, possible obstacles that can arise and suggested solutions to those obstacles can be collected. In this data, it is no longer necessary to maintain ordering in the data provided by the workers. Instead, it is only necessary to maintain a link between the step and the associated obstacles and solutions. To preserve this relationship, our data structure was designed such that each step object has 2 sub-layers; the first contains the obstacles and the next contains the solutions to those obstacles (Figure 4.16). In this way, at every point in the script it will be possible to present the author with suggestions for obstacles that may arise, and solutions to offer the student as they guide them in the process of overcoming the obstacle that is introduced.

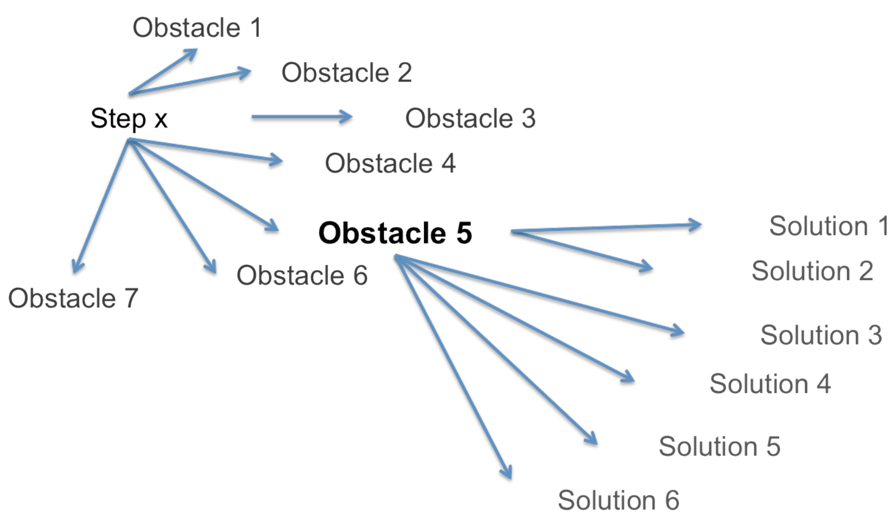


Figure 4.16. Illustration of the data structure/layering.

Data Processing to Facilitate Phase Transitions

In addition to organizing the data that is collected, the back end system was also called upon to do some processing to facilitate the transitions between the collection and classification phases. The data collection process requires that steps, obstacles, and solutions that are potentially similar be extracted from the data so that they may be presented to workers in manageable numbers for classification. This made it necessary to add some basic natural language processing functionality. The Stanford Parser's part-of-speech tagger was used to facilitate this task [Stanford Parser, 2011]. This software is freely available as an easy to use Java Plug-in, and provides part-of-speech tagging with degrees of accuracy that are appropriate for our purposes (around 89% for unknown words).

Phrases describing actions can be characterized based on the verb and noun that they contain. For example, extracting the verb and noun from the phrase “wait in line” (verb: “wait,” noun: “line”) gives a good characterization of the action being described. As a result, it was natural to attempt to filter the steps that were collected based on these two parts of speech. The following procedure was used:

1. All the nouns and verbs in the data were extracted and ranked based on frequency of occurrence.
2. The highest-ranking verb was paired with the highest-ranking noun that it co-occurs with in the data until all the verbs and nouns are paired.
3. The co-occurring verb-noun pairs were ranked based on their frequency of occurrence.
4. All the steps containing a particular pair were extracted, beginning with the highest ranking, until all the steps are grouped.

An example of this process from the data that was collected for a script about

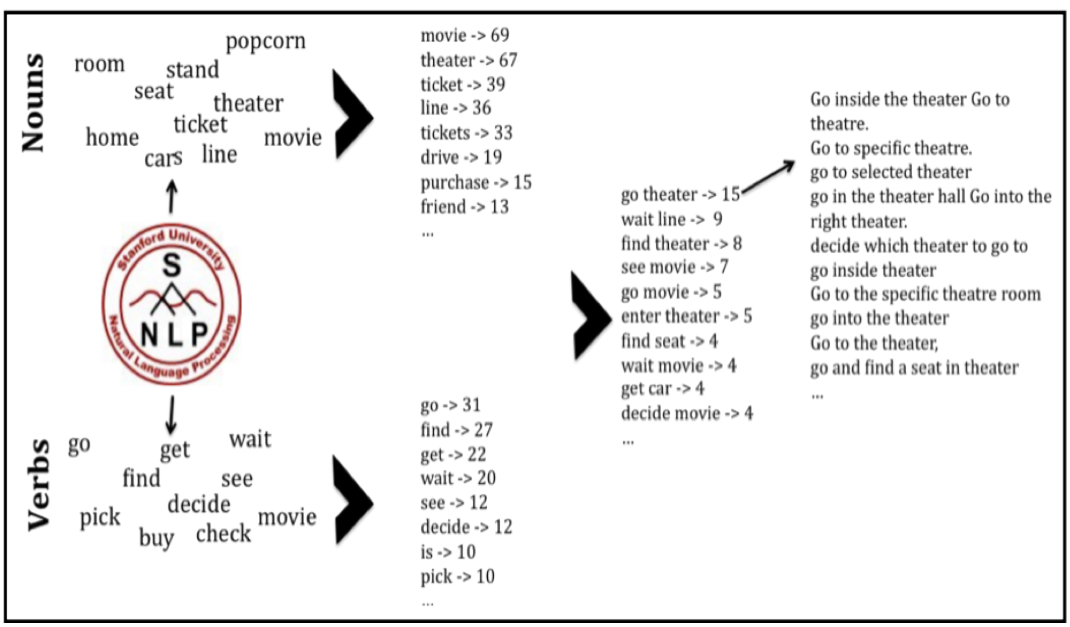


Figure 4.17. Example of the data processing conducted to extract steps to present to AMT workers

"going to a movie with a friend at a theater" can be seen in Figure 4.17. It shows that this approach did indeed provide usable groups in the classification phase. This was also the case in the solution data, as these phrases described actions as well. For the obstacle data, however, filtering the data based on adverb-verb and adjective-noun pairs was more effective, as these phrases tended to be more descriptive (e.g. "I get the wrong change").

The example in Figure 4.17 shows that this approach does produce errors. For instance, the tagger chose to tag the words "purchase" and "drive" as nouns, when in the data they more often appeared as verbs. In addition, words that were spelled in different ways, or incorrectly (eg. "theater" and "theatre") were not seen by the application as being the same. We proceeded with the groupings as the application outputted them, and found that the crowd managed to handle these errors, and good classifications were achieved despite these discrepancies. Another example of the output of this process, but from the restaurant data, was presented earlier in Figure 4.10. In that example it can be seen that the noun was "food" and the verb was "wait."

Data Collection Using AMT

Data was collected on AMT for three social situations:

Scenario 1:

- Task: "eat lunch"
- Location: "a fast food restaurant"
- Start: "when you enter the restaurant"
- End: "when you leave the restaurant"

Scenario 2:

- Task: "going to a movie with a friend"
- Location: "a movie theater"

- Start: “when you enter the theater”
- End: “when you leave the theater”

Scenario 3:

- Task: “going to the doctor because you are not feeling well”
- Location: “doctor’s office”
- Start: “when you enter the waiting room”
- End: “when you get back in your car after you have seen the doctor”

To date, the complete model has only been built for the restaurant scenario. For the movie scenario, step and obstacle data has been collected, and for the doctor scenario only step data has been collected. Based on several pilots, the following numbers were used:

Scripts: 40 workers provided responses, and each was paid $.05 (i.e. 40 scripts were collected).

Obstacles: 10 workers provided suggestions for obstacles for each classified step, and each was paid $.01.

Solutions: 10 workers provided suggestions for solutions for each classified obstacle, and each was paid $.01.

Classification: each preprocessed group of steps, obstacles, and solutions was presented to 2 workers and using a voting approach a third worker was called upon to break any ties. Each worker was paid $.01.

Due to the nature of the data (the quantity increases with each type) over 1000 responses were collected and the cost increased substantially with each phase. While the step data cost under $20 to collect and classify, the solution data cost more than $60 to collect and classify. In total the data to build the entire restaurant model (all 6 phases) cost approximately $110 to collect. While there is certainly room for improvement in terms of efficiency, I was very pleased to find that workers gave quite thorough responses despite the fact that they were being paid very little for each task. I believe that the presentation of the motivation and goals of the research played an important role in this outcome. I requested comments from all the workers, and those comments indicated that workers had acknowledged the goals of the data collection and kept them in mind when they were formulating their responses. This leads me to believe that it may be possible to solicit some or all of this data from workers for free. I intend to explore this possibility and other improvements and modifications to this data collection approach in the future.

To get an idea of the demographics of our workers I randomly selected half of the HIT results and compiled the responses. The information was as follows: the average age was 33.5 and 57% of workers were female. The location information indicated that 45% of workers were from US/Canada, 40% from India, and 6% were from Europe. Most interestingly 11% indicated that they either had experience with autism in their personal or professional life, or had autism or another disability that causes difficulty with social skills themselves (e.g. Asperger's Syndrome).

Evaluation of Script Models using AMT

While the preliminary results of the data collection were promising, I realized that it is necessary to both systematically and objectively evaluate the model, to obtain a preliminary measure of how useful the data would be to someone who is trying to author a social skills instructional module. Before going to the trouble of creating the authoring environment and evaluating it with potential users, I wanted to determine the completeness and utility of the crowdsourced social script models. My experiences, and those documented in the literature [Heer and Bostock, 2010, Kittur, 2008] led me to believe that the crowd could be an effective proxy for a participant in a more traditional user study. Indeed, the nature of the input I was seeking made the crowd a better source than traditional participants because responses could be solicited efficiently from many different people.

Method

The three questions I sought to answer were:

1) Are the steps collected reasonable?

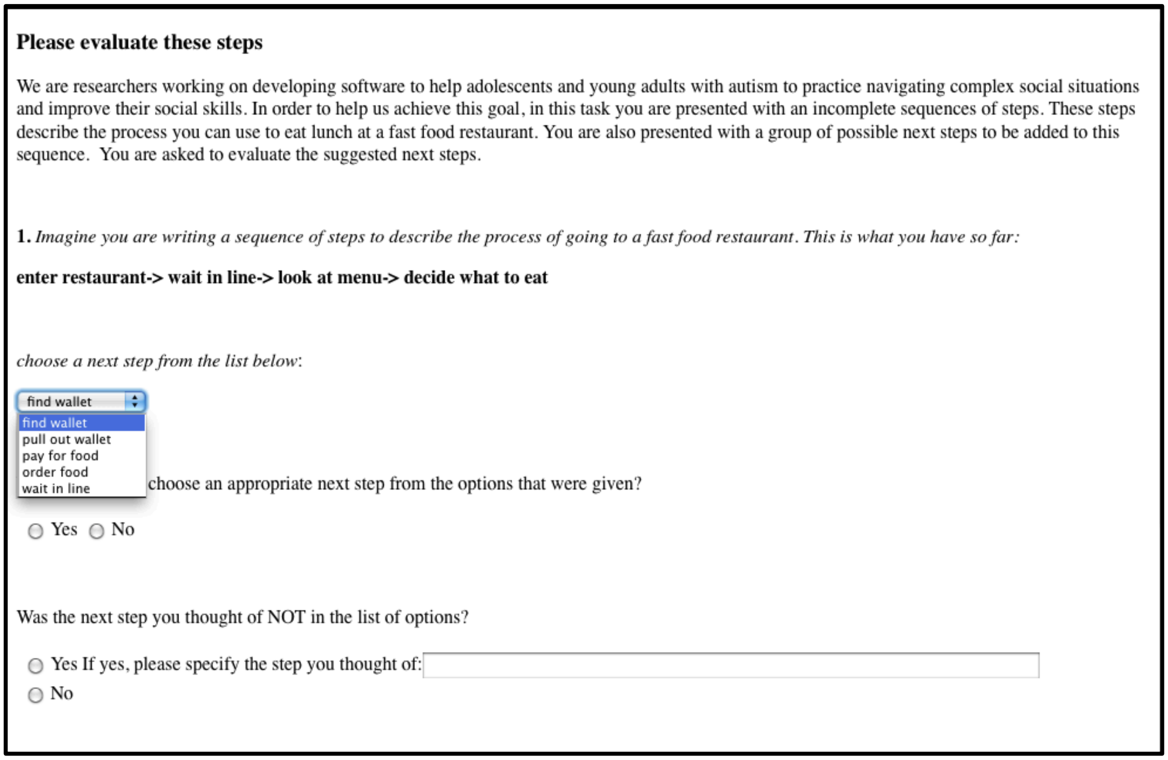
2) Are the obstacles presented reasonable?

3) Are the solutions provided reasonable?

The definition of “reasonable” was operationalized to gauge whether the set of possibilities generated by the crowd represent a set of options that is meaningful and varied enough to be useful to a potential author.

Evaluation of the Step Data

Workers were presented with a randomly generated initial subsequence of steps from our restaurant model and asked to choose a next step for the sequence from a list of available options (Figure 4.18). That list was exactly the list generated from the crowd, as described above. Workers were then asked five questions, the first of which requested that workers indicate whether or not they were able to find an appropriate next step from the options that were given. Workers were then asked if the step that they thought of was not included in the list of options. These questions were meant to get at the completeness of the data in the model. Workers were also asked to provide quantitative subjective evaluations of the data in three dimensions; richness, unexpectedness, and level of inclusion of social interaction in the steps (Figure 4.19). This allowed us to gain additional insight and assess the meaningfulness of the data.



Please evaluate these steps

We are researchers working on developing software to help adolescents and young adults with autism to practice navigating complex social situations and improve their social skills. In order to help us achieve this goal, in this task you are presented with an incomplete sequences of steps. These steps describe the process you can use to eat lunch at a fast food restaurant. You are also presented with a group of possible next steps to be added to this sequence. You are asked to evaluate the suggested next steps.

1. *Imagine you are writing a sequence of steps to describe the process of going to a fast food restaurant. This is what you have so far:*

enter restaurant-> wait in line-> look at menu-> decide what to eat

choose a next step from the list below:

find wallet
find wallet
pull out wallet
pay for food
order food
wait in line

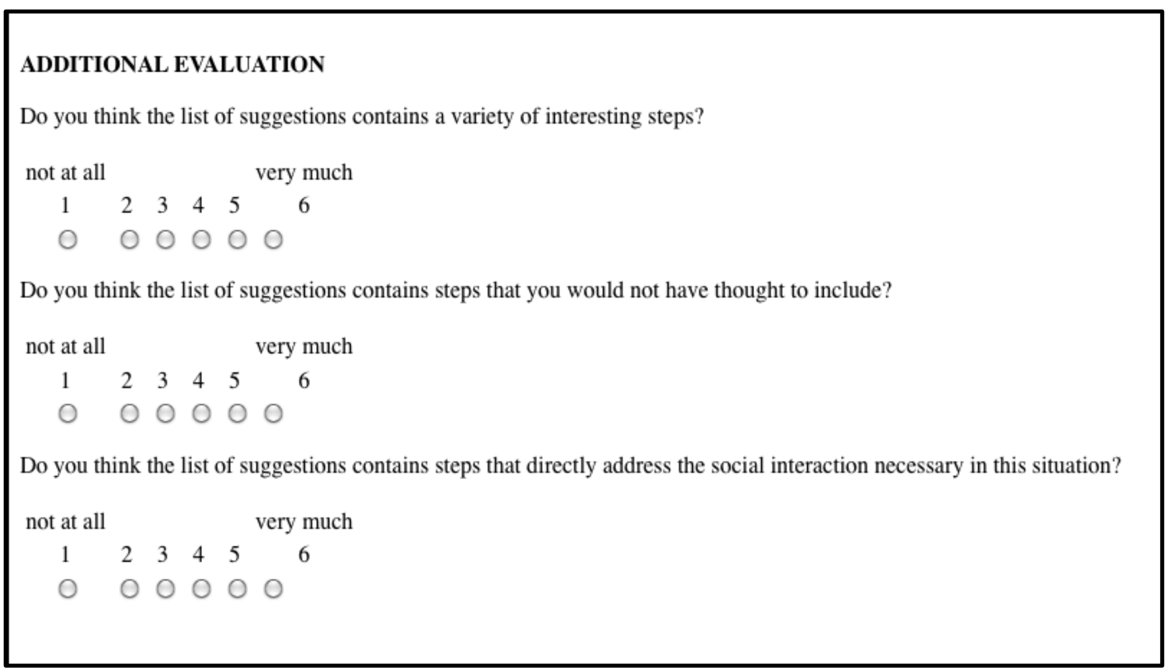
choose an appropriate next step from the options that were given?

Yes No

Was the next step you thought of NOT in the list of options?

Yes If yes, please specify the step you thought of:
No

Figure 4.18. Example of a portion of a step evaluation HIT.



ADDITIONAL EVALUATION

Do you think the list of suggestions contains a variety of interesting steps?

not at all very much
1 2 3 4 5 6

Do you think the list of suggestions contains steps that you would not have thought to include?

not at all very much
1 2 3 4 5 6

Do you think the list of suggestions contains steps that directly address the social interaction necessary in this situation?

not at all very much
1 2 3 4 5 6

Figure 4.19. Example of how the additional evaluation questions appeared in the step HITs.

Five HITs were generated using this template, each depicting a different subsection of the data in the model (5 different randomly generated partial sequences, and the possible next steps for 5 different nodes in the model). Twenty workers were asked to respond to each HIT and each was paid $0.01. In this way, 100 evaluations of the step data were collected for $1.

Evaluation of the Obstacle Data

In order to evaluate the obstacle data, workers were again presented with a randomly generated initial subsequence of steps, and asked to introduce an obstacle from a list of suggestions. They were then asked the same 5 questions appropriately adapted for the type of data being evaluated. Figures 4.20a and 4.20b present an example of one of the HITs created to evaluate the obstacle data in the restaurant model.

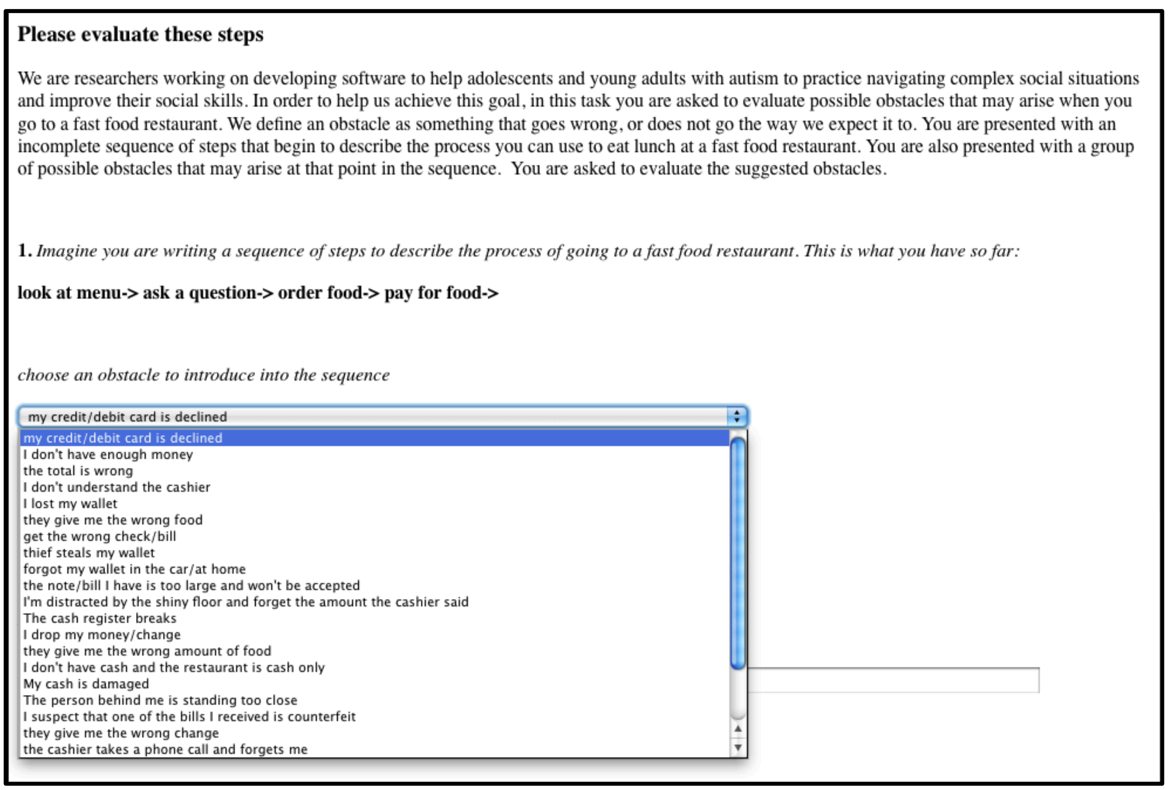
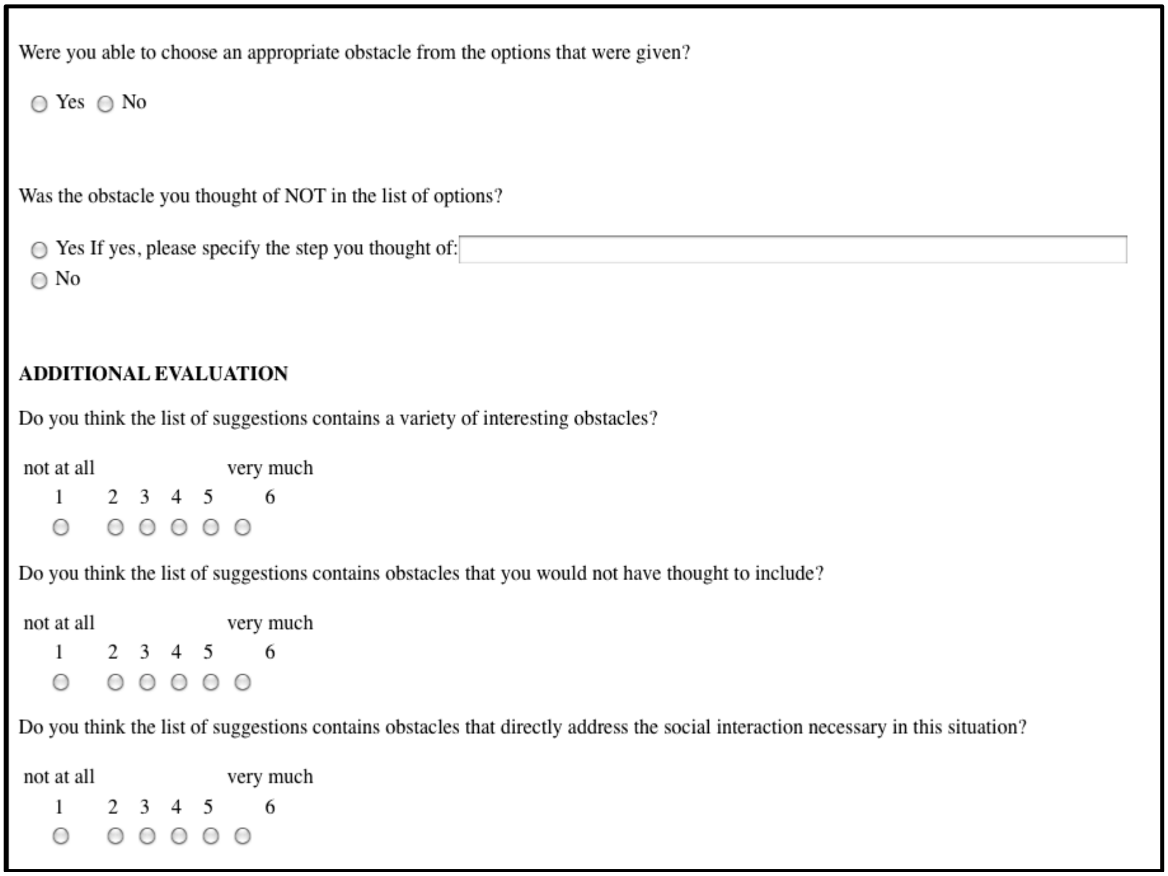


Figure 4.20a. Example of an obstacle evaluation HIT.



Were you able to choose an appropriate obstacle from the options that were given?

○ Yes ○ No

Was the obstacle you thought of NOT in the list of options?

○ Yes If yes, please specify the step you thought of:

○ No

ADDITIONAL EVALUATION

Do you think the list of suggestions contains a variety of interesting obstacles?

not at all very much

1 2 3 4 5 6

Do you think the list of suggestions contains obstacles that you would not have thought to include?

not at all very much

1 2 3 4 5 6

Do you think the list of suggestions contains obstacles that directly address the social interaction necessary in this situation?

not at all very much

1 2 3 4 5 6

Figure 4.20b. Example of the evaluation questions in an obstacle evaluation HIT.

As in the step data, 5 HITs were generated using this template depicting distinct subsets of the obstacle data and a total of 100 responses were collected (20 each).

<u>Evaluation of the Solution Data</u>

Evaluation HITs were created for the solution data using a similar approach. Below is an example of one such HIT:

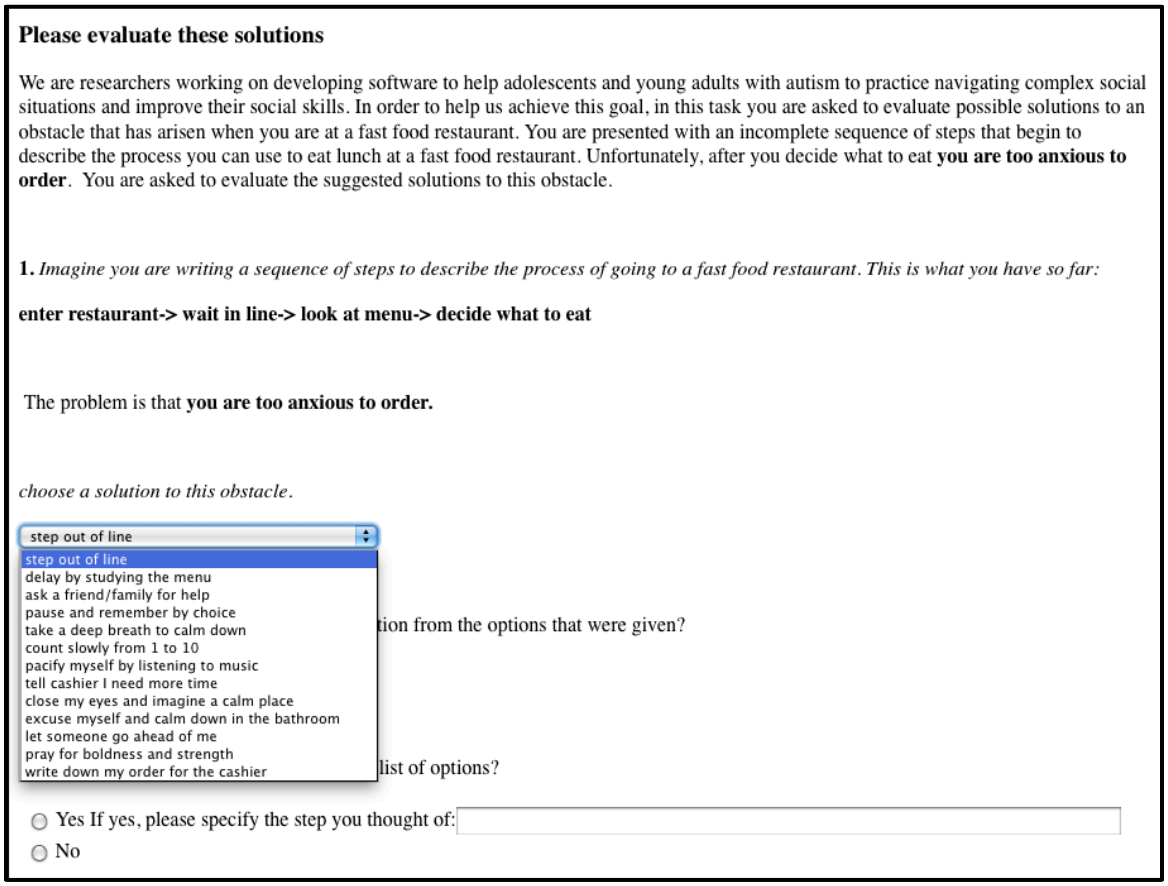


Figure 4.21. Example of the first portion of a solution evaluation HIT.

Again 5 HITs were generated using this template depicting distinct subsets of the solution data, and workers were asked the same five questions about the solution data. In this way 100 participants evaluated each type of data, and a total of 300 evaluations were collected for $3.

Results and Discussion

The results of the evaluation HITs can be seen in Tables 4.2 and 4.3. In response to the first question, 92% of the 300 participants indicated that they were able to find an appropriate option from the suggestions that were provided. In addition, only 5% indicated that the option they thought of was not listed among the suggestions.

Table 4.2. Participant responses regarding their ability to find an appropriate step in the options.

	Was an appropriate step/obstacle/solution available?	**Was the option they thought of not in the list?**
Steps	89/100	6/100
Obstacles	95/100	10/100
Solutions	95/100	4/100
Average	*92%*	*5%*

Table 4.3. Participant quantitative subjective evaluations of the data (out of 6)

	Richness	**Unexpectedness**	**Social Interaction**	***Average***
Steps	4.41	3.22	4.43	*4.02*
Obstacles	5.05	4.3	4.7	*4.68*
Solutions	4.98	4.18	4.91	*4.69*
Total	*4.81*	*3.91*	*4.68*	4.46

The subjective data provided additional insight on the quality of the data. As shown in Table 10, the obstacles received the highest richness scores, averaging to 5.05 out of 6, and the solutions received the highest social interaction score, averaging to 4.91 out of 6.

The dimension that received the lowest score was the unexpectedness of the steps. As can be seen in the examples, the number of suggestions for obstacles and solutions provided by the crowd was much larger than that for steps. It is obvious that there are many more ways that things can go wrong than there are correct ways to navigate a situation. Furthermore, the approach that was used for collecting the obstacle and solution data (i.e., asking for 4 suggestions) exposed more of these possibilities. This outcome is desirable as it gives the author many suggestions to choose from, and consequently many ideas for situations to prepare their child.

Lastly, one concern that autism experts who consulted on this research had was related to the ability of this approach, and in essence the crowd, to come up with obstacles and solutions that are relevant to the autism community. Individuals with autism struggle with most of the same obstacles individuals without autism struggle with. The difference is that they also struggle with additional things that may not affect individuals without autism (e.g., sensory issues related to light and noise). As can be seen from the suggestions above (Figure 4.20a: e.g. “I'm distracted by the shiny floor and forget the amount”, “The person behind me is standing too close” etc.), I found that many relevant autism related obstacles did emerge. The demographic data and comments that were collected in each HIT allowed us to deduce that the workers, mentioned earlier, who had experience with autism in their personal or professional life were the ones who mostly supplied these suggestions.

Conclusion and Contributions

The social world that most of us navigate effortlessly can prove to be a perplexing and disconcerting place for individuals with autism. Currently there are no models to assist authors as they create customized social script-based instructional modules for a particular child. In this chapter, I systematically verify that the data collected through the use of this approach enables the creation of models for complex and interesting social scenarios, possible obstacles that may arise in those scenarios, and potential solutions to those obstacles. I also presented a preliminary evaluation of the data in the model that was created. Overall, workers’ responses were positive and provide confirmation that the model could provide good suggestions for a potential author. I believe that human input is the natural way to build these models, and in so doing create valuable assistance for those trying to navigate the intricacies of a social life.

Now that the approach to building the knowledge base of the system was developed and tested, and a model of the restaurant scenario was built, the next step was to build a prototype of the authoring tool, and put the data into the hands of real authors. The technology design and development is described in the next chapter.

In summary, this portion of the thesis addressed research question two, which was:

What is a mechanism for generating rich models of social knowledge that are consistent with the *obstacle-based branching* approach to problem solving and can be used to provide scaffolding for the authoring of social skills instructional modules?

Contributions: The development of an approach to building models of social knowledge that can be dynamically created and expanded using crowdsourcing.

CHAPTER 5

REACT: FACILITATING THE AUTHORING OF MULTIMEDIA SOCIAL PROBLEM SOLVING SKILLS INSTRUCTIONAL MODULES

REACT is an authoring tool designed to help parents, teachers, and other caregivers to create Refl-ex instructional modules. The goal is to not only facilitate the authoring of branching stories, but also to use the crowdsourcing approach, described in Chapter 4, to provide the authors with support throughout the authoring process. Once the approach to building the system's knowledge base had been implemented and preliminarily verified it was time to develop a prototype of the authoring tool and evaluate it with real authors.

REACT is designed to augment existing tools like Stories2Learn (described in chapter 2) in 2 ways: 1) by facilitating the creation of branching stories in which an obstacle arises and the student is given options for ways to over come the obstacle; and 2) by providing the author with suggestions for steps to include, obstacles to introduce and solutions to suggest in the module they are authoring. In order to evaluate the usability and utility of each functionality effectively, with the help of another student[3], two versions of the authoring tool were created, one that only facilitated the creation of the branching structure, and a smart version that additionally supported the authoring process by providing the author with suggestions. In addition to the two versions of the authoring

---

[3] Husayn Versee

tool, a player had to be created to allow the authors to preview the story they created and the experts to view the module when evaluating it. Following is a description of the user interface of each of these prototypes.

The REACT Prototype

The tool is designed such that the role of the user is to act as the author of the modules, and contribute social skills expertise, and knowledge of the particular needs of the student. I assume that, the authors have social problem solving skills expertise. In REACT the author is provided with an interface that facilitates the creation of the branching structure, and prompts them to input the content. As described earlier (chapter 3), the instructional modules are made up of four different screens; the narrative screens that present the scenario, the decision screens where the student chooses a solution, the rethink screens were the student is explained the consequences of an unproductive or counterproductive solution, and the reflection screen where the student recreates the story. To facilitate the evaluation of a strictly *obstacle-based branching* scenario, in this thesis I did not address the creation of the rethink screens and the reflection screens, therefore the prototype of REACT supports the creation of two types of screens; the narrative screens and the decision screens.

Figure 5.1 shows that a minimalist approach was used when designing the interface of the REACT prototype. Since the participants would be creating a story about going to a fast food restaurant, a generic image was chosen to depict the scene. The author simply has to type their step text in the text area and press the green button to proceed to the next step. To allow for the evaluation of the text content on its own, in this prototype the interface did not allow for the addition of audio or for changing the image.

Figure 5.1. User Interface as it appeared when creating a narrative page.

Figure 5.2. User Interface as it appeared when adding an obstacle.

Once the author has reached a point in the module where they would like to introduce an obstacle, they need only click the "Add an Obstacle" button. The button takes them to a screen where they are prompted to introduce the obstacle (Figure 5.2). As can be seen in Figure 5.2, throughout the authoring process the author is shown their progress in the authoring process (i.e. the story they have written so far) in the left pane of the user interface, we call this the outline. The step they are currently working on is highlighted in yellow.

After they have added the obstacle, the author is immediately taken to the screen in which they are able to create the decision page (Figure 5.3). On this page, the author is asked to provide 3 solutions to the obstacle he or see has introduced. They are reminded of the obstacle both on the main pane and in the outline. The author can choose to input all the solutions and then fill in each of the corresponding branches, or add a solution and fill in its corresponding branch before adding the others. The solutions and their branches can also be added in any order.

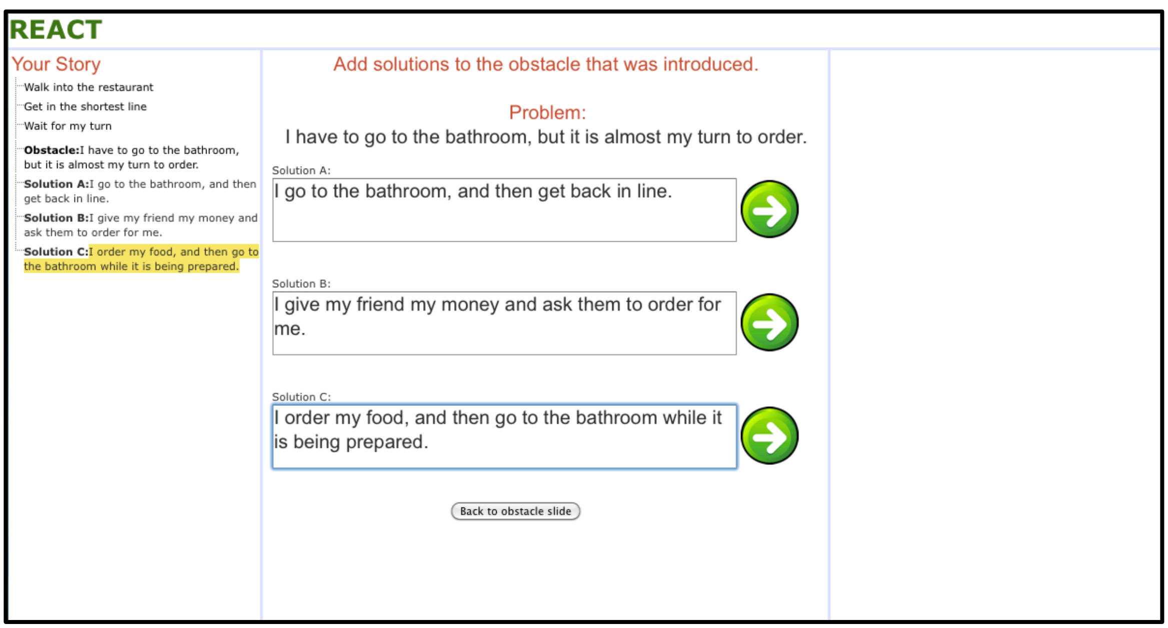


Figure 5.3. User Interface as it appeared when creating a decision page.

When the author chooses to proceed to fill in the branch, they are again asked to input the steps in a text area (Figure 5.4). In order to help the author to keep track of where they are in their story, in addition to highlighting the step being written or edited, the branch that is being populated is expanded and the solution text is featured in red. When the author is done filling in the branch they can either click on the “fill next solution” button, or click on the solution whose branch they want to fill in the outline.

The last functionality that was important to include, was editing capabilities. For this reason, the authoring tool is designed such that throughout the authoring process the author is able to return to any part of their story they would like, and edit or delete it. To return to the previous screen they can use the green back arrow, or they can navigate to any screen in the entire story by simply clicking on the step they want to edit in the outline (Figure 5.5).

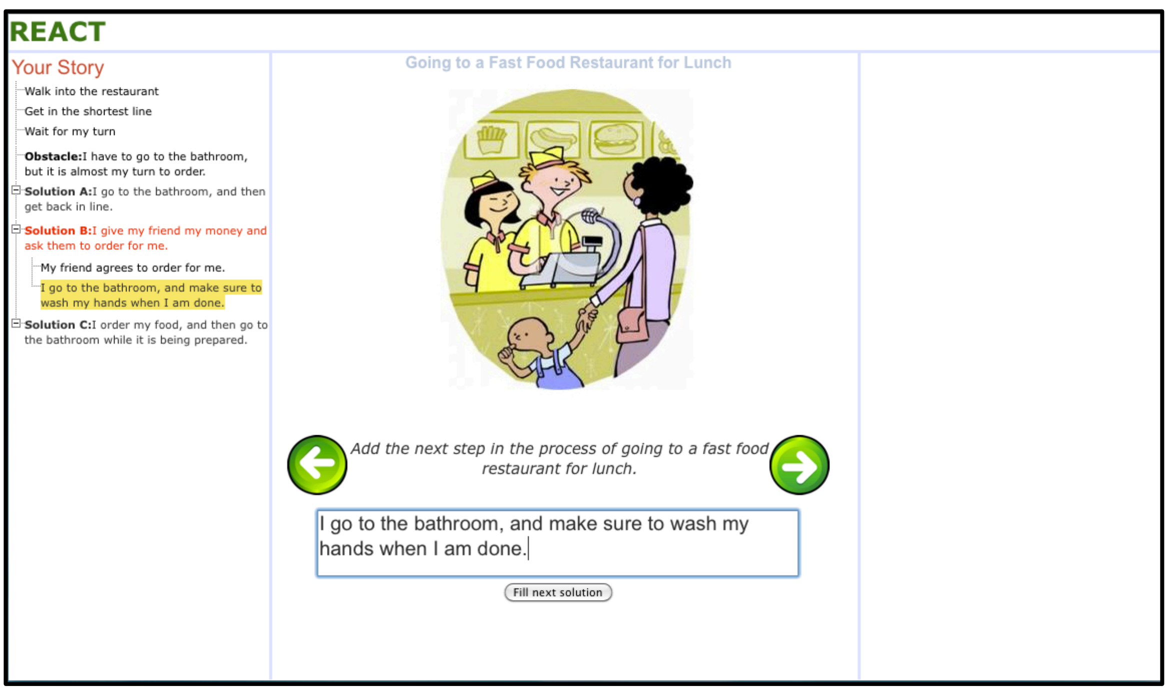


Figure 5.4. User Interface as it appeared when filling in the branch of solution B.

Figure 5.5. User Interface as it appeared when editing the module.

The Smart REACT Prototype

It can be very intimidating for an author to be presented with a blank screen, as in figure 5.1, and asked to input content. In the smart REACT prototype I try to make the process as easy as possible for the author. I do this by offering the author several suggestions for text they can use. The smart prototype generates these suggestions using the models in the knowledge base that were collected from the crowd via Mechanical Turk (Chapter 4). The layered structure of the model that was developed made it possible to easily provide authors with step suggestions, relevant obstacle suggestions at each step, and possible solutions to the chosen obstacle. To enable the evaluation of the suggestions, the interface is identical to the original prototype with the only difference being the addition of a suggestions pane to the right.

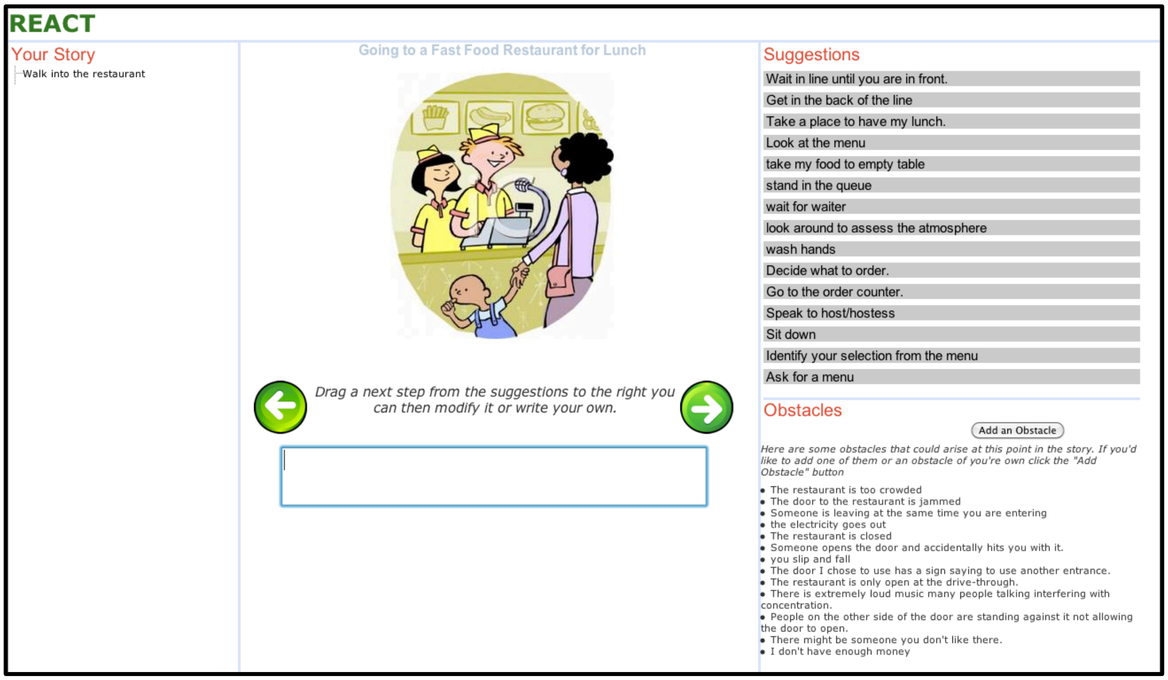


Figure 5.6. User Interface with suggestions for next steps and obstacles.

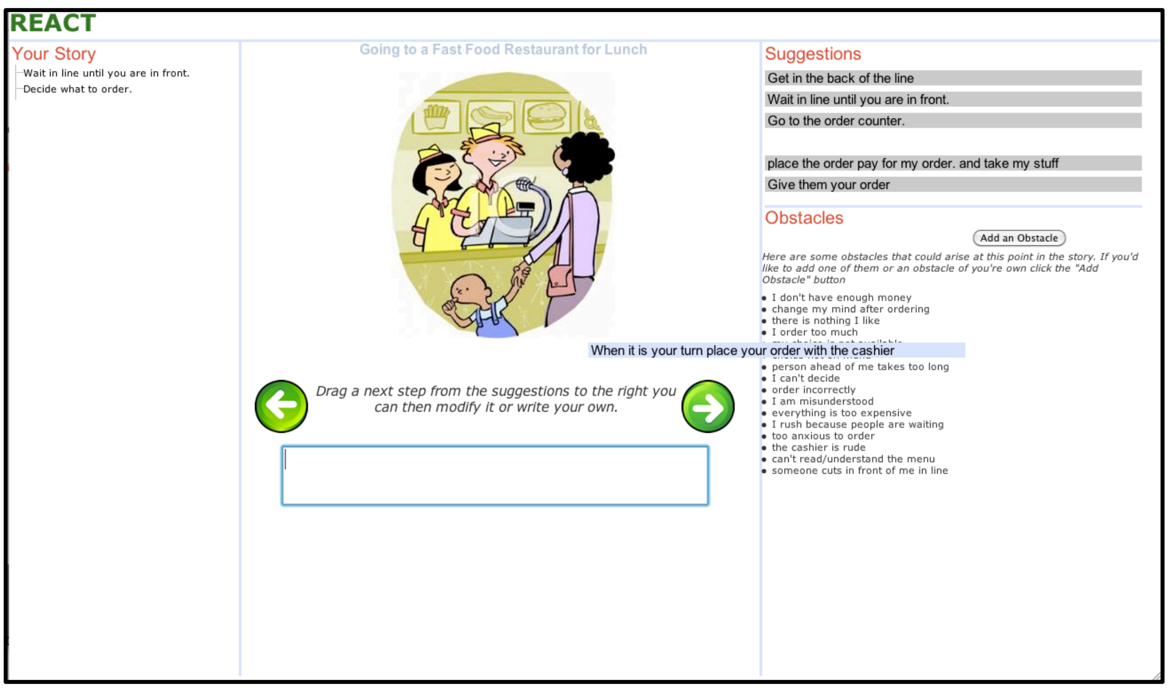


Figure 5.7. Suggested step being dragged to the text area.

At every step in the authoring process the smart prototype provides the author with suggestions for steps to include next and obstacles that could arise at this point in the story (Figure 5.6). The author can choose to either drag the suggestion to the text field and modify it as they see fit (Figure 5.7), or use it as inspiration for their own text. The idea being that the author has intimate knowledge of the needs of the student for whom the story is being created; therefore they can, in theory, choose from and adapt the suggestions such that they become personalized for the target student.

The authors are provided with the suggestions for possible obstacles throughout the authoring process. In this way the author sees more options for possible obstacles, and can use the suggestions to help them think of obstacles he or she would like the student to practice overcoming, or that represent situations they know the student to struggle with. When the author would like to introduce an obstacle, they can click the “Add an Obstacle” button to add the obstacle. On the screen they are taken to (Figure 5.8) the obstacle suggestions become draggable.

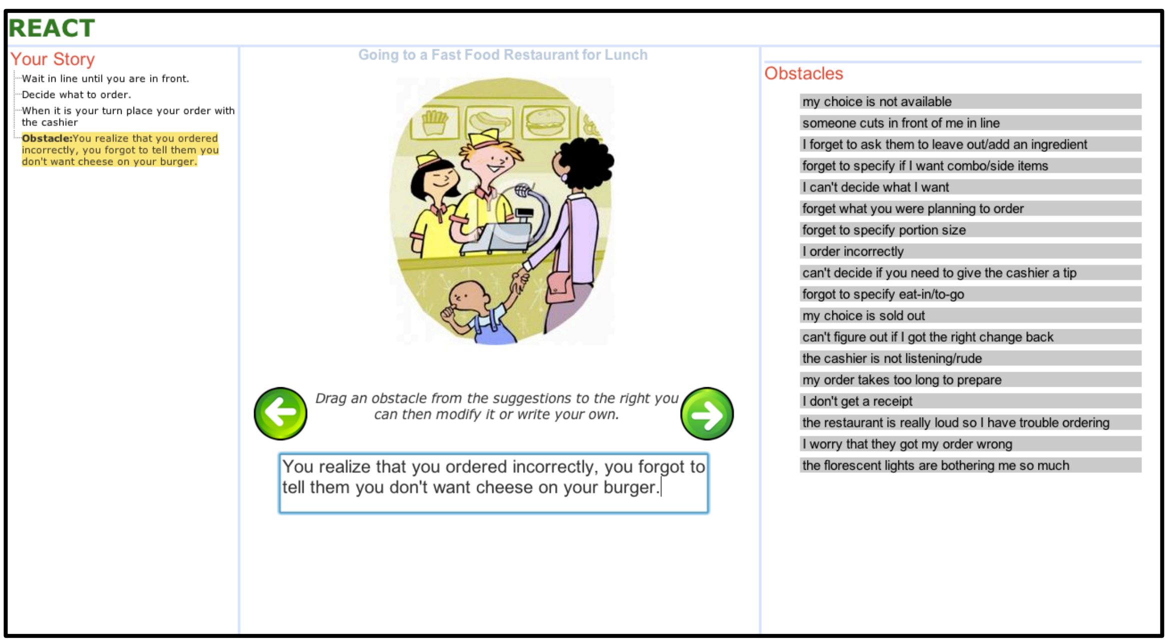


Figure 5.8. User interface providing suggestions when adding an obstacle.

After the obstacle has been introduced the author is taken to the same decision authoring page described earlier. In the smart prototype, however, the system provides suggestions for possible ways to overcome the obstacle that has been introduced (Figure 5.9). Again these suggestions are draggable.

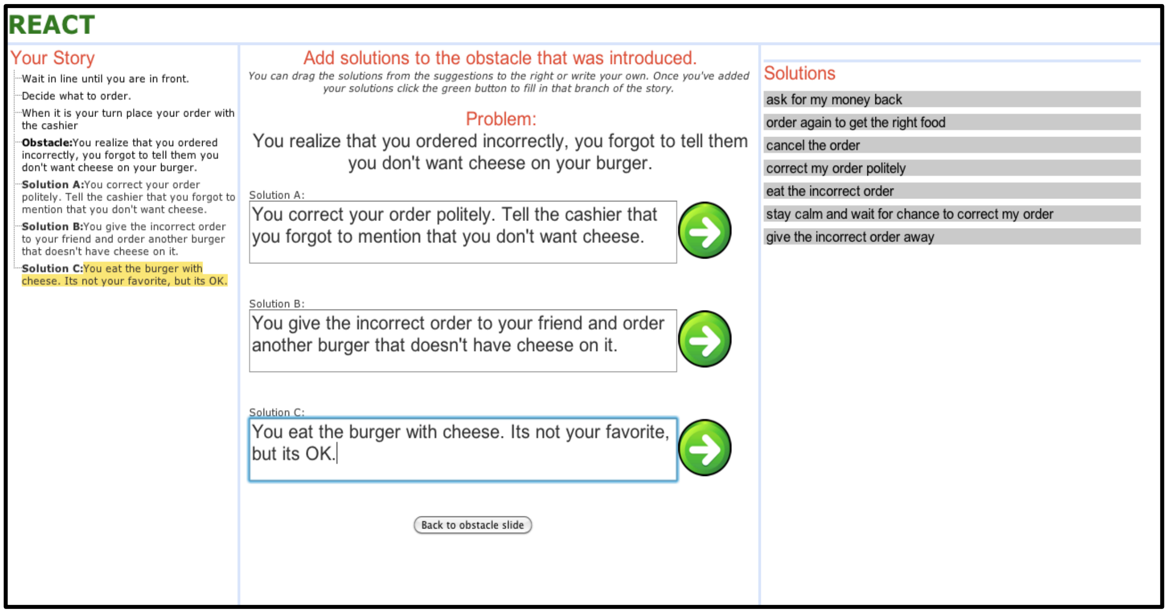


Figure 5.9. User interface providing suggestions for solutions.

Figure 5.10. User interface providing suggestions for steps to include in branch A.

The last difference between the first prototype and the smart prototype is that when the author begins to fill in the branches they are again given suggestions for steps to include in the story (Figure 5.10). These suggestions start with the possible next steps for the last step that was added to the story before the obstacle was introduced.

Figure 5.11. Introductory narrative page in the REACT player.

The REACT Player

It was also necessary to develop a player to allow for the completed module to be viewed by the author and the expert that would later be evaluating the module. To this end, software was created that played the author's story. Again the interface was straightforward, the major change being than the text was no longer editable (Figures 5.11- 5.14). In order to allow for easy viewing of the entire story, the player allows the user to show the outline in the left pane as it does in the authoring tool (Figure 5.14). The outline is hidden by default (Figures 5.11- 5.13).

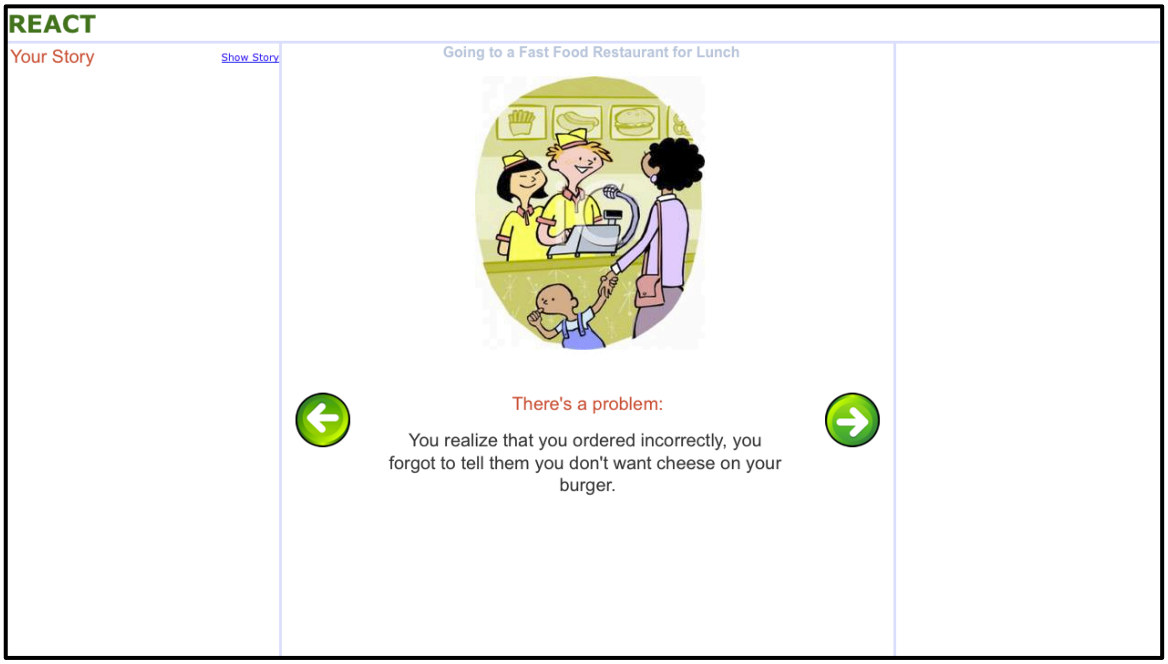


Figure 5.12. Introduction of the obstacle in the REACT player.

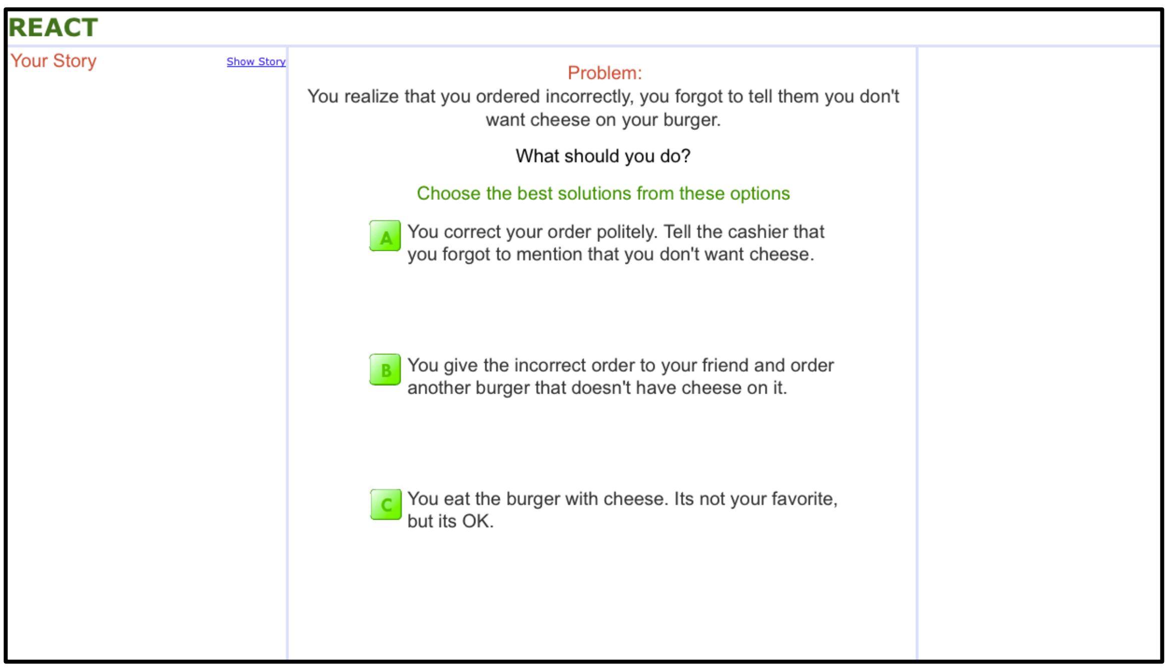


Figure 5.13. The decision page in the REACT player.

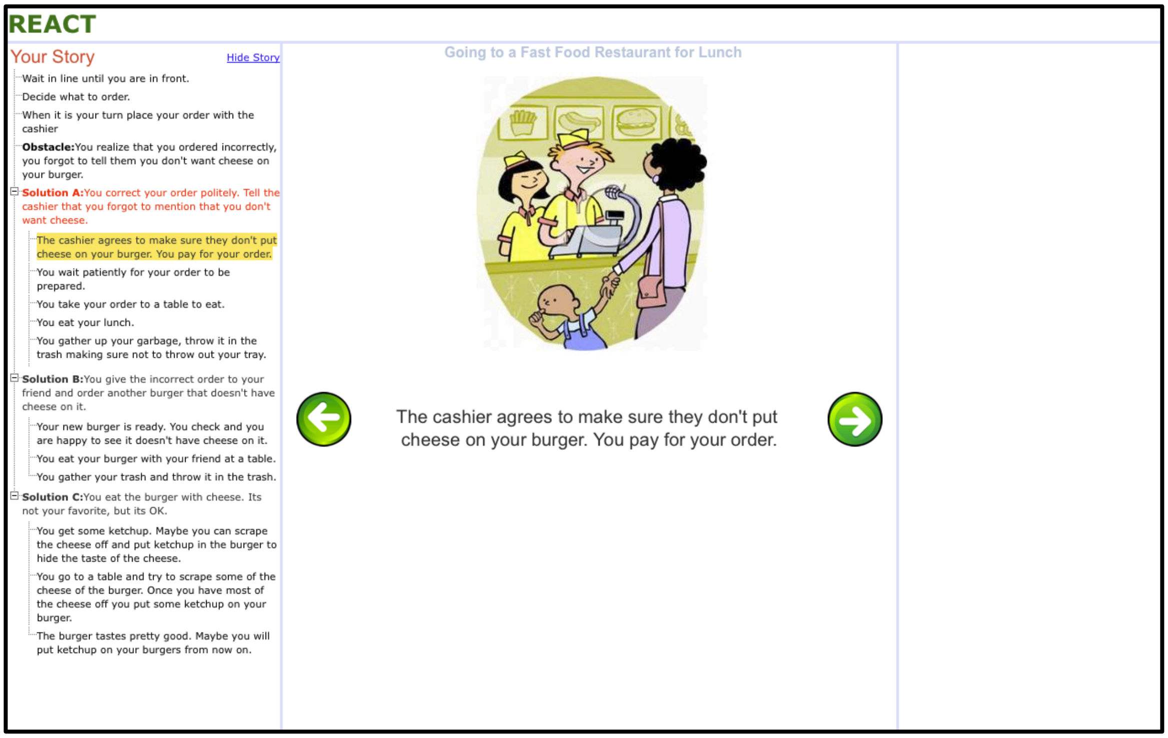


Figure 5.14. Narrative page in a branch in the REACT player (with outline shown).

Conclusion

In this chapter the prototypes that were developed of the REACT system were described. It can be seen that the smart REACT system is designed to collaborate with the author by providing suggestions throughout the process of authoring an instructional module, thereby enabling the creation of a customized module. Before real claims could be made about the usefulness of the tool, however, the system had to be evaluated in two ways. First the tool had to be put in the hands of a group of potential authors, to determine its usability. Once the system's usability had been evaluated, the modules created by these potential authors (the output of the authoring tool) had to be evaluated by experts to determine how useful and effective they would be to the child they were written for. A description of the design and results of these two studies is presented in the next chapter.

CHAPTER 6

EVALUATION OF THE REACT PROTOTYPES

The goal of REACT is to enable parents, teachers, and other interested stakeholders to develop customized social skills instructional modules for individuals with autism. For this reason, the evaluation of the REACT prototypes involved two studies; a usability study to evaluate the interaction design used in REACT, and a study to evaluate the output of the system. Following is a description of each of the studies.

The Parent Study

How effective the authoring tool is in aiding authors to create instructional modules will be impacted by how easily they can use it. For this reason, the first study I conducted to evaluate REACT was a usability study.

Participants and Recruitment

Participants were recruited from the target author population; parents or other caregivers of individuals with autism. These participants were recruited by word of mouth. Since I have shared my research with many in the autism community in Atlanta, it was important to recruit participants who had not interacted with Refl-ex in the past to ensure no confounding of the study results.

Procedure

Before interacting with the software, demographic information was collected via a questionnaire (Appendix C). This information included the participants' highest level of education, the social skills techniques they use with their child and their experience writing and using social stories with their child. The participants were also asked to describe their child, including his or her level of functioning, language abilities (reading and receptive), situations he or she struggles with and the problem behaviors that need to be addressed.

After a brief presentation of the Refl-ex instructional modules, the participants were asked to create their own modules. Each participant was asked to create 2 modules about having lunch at a fast food restaurant, one using the REACT prototype, and a second using the smart REACT prototype. The order of using the systems was such that it counterbalanced any learning effects (half of the participants used REACT first, and the other half used smart REACT first).

After the participants created their modules, they were asked to complete a questionnaire that asked them to assess their experience using the software (Appendix D). There were four types of questions in the questionnaire:

1. Questions about the user interface and interaction with it;
2. Questions about the suggestions;
3. Questions about module customization and structure; and
4. Overall evaluation.

In this way, this study allowed for the evaluation of the usability of the system, the evaluation of the relevance of the suggestions, and the creation of modules that could be used in the expert study to evaluate the output of the tool.

Results and Discussion

Participant Information

Initially the goal was to recruit five parents to interact with the system. In the end, however, nine parents participated in the study creating a total of 18 stories. Table 6.1 shows the demographic information of the parent participants. It can be seen from the Table that the largest number of participants indicated that they were between the ages of 45 and 54. Most of the participants were mothers of children with autism, but the participants did include 2 fathers and one grandmother. Lastly, all of the parents had only one child with autism.

Table 6.1. Demographic information from the parent study.

	Age	Relationship to the child with autism	Marital Status	Number of Children (with autism)	Education
1	45-54	Mother	Married	2 (1)	Post-graduate
2	45-54	Mother	Married	2 (1)	Post-graduate
3	55-64	Father	Divorced	1 (1)	College
4	45-64	Mother	Married	3 (1)	College
5	65 or above	Grandmother	Married	1 (1)	Post-graduate
6	35-44	Mother	Married	1 (1)	High-school/GED
7	45-54	Mother	Married	2 (1)	College
8	35-44	Mother	Married	2 (1)	Post-graduate
9	25-34	Father	Married	2 (1)	College

In addition to providing information about themselves, the participants were asked to answer a series of questions about their child. These questions included the child's age, actual and functional grade-level, level of functioning and problem behaviors. A summary of the child information can be seen in Table 6.2. All but one of the children were between the ages of 12 and 16, with the remaining one being eight years old. Parents indicated that their children struggled with such situations as; turn taking, patience, obsessing on thoughts, making friends, giving up control, and many others. Parents also indicated such problem behaviors as respecting authority, violent outbursts, and understanding rules. Some details of the child information can be seen in Table 6.2.

Table 6.2. Child Information.

	Age	Gender	Diagnosis
1	15	Male	ADHD with Autism
2	15	Male	Asperger's Syndrome
3	13	Male	ADHD, maybe Asperger's Syndrome
4	16	Male	PDD-NOS
5	14	Female	Asperger's Syndrome
6	12	Female	Asperger's Syndrome
7	16	Female	Autism and Spina Bifida
8	12	Male	High Functioning Autism
9	8	Male	High Functioning Autism and Generalized Anxiety Disorder

The anonymized answers to all of these questions were used to create a profile of each child that could be presented to the experts (all the details can be seen in Appendix E). This enabled the experts to consider the needs of the particular child when evaluating the modules. It was clear that each of the children struggle with different situations and exhibit different problem behaviors. Since the parents created the social problem solving skills module, however, they were able to keep these needs in mind throughout the authoring process.

Finally, in the pre-questionnaire participants were also asked about their current social skills practices. Two out of the nine participants indicated that they do not engage in any social skills approaches at home and that they rely on school to provide the instruction. The most commonly mentioned approaches were prompting and priming. Five out of the nine participants indicated that they both prompt the child to exhibit appropriate behavior in the situation and prepare the child in advance for situations (priming). The modules that REACT helps parents author can be used for priming, as they allow the child to practice situations in advance, but they can also be used for reflecting on situations after they happen. This is important, as four of the participants also indicated that they reflect on situations with their child after the fact to help them understand what happened.

Evaluation of the Interface

After the participants created their two modules, they were asked a series of questions to determine their perceptions of the system. The first two questions were related to how easy to use and understand the interface was. Figures 6.1a and 6.1b indicate that all but one of the participants gave the interface a four or a five rating for ease of use. The ease of understanding ratings only received one fewer 4. In conversation with the participants, the participants that gave ratings of 3 explained that they would have liked to be given a tutorial on how to use the system. Overall, however, both criteria on average received quite good ratings of above four on a scale from one to five (ease of use = 4.33, ease of understanding = 4.22). This confirms the system was perceived as "easy to use" by participants in a walk up and use situation. I believe that this rating would only improve as the parents got more opportunities to interact with the system.

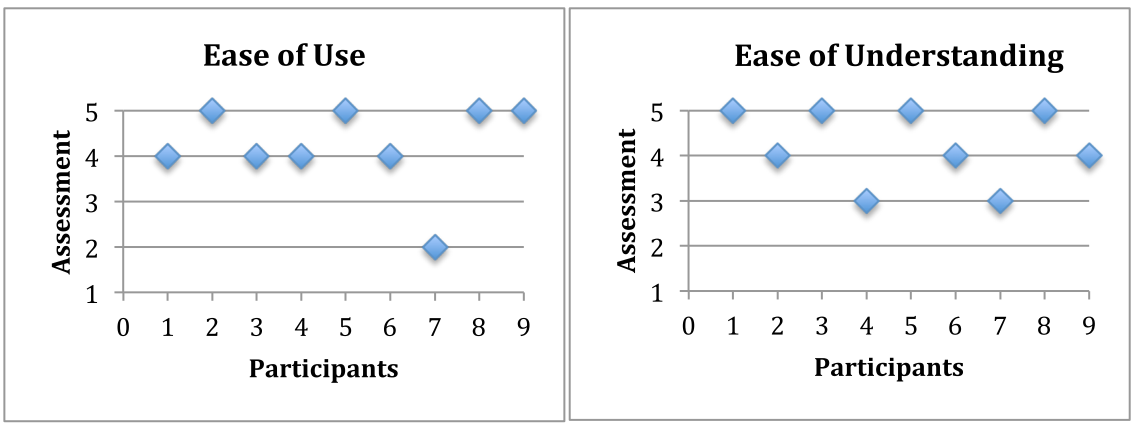


Figure 6.1a-b. (a) Ease of use assessments. (b) Ease of understanding assessments.

<u>Evaluation of the suggestions</u>

Participants were asked to evaluate the relevance of the suggestions to their child. Figures 6.2 a-c present the detailed data from this evaluation. Five of the nine participants rated all three types of data as "very relevant" (5) to their child. The remaining four participants had mixed responses. However, overall, the obstacle suggestions were on average perceived to be the most relevant of the suggestions (4.33/5), and the step suggestions received the lowest ratings (3.88/5). One outlier in the data is Participant 4's assessment of the step suggestions. In conversation with this participant it became apparent that her reason for this assessment was because of the lack of emotional content in the suggestions, saying that:

"*You need to address how the child feels in the particular situation.*" (Participant 4)

This participant's child attends a school whose educational approach strongly emphasizes the discussion of emotions. While this is good feedback, the intention in REACT is that the author will supply this information; it would be very difficult for the crowd to provide this information as every person reacts in their own way to particular situations.

In addition to asking for ratings, participants were also asked what they liked most, and what they liked least about the suggestions. The positive comments included such statements as:

"*They speed the effort, and help you think through the problem."* (Participant 2);
"*gives you ideas in case you are having trouble generating your own*" (Participant 4); and
"*helps initiate story design rather than requiring complete story genesis*" (Participant 9).

The negative comments, however, were more interesting. Two participants indicated feeling like they were relying too heavily on the suggestions saying:

"*it's like leading the witness*" (Participant 2); and
"*may box in some users and prevent them from branching out and truly customizing*" (Participant 9).

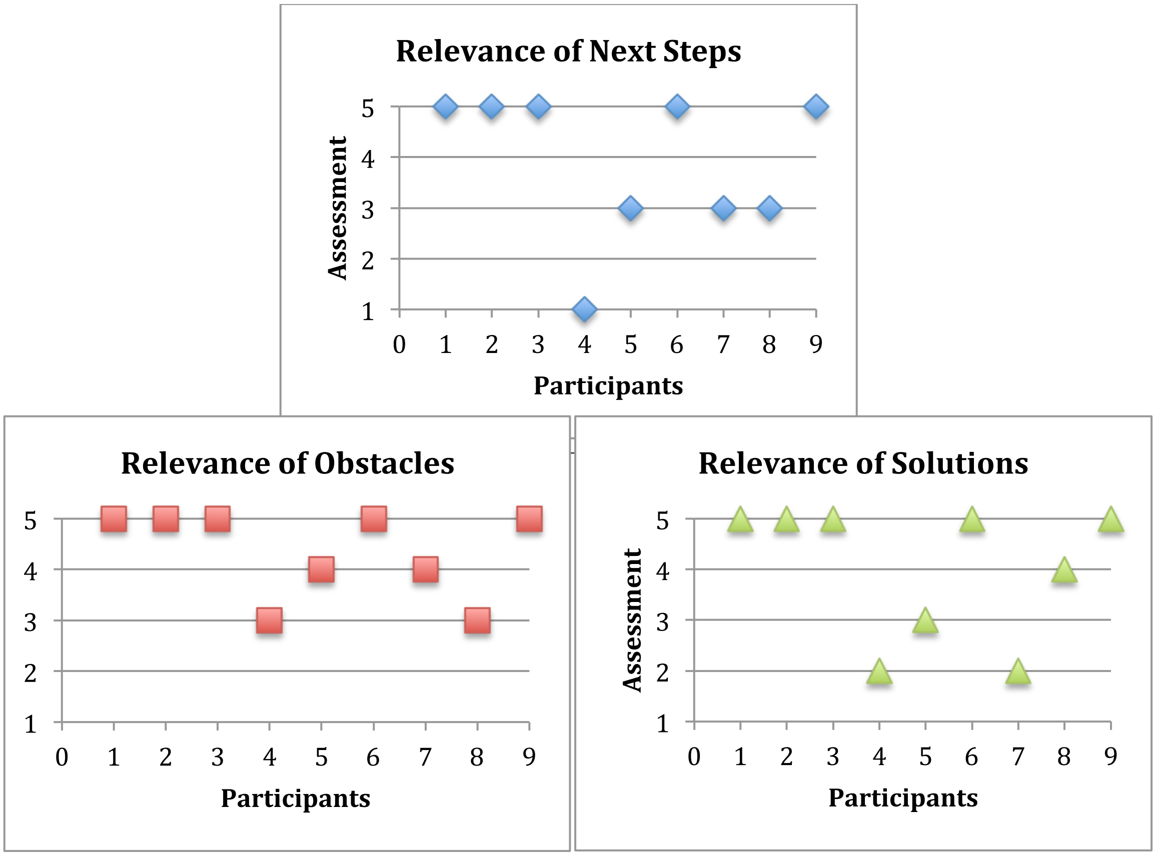


Figure 6.2a-d. (a) Relevance ratings for the next step suggestions. (b) Relevance ratings for the obstacle suggestions. (c) Relevance ratings for the solution suggestions

By far the most useful and interesting comment came from the participant that liked the suggestions least, she stated:

> "*[the suggestions] are not categorized, [it] might be better for them to be themed. For example, 'noise' oriented.*" (Participant 4).

I met this participant fairly early in the process, so I observed the subsequent participants closely to specifically determine what kind of obstacles they were looking for. It appeared that indeed many of them would have benefitted from having the suggestions presented not based on what could happen next, but instead on the type of problem they wanted to introduce and have their child practice. This is a finding that I plan to address in future research.

The last question that was asked about the suggestions was related to the design decision made; that of making the suggestions draggable. Participants were asked how easy they thought this interaction was. The average user response was 4.33 on a scale from one to five, with five of the participants giving a score of five, and no scores below a three being given. Therefore the user response to the draggable suggestions was positive.

Module Customization and Structure

Participants were asked what they thought of the structure of the module they were authoring. In particular, participants were asked what they thought of providing the student with options for ways to overcome the obstacle. Responses to this question were positive with the average of the responses coming to 4.44. Despite liking the idea of giving the child options, however, some the participants did feel that there were instances when coming up with three distinct options was difficult and not necessarily appropriate.

Table 6.3. How often parents would create custom modules for their child.

How Often	# of Participants
Every other month	1
Multiple times a month	4
Once a week	3
Multiple times a week	*1*

Participants were also asked how useful they thought their ability to create custom stories would be to their child, and how often they would create them. The average of the participants' responses was 4.33, implying that they thought it would be quite useful.

Table 6.3 shows that eight out of the nine participants said they would create stories at least a few times a month.

Overall Evaluation

The final set of questions allowed for a high level evaluation of the authoring tool and the modules it allowed the participants to create. A summary of the participants' quantitative responses can be seen in Table 6.4.

Table 6.4. Overall evaluation questions.

Question	Average of responses
How does the system compare to your current social skills practices? (1= less effective – 5 = more effective)	4.11
Overall, how effectively do you think system will help you meet your child's needs? (1 =not at all – 5=very effectively)	4

In addition to the quantitative questions, participants were asked three open ended qualitative questions. Participants were asked what they liked best and least about the system, and which feature they thought was most effective in meeting their child's needs. Not surprisingly, the structure of the module, in that it presented an obstacle and provided options for solutions, and the ability to create customized social skills instructional material were the responses most often given by participants to the last question.

Only five of the participants had answers to the questions of what they liked least about the system and all the responses took the form of suggestions for improvements. One of the participants in the study was disabled, so he indicated that he would prefer that the tool have less mouse interaction. The other four responses were: 1) being able to develop more scenarios; 2) adding more emotional content to the suggestions; 3) allowing for picture options versus written options for solutions; and finally 4) giving participants additional feedback about their progress through the authoring process through an alternate progress bar. These suggestions are important to consider in future versions of the tool.

Overall, the feedback received during the parent study was positive and informative. The study confirmed not only that the current prototype can easily be used by parents to create custom modules for their child, but also that it can be integrated with the models collected from the crowd in a way that is usable by potential authors. In addition, a great deal of interesting information was provided that will be used to inform the design of the complete authoring tool.

The Expert Evaluation

In addition to evaluating the interaction design, it was important to evaluate the output participants were able to achieve when using the authoring tool; the modules. In order to evaluate the output, I used the modules that the participants created in the parent study and asked experts to evaluate them. As mentioned in the previous section, the parents were asked to author 2 modules each; one with REACT and one with the smart REACT system. The former provided no suggestions to the author. This allowed me to attempt to isolate the impact of the suggestions on the quality of the modules.

Participants and Recruitment

As in the Branching Validation Study presented in Chapter 3, the participants in this study were individuals who have experience working with students with autism and other developmental disabilities. In particular, I recruited participants who have experience in social skill instruction. In fact, two of the participants in this study also participated in the Branching Validation Study.

Procedure

Each interaction with a participant was conducted one-on-one, as an informal semi-structured interview. In order to obtain feedback systematically I began with a questionnaire requesting demographic information, and in particular: degrees achieved, certifications and other qualifications, the extent of their experience working with children with autism, and their experience teaching social skills (very similar to the one used in the Branching Validation Study, see appendix A).

Participants were then presented with the modules created by the participants in the parent study (randomized so all participants did not see them in the same order), along with the profile of the child the module was created for (appendix E). The participants were asked to fill out a short evaluation sheet for each of the modules (appendix F). These sheets were used to guide the discussion with the participant regarding the strengths and weaknesses of the individual modules.

In the evaluation sheet, participants were first asked to indicate how familiar they are with the student described in the profile. This is important, as students with autism have unique needs. The experts were then asked to rate each of the modules on their appropriateness, usefulness, and potential effectiveness for the target student. For instance, the evaluation sheet asked:

1. Is the situation presented with an appropriate level of detail?
2. Is the situation presented the appropriate length?
3. How well does the module enable learning appropriate behavior?

4. Overall rating of the module.

The experts were also asked which of the two modules created for the child they thought was better and why. This helped to enable me to evaluate the impact the suggestions provided in the smart REACT prototype had on the expert evaluations.

Participant Information

Five experts were recruited to evaluate the modules. As mentioned earlier, two of the five participants also participated in the branching validation study. The detailed participant information can be seen in Table 6.5.

Table 6.5. Expert participant information.

	Age	Relationship to the child with Autism	Number of Children Interacted with	Grade/Age Level	Years of Experience
1	25-34	Teacher	11 or more	High School up to age 21	4
2	55-64	Teacher and Therapist	11 or more	Elementary school and Young adults	15
3	45-54	Therapist	11 or more	High School up to age 22	26
4	25-34	Teacher	11 or more	4-20	12
5	45-54	Teacher	6-10	Elementary School	10

As can be seen, the participants worked with students at all age-levels (elementary school – high school), and all the participants worked with at least six students with autism. Not included in the table was the fact that all the participants had postgraduate degrees. Most importantly, four of the five participants had at least ten years of experience working with students with autism.

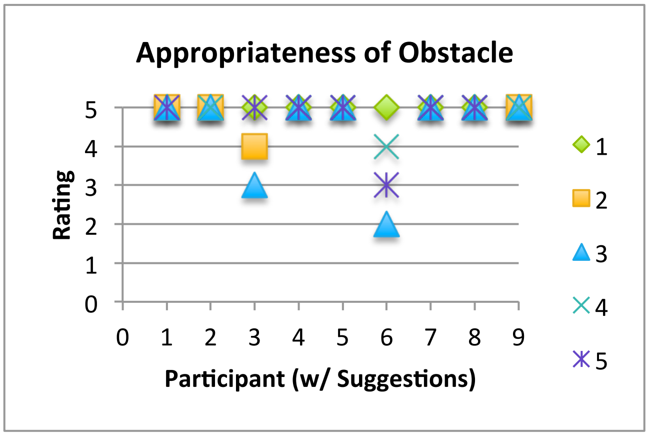
Results and Discussion

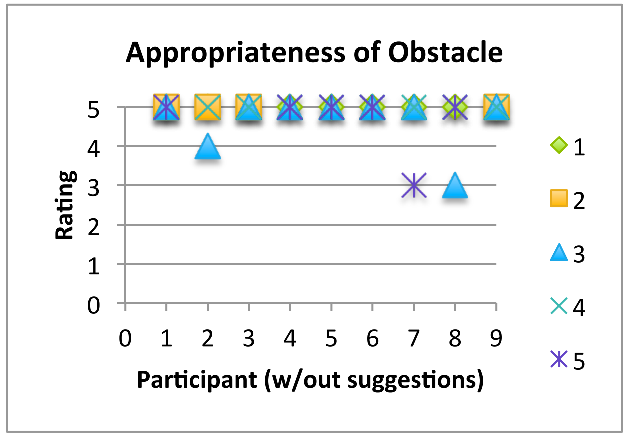
Module Evaluations

Each of the modules was evaluated by at least four experts. The modules that were evaluated by each expert were chosen based on two criteria; first there were a couple of expert participants who were familiar with some of the parents who created the modules, for this reason they were not allowed to review those modules, second, all the interactions with the participants were capped at one hour, in other words, they were not allowed to start the evaluation of a new pair of modules after the one hour mark. Based on these criteria, two participants evaluated all eighteen modules, two evaluated fourteen modules, and one evaluated eight modules.

Figures 6.3 a-b show the averages of the expert assessments of the appropriateness of the obstacles introduced into the modules by the authors. The assessments of the modules created with suggestions have been separated from the assessments of those created without suggestions. On the questionnaire a rating of 5 indicated that the expert thought the obstacle was “very appropriate” for the child the module was created for.

Figure 6.3 shows that 7 of the 9 modules created with suggestions and 6 out of the 9 modules created without suggestions received perfect obstacle appropriateness scores. The experts’ concerns about the obstacle introduced in module m3 with suggestions and m2, m7, and m8 without suggestions was that they were not as likely to occur. Module m6 with suggestions received lower ratings than all the rest. In further discussion, the experts indicated that they thought that, the obstacle in m6 with suggestions was not appropriate for the child. The obstacle introduced in that module was "you only have a credit card, and the restaurant is cash only." Experts felt that a child that age would not likely be given a credit card. This was an example of a case when the parent did not appropriately choose from the suggestions provided.

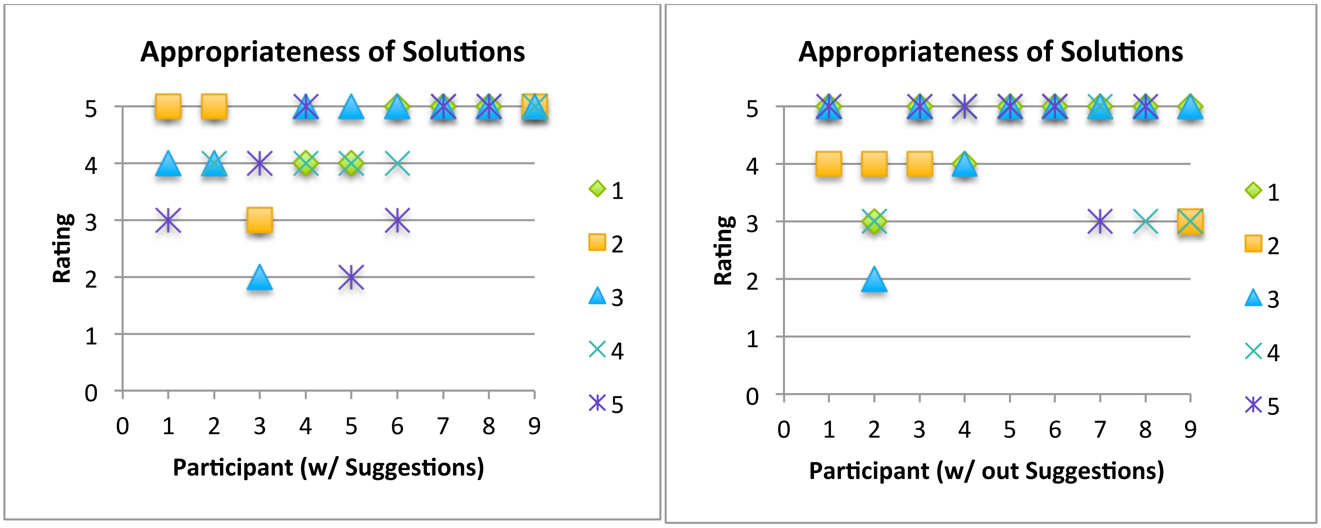




(a) **(b)**

Figure 6.3. a) Appropriateness of the obstacles in the modules with suggestions b) Appropriateness of the obstacles in the modules without suggestions.

Figures 6.4a and b indicates that, while the solution assessments were not as high as the obstacles assessment, three of the modules with suggestions and two without suggestions did receive perfect scores. The remaining modules received positive ratings, with the exception of two outliers, m3 with suggestions and m2 without suggestions. In the case of m3, experts indicated that two of the three options offered were essentially the same, so the child was not truly given three options. In addition, the experts indicated that the options were not appropriate. As mentioned in the discussion of the findings of the parent study, some of the parent participants indicated that they had experienced difficulty coming up with three distinct options, even with the suggestions, and that the system should allow for the option of only providing two.



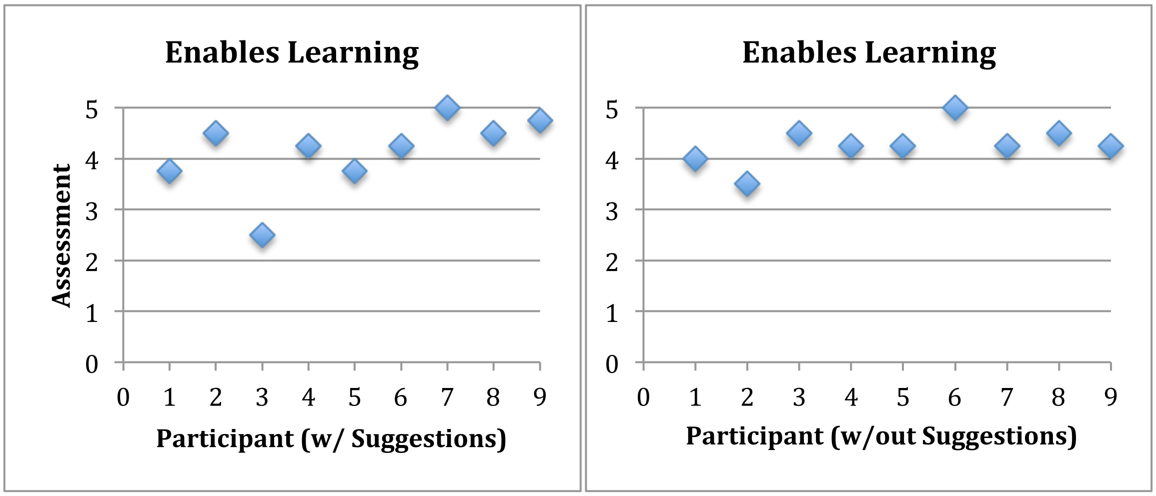
(a) (b)

Figure 6.4. a) Appropriateness of the solutions in the modules with suggestions b) Appropriateness of the solutions in the modules without suggestions.

In the case of m2 without suggestions, the experts stated that two of the options given were not realistic. In this module the obstacle was that "you do not have enough money to pay for your order." One of the options given was to ask the cashier to let you eat the food now and come back later with money, and the other was to call your parents to bring more money and eat in the meantime. Both of these depended on the cashier agreeing to let you do this, which is not realistic.

These findings imply that parents are not always the best judges of what is appropriate for their child. The experts indicated that they had experienced this in their interaction with parents. There were only 2 instances where this was the case, however.

Experts were also asked how well the modules enabled the child to learn the appropriate behavior for the situation being presented. The responses to this question can be seen in Figures 6.5a and b. Only one module in each group received a perfect score on this measure. The module that received the lowest ratings was m3 with suggestions. As mentioned earlier this module was the one that experts felt had redundant options that were not all appropriate.

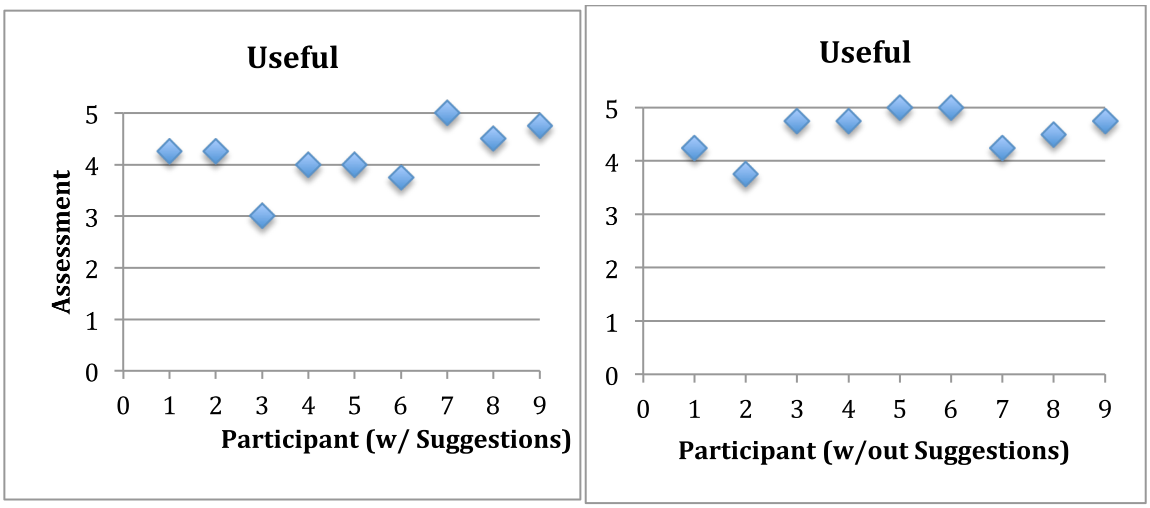


(a) (b)

Figure 6.5. Expert assessment of how well the modules enable learning. A) modules with suggestions b) modules without suggestions.

Participants were also asked to evaluate each of the modules on three high level criteria: 1) how useful the module is to the child it was created for (5 = very useful); 2) how appropriate the module is for the child it was created for (5 = very appropriate); 3) overall, how they would rate the module (1= poor – 5 = excellent). The average of the participants' responses to each of these questions for each of the modules can be seen in Figures 6.6a-b, 6.7a-b and Table 6.6.

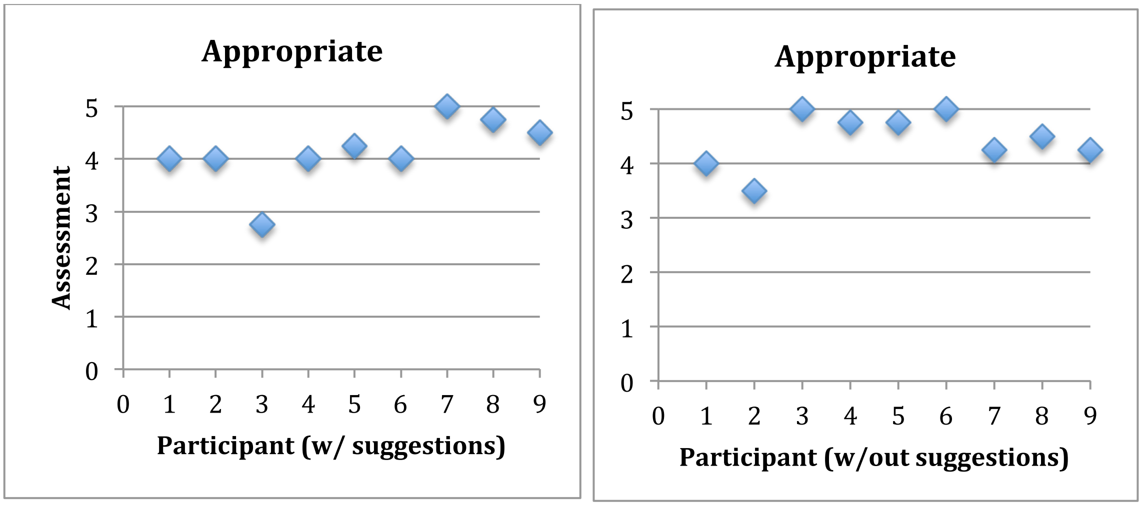
As can be seen from Figure 6.6, once again module m3 with suggestions received the lowest usefulness rating across all 18 modules. This module was the one that experts thought did not truly have three distinct appropriate solutions. The remaining modules performed well with all but m6 with suggestions, and m2 without suggestions receiving average ratings of at least 4 (4 = useful). Module m6 with suggestions was the module where the obstacle was not seen as age appropriate ("the restaurant is cash only, and you only have a credit card") and module m2 was the module mentioned earlier where experts felt the suggestions were not realistic because they were contingent on the cashier agreeing to let the child eat without paying for their meal first.



(a) (b)
Figure 6.6. Expert assessment of how useful the module is to the child it was created for. a) modules with suggestions b) modules without suggestions.

Similarly, as can be seen in Figure 6.7 only two of the modules received average appropriateness assessments lower than 4. As expected, those two modules were module m3 with suggestions and module m2 without suggestions. Overall, the experts appeared to concur that the majority of the modules the parents created were both useful and appropriate for the child they were created for.

One of the things I was concerned about was that, in the cases where the content the parents created was not appropriate, the child might learn inappropriate behavior. For this reason, I asked the experts what the effects of the less appropriate modules would be. All of the experts confirmed that, in their opinion, the worst that could happen is that the child wouldn't learn anything. There was no risk of the child learning inappropriate behavior.



(a) (b)

Figure 6.7. Expert assessment of how appropriate the module is for the child it was created for. a) modules with suggestions b) modules without suggestions.

Finally, participants were asked to provide an overall rating for each of the modules. The average of all the participants' ratings for each of the modules can be seen in Table 6.6. The same two modules that have been receiving the lowest ratings throughout were given the lowest overall ratings (p2's without suggestions and p3's with suggestions). The remaining modules all received overall ratings of at least "good."

After they had finished evaluating each pair of modules the participants were asked which of the two they thought was better. They were also allowed to indicate that they thought the modules were equal. In the verbal assessment, the vote was exactly equal, in four of the pairs the modules with suggestions were rated as better, in four other pairs the modules without suggestions were rated as better, and one pair was rated as equal. Interestingly, these verbal responses corresponded exactly with the overall assessments given in writing. The module receiving the highest score in each pair is bolded in the table.

Table 6.6. Overall expert evaluation of modules.

Participant	Overall Expert Assessment	
	Module (with suggestions)	Module (without suggestions)
P1	3.75	3.75
P2	**3.5**	2.75
P3	2.25	**4.25**
P4	3	**4.5**
P5	3.75	**4.75**
P6	3.75	**4.5**
P7	**5**	4
P8	**4.5**	4.25
P9	**4.5**	4.25

Conclusions and Contributions

The REACT prototype enabled the parents to quickly and effectively create the modules for their child. Further, in their evaluations the parents indicated that the tool was easy to use and understand, and that the suggestions the tool provided were relevant to their child. This study also revealed interesting insights for ways to improve the system that I intend to pursue in future work.

In the second part of the study, the experts responses overall indicated that indeed the parents were able to create good instructional modules for their child. The experts' evaluations largely showed that the modules the parents created would be appropriate and potentially useful to the child. It was found, however, that the modules with suggestions were rated fairly equally to those without suggestions. There are several explanations for this finding:

1) There was a conflict between the parents' expectations for the child and those of the experts;
2) The parents struggled with coming up with the required number of solutions to the obstacle that was introduced;
3) The modules were created about going to a fast food restaurant, a scenario which the parents were likely very familiar with, and therefore they may not have needed the suggestions as much as they might have for a different task or if they were asked to create more than two stories;
4) The parents did not act as filters of the crowdsourced data as effectively as was expected in that:
    a. They did not always choose the most appropriate suggestions.
    b. They rarely modified the suggestions once they added them to their module.
5) It was not possible to isolate the benefit of the authoring tool itself. The REACT tool enabled the creation of *obstacle-based branching* stories in both cases, therefore the value this added to the modules may have confounded the influence of the suggestions.

These findings speak to several factors to consider in future work. For instance, allowing for a flexible number of solutions to be provided, and studying how the authoring process and modules would be different if the parents were asked to create modules about a less familiar situation.

Further, each of the expert participants indicated a desire to be allowed to use the tool themselves to create stories for their student. Seeing as they are the ones most often providing social skill instruction, it is my belief that they would have great success creating modules using REACT. It is my goal to put the tool into the hands of these users as soon as possible.

To conclude, the goal of the REACT system is to enable authors to create customized social skills instructional modules. The two studies presented in this chapter address research question 3. Through a study with parents of children with autism the prototype of the authoring tool was found to be easy to use and effective at allowing parents to create custom modules for their child. The modules created by the parents were then presented to experts for evaluation. Through the expert study it was seen that parents could indeed create good social skills instructional modules for their children. Overall, the experts found that almost all of the parents' modules were useful and appropriate for the children they were created for.

Contributions:

The development of an authoring tool that: 1) is easy to use; 2) empowers parents and other caregivers to easily create customized social skills instructional modules for their children and students; and 3) has been confirmed by experts to enable authors to create good individualized social skills instructional modules.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

The goal of this research was to explore the possibility of developing a system that can help parents and other caregivers author instructional modules that individuals with autism can use to practice social problem solving skills. To this end, two prototype systems were developed: Refl-ex, which is a collection of multimedia instructional modules designed to enable adolescents with autism to practice social problem solving skills; and REACT, a system to facilitate the authoring of a wider variety of instructional modules.

Conclusions

What separates the Refl-ex modules from current approaches to social skills instruction is that the students are presented with a social scenario in which an obstacle arises, and they are taken through the process of overcoming that obstacle. The decision points that the student must navigate in response to the obstacle create a branching structure that is unique and that experts have confirmed is an improvement to current practices. We call this structure *obstacle-based branching.*

While this branching structure did prove to be beneficial, it makes the process of authoring the modules more difficult and time consuming [Bruckman, 1990; Riedl and Young, 2005]. The varied and vast nature of the potential content that is necessary to prepare an individual to navigate our social world is apparent. This motivated the need for an authoring tool that could help authors create the branching scenarios and give them ideas for potential content.

In order to provide these suggestions for content, I turned to crowdsourcing. The nature of the social knowledge that was necessary to provide authors with appropriate suggestions for content was such that computational techniques alone would not suffice; the knowledge needed to come from *people* who understand our highly complex social world. To effectively collect this data I developed a 6-phase process for querying the crowd that enables the creation of models of social knowledge that contain not only information related to how to successfully complete particular social tasks, but also obstacles that may arise, and ways to overcome those obstacles.

The rich models of social knowledge developed using the crowdsourcing approach were used to provide suggestions to the authors as they create customized instructional modules for a particular child. The suggestions included possible next steps in the social situation, obstacles that may arise at each step, and solutions to those obstacles. In this way, the authoring of the interactive software is facilitated with the aid of models of social knowledge.

To address the various facets of this work, my thesis was that:

> An authoring tool can be developed that uses crowdsourced models of social knowledge to help parents easily author individualized *obstacle-based branching* instructional modules, a structure experts evaluate to be an improvement to current approaches to social skills instruction for children with autism.

In particular, I addressed the following research questions:

- *RQ1a: How can software modules be developed to help a student with autism prepare for various social contexts?*

- *RQ1b: What value do experts perceive in obstacle-based branching scenarios, and how do they compare to the current approach of using sequential stories to teach social skills?*
- *RQ2: What is a mechanism for generating rich models of social knowledge that are consistent with the obstacle-based branching approach to problem solving and can be used to provide scaffolding for the authoring of social skills instructional modules?*
- RQ3: *How will parents use a tool that employs rich models of social knowledge to facilitate authoring, and will the tool enable them to produce good instructional modules?*

To address these questions a series of studies was conducted and 2 high-fidelity prototypes were developed. The first two studies addressed research questions 1a and 1b. In an exploratory study with adolescents and young adults with autism, Refl-ex modules enabled students to successfully navigate the social scenarios in which an obstacle had arisen by providing sufficient scaffolding and guiding the student through the social problem solving process. This study was followed by a study with experts (i.e., individuals who are experienced at providing social skills instruction to students with autism) in which the *obstacle-based branching* structure was evaluated against current approaches to social skills instruction. In this study experts confirmed that the branching approach was an improvement to current approaches.

Once these studies were completed work was done on developing and validating the approach to producing the models of social knowledge. Through two studies, the possibility of collecting social knowledge from the crowd in a manner that is effective and useful was explored and confirmed. My approach is innovative and unique in several ways.

1) It uses the crowd to generate new data/content that is best provided by people.
2) It also uses the crowd to process the data that is generated, utilizing the common sense and semantic knowledge that is again unique to human intelligence.
3) It uses an approach to aggregating the data that maintains the richness and variety of the responses while making it useful and usable.

In other words, in this work humans were used both as producers and processors of data, and were able to aggregate the data successfully not to a single correct response but to a model containing as many distinct correct responses as possible. Once the model had been produced the crowd was used again; this time it confirmed that the model would be useful and usable by a potential author.

Finally, two versions of the authoring tool were developed that enabled the creation of the obstacle-based instructional modules. The only difference between the two versions was that one used a crowdsourced model to provide suggestions to the author during the authoring process, and the other did not. A two-part study was conducted to evaluate the tool. First, parents of children with autism were recruited to create two modules for their child, one with each version of the tool, and to evaluate the tool. Once this part of the study was completed, social skills experts were recruited and asked to evaluate the modules the parents had created.

In the first part of the study, the parents were able to quickly and effectively create the modules for their child. Further, in their evaluations the parents indicated that they thought the tool was easy to use and understand, and that the suggestions the tool provided were relevant to their child. This study also revealed interesting insights for ways to improve the system that I intend to pursue in future work.

In the second part of the study, the experts' responses overall indicated that indeed the parents were able to create good instructional modules for their child. The experts' evaluations largely showed that the modules the parents created would be appropriate and potentially useful to the child. It was found, however, that the modules with suggestions were rated fairly equally to those without suggestions. There are several explanations for this finding:

1) There was a conflict between the parents' expectations for the child and those of the experts;
2) The parents struggled with coming up with the required number of solutions to the obstacle that was introduced;
3) The modules were created about going to a fast food restaurant, a scenario which the parents were likely very familiar with, and therefore they may not have needed the suggestions as much as they might have for a different task or if they were asked to create more than two stories;
4) The parents did not act as filters of the crowdsourced data as effectively as was expected in that:
    a. They did not always choose the most appropriate suggestions.
    b. They rarely modified the suggestions once they added them to their module.

5) It was not possible to isolate the benefit of the authoring tool itself. The REACT tool enabled the creation of *obstacle-based branching* stories in both cases, therefore the value this added to the modules may have confounded the influence of the suggestions.

Despite some of the unexpected findings, the authoring tool indeed enabled parents to easily create individualized obstacle-based branching modules for their child that experts perceived to be valuable. Further, this study provided several ideas for actionable ways to improve the system as I work toward creating a fully functioning authoring tool.

In summary, the contributions of this work were:

1) The development of interactive software modules that adolescents with HFA can use to independently practice social skills.
2) The creation of a branching approach to developing interactive social skills instructional modules that experts confirm is an improvement to current practices.
3) The development of an approach to building models of social knowledge that can be dynamically created and expanded using crowdsourcing.
4) The development of an authoring tool that: 1) is easy to use; 2) empowers parents and other caregivers to easily create customized social skills instructional modules for their children and students; and 3) has been confirmed by experts to enable authors to create good individualized social skills instructional modules.

Future Work

Many interesting ideas for future work emerged from the findings of this thesis. First, one of the most exciting parts of the work was the evolution of the crowdsourcing approach for developing the models of social knowledge. There are several avenues that ought to be explored related to this approach. First, there are approaches I would like to attempt for making the data collection process more efficient and cost effective. In chapter 4 the idea of creating a philanthropic crowdsourcing platform was presented. People would respond to questions and complete tasks not because they are being paid a few cents, but because they care about the cause and want to help. Another approach could be to somehow enable the workers to receive community service credits, like those that are often required by social organizations like sororities and fraternities, and some schools.

In addition to improving the process, the parent study revealed that it would be useful to include additional information about the data in the model. This idea emerged when a parent suggested that she would have liked to have the obstacle data organized based on themes (e.g. obstacles about noise). I believe there is an opportunity to again employ the crowd to perform this categorization. This would enable the authoring tool to help the parent to more easily individualize the modules. For instance, if a parent knows that their child struggles with waiting and other time-related social behaviors, they could filter the suggestions to show only time related obstacles. I am currently working on exploring ways in which this might be accomplished.

In future work, it is also necessary to address how to effectively follow the author's path through the complex social models. Introducing obstacles, and solutions to those obstacles, makes returning to the original path in order continue the story once the obstacle is overcome somewhat of a challenge. I believe that in addition to taking direction from the planning literature [Draper et. al., 1993; Yang and Tenenberg, 1990; Yokoo, 1994], this challenge may also be amenable to being tackled with a crowdsourcing approach.

The version of REACT that was developed in this work was a prototype. Work remains to be done before REACT will truly enable the authoring of Refl-ex modules. This includes facilitating the creation of multiple decision points, allowing the inclusion of inappropriate or counterproductive solutions, and enabling the creation of the reflection section (Chapter 3). As was shown in the evaluation studies described in chapter 6, it will be important to give the authors the flexibility to include a variable number of solutions to the obstacle, it is my intuition that this will also apply to the overall Refl-ex structure. In other words, allow the author to decide how many decision points to include, and how many branches or solutions to offer at each decision point.

Beyond expanding the structure, the prototype authoring tool only facilitated the creation of the text content of the Refl-ex modules. The Refl-ex modules also had imagery and audio narration (chapter 3). Facilitating the incorporation of this functionality also needs to be addressed.

The user study also revealed that the authors might have benefitted from having more guidance or feedback while creating their modules. This could include prompting about instances in which explanations of the emotions the child is feeling might be beneficial, and where more detail might be required to help the child understand the scenario they are being presented. This would potentially improve the efficacy of the modules, and provide the critiquing portion of the REACT tool that was not addressed in this thesis.

The ultimate goal is to empower authors to create customized modules about any scenario they believe their child needs to practice. In addition to working towards a complete authoring tool, there are several other areas where there are opportunities to study the authors and use this knowledge to work towards this goal. First, it would be interesting to explore how parents will respond if they are given the tool and allowed to use it for an extended period of time. If left on their own to use the tool however they like, how will they use it? What scenarios will they build models for? Will they be willing to share their stories with others? It is also important to get the tool into the hands of other classes of authors, including teachers and therapists. Given that these users should have extensive knowledge of social skills instructional strategies, how different will their modules be from the parents'? Will they use the tool differently? All of these are questions that would be useful to explore in future work.

It is important to empower the individuals themselves whenever possible. I believe there is an opportunity to allow the individual with autism himself or herself to author the module. Given the right suggestions and feedback, the authoring process could potentially be as beneficial as practicing with the completed modules to the acquisition of appropriate social problem solving skills. To this end, I would like to attempt to allow individuals with autism who are able, to write their own modules and explore the impact of the authoring process.

Finally, I would also like to test the tool and the modules with other populations of target students with similar cognitive profiles to students with autism. I have been approached by researchers working with other populations (e.g. children with brain injuries) who have exhibited challenges with social skills and problem solving. It would be interesting to see if these populations would benefit from Refl-ex and REACT, and if and how the two systems may have to be adapted to better support these populations' needs.

In summary, several interesting areas for future work emerged from this thesis. They include:

1) Improving the speed and cost effectiveness of the crowdsourcing approach.

2) Exploring opportunities for categorizing the data using crowdsourcing.

3) Developing approaches for tracking the author's progression through the social model to more effectively support the authoring process.

4) Expanding the tool to facilitate the authoring of complete Refl-ex modules including imagery and audio narration, and building in more flexibility in the structure.

5) Exploring the impact of providing feedback to the author related to the inclusion of emotional content and other relevant context.

6) Studying the usage of the tool in different scenarios including:

a. Allowing parents to use it for an extended period of time.

b. Giving the tool to different classes of authors, including teachers and therapists.

c. Giving the authors more models to get suggestions from.

d. Allowing authors to attempt to create the models themselves.

e. Providing authors with the ability to share their modules and use others' modules as starting templates for their own.

7) Allow the individual with autism himself or herself to use the tool to author modules for themselves, and explore the impact of the authoring process on their acquisition of social problem solving skills.

8) Exploring the potential of using Refl-ex and REACT to benefit other populations with similar cognitive profiles.

Conclusions

The population of young adults with autism is growing. Many of these individuals can function effectively and autonomously, but need assistance to handle the complexities of society. In this dissertation, I presented a way in which technology may provide some assistance. Two notable accomplishments were made. First, I introduced *obstacle-based branching* as a pedagogical tool. Next, I developed a human computation approach to empowering authors to create customized *obstacle-based branching* instructional modules. Human input is the natural way to facilitate the authoring of social skills modules, and in so doing empower and assist those trying to navigate the intricacies of a social life.

REFERENCES

[Aleahmed et.al., 2010] Aleahmad, T., Aleven, V., Kraut, R. Automatic Rating of User-Generated Math Solutions. Educational Data Mining, Pittsburgh, PA, 2010.

[Aleven et. al., 2009] Aleven, V., McLaren, B. M., Sewall, J., Koedinger, K.R. A New Paradigm for Intelligent Tutoring Systems: Example-Tracing Tutors. Int. Journal of Artificial Intelligence in Education,19: 105-154. 2009.

[Alonso et. al., 2008] Alonso, O., Rose, D.E. Stewart, B. Crowdsourcing for Relevance Evaluation. In Proc. of the ACM Special Interest Group for Information Retrieval Forum 2008. 2:9-15, 2008.

[Anderson et. al. 1995] Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. The Journal of the Learning Sciences, 4 (2) 167-207.

[APA, 2000] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th Edition, text-revised). Washington, DC. 2000.

[Asperger, 1944] Asperger, H. Die "autistichen psychopathen" in kindasalter. Archiv fur Psychiatrie and Nervenkrankheiten. 117: 76-136. 1944.

Asperger, H. The Autistic Psychopathy of Childhood. Archive for Psychiatry and Nervous Diseases, 117:76-136. 1944.

[Banda and Robinson, 2011] Banda, N. and Robinson, P. Multimodal Affect Recognition in Intelligent Tutoring Systems. Affective Computing and Intelligent Interaction. 2011.

[Baron-Cohen, 2008] Baron-Cohen, S. Autism and Asperger's Syndrome. 160. Oxford University Press, September 2008.

[Beaumont and Sofronoff, 2008] Beaumont, R. and K. Sofronoff. A Multi-Component Social Skills Intervention for Children with Asperger's Syndrome: The Junior Detective Training Program. The Journal of Child Psychology and Psychiatry, 49(7): 743-753. 2008.

[Bellini et. al., 2007] Bellini, S., Peters, J.K., Benner, L., Hopf, A. A Meta-Analysis of School-Based Social Skills Interventions for Children with Autism Spectrum Disorders. Remedial & Special Education, 28: 153-162. 2007.

[Bernard-Opitz et. al., 2001] Bernard-Optiz, V., Sriram, N., and Nakhoda-Sapuan, S. Enhancing Social Problem Solving in Children with Autism and Normal Children Through Computer-Assisted Instruction. Journal of Autism and Developmental Disorders, 31(4): 377-384. 2001.

[Bernstein et. al, 2010] Bernstein, M.S., Tan, D., Smith, G., Czerwinski, M., and Horvitz. E. Personalization via Friendsourcing. ToCHI 2010.

[Blessing et. al. 2006] Blessing, S.B., Gilbert, S., Ritter, S.: Developing an Authoring System for Cognitive Models within Commercial-Quality ITSs. In Proc. of the Nineteenth Int. FLAIRS Conf., 497–502. AAAI Press, Melbourne. 2006.

[Blessing and Gilbert, 2008] Blessing, S.B. and Gilbert, S. Evaluating an Authoring Tool for Model-Tracing Intelligent Tutoring Systems. In Proc. of Intelligent Tutoring Systems, 204–215, 2008.

[Bosseler and Massaro, 2003] Bosseler, A. and Massaro, D. Development and Evaluation of a Computer-Animated Tutor for Vocabulary and Language Learning in Children with Autism. Journal of Autism and Developmental Disorders, 33(6): 653-672. 2003.

[Boujarwah et. al., 2010] Boujarwah, F. A., Hong, H., Isbell, J., Arriaga, R.I., Abowd, G.D. Training Social Problem Solving Skills in Adolescents with High-Functioning Autism. In Proc. of the 4th Int. Conf. on Pervasive Computing Technologies for Healthcare (Pervasive Health). 2010.

[Boujarwah et. al., 2011] Boujarwah, F. A., Kim, J. G., Abowd, G. D., Arriaga, R. I. Developing Scripts to Teach Social Skills: Can the Crowd Assist the Author? Presented at the Human Computation Workshop at AAAI 2011, 8-11 Aug 2011, San Francisco, CA.

[Boujarwah et. al., 2012] Boujarwah, F. A., Abowd, G. D., Arriaga, R. I. Socially Computed Scripts to Support Social Problem Solving Skills. To be presented at The International ACM SIGCHI Conference on Human Factors in Computing Systems 2012 (CHI '12) , 5-10 May 2012, Austin, Texas.

[Brabham, 2008] Brabham, D. C. Crowdsourcing as a Model for Problem Solving: An Introduction and Use Cases. Convergence: The International Journal of Research into New Media Technologies, 14(1):75-90, 2008.

[Bruckman, 1990] Bruckman, A. The Combinatorics of Storytelling: Mystery Train Interactive. Unpublished paper, MIT Media Lab.1990.

[Bryant et. al., 2005] Bryant, S. L., Forte, A., Bruckman, A. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In Proc. of GROUP 2005, Sanibel Island, FL, November 2005.

[CastingWords, 2001] CastingWords: http://castingwords.com/. last accessed: 01/19/2012.

[Cognitive Models Study, 2010] Cognitive Models Study: http://research.hsi.gatech.edu/reflexstudy/CognitiveModels/hello.html. last accessed 01/20/2012.

[Cooper et. al., 2010] Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popvic, Z. Predicting Protein Structures with a Multiplayer Online Game. Nature 466: 759-760, August 2010.

[Corneli, 2010] Corneli, J. Crowdsourcing a Personalized Learning Environment for Mathematics. Probation Report, 2010.

[Draper et. al., 1993] Draper, D. Hanks, S. Weld, D. Probabilistic Planning with Information Gathering and Contingent Execution. Technical Report 93-12-04. 1993.

[D'Zurilla et. al., 2002] D'Zurilla, T.J., Nezu, A.M., Meydeu-Olivares, A.M. Social Problem Solving Inventory-Revised. The Cognitive Center Foundation. 2002.

[Eckert et. al., 2010] Eckert, K., Niepert, M., Niemann, C., Buckner, C., Allen, C., Stuckenschmidt, H. Crowdsourcing the Assembly of Concept Hierarchies. In Proc. of the Joint Conf. on Digital Libraries. Gold Coast, Queensland, Australia. June 2010.

[Gagnon, 2001] Gagnon, E. Power Cards: Using Special Interests to Motivate Children and Youth with Asperger's Syndrome and Autism. Autism and Asperger's Publishing Company, Shawnee Mission, KS. 2001.

[Gordon et. al., 2010] Gordon, J., Van Durme, B., Schubert, L. K. Evaluations of Commonsense Knowledge with Mechanical Turk. In Proc. of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Los Angeles, CA. June 2010.

[Gray, 1995] Gray, C. Social Stories Unlimited: Social Stories and Comic Strip Conversations. Jenison, MI: Jenison Public Schools, 1995.

[Gresham, 1986] Gresham, F. M. Conceptual and Definitional Issues in the Assessment of Children's Social Skills: Implications for Classification and Training. Journal of Clinical Child Psychology, 15(1), 3–15. 1986.

[Griswold et al., 2002] D.F. Griswold, G.P. Barnhill, B.S. Myles, T. Hagiwara, and R.L. Simpson. Asperger Syndrome and Academic Achievement. Focus on Autism and other Developmental Disabilities, 17: 94-102. 2002.

[Heer and Bostock, 2010] Heer, J. and Bostock, M.. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. CHI 2010.

[Kittur, et. al., 2008] Kittur, A. Chi, E.H. and Suh. B. Crowdsourcing User Studies with Mechanical Turk. CHI 2008.

[Heflin and Alberto, 2001] Heflin, L.J. and Alberto, P.A. Establishing a Behavioral Context for Learning with Students with Autism. Focus on Autism and Other Developmental Disabilities, 16, 93-101. 2001.

[Hopkins et. al., 2011] Hopkins, I.M., Gower, M.W., Perez, T.A., Smith, D.S., Amthor , F.R., Wimsatt, F.C., Biasini, F.J. Avatar Assistant: Improving Social Skills in Students with an ASD Through a Computer-Based Intervention. Journal of Autism and Developmental Disorders, February 2011.

[Howe, 2006] Howe, J. The Rise of Crowdsourcing. WIRED Magazine, (14.06), 2006, last accessed 12/29/2010.

[Howlin, 2003] Howlin, P. Outcome in High-Functioning Adults with Autism with and without Language Delays: Implications for the Differentiation Between Autism and Asperger's Syndrome. Journal of Autism and Developmental Disorders, 33: 3-13. 2003.

[InnoCentive, 2012] InnoCentive: http://innocentive.com. last accessed 01/18/2012.

[Intellitools, 2012] Intellitools Classroom Suite: http://www.intellitools.com/ last accessed 03/29/2012.

[Kanner, 1943] Kanner, L. Autistic Disturbances of affective content. The Nervous Child, 2: 217-250. 1943.

[Klin and Volkmar, 2003] Klin, A. and Volkmar, F.R. Asperger Syndrome and External Validity. Child and Adolescent Clinics of North America, 12(1): 1-13. 2003.

[Koedinger et. al. 1997] Koedinger, K.R., Anderson, J.R., Hadley, W.H., & Mark, M.A. (1997). Intelligent Tutoring Goes to School in the Big City. Int. Journal of Artificial Intelligence in Education, 8(1), 30-43.

[Laffey et. al., 2009] Laffey, J., Schmidt, M., Stichter, J., Schmidt, C., Goggins, S. iSocial: A 3D VLE for Youth with Autism. In Proc. of the Conference on Computer Supported Collaborative Learning 2009, Rhodes, Greece.

[Little et. al., 2009] Little, G., Chilton, L. B., Goldman, M., and Miller, R. C. Turkit: tools for iterative tasks on mechanical turk. KDD-HCOMP '09.

[Mechanical Turk, 2012] Mechanical Turk: http://mturk.com. last accessed 01/29/2012.

[Munro et. al., 2010] Munro, R., Bethard, S., Kuperman, V., Tzuyin Lai, V., Melnick, R., Potts, C., Schnoebelen, T., Tily, H. Crowdsourcing and Language Studies: The New Generation of Linguistic Data. In Proc. of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Los Angeles, CA. June 2010.

[Ohlsson 1992] Ohlsson, S. (1992) Constraint-Based Student Modeling. Journal of Artificial Intelligence and Education, 3(4), 429-447.

[Orkin and Roy, 2009] Orkin, J. and Roy, D. Automatic Learning and Generation of Social Behavior from Collective Human Gameplay. Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Budapest, Hungary, 2009.

[Parson et. al. 2004] Parsons, S. Mitchell, P. and Leonard, A. The Use and Understanding of Virtual Environments by Adolescents with Autistic Spectrum Disorders. Journal of Autism and Developmental Disorders, 34(4): 449-466. 2004.

[Putman and Chong, 2008] Putman, C. and Chong. L. Software and technologies designed for people with autism: What do users want? In Proc. of the 10th ACM Conf. on Computers and Accessibility (ASSETS). 2008.

[Rao et. al, 2008] Rao, P.A. Beidel, D.C. and Murray, M.J. Social Skills Interventions for Children with Asperger's Syndrome or High-Functioning Autism: A Review and Recommendations. Journal of Autism and Developmental Disorders, 38(2): 353-361. 2008.

[Refl-ex, 2012] Refl-ex Prototype: http://research.hsi.gatech.edu/reflexstudy/index.php, last accessed 01/15/2012.

[Reynhout and Carter, 2009] Reynhout, G. and Carter, M. The Use of Social Stories by Teachers and their Perceived Efficacy. Research in Autism Spectrum Disorders, 3: 232-251. 2009.

[Riedl and Young, 2005] Riedl, M. and Young, M. From Linear Story Generation to Branching Story Graphs. In Proc. Of the First Artificial Intelligence and Interactive Digital Entertainment Conference. 2005.

[Riedl et. al., 2008] Riedl, M., Stern, A., Dini, D., and Alderman, J. Dynamic Experience Management in Virtual Worlds for Entertainment, Education, and Training. Int. Transactions on Systems Science and Applications, 4(2), 2008.

[Rosa and Eskenazi, 2011] Rosa, K. D. and Eskenazi, M. Impact of Work Sense Disambiguation on Ordering Dictionary Definitions in Vocabulary Learning Tutors. AAAI, San Francisco, CA, 2011.

[Schank and Abelson, 1977] Schank, R.C and Abelson, R.P. Scripts. In Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures. Lawrence Erlbaum Associates Publishers, Hillsdale, NJ. 1977.

[Singh et. al., 2002] Singh, P., Lin, T., Mueller, E.T., Lim, G., Perkins, T., and Zhu, W.L. Open Mind Common Sense: Knowledge Acquisition from the General Public. ODBASE 2002.

[Stanford Parser, 2011] Stanford Parser: http://nlp.stanford.edu:8080/parser/, last accessed 10/20/2011.

[Stories2Learn, 2012] Stories2Learn: http://itunes.apple.com/us/app/stories2learn/id348576875?mt=8, last accessed 01/29/2012.

[Storymovies, 2012] Gray, C. Shelley, M. The Special Minds Foundation. Storymovies: http://www.dttrainer.com/products/storymovies/, last accessed 01/24/2012.

[Suppes and Morningstar, 1965] Suppes, P. and Morningstar, M. Computer-Assisted Instruction. Science, 166:343-350. 1969.

[Surowieki, 2005] Surowieki, J. The Wisdom of Crowds. 336. Anchor, August 2005.

[Tataro and Cassell, 2008] Tartaro, A. and Cassell, J. Playing with Virtual Peers: Bootstrapping Contingent Discourse in Children with Autism. In Proc. of the Int. Conf. of the Learning Sciences (ICLS). 2008.

[Trillingsgaard, 1999] Trillingsgaard, A. The Script Model in Relations to Autism. Journal of European Child and Adolescent Psychiatry, 8:45-49. 1999.

[Volden and Johnston, 1999] Volden, J. and Johnston, J. Cognitive Scripts in Autistic Children and Adolescents. Journal of Autism and Developmental Disorders, 29(3):203-211. 1999.

[von Ahn, 2005] Human Computation. Carnegie Mellon University, Pittsburgh, PA. December 7, 2005.

[von Ahn et. al., 2003] von Ahn, L., Blum, M., Hopper, N. J., Langford, J. CAPTCHA: Using Hard AI Problems for Security. Advances in Cryptography, 2656/2003(646). 2003.

[von Ahn et. al., 2006] von Ahn, L., Ginosar, S., Kedia, M., Liu, R., Blum, M., Improving Accessibility of the Web with a Computer Game. In Proc. of CHI 2006. Montreal, CA, April 2006.

[von Ahn et. al., 2008] von Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M. reCAPTCHA: Human-Based Character Recognition via Web Security Measures. Science Magazine, 421: 1465-1468. 2008.

[Wightman, 2010] Wightman, D. Crowdsourcing Human-Based Computation. In Proc. of NordoCHI 2010, Reykjavik, Iceland, October 2010.

[Wikipedia, 2012] Wikipedia: http://en.wikipedia.org/wiki/Wikipedia. last accessed 03/22/2012.

[Willams and Wright, 2002] Williams, C. Wright, B., Callaghan, G., Coughlan. B. Do Children with Autism Learn to Read more Readily by Computer Assisted Instruction or Traditional Book Methods? Autism, 6(1): 71-91, 2002.

[Yahoo! Answers, 2012] Yahoo! Answers: http://answers.yahoo.com/. Last accessed 01/29/2012.

[Yang and Tenenberg, 1990] Yang, Q. and Tenenberg, J.D. ABTWEAK: Abstracting a Nonlinear, Least Commitment Planner. AAA1 1990.

[Yokoo, 1994] Yokoo, M. Weak-commitment Search for Solving Constraint Satisfaction Problems. AAAI 1994.

[Zhang et. al., 2011] Zhang, H., Horvitz, E, Miller, R.C., and Parkes, D.C. Crowdsourcing Human Computation. The International ACM SIGCHI Conference on Human Factors in Computing Systems 2011 (CHI '11) , 7-12 May 2011, Vancouver, BC.

APPENDIX A

BRANCHING VALIDATION STUDY - PRE-QUESTIONNAIRE

Please answer the following questions.

1. What is you're age?

   18-24   25-34   35-44   45-54   55-64   65 or above

2. Gender (circle one):

   Male   Female

3. What is your ethnicity? (Circle all that apply.)

   African-American   Asian/Pacific Islander   Hispanic/Latino

   White/Caucasian   Other______________________________

4. What is your highest level of education completed? (*Circle one*.)

   Elementary   Middle-school   High-school/GED

   College   Post-Graduate

5. If you have earned a college degree, what is your degree in? ______________________________

   ____

   ____

6. What additional certifications or training have you completed? (e.g. Social Stories, ABA, etc)

   ____

   ____

7. Which of the following apply to you:

   I teach individuals with Autism   I work with individuals with Autism as a therapist

   Other ____

8. How many individuals with Autism do you interact with regularly?

   1-5   6-10   11 or more

9. How old/what grade level are these individuals? ______________________________

   ____

10. How would you describe their level of functioning (verbal communication skills, academics, level of independence)? ____

____

____

11. Do you provide these individuals with social skills instruction (e.g. priming)?

Yes **As needed** **No**

12. If yes, what approaches do you use to teach social skills? ________________________________

____

____

13. Do you use Social Stories? ____

____

14. Do you use software to provide social skills instruction? If yes, which software? ___________

____

____

15. Do any of your students/patients have other developmental impairments? If yes, which impairments?____

____

16. If yes, do provide social skills instruction to the students/patients with other impairments?

____

____

17. If yes, do you use the same or different approaches than those you use with your students/patients with autism? ____

____

____

APPENDIX B

BRANCHING VALIDATION STUDY - POST-QUESTIONNAIRE

1. What did you like about format 1?

____

____

2. What did you dislike about format 1?

____

____

3. How useful do you think format 1 would be for your students?

   Not useful at all — Very useful

   1 2 3 4 5 6

4. How appropriate do you think format 1 is for your students?

   Not appropriate at all — Very appropriate

   1 2 3 4 5 6

5. What did you like about format 2?

____

____

6. What did you dislike about format 2?

____

____

7. How useful do you think format 2 would be for your students?

   Not useful at all — Very useful

   1 2 3 4 5 6

8. How appropriate do you think format 2 is for your students?

Not appropriate at all — Very appropriate

1 2 3 4 5 6

9. What did you like about format 3?

____

____

10. What did you dislike about format 3?

____

____

11. How useful do you think format 3 would be for your students?

Not useful at all — Very useful

1 2 3 4 5 6

12. How appropriate do you think format 3 is for your students?

Not appropriate at all — Very appropriate

1 2 3 4 5 6

13. How useful do you find social stories?

Not useful at all — Very useful

1 2 3 4 5 6

14. How often do you create custom stories?

Multiple times a day   once a day   multiple times a week

once a week   multiple times a month   once a month

other____

15. How long does it take for you to create a customized story?

____

16. Would you want to create social stories more often? **YES** **NO**

17. Do you think it makes a difference for a story to be customized?

No difference at all A Significant Difference

1 2 3 4 5 6

18. What types of behaviors do you use social stories to address? (select all that apply)

Social interaction skills Conversation skills

Address/introduce changes/new routines Address/reduce inappropriate behaviors

What other behaviors would you use social stories to address? ______________________

____

19. What settings would you use social stories in?

Classroom playground at home

What other settings would you use social stories in? _______________________________

____

20. Which format did you think was the best?

Format 1 Format 2 Format 3

21. Which did you think was the worst?

Format 1 Format 2 Format 3

22. How useful did you think the introduction of the obstacle was?

Not useful at all Very useful

1 2 3 4 5 6

23. How useful did you think that incorporating interaction was?

Not useful at all Very useful

1 2 3 4 5 6

APPENDIX C

PARENT STUDY - PRE-QUESTIONNAIRE

Your Information

Please answer the following questions about yourself.

1. What is your age?

   18-24 **25-34** **35-44** **45-54** **55-64** **65 or above**

2. Gender (circle one):

   Male **Female**

3. What is your marital status? ______________________________

4. What is your highest level of education completed? (*Circle one*.)

   Elementary **Middle-school** **High-school/GED**

   College **Post-Graduate**

Child's Information

Please answer the following questions about your child(ren).

1. How many children do you have? ______________________________
2. Age(s): ______________________________
3. Grade(s)/Academic Level(s): ______________________________

   ____

4. What is/are your child(ren)'s diagnosis? ______________________________

   ____

5. How would you describe your child(ren)'s level of functioning: verbal communication skills, academics, level of independence, etc? (if you have more than one child please describe each individually)

   ____

   ____

   ____

   ____

6. What are some situations your child(ren) struggle(s) with?

   ____

   ____

____

____

7. Do(es) your child(ren) exhibit problem behaviors? If yes, please describe them.

   ____

   ____

   ____

8. Do you practice social skills with your child? If yes, what approaches do you use?

   ____

   ____

   ____

   ____

9. Do you use any technology or software to help your child with social skills? If yes, please describe them.

   ____

   ____

   ____

10. Do you use social stories with your child? ________________________________

    a. How often do you create stories for your child?

Multiple times a day	Once a day	Multiple times a week
Once a week	Multiple times a month	Once a month
Other: ____		

    b. How long does it take you to create a story? _________________________

    c. If it was easier/faster would you create stories for your child more often?

    ____

APPENDIX D

PARENT STUDY - POST-QUESTIONNAIRE

Please answer the following questions

1. What do you believe was the main purpose or goal of the system you just used?

   ____

   ____

2. How would you describe the system:

Difficult to use				Easy to use
1	2	3	4	5

3. How would you describe the layout of the system:

Difficult to understand				Easy to understand
1	2	3	4	5

4. Which features(if any) would you add to the current system?

   ____

   ____

5. Which features(if any) would you exclude/remove from the current system?

   ____

   ____

6. How relevant do you think the suggestions for next steps were to your child?

Not relevant				Very relevant
1	2	3	4	5

7. How relevant do you think the suggested obstacles were to your child?

Not relevant				Very relevant
1	2	3	4	5

8. How relevant do you think the suggested solutions to obstacles were to your child?

Not relevant				Very relevant
1	2	3	4	5

9. How would you describe the approach used to incorporate suggestions (i.e. dragging them from the suggestion box to the text input box):

Difficult to use				Easy to use
1	2	3	4	5

10. What do you like about the suggestions?

____

____

____

11. What do you dislike about the suggestions?

____

____

____

12. How useful would your ability to create custom stories be for your child?

Not useful				Very useful
1	2	3	4	5

13. How often would you create custom story?

Multiple times a day	Once a day	Multiple times a week
Once a week	Multiple times a month	Once a month
Other: ____________________________________		

14. What types of behaviors would you use the story to address?

Social Interaction Skills		Address/introduce changes/new routines
Address/reduce inappropriate behaviors		Conversation skills
Other: ____________________________________		

15. One of the features of the system is to provide the child with solutions to the obstacle s/he encountered. How useful do you feel the choices the child is given are?

Not useful				Very useful
1	2	3	4	5

16. What would you add to this feature?

____

17. What would you remove from this feature?

18. Which features do you feel were most useful in helping you meet your child's needs?

19. What did you like best about the system?

20. What did you like least about the system?

21. How does the system compare to your current social skills practices?

Less effective				More effective
1	2	3	4	5

22. Overall, how effectively do you think system will help you meet your child's needs?

Not at all				Very effectively
1	2	3	4	5

Thank you for your participation!

APPENDIX E

EXPERT STUDY – CHILD PROFILES

Child Profile — Module: ___1 & 2 ___

Age : ____15_____ **Diagnosis:** ____ADHD with Autism____
Grade level: ____9_____ **Academic level:** ______4th Grade______
Level of Functioning as described by the parent: __High-functioning as regards his autism, he can communicate his needs. He can mimic anything, but sometimes his ability to communicate and negotiate his free thoughts is limited. Unless he feels strongly about something, then he is very outspoken to the point of being bossy. He definitely knows what he wants and can be strong-willed.
Situations the child struggles with: _waiting in line, turn taking, conversation skills (e.g. jumping into conversations), patience, giving up control, playing with others, when things don't go the way he expects (the video game he wants is not available at the store), communicating his feelings (e.g. if he wants something he will say his brother wants it instead)
Problem Behaviors: __you have to explain to him more than the ordinary of the order of how things were done or how things happen. So that he understands why.________________________

Child Profile — Module: ___3 & 4 ___

Age : ____15____ **Diagnosis:** ____Asperger's Syndrome____
Grade level: _10_ **Academic level:** on level, but doesn't like anything other than Science
Level of Functioning as described by the parent: __fully-verbal but with impaired social skills. He has normal cognitive functioning (average IQ), academically he can do "on level" but he doesn't like it except for science usually. Independence is mixed; we usually supervise him or are at home with him. We think consciously about minimizing leaving him alone, we also tend to wait on him "to keep the peace" but that might be a teenager thing.
Situations the child struggles with: _making friends, doing anything other than playing video games, homework, new experiences, obsessing on thoughts, PE day at school (because it's a change in routine and it is unstructured and it is also physically exerting).
Problem Behaviors: __interrupting, anger management and overreacting, responding appropriately, understanding rules, respecting authority, doesn't understand what you can and can't say to people.

Child Profile — Module: ___5 & 6 ___

Age : ____13____ **Diagnosis:** _ADHD maybe Asperger's Syndrome
Grade level: ____8th_____ **Academic level:** ___on level_________
Level of Functioning as described by the parent: __high verbal functioning and very highly functioning academically. Medium in independence when it interests him.
Situations the child struggles with: _turn taking, interacting with peers his own age and older, if its not video game related, or related to history he has a hard time interacting. He also tends to talk a lot about things he is interested in without seeking or allowing the other person to contribute.
Problem Behaviors: __periodic meltdowns, but few and far between as he gets older, getting angry when we reduce his "screen time" (time with the TV, video games, computer, iPad).

Child Profile Module: ___7 & 8 ___

Age : ____**16**____ **Diagnosis:** ____PDD-NOS___________
Grade level: _high school_____ **Academic level:** ______not on level___
Level of Functioning as described by the parent: __highly verbal and can express full range of emotions. Not on level for academics. Had trouble understanding more abstract themes. Anxiety and phobia driven so independence is hard.
Situations the child struggles with: _group settings, sitting and waiting, handling surprises and sudden noises, understanding fast paced speech.
Problem Behaviors: __can be reactive and extreme when upset; can lash out at others and fall into sobbing and wanting to isolate/retreat.

Child Profile Module: ___9 & 10__

Age : ____**14**____ **Diagnosis:** __Asperger's Syndrome___
Grade level: ____8th grade_____ **Academic level:** __on level (all A's)___
Level of Functioning as described by the grandparent: __She is very high functioning, and has no problems communicating with people she knows well. She gets all A's in her regular academic classes. She has some independence, but it is very hard to get her to try new things, or meet new people.
Situations the child struggles with: _bullying in the school cafeteria. Changes in routine, no matter how small cause her to act out or meltdown.
Problem Behaviors: __Meltdowns, hitting others, biting (rarely)_________________________

Child Profile Module: __10 & 11__

Age : ____**12**____ **Diagnosis:** ______Asperger's Syndrome_______
Grade level: ____7th grade_____ **Academic level:** ______on level_____
Level of Functioning as described by the parent: __Able to communicate. Is independent. Has some anxiety- especially related to time (being late). Mostly social issues.
Situations the child struggles with: _Anxiety with running late. Difficulty with one-on-one conversations
Problem Behaviors:______________________________none_________________________________

Child Profile Module: __12 & 13__

Age : ____**16**____ **Diagnosis:** _Autism and Spina Bifida (wheel-chair bound)
Grade level: __11th grade___ **Academic level:** anywhere from 1-4 yrs below grade-level
Level of Functioning as described by the parent: __Communicates well verbally- doesn't volunteer much information, but will answer questions. Academics below grade-level (anywhere from 1-4 years). Not very independent in self-care, problem solving, etc. (partly because of wheel chair use/physical disabilities.
Situations the child struggles with: _making friends, getting independence from parents, making decisions.
Problem Behaviors: __occasional emotional outbursts (adolescence + frustration + hormones)__

Child Profile

Module: __14 & 15__

Age : ____**12**____ **Diagnosis:** ________high-functioning autism__________

Grade level: ____7th grade_____ **Academic level:** ______grade-level with para-pro_______

Level of Functioning as described by the parent: __verbal, intelligent, regular classes with para-pro, requires constant cues, reminders for things related to hygiene, doing homework, organization, not provoking his brother. He has limited interests and is on the computer a lot.

Situations the child struggles with: _instigating/making brother angry, he is a sore loser, high frustration level- when told to do something he doesn't want to do. (i.e. homework, some classwork)

Problem Behaviors: __tantrums in the classroom, outbursts, fighting with his brother constantly______

Child Profile

Module: __17 & 18__

Age : ____**8**____ **Diagnosis:** high-functioning autism and a generalized anxiety disorder

Grade level: 2nd grade **Academic level:** _______on-par in most academic areas_______

Level of Functioning as described by the parent: __highly verbal, on par with most academic areas of study (current reading comprehension deficit, especially when expected to make inferences) fairly independent for his age level, though requires prompting for certain tasks.

Situations the child struggles with: _deviations from expected outcomes such as the day's schedule not matching with his previous expectations

Problem Behaviors: __verbal stimming, verbally inappropriate at times. Lack of empathy at times._____

APPENDIX F
EXPERT STUDY – EVALUATION SHEET

Please answer the following questions.

1. How familiar are you with the type of student described in the profile?

Not familiar at all				Very familiar
1	2	3	4	5

2. Given the profile of the child, the level of detail with which the situation is presented is:

Too little		Just right		Too much
1	2	3	4	5

3. Given the profile of the child, the length of the module is:

Too short		Just the right length		Too long
1	2	3	4	5

4. Given the profile of the child, the obstacle that is introduced is:

Not at all appropriate				Very appropriate
1	2	3	4	5

Why? ____

5. Given the profile of the child, the solutions that are offered for overcoming the obstacle are:

Not at all appropriate				Very appropriate
1	2	3	4	5

Why? ____

6. The module enables learning appropriate behavior for the described situation.

Strongly disagree				Strongly agree
1	2	3	4	5

Why? ____

7. What do you like about this module? ________________________________

____

____

8. What do you dislike about this module?____________________________________

____

____

9. Overall, how useful do you think this module is to its intended student?

Not useful				Very useful
1	2	3	4	5

Why? ____

10. Overall, how appropriate do you think this module is for its intended student?

Not appropriate				Very appropriate
1	2	3	4	5

Why? ____

11. Overall, how would you rate this module:

Poor	Fair	Good	Very Good	Excellent
1	2	3	4	5

12. What would you change in the module? And Why?

____

____

____