Eyes Up: Influencing Social Gaze Through Play

by

Micah Rye Eckhardt

B.S., University of California San Diego (2006)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author____
Program in Media Arts and Sciences
June 11, 2010

Certified by____
Rosalind Picard
Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by____
Prof. Pattie Maes
Associate Academic Head
Program in Media Arts and Sciences

Eyes Up: Influencing Social Gaze Through Play

by

Micah Rye Eckhardt

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on June 11, 2010, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

Autism can be a debilitating condition that affects a person's personal and social affairs throughout their lifetime. With 1 in 110 people diagnosed with an Autism Spectrum Disorder (ASD) [49], it is important that we develop assistive and learning technologies to help them achieve their potential. In this work I describe the development of a new technology-mediated therapeutic game, *Frame It*, and the subsequent use of *Frame It* in an intervention, called *Eyes Up*, with children diagnosed with autism. *Eyes Up* requires the player to attend to details of the human face in order to correctly construct puzzles of people's eyes and then assign an expression label to them. The intervention is intended as a play-centered activity with the goal of increasing attention to other people's face and eyes region and improving expression recognition abilities. Through the application of user-centered design principles and special considerations to our participants we have been able to develop an engaging game that sustains interest. Using an eye-tracking system in conjunction with specifically designed experiments, we have been able to test the system's ability to influence gaze behavior and expression recognition. Analysis of of pre- and post- experimental measures reveals statistically significant increases in attention to the face and eyes and increases in expression recognition abilities.

Thesis Supervisor: Rosalind Picard
Title: Professor of Media Arts and Sciences, Program in Media Arts and Sciences

Eyes Up: Influencing Social Gaze Through Play

by

Micah Rye Eckhardt

The following people served as readers for this thesis:

Thesis Reader____

Mitchel Resnick
Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Thesis Reader____

Pattie Maes
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Acknowledgments

There is no such thing as the success of an individual. Success is a gift, a reward that may often be embodied by one, but which is always dependent on many and supported by a myriad of factors. Success emerges from the interaction of many working in concert, whether loosely or tightly knitted; it is the product of community. We might call this community a family or a team, or it may take on some larger more abstract sense, such as a society. Whether small or large it is this community that gives rise to success and for which the individual must pay homage. The success of an individual is nothing more than the embodiment of effort given forth by a community, symbolized through one. Success is truly a reflection of community.

Realizing this I, too, must pay homage to those who have helped me succeed and have provided me with community. Let me first thank my partner, who for many years has provided me with encouragement, care, and support, thus making even gruelling years more joyful. My sister and family for helping make me, me; my friends who have always been family (the Hebdons, Karuna, Geba, Petey, Bird and Meui) your influences have been substantial in my life; Joseph Fucciello, without risks we would never have made it this far; Gil Henry, not only for being a great mentor, but a great friend; all the teachers that have taken the time to care and teach, from Palomar to UCSD to MIT. I would also like to acknowledge all the great folks at the MPLab, particularly Javier Movellan and Ian Fasel.

I would like to give a special thanks to Matthew Goodwin for his advice and help, I truly could not have done this work without your help. I appreciate your years of dedication to the children you help. I know both they and I have benefited from your efforts. Likewise, I want to recognize the truly amazing help given from the Groden Center and especially Daniella Aube and Eliza Lane who helped make everything run smoothly. I also would like to thank all of the participants of my study, who have taught me much.

I would like to express my gratitude to MIT for providing me so many opportunities and the Media Lab for giving me a home to continue my crazy ways. I also would like to thank my lab mates and the Media Lab community for providing insights, inspiration and plain old fun from time to time. In addition, I would like to thank Amit Zoran and Santiago Alfaro for helping me become a better designer and providing me with interesting conversations.

I would also like to acknowledge Ami Klin and Robert Joseph for taking time to consult with me. In addition, I would like to give special thanks to my thesis readers Patie Maes and Mitchel Resnick. Your expertise and suggestions have made this work better.

Last, but far from least, I would like to recognize and thank my advisor, Rosalind Picard. Your years of hard work and high standards have helped create a space for me to thrive, I truly value the opportunity you have provided me. Thank you for your guidance, support, and encouragement. I hope to pay you back with continued hard work, insights, and furthered goals.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Autism is a complex and heterogeneous condition that is often typified by a significant difficulties with respect to social and emotional cue recognition and expression. These social-emotional processing differences make forming relationships with others, expressing internal affective states, and recognizing subtle social-emotional cues particularly difficult for persons with an Autism Spectrum Disorder (ASD) [2]. In particular, people diagnosed with ASD typically perform worse then neurotypical (NT) people recognizing facial expressions [14], especially when recognition is dependent solely on information surrounding the eyes [6, 9]. It has been shown that both children and adult autistics attend more to information in the lower part of the face than to information surrounding the eyes [44, 33, 30, 41]. This may be due, in part, to an aversion to eye contact, atypical scanning behavior of the face region, and/or even language related factors.

Efforts to improve communicative and social-emotional abilities in ASD are regarded as a critical priority [47] and extensive therapy is dedicated to helping to achieve this goal. Research indicates that early, intensive, and extended therapy can have significant positive impact, with recommendations of a minimum of 25 to 40 hours of therapy per week [40].

Unfortunately because of significant cost [24] and time constraints it is often difficult for an individual, family, school, or therapeutic service to provide this minimum level of service. Because of these needs and constraints new technologies have a tremendous opportunity to provide real benefits to individuals and families.

The eye region is commonly understood to be the most important facial region for conveying social-emotional information. It is also well known that those diagnosed with ASD, like many people with nonverbal learning disabilities, perform poorly at recognizing such social signals [7]. Furthermore, It has been suggested that interpretation of gaze and expression plays an important role in normal functioning of theory of mind (ToM) [3]. Therefore, it is important that we address these difficulties in recognizing such valuable information by developing methods and tools to help people better recognize and understand these social-emotional cues.

This work presents the development of a new technology, *Frame It*, and a subsequent intervention *Eyes Up* using *Frame It*. The *Frame It* game platform is intended for any person as a play-centered interactive device. The *Frame It* game is an interactive tangible puzzle game that blends digital and physical elements. For the purpose of this work we have focused on its use as a play-centered teaching and therapeutic tool for young severely challenged children diagnosed with ASD.

1.2 Purpose of Research

The challenges presented by ASD are many and require a broad array of solutions. Increasingly the application of technology to ASD has provided meaningful results [4, 5, 45, 50]. Despite the current awareness of autism, there is significantly less ASD literature focused on individuals who are non-speaking, exhibit echolalia, are routine oriented, experience anxiety when encountering unfamiliar people, exhibit restricted or repetitive behaviors, or otherwise are less-compliant. In this work we look to further our understanding and ability to create helpful technologies for young and challenged children diagnosed with ASD. In particular, we aim to create a helpful technology centered around playful interaction that results in behavioral changes and learning.

This work aims to determine if the designed system, *Frame It*, can be used in play-centered activities to influence the gaze behavior and expression recognition ability of children diagnosed with ASD. Though there have been many studies that have explored gaze behavior and/or expression recognition ability of persons diagnosed with ASD we know of none that have explored both with young, severely challenged children. Furthermore, we know of no work that has attempted an intervention with such children to determine if gaze behavior and expression recognition can be positively influenced. In part, this is because of the difficulty of working with this portion of the population. As such, we suggest methods to help facilitate study procedures through design, experimental considerations, and intervention procedures. In sum, the goal of this work is to provide practical help for these children. To this aim we hope that we have contributed.

1.3 Definition of Abbreviations

ASD: Autism spectrum disorder is a spectrum of psychological conditions characterized by abnormalities of social interactions and communication, as well as severely restricted interests and highly repetitive behavior [42].

HFA: High functioning ASD is a term commonly used to describe people diagnosed with ASD with improved ability in some area of functioning compared with others diagnosed with an ASD. Typically, HFA have normal to above normal intelligence scores, perform well on verbal comprehension test and exhibit less atypical physical behavior.

NT: Neurotypical refers to non-autistic individuals that have neurological development and states that are consistent with what most people would perceive as normal, particularly with respect to their ability to process linguistic information and social cues

IQ: Intelligence quotient is a score derived from a standardized test designed to assess a person's intelligence. There are many different tests, but most follow a standard scoring metric. An IQ score follows a Gaussian distribution with the mean value of 100 and standard deviation of 15. IQ scores two standard deviations below the mean is an indication of mental retardation ($IQ < 70$).

ToM: Theory of mind is the ability to attribute mental and emotional states to oneself and others and to understand that others have beliefs, desires, and intentions that are different from one's own [46].

HCI: Human computer interaction is the study of interaction between people and computers. This discipline is concerned with the design, evaluation, and implementation of computational systems centered around human interaction.

CCI: Child computer interaction is similar to HCI, but focuses on child specific design and interaction issues.

ROI: The region of interest denotes a specific defined area of the face for which gaze fixation data is analysed.

RFID: Radio-frequency identification refers to the use of RFID-readers and RFID-tags used for the purpose of the detection of physical tags by radio transmissions.

1.4 Organization of Thesis

Chapter 1 presents the problem space of this work and discusses the purpose of this research and defines terms used throughout the thesis. Chapter 2 presents prior relevant and prominent works related to the development of technologies for ASD, specifically for gaze behavior and expression recognition. The design of the new *Frame It* system is presented in Chapter 3. The methodologies used for the *Eyes Up* intervention are discussed in Chapter

4. Chapter 5 presents an analysis of results from the intervention. Chapter 6 contains a discussion of the analysis, a summary of findings, and future directions of this work. A list of references and appendix follow the final chapter.

Chapter 2

Review of Literature

This work presents two interrelated themes, the development of new technology and the subsequent evaluation of this technology as a educational and therapeutic tool. To better understand the full scope of this work we present a review of works related to the development of technologies for persons diagnosed with ASD, and a further review of studies that have investigated gaze behavior and/or expression recognition in individuals diagnosed with ASD.

2.1 Design Literature Review

In recent years there has been increased awareness of ASD by engineers and technology developers. Researchers continue to explore the use of technology to help detect, treat and enhance the ability of those with ASD [52, 32, 17, 31]. In particular, interactive software packages, video games, robotics, augmented objects, tangible user interfaces (TUIs), and interactive surfaces have been developed and researched for their efficacy.

2.1.1 Interactive Software Packages and Video Games

Treatment technology has primarily focused on the development of interactive software packages and video games that explore facial expressions and social scenarios. Notable software packages are *Mind Reading: The Interactive Guide to Emotions* [4] and *Transporters* [5]. Both software packages allow for the user to explore different emotions and their corresponding facial expressions. While the *Mind Reader* software is an interactive guide to emotions, the *Transporters* software is a rich animation based story telling package.

There is also a great deal of interest in the development of video games that focus on helping autistic children. Within this domain, games are focused on matching facial expressions or other social-emotional cues. The *Facesay* game [28] is an example of a game that augments facial features with a focus on matching expression types between different people. This type of game tries to help children recognize and generalize facial expressions. A smaller set of games focuses on creating specific styles of interaction to help keep the player on task while other experimental trials are carried out. These types of games are used as a research tool to investigate people's cognitive and perceptual skills [11].

2.1.2 Robotics

With the continued development of robotics there has been an increased interest in Human Robotic Interaction (HRI) and the use of robotics as teaching and therapeutic companions. Research has explored the use of robots as teaching aids for young children [38, 39] and explored their use for therapeutic purposes for those diagnosed with ASD [50, 16, 53]. Though there may be significant positive advantages for using robots as therapeutic devices for people diagnosed with ASD there are currently significant technical challenges, reliability, and cost concerns that must first be overcome before such approaches become practical for everyday unsupervised use.

2.1.3 Augmented Objects and Tangible User Interfaces

Augmented objects include toys that have been equipped with sensors for the purpose of capturing and recording interaction, while TUIs allow people to interact with digital information through physical objects. Though there has been a great deal of research in augmented objects and TUIs there has been relatively little work exploring the use of such items as therapeutic tools for those diagnosed with ASD.

Toys that are capable of detecting and recording different types of play have been developed to detect signs of ASD from toddlers' play [59]. While the aforementioned work allows for the capture of free and undirected play, and may prove helpful in detection of ASD, it is not intended to be interactive or therapeutic.

The work of Blocher et al. explores the use of multiple physical characters (plush dwarf dolls) that each embody a particular emotion [12]. The characters are used in an interactive manner with a digital story telling component. In the default case, the system starts with a video clip displaying a scene with a primary emotion for the child to identify and match with the appropriate doll. In this case, the dolls represent physical answers that the child can interact with when answering the emotion related questions of the digital story.

Paiva et al. [43] explore the use of a TUI doll (SenToy) that can be manipulated to change the emotional state of a corresponding video game character. The doll allows for different body gestures to be mapped to emotional states. This allows the child to explore the connection between body gestures and emotional states. Though the SenToy allows for interesting interactions and exploration of body gestures it does not include facial gestures, which are commonly understood to be the most important gestures for social-emotional signals, and a key problem in ASD.

Although the *Smart Jigsaw Puzzle* [13] is not intended as a therapeutic device, we felt that it should be mentioned since it is the most similar technology to *Frame It.* The *Smart*

Jigsaw Puzzle combines tangible and digital elements of interaction in the form of jigsaw puzzle construction. The work is not intended as a therapeutic or teaching platform, but simply explores new forms of interaction with a TUI. The relationship between the physical and digital representations are static, and they simply employ RFID technology to allow the user to scan puzzle pieces, thereby locating the next jigsaw piece to add to the puzzle.

2.1.4 Interactive Surfaces

Several works have explored the use of interactive surfaces [26, 45, 10]. These works have typically taken advantage of the DiamondTouch table [18], a multi-user touch sensitive tabletop with top-projection display. This technology offers many interesting possible uses and researchers have explored its use as a platform for encouraging social engagement with others. In particular, Piper et al. [45] have developed games that require two or more people to work together to achieve game objectives, thereby encouraging social interaction. Currently, a DiamondTouch table cost approximately $10,000, making it unattainable to most people.

2.2 Gaze behavior and Expression Recognition Literature Review

The gaze behavior and expression recognition abilities of people diagnosed with ASD have been well studied [7, 29, 33, 41, 44, 55, 56]. Unfortunately, these studies have not typically included young and severely challenged children. Furthermore, we know of no studies that have implemented interventions to study changes in gaze behavior and expression recognition in this population. The following details several of the relevant and prominent studies.

2.2.1 Reading the Mind in the Eyes

The influential work of Baron-Cohen et al. [7] studies the differences between HFA and NT adults. Participants across all comparison groups had an average age above 20 years and all had normal IQ scores. Participants were presented with a series of 25 photographs of the eye-region of the face of different actors/actresses and asked to choose one of four words that best described what the person in the picture was thinking or feeling. Images presented depict the six "basic emotions"[1] [22], and complex mental and emotional states such as bewildered or ashamed. Their findings show that HFA score significantly lower at recognizing other's internal states from information surrounding the eyes than NT individuals. It should be noted that a prior work of Baron-Cohen et al. [8] has shown that HFA are able to detect basic emotions in the whole face.

2.2.2 Visual Scanning of Faces

The work by Pelphrey et al. [44] represents the first work to explore the visual gaze behavior of autistic and nonautistic individuals in response to face stimuli. The study was comprised of HFA and NT individuals with normal IQ scores. Participants in the study were shown images from the well known set of Ekman and Friesen black and white pictures [23] of facial affect, during which eye movements were recorded. In addition, during one phase of the study participants were asked to determine what expression the person in the image was depicting. Results found a significant difference in gaze behavior, with autistic individuals devoting less time to core features of the face and the eyes in particular. There was no difference found in expression recognition, similar to other studies that have compared HFA and NT expression recognition of "basic emotions" [2, 8, 58].

2.2.3 Naturalistic Social Scenes

Recent investigations [33, 41] of gaze behavior have focused on "naturalistic social scenes." These works have used segments of social interaction from the 1967 film "Who's Afraid of

[1]The six basic emotions are fear, sadness, anger, surprise, disgust and happiness

Virginia Wolf?" or have made their own social scenario films. The work by Klin et al. [33] was the first work to investigate gaze behavior of HFA and NT individuals when viewing videos of social scenarios. Similar to works with static images, their work found significant differences between the HFA and NT groups with the HFA group viewing the eyes significantly less than the NT group.

Norbury et al. [41] have extended the above work by investigating the relationship between language impairment, ASD and gaze behavior. They implement a similar methodology to [33], using an eye-tracking system to record eye-movements while viewing social video stimuli. They conclude that the HFA without language impairments group spent more time fixating on the mouth region than the NT group. In contrast, there were no differences in gaze behavior between the HFA with language impairments group and their matched NT peer group.

Their data set demonstrates that differences in gaze behavior is aligned with language abilities, with increased attention to the mouth region by those with communicative competence.

2.2.4 Face Processing

The work of Volkmar et al. [58] does not investigate gaze behavior or expression recognition, but it does share some similarities to this work. Their study examines the ability of ASD participants to use the human face as a source of information. All participants in their study met DSM-III criteria for autism, and ranged in age from 9 to 18 yrs ($m = 13.7$, $s.d. = 3.2$). Full scale IQs of the participants were determined using several standard assessment instruments; Kaufman Asessment Battery for Children, Wechsler Intelligence Scale for Children, and Wechsler Adult Intelligence Scale. The mean IQ score was determined to be 36.8 (15.6). Participants were age, sex, IQ matched with individuals without ASD.

Their experimental task was designed to be "...administered to even low functioning subjects." They chose to investigate the above using puzzles because of the general observation that autistic individuals perform best on tasks involving object-assembly. Puzzles were constructed from images of human faces and differed with respect to familiarity of the face, configuration (normal vs. scrambled face) and were of varying complexity. They concluded that individuals diagnosed with ASD do process and use information of the human face.

Chapter 3

Design

Designing therapeutic technologies for children diagnosed with ASD presents many unique challenges. One not only has to consider that children often act, react and communicate differently then adults, but that children diagnosed with ASD often act, react and communicate differently than NT children. Our study aims to develop a therapeutic technology for children diagnosed with ASD, that is not limited to those with good verbal skills or normal IQ. This study has targeted young, challenged children. It was common for participants to be routine oriented, experience anxiety when encountering unfamiliar people, not respond in typical manners both physically and verbally, exhibit restricted or repetitive behaviors (stereotype), or require assistance from teachers or therapist to perform certain tasks. These and related factors make data gathering and intervention execution challenging. In part, these difficulties have led to a bias in ASD research towards the use of so-called "high" functioning individuals. It is important that we meet the challenges posed by the complex nature of ASD and develop methods and practices that facilitate understanding and helping the full spectrum of ASD.

To help achieve this goal we have developed *Frame It*, a blended tangible-digital puzzle game. We have taken a user-centered design approach with the inclusion and assistance of other stakeholders including teachers, therapists, psychologist, and educational technology researchers. The resulting system can be used for many types of puzzle games by children with a vast range of abilities. In the following section we describe the overall system design and subsequent usability testing.

3.1 Overall Design Considerations

The heterogeneous nature of ASD results in a wide range of cognitive and physical conditions that must be considered. Along with social-emotional difficulties, people diagnosed with ASD often have cognitive, vision, hearing, and motor impairments [35]. Additionally, it is important to consider the overall safety of the game, the ability of the game to engage and maintain a child's interest over many sessions of use, and the practicality of the system. Below we discuss design considerations in terms of the user, and the physical and digital components of *Frame It*.

3.2 User Design Considerations

A design objective for the *Frame It* system is to be usable by children 5 and older, since early therapeutic intervention in autism has been shown to have significant positive influences [51]. Given this goal the design must account for the differences between children and adult HCI [48] and the complex nature of ASD. The following represent primary considerations for users of *Frame It* and were used to guide the overall design of the system.

- **Cognitive Load:** The game should leverage prior knowledge and minimize cognitive load
- **Vision:** Both physical and digital representations must be large enough for people with poor eye sight
- **Hearing:** The audio system must account for a broad range of hearing abilities including audio sensitivities

- **Motor:** Physical and digital components need to account for reduced fine and gross motor skills
- **Personalization:** Modification of game elements and actions should allow for special interests and personal sensitivities.

3.3 Frame It Design Considerations

The *Frame It* system was designed as a computer accessory, in part to account for those that have little experience with computers, or have not previously shown interest in using computers. We felt that for some, the *Frame It* system could actually function as an introduction to computers and more complex technologies. In addition, the accessory model allowed us to meet a particular desire to meld physical and digital elements with a high-resolution display. The display is important for allowing us to expose the children to large, vivid images of peoples eyes. We felt the inclusion of a such a display into the *Frame it* platform would have reduced the durability of the system, increased safety concerns, and added cost.

Because of the wide range of physical and cognitive abilities of children with ASD it is important to emphasize game safety. Participants may put game components in their mouth and chew on them, throw game pieces, engage in restricted or repetitive behaviors while playing, drop game objects, step on game objects, spill liquids on game parts, or otherwise interact with the game in ways that could be undesirable for the child or the game components. In addition to safety concerns, we feel that any practical technologies to mediate therapeutic effects for children with ASD must be within financial means of as many people as possible. With this in mind our design considered the following:

- **Safety:** All puzzle components must be non-toxic, large to prevent swallowing, and without sharp edges
- **Durability:** All puzzle components must be rugged and water resistant

- **Replicability:** Puzzle pieces should be able to be made or replaced at little cost or effort

- **Customization:** Physical and digital elements should allow for customization

- **Cost:** Puzzle components should be inexpensive

3.4 Details of Design Considerations

Given the above design considerations, field observations, prior design work on other assistive and learning technologies [39, 38, 34], and preliminary consultations with an occupational therapist who works with children on the ASD spectrum, we developed the initial version of the *Frame It* game as an integrated tangible-digital interface (See Figure 3-1). The physical game board and puzzle pieces are sturdy and allow for physical free play while containing sensors that establish real-time connections to the digital representation. During play the game requires the user to correctly match and construct a target puzzle and then choose the appropriate expression label for it. The following sections describe how we accounted for the design considerations, a detailed description of the physical and digital components, explanation of game data, and an overview of game play.

Figure 3-1: *Early prototype of physical and digital components. The physical puzzle pieces and game board (left) and corresponding digital elements (right).*

3.4.1 Reducing Cognitive Load

The acquisition of new knowledge can be limited by cognitive load [57], the process of overwhelming working memory during a learning task. This overload can occur because of excess information in the form of learning material or game interface. For this reason we have attempted to leverage the child's prior knowledge and create a simple interface.

Observations from parents, teachers, therapists, and researchers suggest that many children diagnosed with ASD enjoy and are skilled at puzzle construction [58, 45]. Independently, puzzle construction and facial expression recognition tasks are a common part of the learning and therapy curriculum for our targeted population. For these reasons we decided to implement a puzzle game based on recognizing expressions from information surrounding the eyes.

The puzzle was implemented in a tangible form to capitalize on the children's familiarity with puzzles and allow for engaging, multi-modal interaction. The use of tangibles can help people leverage prior skills they have developed [58] and reduce the cognitive load of a task. The physicality of the system also accounts for the fact many participants, and people diagnosed with ASD, have little or no previous experience using a computer, and therefore, a purely digital interface would increase cognitive load. Additionally, we attempted to keep the physical and digital interfaces as simple as possible to reduce the control-interface option space.

3.4.2 Vision Motor and Hearing

Design motivations for tangible elements were also influenced by the need for ease-of-use by children with poor vision and motor skills. Since the game requires the player to select the correct puzzle pieces from many incorrect puzzle pieces a purely digital representation would force the use of a limited amount of monitor display space and require the pieces to be very small or would require a scrolling interface to view all possible pieces. Additionally, it has been shown that young children and people diagnosed with ASD have difficulty using the mouse as a computer interface [25, 22]. By having the puzzle pieces be physical we were able to make the pieces large; this made viewing the pieces easier for people with poor vision, and allowed for easier manipulation. In addition, consideration was given for children with sound sensitivities and all sounds are configurable and/or optional.

3.4.3 Personalization Through Customization

Many children have a particular special interest which can be very motivating for them [37]. Additionally, it is not uncommon for children diagnosed with ASD to have sensitivities or aversions to particular stimuli. For these reasons we have attempted to make the physical and digital components of the game easy to modify and customize to the personal interest of the user. In particular, we found it important to allow the children to choose a reward character of their liking to accompany game interaction (See Figure 3-2).

3.4.4 Safety and Durability

As discussed above, consideration to safety should be part of any design methodology for devices for children. This is particularly important when developing for children with cognitive and physical impairments. The game board and puzzle pieces (See Figure 3-1) are made from wood for its combination of light weight, durability, water resistance, and non-toxicity. The game board houses all electronic hardware (See Figure 3-3) and is sealed with a non-toxic glue. Any sharp edges on the game board or puzzle pieces were rounded and made dull. All puzzle images where printed on paper and attached to the wooden puzzle pieces using a non-toxic glue.

3.4.5 Replicability and Cost

Because of the great need for practical solutions we feel it is important to consider the cost of therapeutic technologies. Since cost can ultimately reduce efficacy by making treatment

Figure 3-2: *Before game play starts users are able to select a special interest reward character from a screen of characters that interest them.*

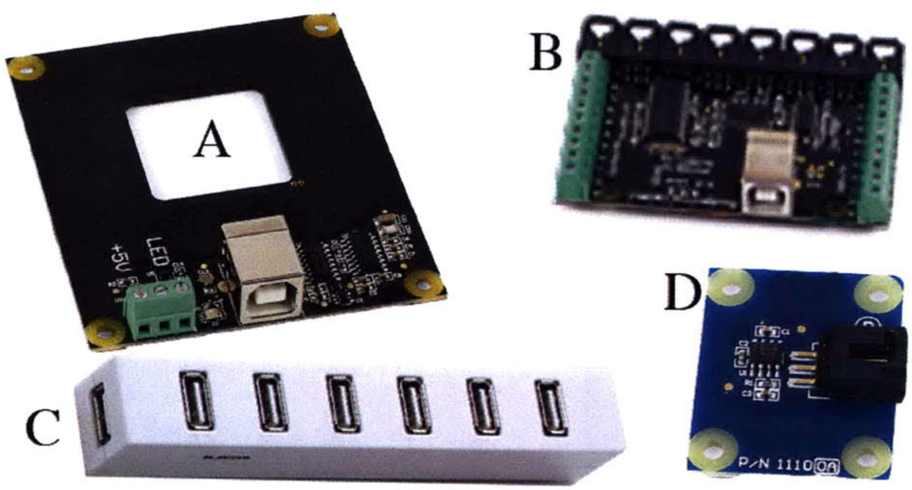

Figure 3-3: *Hardware components used for game board. Three Phidget RFID readers (A) one Phidget 8/8 interface board (B) and four Phidget capacitive touch sensores used, along with one USG hub (C).*

unavailable to wide segments of the population we feel any practical solution will have to consider cost.

The final design required approximately $300 for all physical components and ultimately could cost much less. We recognized that the puzzle pieces themselves were the most likely part of the system to be lost or broken, so we made each puzzle piece inexpensive, less than one US dollar and easy to replace if lost or broken. Puzzle pieces can be remade by simply taping an inexpensive passive RFID tag to a cutout piece of cardboard and gluing the appropriate image to it.

3.5 Details of Game Board and Puzzle Pieces

The *Frame It* game board and puzzle pieces are made from wood because of its low cost, durability, water resistance and aesthetic qualities. The game board is constructed from 1/4 inch plywood with a joint and layer system for structural integrity and to house electrical hardware components. The 3-D CAD software SolidWorks was used for the design of the game board (See Figure 3-4). This allowed for modeling the design and considering the internal arrangement of hardware components.

From the CAD design the parts were fabricated using a laser cutter to make precision cuts. The construction of the game board was sealed with a non-toxic glue to increase strength and create a water resistant surface. The final game board design measures 9x11x1 inches. The puzzle pieces were constructed from 1/8 inch wood and are rectangular in shape.

3.5.1 Hardware

The *Frame It* hardware architecture is composed of: three radio-frequency identification (RFID) readers, four capacitance touch sensors, one digital/analog interface, one USB hub, and RFID passive tags (See Figure 3-3)

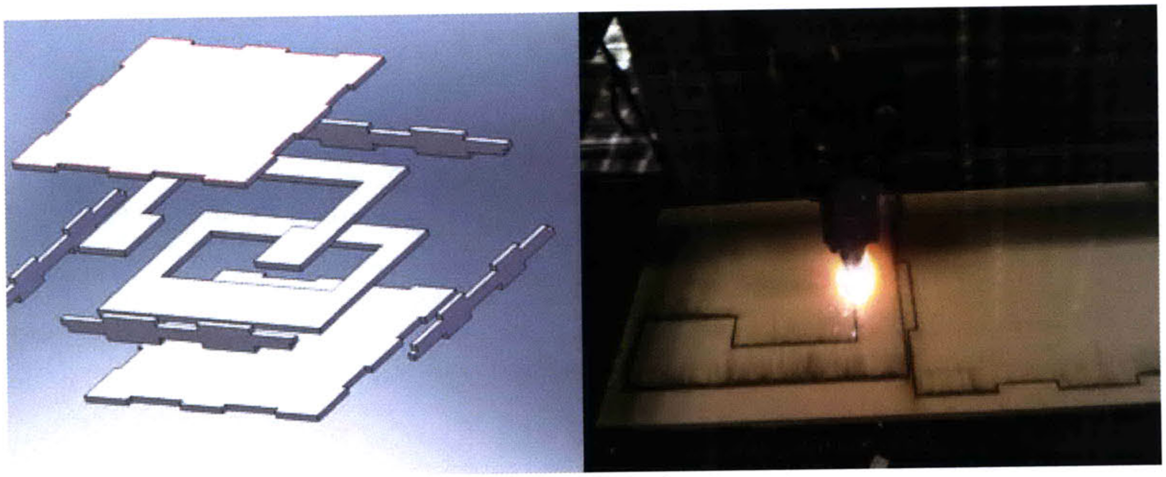

Figure 3-4: *3-D view of CAD design used for construction of game puzzle board (Left), with laser cutting of wood used for construction of game board (right).*

The RFID readers, digital/analog interface, touch sensors, and USB hub are housed inside the game board, while each puzzle piece contains a unique RFID passive tag. The RFID readers are able to detect the unique identification of a RFID passive tag. Because the RFID readers are used in close proximity it is necessary to reduce the signal strength of the radio signal and modulate their power. If this is not done the RFID readers will interfere with one another and there will be false positive detections for relative tag localization and a general lack of system robustness. With the current system there is a high degree of localization accuracy and a maximum detection delay of 200ms due to RFID power modulation.

To allow for a direct interface to the digital components of the game the physical game board was outfitted with touch sensitive buttons. To reduce external parts, and the possibility of breakage, four capacitive sensors mounted within the game board were used for detecting the users touch. The four touch sensitive buttons allow for a mapping between the physical and digital.

The USB hub and digital/analog interface allow for the internal hardware components to be connected, powered, and interfaced to an external computer.

3.6 Details of Software

The software architecture comprises two main components. The first is the hardware interface. This interface communicates with and controls the hardware system. The second component manages the game and renders the digital representation of the physical interaction. All physical puzzle pieces have been represented in digital form and are dually constructed as the physical puzzle is constructed. The system runs on Linux, Macintosh, or Windows XP/Vista operating systems.

3.6.1 Game Data

Each puzzle is composed of three equal sized parts (See Figure 3-5) from which a complete puzzle is constructed. The completed puzzle measures 8×3 inches, a size deliberately chosen to be slightly larger than a typical adult face. Since expression labelling is part of the game, and a focus of the subsequent intervention, we used images of people showing specific expressions. Pre-existing data sets [4] of labelled eye regions exist, but were low resolution and could not be reproduced in life size. For this reason it was necessary to construct a new data set of high resolution color images (See Figure 3-6) to produce realistic eye puzzle pieces. The new data set was collected over a three month period and consists of over 1000 high resolution color images (See Figure 3-7). Individuals were asked to demonstrate one of twelve expressions: fear, anger, disgust, happiness, sadness, surprised, thinking, disagreeing, interested, concentrating, agreeing and confused. The first six expressions makes up the six culturally common expressions [22] and the second six are common mental states [4, 19].

To determine if a photographed expression was a “true” example of desired expression we conducted independent coding of all images. Extensions to the online video meta-labelling program VidL [20] were developed to allow for the meta-labelling of still images. Using VidL we recruited 10 NT individuals to code all images. Images were cropped such that only the eye region used in the puzzle was presented to coders. Coders were presented with four label choices for each image. The expression choices included the name of the expression the photographed person was asked to poise, and three other randomly chosen expressions from the remaining five expressions for the particular category (basic expression vs mental state). Expression choices were randomized programatically . All coders were presented with the same images and expression choices. Only those images with 90% or greater inter-coder agreement were used for game puzzles.

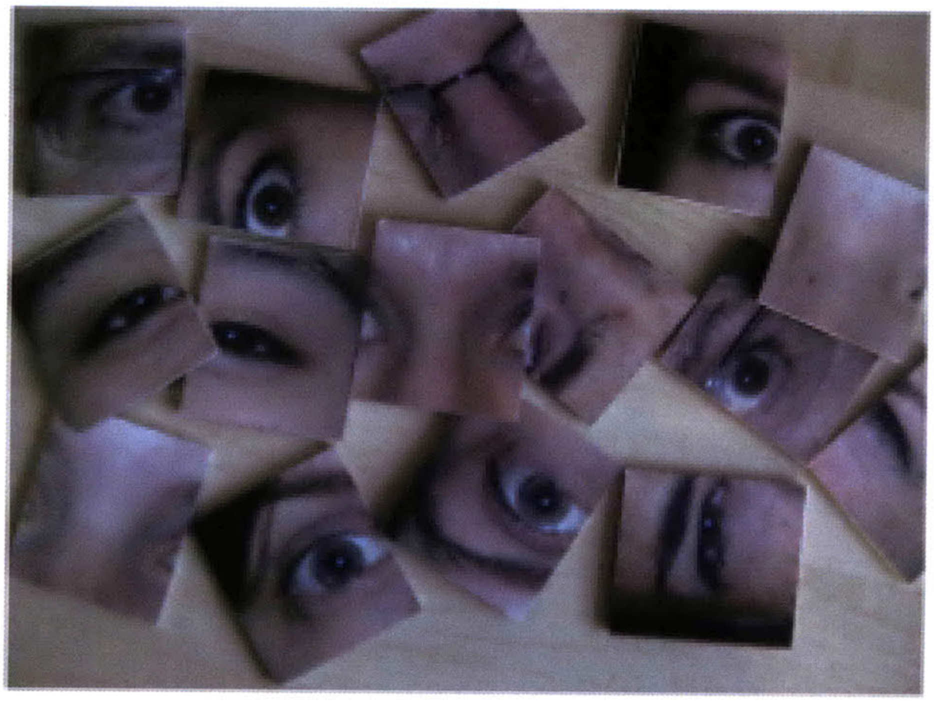

Figure 3-5: *Puzzle pieces from several eye region puzzles*

3.7 Overview of Game Play

Game play is controlled by both the physical and digital components of the game. The software controls the flow of play, provides feedback, and records all game related performance data, while the physical pieces control digital visual objects and player controls.

When *Frame It* is launched the player can login by selecting a picture of themselves (See Figure 3-8) , after which they can select a game reward character of their own choosing (See

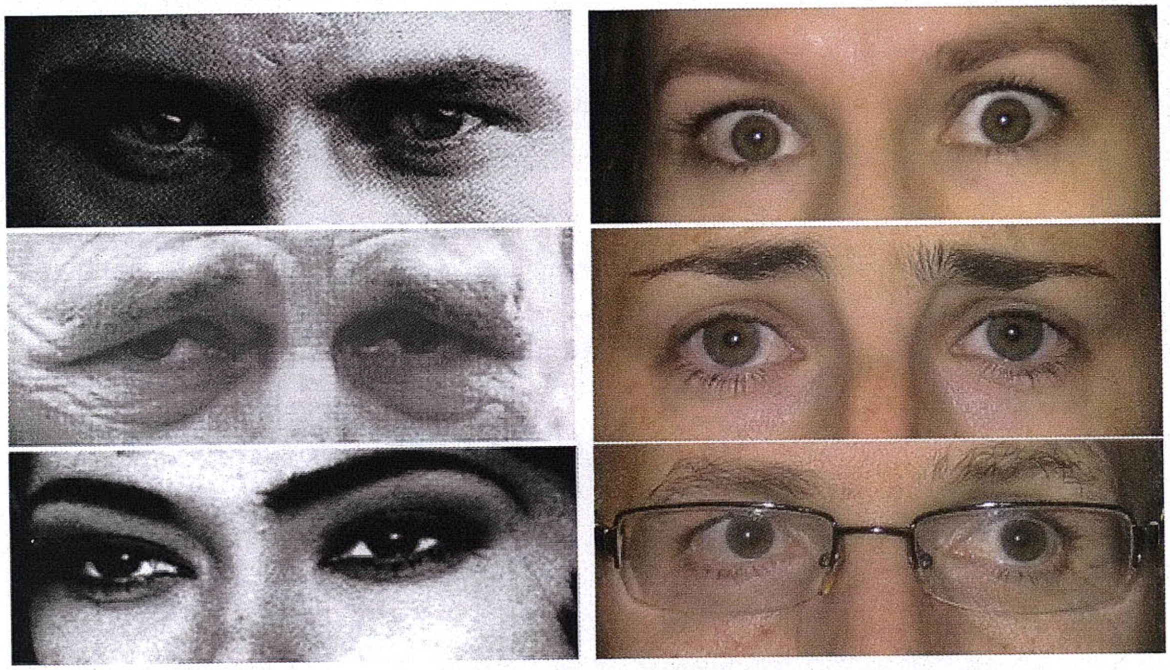

Figure 3-6: *Left column images are low resolution and become pixelated when scaled to life size. Right column images are high resolution images collected for this work.*

Figure 3-7: *Example of images gathered for making the eyes data set. All images were subsequently cropped to reveal only the eyes.*

Figure 3-8: *User login selection screen. The above are generic login icons that can be replaced with specific images of users.*

Figure 3-2) that will be displayed when the user correctly constructs a puzzle, or chooses the correct puzzle facial expression label. Once this has been done the puzzle game starts. The player is presented with a digital image of a particular eye region, i.e. the target puzzle (See Figure 3-9). This is used as a reference, which the child may consult while trying to find the physical puzzle pieces that correspond to the target puzzle. Once a complete puzzle has been constructed the facial expression label buttons become active and the player can select an appropriate label. Although there can be many possible selections, the player is limited to four possible selections–each one corresponding to a particular physical button on the game board. The corresponding digital button matches the physical buttons color and contains the word or image associated with the choice. If the puzzle has been constructed and labelled correctly a reward animation is played with the chosen reward character. After completion of the reward animation the next puzzle is presented.

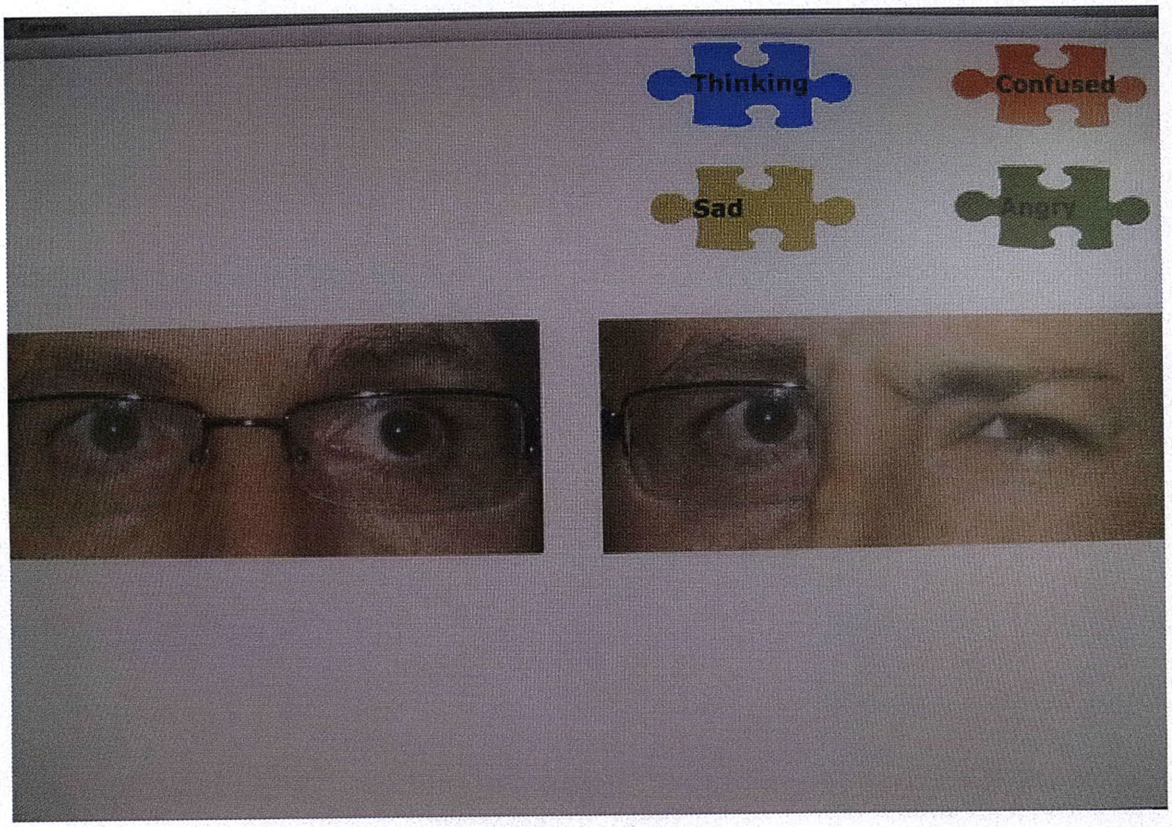

Figure 3-9: *Early prototype of digital interface with target puzzle. In this version of the sytem the target puzzle was presented on the left and the constructed puzzle appeared on the right. Notice that the constructed puzzle is not correct.*

3.7.1 Feedback

The system provides visual and auditory feedback as the user plays. As the player places the physical pieces on the game board they are then visually represented in the corresponding location in the digital interface presented on the computer display. In addition, there are automatic visual and auditory responses corresponding to the correctness of the puzzle piece used. If the piece used is the correct part in the correct location there will be a green bar placed beneath the puzzle piece image and a specific bell sound. On the other hand, if the piece is not correct or in the wrong location there is a red bar placed beneath the puzzle piece image and a specific bell sound.

When a button is touched to select an expression choice there is a ding sound to let the user know that their selection was recognized. Following the ding sound, the corresponding word that the button selection represents is played aloud. All hints, rewards and encouragements are presented with accompanying visual and or audio components. All visual and audio feedback features can be disabled to make play more difficult.

3.7.2 Recorded Play Data

During a game the player can construct many puzzles. The system records the following data for each puzzle: the correct puzzle facial expression label, the player selected facial expression label, puzzle start and stop time, total number of puzzle pieces placed on board, number of correct puzzle pieces placed on board, the correct puzzle configuration, and the player's final puzzle configuration. All data is saved to a text file. From these, data inferences can be made as to how quickly and accurately puzzle construction and labelling is taking place for a user.

3.8 Usability Testing

Because we are particularly interested in using *Frame It* as a tool for investigating ASD we paid special attention to design and usability issues specific to this community. It is our feeling that these highlighted aspects not only make *Frame It* more accessible to our target group, but to all players. As part of the design and usability development we conducted tests with a group of children diagnosed with ASD and a group of NT children. The following details our usability testing.

3.8.1 ASD Usability Testing

In total eight people with a diagnosis of ASD took part in our study. There were six male and two female participants. The average age of participants was 12 years and ranged from 7 to 20 with a standard deviation of 4.3 years. Only two of the participants performed at their age on educational evaluations for communication, visual and motor skills. The remaining six participants had performance scores, similar to 2-7 year olds, on communication, visual and motor skills. No participant performed above the 8 year old level on a developmental test. This difference between developmental and chronological age must be accounted for in the design process and was a factor in our decision to follow principles from CCI [8, 21]. In addition, comorbidity was present in all but two participants. Conditions included: anxiety disorder, restricted and repetitive movement, attention deficit hyperactivity disorder, obsessive compulsiveness, global development delay, bi-polor disorder, and mild metal retardation.

Tests were conducted at the Groden Center, a school in Rhode Island that provides services to people diagnosed with ASD and other developmental disabilities. Tests were administered in a therapy room that the participants were familiar with. Each participant was accompanied by his or her familiar teacher or therapist throughout the usability test. The author of this thesis and a staff therapist that the participants knew well performed the usability session.

Sessions were held during the school day and required a teacher or therapist to accompany the student. For this reason usability sessions were limited to approximately twenty minutes. Each testing session included a short introduction between the participant and researcher, during which an explanation of the game was given to the participant and their companion. After the introduction the participant was allowed to play *Frame It* while observations of the session were recorded. Participants played for approximately fifteen minutes, though two participants insisted on playing for nearly thirty minutes. After the session a post-use interview was conducted to further explore the overall session.

Because many children on the spectrum have low language skills or otherwise atypical communication patterns it is necessary to either know the person well or have persons that know the child well to facilitate communication. For many of the children in this study their teacher or therapist acted as a translator between the child and the researcher during post-use interviews, thereby facilitating the communication of the child and relating the experience of the child to the researcher. This is an imperfect method, but was the only option that we had for this population.

During usability test a set of criteria (See Table 3.1) were used to help guide observation towards specific design considerations. In addition, during game play *Frame It* recorded user play data (See Table 3.2), which was used to make further inferences of the overall usability of the system.

3.8.2 NT Usability Test

Although the goal of this work is to make a practical therapeutic game for people diagnosed with ASD, we recognize that the game platform may have general benefits to others. In addition, there are many similarities between children, regardless of mental and physical abilities, and we felt that we could gain further insight into the usability of *Frame It* by including NT children in our study. It also provides us with a base reference of NT interaction and performance for making further comparisons.

Both ASD and NT children may experience similar issues using the game, but the NT children may be able to express these issues more effectively. Usability sessions with the NT group were conducted at MIT and at the personal residences of several participants. Detailed observations where taken during usability sessions with *Frame It*, game performance data was recorded, and post-use interviews were conducted with all participants.

In total eight NT individuals took part in our study, four male and four female participants. The mean age of the participants was 11 years and ranged from 6 to 18 years of age with a standard deviation of 5.1 years. All eight participants were enrolled in standard or advanced classes, had no known physical or psychological conditions and were assumed to have normal or above normal intelligence.

The author of this thesis conducted all usability sessions with the NT group. Sessions were approximately 20 minutes long and included a short introduction and explanation of *Frame It*, after which the participant was allowed to play Frame It while observations of the session were recorded. After free play with *Frame It* a post-use interview was conducted to further explore the overall session.

The same observational criteria and game play data recording used for the ASD group was used for the NT group (See Table 3.1 and 3.2).

3.9 Usability Results

The physicality of *Frame It* was found to be particularly engaging for participants, and unlike most physical games it allows for the ability to track game performance, providing an important quantitative measure for assessment. Participants in both groups reported that they enjoyed playing the game and liked the connection between physical and digital. The following is an overview of the qualitative and quantitative data gathered during testing.

3.9.1 Qualitative Results

During each session observation notes were taken and a post session interview was conducted to assess the usability of the system. Table 3.1 reports the percentage of respondents for which the response was true. Overall, users in both groups enjoyed being able to choose a reward character and found the game experience enjoyable. Notice that 63% of the ASD group required assistance constructing the puzzle. This assistance was typically only for the first three puzzles or when trying to find the nose region, which both groups had difficulty with. Additionally, several participants in the ASD group did not provide clear feedback for each criteria and were classified as not applicable. Several participants in the ASD group became fixated with the textured surface of the game board and teachers felt that extra textures and visual stimulus would be problematic for some students that easily fixate on such items. Observations also revealed both groups found the horizontal orientation of target puzzle and constructed puzzle to be unintuitive.

3.9.2 Quantitative Results

During each session game metrics were recorded (See Table 3.2) which provided insight into the usability of *Frame It* and revealed differences between the ASD and NT groups. In particular, independent-samples t-tests were conducted to compare the differences between groups with respect to total number of puzzle pieces placed per puzzle, the correct number of puzzle pieces placed per puzzle, and the correct labelling of expression for all puzzles played.

There were statistically significant differences in the total number pieces used per puzzle between the ASD group, $m_{asd}8.1$ (1.9), and the NT group, $m_{nt} = 4.5$ (1.3), with $t(89) = 6.27$, $p < 0.01$. From observational data the ASD group tended to place pieces on the game board and then use the digital representations to check correctness, while the NT group would often consider the correctness of a puzzle piece before placing it on the game board.

There were statistically significant differences in the number of correct puzzle pieces placed per puzzle between the ASD group, $m_{asd} = 4.4$ (1.3), and the NT group, $m_{nt} = 3.0$ (0.7), with $t(89) = 3.83$, $p < 0.01$. Recall that each puzzle required three pieces to complete, and if the wrong pieces were used the puzzle would have fewer than three correct pieces used. These results suggest that the ASD group required more movements of the puzzle pieces to find the correct puzzle configuration.

There were also statistically significant differences in the total number of correct expression labels per session between the ASD group. The mean percentage correct for the ASD group was $m_{asd} = 0.53$ (0.31), and the NT group, $m_{nt} = 0.75$ (0.18), with $t(89) = -2.20$, $p < 0.05$. These results are consistent with other research that shows individuals with ASD have more difficulty than NTs correctly labelling expressions from peoples eyes region [6].

Another interesting result was the similarity in time to complete a puzzle between both groups, $m_{asd} = 1.5min$ (0.5), $m_{nt} = 1.6min$ (1.3). This suggest that object manipulation by the ASD group was not a problem, since they placed more pieces in less time then the NT group. It also suggest that the two groups used different strategies for solving the puzzles. The ASD group tended to use the digital representations to make selection decisions, while the NT group considered the correctness of pieces before placing them on the game board.

3.10 Summary of Frame It System

As a result of usability sessions and multiple prototypes we addressed many early shortcomings and developed a better system. In particular there were changes between early prototypes and the final system that made interaction easier and better facilitated users interactions. In particular, the game board was redesigned and aesthetic etching was removed to reduce tactile fixations (See Figure 3-10). Changes to the digital interface were also implemented, with the target puzzle and construction area now aligned vertically instead of horizontally and expression choices were not made visible until after the puzzle was constructed (See Figure 3-12). In addition, there were changes to the sounds associated with the game and added visual cues for hinting related to the puzzle's correctness.

Table 3.1: *Usability criteria and responses used for qualitative assessment of interaction. Values reported are the percentage of respondents that the question was true for.*

Criteria	**True**	
	ASD	**NT**
Liked choosing companion character	1.00	.88
Enjoyed puzzle pieces appearing on screen	0.63	1.00
Understood puzzle play by third trial	0.88	1.00
Understood expression labelling by third trial	0.63	1.00
Required help constructing puzzle	0.63	0.00
Scanned between physical and digital representations	0.75	1.00
Difficulty manipulating pieces	0.13	0.00
Responded positively to sound elements	0.75	1.00
Able to play with upto 30 puzzle pieces	0.50	1.00
Distracted by interface	0.38	0.00
Difficulty with controls	0.13	0.00
Enjoyed overall experience	0.63	1.00

Through inclusion of children, teachers, therapists, psychologists, and educational technology researchers, along with special attention to design considerations, we have been able to develop and explore the usability of this technology. Usability sessions revealed that both participant groups were able to use *Frame It* and enjoyed the game. Furthermore, our exploration of usability with the ASD and NT groups revealed both commonalities and group specific issues, while consultations with stakeholders helped us identify more specific usability concerns for our ASD group. Changes made resulted in a final version that was easy to use and engaging for a wide range of children (See Figure 3-12). Overall our approach of including a wide range of stakeholders and diverse groups has helped us refine the systems design and ready it for further gaze behavior and expression recognition studies.

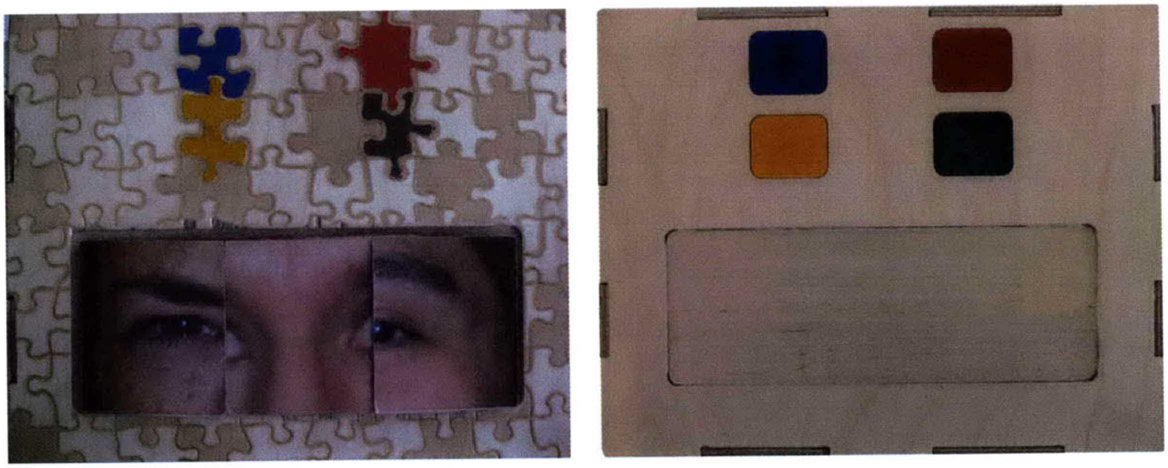

Figure 3-10: *Early prototypes of the game board (left) included textured etching that was found to be distracting to some children, a later design (right) has no etching.*

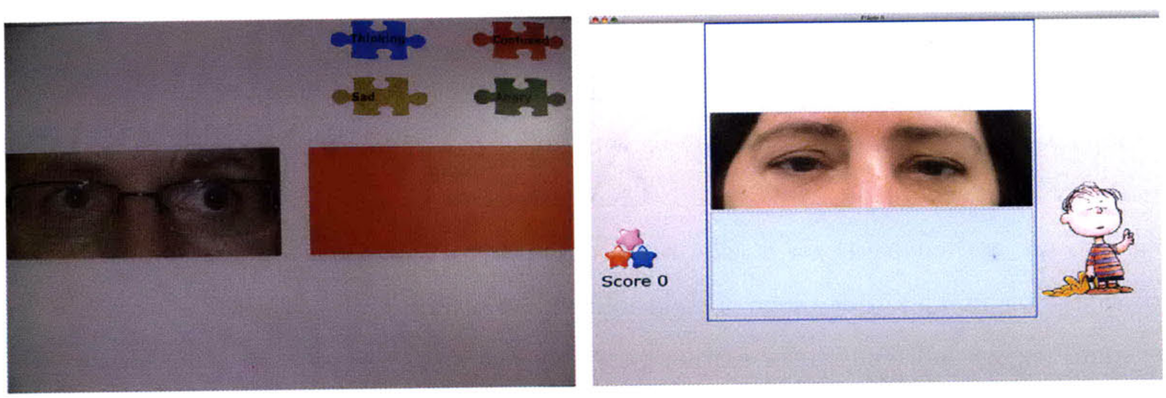

Figure 3-11: *Early prototypes of the digital interface (left) included horizontal alignment of target puzzle and construction area that was found difficult for some users, a later design (right) uses vertical alignment. In addition, expression labels only appear after puzzle construction in the new interface and text and picture icons appear for label choices.*

Table 3.2: *Usability performance metrics used for quantitative assessment of interaction. Mean and (standard deviation) reported for usability session performance fro group of size* $N = 8$

Performance Metric	ASD	NT
Number of puzzles completed	5.9 (3.5)	5.5 (1.3)
Number of puzzles completed in correct spacial order	4.9 (2.9)	4.9 (1.6)
Number of puzzles correctly labelled	3.1 (1.8)	4.1 (1.0)
Number puzzle pieces used	8.1 (1.9)	4.5 (1.3)
Number correct puzzle pieces used	4.4 (1.3)	3.0 (0.7)
Puzzle completion time in minutes	1.5 (0.5)	1.6 (1.3)

Figure 3-12: *The final design of the puzzle board and a mixture of puzzle pieces for several puzzles (left) and digital representations (right) of target puzzle (top) and corresponding constructed puzzle (bottom). This design resulted from several iterations of prototypes.*

Chapter 4

Hypothesis and Methods

With the successful development of the new *Frame It* system our aim is to evaluate its use as a play-centered educational and therapeutic tool. The following sections present our hypothesis, the methodology of our intervention, *Eyes Up*, and describe the experimental procedures.

4.1 Hypothesis

Engagement in play-centered activities centered around the construction and labelling of people's expressive eyes region will increase attention to other people's face and eyes region and will improve expression recognition ability.

4.2 Methodology Overview

The *Eyes Up* intervention seeks to explore how playful interaction with puzzles of people's eyes affects the participant's social gaze behavior and ability to recognize other people's facial expressions. A new data set composed of high resolution images of people's eyes was constructed for this work (See Section 3.6.1). This allowed us to make slightly larger than life-size puzzles of people's eyes. We felt that play-centered interaction with these large puzzles would help desensitize people to eyes, especially children diagnosed with ASD, who were averse to social eye contact with others. Furthermore, the play involved labelling the eyes with an appropriate expression. This aspect of the game was intended to help players become better at recognizing expressions from information surrounding the eyes.

The intervention was conducted with 10 children, all previously diagnosed with ASD, at the Groden Center in Providence, Rhode Island. No children in the study took part in the design or usability study, or otherwise had any previous interactions with the *Frame It* system. Before participation, all individuals and their parents gave informed written consent (See Appendix A-2, A-3). The study was approved by the Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects (See Appendix A-1). Additionally, the Institutional Review Board of the Groden Center approved all study procedures (See Appendix A-2).

As a goal of our work we have sought to work with children not typically included in the literature related to social gaze behavior and expression recognition. In part, this group is generally underrepresented throughout the autism literature, as the majority of studies have been conducted with HFA. The children in this study exhibited complex and challenging cognitive and behavioral conditions. It was common among participants to have impaired verbal ability, exhibit echolalia, be routine oriented, experience anxiety when encountering unfamiliar people, not respond in typical manners verbally or physically, exhibit restricted or repetitive behaviors, and/or require assistance from teachers or therapist to perform certain tasks. Because of these conditions, extra measures were taken when collecting data and working with the children to insure data validity and participant well-being. The following sections describe participants, pre- and post- measures and the intervention procedures.

Table 4.1: *Pre-intervention Intelligence and Autism Rating Scores*

Test	Mean	Standard Deviation
KBIT-verbal	44.90	9.37
KBIT-non-verbal	54.90	18.69
KBIT-composite	47.40	10.33
CARS	34.00	4.05

4.2.1 Participants

The study group comprised 10 children, all previously diagnosed with ASD, 8 male and 2 female. The average age was $m = 12.3yrs$ $(3.9yrs)$, with an age range from 7-18 years. Prior to the start of the intervention the Kaufman Brief Intelligence Test (KBIT) [27] was administered to all participants. An IQ score follows a Gaussian distribution with the mean of 100 and standard deviation of 15. An IQ score two standard deviations bellow the mean indicates mental retardation ($IQ < 70$). Average and standard deviations for the sample are reported as follows: KBIT-verbal $m = 44.90$ (9.37), KBIT-non-verbal $m = 54.90$ (18.69), KBIT-composite $m = 47.40$ (10.33) (See Table 4.1). In addition, the Childhood Autism Rating Scale (CARS) [54] test was administered. The average and standard deviation for the CARS test were $m = 34.00$ (4.05) (See Table 4.1). CARS scores of 15-29 are considered non-autistic, 30-36 is mild to moderately autistic and 37-60 is severely autistic.

4.2.2 Pre- and Post- Measures

To assess the efficacy of the *Eyes Up Intervention* we conducted pre- and post- test to determine if children changed their social gaze behavior and improved their expression recognition ability. We used the Tobii eyetracking system [1] to record the gaze behavior of children when exposed to images of people expressing different facial expressions. In addition, we used a number of qualitative measures to gain a better understanding of the children's ability, engagement, and progress.

Of the 10 participants 8 took part in pre- and post- measures involving eye-tracking and expression recognition. The remaining two had to be excluded from this part of the study because of difficulty obtaining data. Each of the two excluded participants had physiological and cognitive issues that made data gathering difficult. Both had eye issues that made recording eye movement difficult, they engaged in excessive movement, and were generally less-compliant. Although they were not included in pre- and post- measures they did take part in the subsequent intervention. All pre-test measures were collected over a two-week period prior to starting the intervention. Similarly, all post-test measures were collected over a two-week period following the end of the intervention. The following sections address special accommodations used and describe in more detail the quantitative and qualitative measures used.

Special Accommodations

It was necessary to make special accommodations for participants in the study to facilitate their participation. The most critical consideration was "taking the lab to them." This approach allowed the children to stay in an environment they were familiar and comfortable with. This is particularly important for this population as non-familiar environments and people can be highly stimulating and distracting, making data collection and assessments of true performance difficult. In addition, we found it very helpful to run pre-trials with children of similar ages and abilities that were not part of the study to determine possible issues that would be encountered. Teachers and therapist were also consulted for their expertise in working with and facilitating the children's abilities.

From these preliminary steps, a number of additional helpful approaches were adopted. The use of timers helped some children remain still and reduce anxiety by providing them with specific feedback as to how much longer they needed to maintain attention to the task. Breaking the overall test into segments of five minute or less allowed for short breaks and recalibration of the eye-tracking equipment. The use of special interest objects, such as calendars, facilitated participation. Music was even used for one participant who enjoyed the Greatful Dead, which calmed the child and allowed for better attention to the task. It was also important to have the teacher or therapist help administer the test, as they knew how best to communicate with the child.

Details of Quantitative Measures

To record participant's gaze behavior we recorded their eye movements while viewing still images of people displaying different facial expressions. Eye movements were recorded using the Tobii T60 eye-tracking system. The eye-tracking hardware is integrated into a 17" TFT flat panel monitor with $1280x1024$ pixel resolution, thus the system was unobtrusive for the participants. Eye tracking optics are capable of binocular tracking at $0.5°$ accuracy and sampling at 50Hz. Five point calibration was used for all participants. Participants were situated approximately 80cm from the screen. Similar to Pelphrey et al [44] we presented images from the well known set of Ekman and Friesen black and white images [23] of facial affect. Images are grey scale and consist of posed facial expressions of the six "basic emotions:" fear, anger, disgust, surprise, sadness, and happiness [21] .

We felt it was important that the children be familiar with the testing environment to reduce possible anxiety and distraction. For this reason all recordings were conducted at the Groden Center during normal school hours. Children were accompanied to all sessions by a teacher or therapist. All recordings took place in a quite room used for running experiments and administering tests, in which the children had previous exposure to and were comfortable with. The first author of this paper was the only unfamiliar person that the child associated with. To account for the children's physical and cognitive needs, all tests were broken into five minutes or less subsets; this allowed us to let children take breaks and to recalibrate the eye-tracking system if the child moved excessively. Following is a description of the three phases used for testing social gaze and expression recognition. Eye-movement data was recorded for all phases. Since there are multiple examples of each expression presented in all phases and because phase II and III use the same images we have randomized the image presentation sequence programatically. All images used for pre- and post- tests are from the Ekman and Friesen image set and the image presentation orders are randomized.

- **Phase I:** Participants were shown 12 images, with 1 male and 1 female face shown for each of the 6 basic expressions. Participants were instructed to look at the photograph in any manner they wished. Each image was presented for two seconds followed by a two-second interstimulus interval.

- **Phase II:** Similar to Pelphrey et al. [44] we showed participants 24 images, selected from the Ekman and Friesen images that had at least 90% inter-coder agreement. The 24 images were balanced for male and female, with 2 male and 2 female examples of each of the 6 basic expressions. Unlike Pelphrey et al., we presented images in four sets of six images. Each image was presented for a variable length of time followed by a two second interstimulus interval. Participants were asked to identify the portrayed expression from one of three possible selections (i.e. "how does this person feel, happy, angry or sad?"). All 24 selection sets contained one correct expression and two randomly chosen expressions from the remaining five false possibilities. Showing each image for a variable amount of time allowed us to take into consideration each child's communication differences. Additionally, this approach allowed us to check and recalibrate the system as needed.

- **Phase III:** Phase III is the same as phase II with two differences. First, all images used in phase II have been randomized as discussed above. Second, all images have been altered with an occlusion mask. This was done to determine if this would induce more fixations to the eyes region and if there would be differences in expression recognition between phase II and III. For each image the eye region, defined as the region of size $250x100$ pixels centered on the nose ridge and including the top of the eyebrows, has not been altered. An imagemask of color 0xFEFEFE (hexadecimal) and opacity 50% was applied to the non-eye region portion of the image, this reduced high frequency information while still allowing the entire face to be seen. (see Figure 4-1).

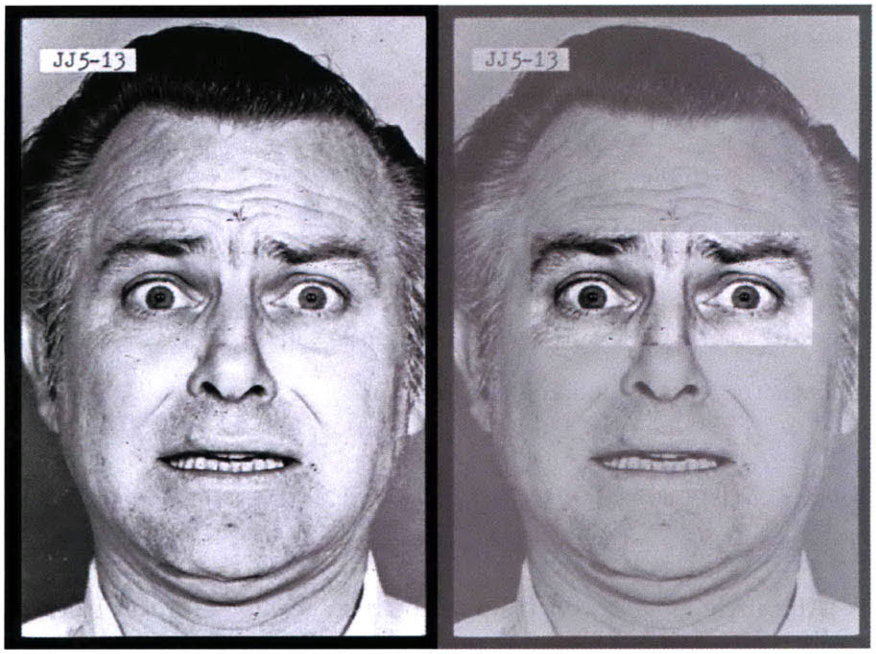

Figure 4-1: *Examples of images from the Ekman and Friesen image set used for phase II (left) and phase III (right).*

4.2.3 Intervention Procedures

After running all pre-tests we waited one week prior to starting the *Eyes Up* intervention. All intervention sessions were conducted at the children's school, the Groden Center, and held in the same room used for pre- and post- tests. Prior to the start of the study, teachers were consulted to determine the best schedule for running the intervention as part of the children's regular school day. It was determined that we should attempt to see each child three to four times a week for 10-15 minutes per session. This schedule was both not overly intrusive to the normal day of the teacher and student and allowed flexibility for when children were unavailable. In total, 11 sessions per child were carried out for each of the children over a 5 week period. The following section is an overview of how each session was conducted.

Session Details

Each session was run with one child, the child's teacher, and the experimenter in the session room. The experimenter would make sure that the *Frame It* game was restarted and that the correct user ID was selected. Once the user ID had been selected, the child was allowed to choose a reward character from a selection window (See Figure 3-2). After choosing their reward character the puzzle game started.

The puzzle game started by saying "let's puzzle" and showing the target puzzle (See Figure 4-2). For each session there were a total of 6 puzzles to construct, making up a total of 18 puzzle pieces. All puzzle images were from the dataset gathered for this work and described in section 3.6.1. Sessions were facilitated by the child's teacher, while the experimenter observed and took notes. All teachers were consulted prior to the start of the intervention and told that they should allow the child to try their best to construct the puzzle and then label it. If the teacher felt the child was struggling or otherwise getting frustrated, the teacher was instructed to facilitate the child's abilities towards solving the puzzle. This included reducing the number of puzzle pieces, pointing in the general direction of were a puzzle piece was, talking to the child about particular scenarios that involved the current expression, or making a similar expression and asking the child if they know how they felt. Once the puzzle was correctly constructed, the child was presented with four options for labelling the facial expression. The options always included one correct answer and three incorrect options chosen randomly from the remaining five possible expressions. The location of the correct puzzle label was randomly chosen each time. The options were presented in both text and picture (See Figure 4-2), using Boardmaker images [36]. When a particular selection was made the selected word was played aloud. If the selection was incorrect the game notifies the player,"almost, try again." When the child selects the correct response the word was played aloud and there was a short animation of the child's selected reward character. Only when the puzzle was constructed correctly and labelled correctly was the next puzzle presented. This process repeated itself until the session was over, typically lasting 10-15 minutes. Several times over the intervention a new set of puzzles were introduced so that children would not memorize images and their corresponding label. All game interaction for each student was recorded, for more detailed overview of game play see chapter 3.

Figure 4-2: *When a new puzzle is first presented only the target puzzle is shown (left). During construction the physical pieces placed on the board appear in the corresponding location in the digital interface (center). When correctly constructed the expression choices are presented with text and picture icons for choosing the expression label (right).*

Chapter 5

Results

The results of the methodologies outlined in chapter 4 are discussed in this chapter. We report results for all participants as a group and as individuals. Analysis of individual results reveals subgroups which are further investigated. Results for gaze behavior and expression recognition are presented first followed by game play statistics and correlation analysis. Throughout this chapter we report mean, standard deviation and t-test results. In addition, we use Cohen's d value, $d = \frac{M_1 - M_2}{\sigma_{12}}$, to relate the effect size for fixation data [15]. Values of 0.2, 0.5, 0.8 are considered small, medium and large effect sizes for Cohen's d value. Finally, correlation analysis is performed to investigate the relationship between KBITS and CARS scores and gaze behavior and expression recognition ability.

5.1 Gaze behavior and Expression Recognition Analysis

5.1.1 Group Analysis of Gaze behavior

As outlined in Chapter 4, 8 of the 10 participants took part in the eye-tracking and expression recognition test. Multiple attempts at data collection were conducted with both excluded participants. Unfortunately, a combination of physiological and behavioral issues made data collection difficult and we were unable to collect a sufficient amount of eye-tracking data for either person. For all others, tests were conducted in three general phases. During phase I participants were asked to view 12 images as they wished. During phase II participants were asked to view 24 images and report the facial expression of the person in the image. During phase III participants were asked to view the 24 images from phase II, which had been modified with an occlusion mask over everything but the eyes region (See Figure 4-1), and asked to report the facial expression of the person in the image. For all phases we investigate fixation behavior. There is not an agreed upon definition for what constitutes a visual fixation. Typically, a fixation is defined as the point of visual focus confined to a region of 30-50 pixels in diameter for at least 100 milliseconds. For this study a point of visual focus confined to 30 pixels in diameter for at least 100 milliseconds is used to determine a fixation. The total facial feature fixation time was calculated both as an absolute time (Face Fixation Time) in seconds and as a percentage of all image fixation time (Face Fixations). In addition, the relative percentage of fixation time to each predefined ROI is reported. Because the image display time during phase 2-3 was variable to allow for participant to respond to facial expression questions we have also included the image display time in seconds (Image Display Time).

Our approach to defining the ROI was similar to [33, 41], though we have chosen to divide the face into three main regions. Each ROI was predefined for each image using the Tobii eye-tracker ClearView software. The ROI are the upper face, eyes, and lower face. The eyes region is the region from each side of the face horizontally and from the tip of the nose to the top of the eyebrows vertically. The remainder of the face above the eyes region is the upper face region, and the face region below the eyes region is the lower face region (See Figure 5-1).

To analyse participant's gaze behavior for phase I the mean percentage of fixations to any ROI is calculated from the collected data for each image in the entire set. This results in eight distinct values for each ROI, for both the pre- and post- test. Review of these measures revealed no statistical significance, however we did see a general trend towards viewing the face region more, pre-test m = 41.20% (11.22%) and post-test m = 45.67% (17.99%). In particular we see participants viewing the upper and lower face region less, while increasing attention to the eyes region (See Table 5.1).

Figure 5-1: *Example of defined regions of interest for a image, upper face (orange), eyes (blue) and lower face (green). Each image had its regions of interest defined by hand prior to use.*

Unlike phase I, phase II required the participants to view and choose the person's expression from one of three possible expressions for each image. All choices include one correct answer and two randomly chosen false choices from the remaining five possible expressions. To analyse this phase of the tests the mean percentage of fixations to any ROI is calculated from the collected data for each image in the entire set. To accommodate the needs of the children this phase of the tests was broken into four subsections, each consisting of six images, for each of the eight participants. The 4 subsections for each of the 8 participants resulted in 32 distinct values for each ROI for both the pre- and post- tests. Analysis of the data revealed a statistically significant increase in the amount of time spent fixating on

Table 5.1: *Analysis of pre- and post- measures of gaze behavior across participants (N=8) for the phase I condition. Test image set consisted of 12 images. Mean and (standard deviation) of the percentage of fixation time to a particular ROI are given, the mean absolute time of all face fixations in seconds for the image set, and the mean image display time in seconds for the image set are also presented. In addition, t-test results and Cohen's d value are presented.*

Phase I Fixation Analysis

Region	Pre-test	Post-test	T-test	Cohen's d
Upper Face	8.58 (13.04)	5.68 (8.96)	$t(14) = 0.52$, $p = 0.62$	0.26
Eyes	56.70 (31.44)	64.66 (24.02)	$t(14) = -0.57$, $p = 0.58$	0.29
Lower Face	30.22 (31.06)	23.64 (18.90)	$t(14) = 0.51$, $p = 0.62$	0.26
Other	4.51 (11.53)	6.03 (12.60)	$t(14) = -0.25$, $p = 0.80$	0.13
Face Fixations	41.20 (11.22)	45.61 (17.99)	$t(14) = -0.60$, $p = 0.56$	0.31
Face Fixation Time	9.89 (2.69)	10.97 (4.32)	$t(14) = -1.04$, $p = 0.32$	0.52
Image Display Time	24.01 (0.01)	24.05 (0.11)	$t(14) = -1.08$, $p = 0.30$	0.53

facial features (See Table 5.2). In addition, though not statistically significant, $p = 0.08$, we see a marked effect size for increased attention to the eyes. It is possible that given more participants such conditions would prove statistically significant. It should also be noted that while the face fixations appear to be lower for phases II-III, in comparison to phase I, the images were shown for longer amounts of time to help facilitate the participants expression choice.

Phase III uses the images from phase II, though the order of display has been randomized and an occlusion mask applied to the images (See Figure 4-1). Like phase II, participants were asked to view the image and chose an expression from one of three possible expressions for each image. The two false choice options have been re-randomized from phase II. Analysis procedures were identical to phase II. Group results of pre- and post- measures of phase III did not reveal any clear trends or statistical significance, though the average amount of time on face fixations did increase (See Table 5.2).

In designing phase III, we expected to see an increase in attention to the eyes region.

Table 5.2: *Analysis of pre- and post- measures of gaze behavior across participants (N=8) for the phase II condition. Test image set consisted of 24 images. Mean and (standard deviation) of the percentage of fixation time to a particular ROI are given, the mean absolute time of all face fixations in seconds for the image set, and the mean image display time in seconds for the image set are also presented. In addition, t-test results and Cohen's d value are presented.*

Phase II Fixation Analysis

Region	Pre-test	Post-test	T-test	Cohen's d
Upper Face	13.59 (20.95)	7.85 (13.02)	$t(62) = 1.32$, $p = 0.19$	0.33
Eyes	40.20 (21.94)	48.87 (16.79)	$t(62) = -1.78$, $p = 0.08$	0.44
Lower Face	43.88 (28.57)	39.24 (17.91)	$t(62) = 0.78$, $p = 0.44$	0.20
Other	2.32 (3.02)	4.03 (5.14)	$t(62) = -1.62$, $p = 0.11$	0.40
Face Fixations	19.70 (12.41)	28.31 (13.75)	$t(62) = -2.63$, $p = 0.01^*$	0.63
Face Fixation Time	65.55 (33.97)	71.81 (35.24)	$t(62) = -0.36$, $p = 0.19$	0.25
Image Display Time	352.50 (46.79)	271.97 (98.79)	$t(62) = 2.08$, $p = 0.06$	0.94

Phase III Fixation Analysis

Region	Pre-test	Post-test	T-test	Cohen's d
Upper Face	8.50 (13.73)	7.68 (11.65)	$t(62) = 0.26$, $p = 0.80$	0.07
Eyes	55.08 (23.11)	53.99 (22.83)	$t(62) = 0.19$, $p = 0.85$	0.05
Lower Face	33.85 (25.52)	35.12 (17.85)	$t(62) = -0.23$, $p = 0.82$	0.06
Other	2.58 (3.26)	3.22 (5.20)	$t(62) = -0.59$, $p = 0.56$	0.15
Face Fixations	19.63 (13.45)	25.76 (16.52)	$t(62) = -1.63$, $p = 0.11$	0.40
Face Fixation Time	66.77 (32.50)	62.52 (45.01)	$t(62) = 0.22$, $p = 0.83$	0.12
Image Display Time	367.72 (82.89)	264.82 (104.04)	$t(62) = 2.18$, $p = 0.05^*$	0.98

$^*p < 0.05$, $^{**}p < 0.01$

Comparison of phase II and phase III reveals that for the pre-test there is a statistically significant difference, $t(62) = -2.64$, $p = 0.01$, in time viewing the eyes region and a rather large effect size, $d = 0.63$. Comparison of the post-test does not reveal a similar trend, $t(62) = -1.02$, $p = 0.31$, $d = 0.25$.

Table 5.3: *Analysis of pre- and post- measures of gaze behavior across participants (N=8) for phase I-III. Mean and (standard deviation) of the percentage of fixation time to a particular ROI are given, the mean absolute time of all face fixations in seconds for the image set, and the mean image display time in seconds for the image set are also presented. In addition, t-test results and Cohen's d value are presented.*

Phase I-III Fixation Analysis

Region	Pre-test	Post-test	T-test	Cohen's d
Upper Face	10.77 (17.24)	7.54 (11.90)	$t(142) = 1.31$, $p = 0.19$	0.22
Eyes	48.64 (24.47)	52.90 (20.76)	$t(142) = -1.13$, $p = 0.26$	0.18
Lower Face	37.90 (27.67)	35.67 (18.34)	$t(142) = 0.57$, $p = 0.57$	0.09
Other	2.68 (4.71)	3.89 (6.30)	$t(142) = -1.31$, $p = 0.19$	0.22
Face Fixations	22.06 (14.32)	29.10 (16.43)	$t(142) = -2.74$, $p = 0.01^{**}$	0.45
Face Fixation Time	47.41 (37.54)	48.43 (41.81)	$t(142) = -0.09$, $p = 0.93$	0.36
Image Display Time	248.07 (170.27)	186.94 (141.83)	$t(142) = 1.35$, $p = 0.18$	0.39

$^{*}p < 0.05,^{**} p < 0.01^{**}$

Lastly, we consider the combination of phase I, II, and III eye-tracking data. Results show a statistically significant increase in the time spent fixating facial features, $t(62) = -2.74$, $p = 0.01$, with $d = 0.45$. In addition, though not statistically significant, we see a modest increase in fixation time to the eyes region, and modest decreases in fixations to above and below the eyes.

5.1.2 Group Analysis of Expression Recognition

During phase II and III, participants were asked to identify the expression of the person in each image they were shown. There were three expression choices associated with each image, one correct and two incorrect. Incorrect expression choices were randomly chosen from the remaining five incorrect options as discussed previously. The same images were used for phase II and III, with the order of images and the incorrect expression choices randomized programatically.

During phase II and III each of the eight participants is shown a total of eight examples of each expression. This results in 64 distinct values for each expression, for both the pre- and post- tests. In addition, we have combined all six expressions for each of phase

Table 5.4: *Analysis of pre- and post- measures of expression recognition responses for all participants for all expressions individually and combined. Mean and (standard deviation) of the percentage of correct responses for each expression are given along with t-test results and Cohen's d value.*

Group Expression Recognition

Expression	**Pre-test**	**Post-test**	**T-test**	**Cohen's d**
Afraid	35.95 (24.49)	54.69 (30.57)	$t(126) = -2.15$, $p = 0.03^*$	0.66
Angry	50.00 (34.07)	59.38 (29.69)	$t(126) = -1.06$, $p = 0.29$	0.30
Disgusted	40.62 (28.15)	56.25 (21.13)	$t(126) = -1.78$, $p = 0.08$	0.61
Happy	71.88 (35.83)	82.81 (18.82)	$t(126) = -1.48$, $p = 0.14$	0.39
Sad	40.62 (31.16)	64.06 (27.09)	$t(126) = -2.71$, $p = 0.01^{**}$	0.76
Surprised	43.75 (37.20)	62.50 (29.12)	$t(126) = -2.15$, $p = 0.03^*$	0.56
Phase II	48.96 (17.22)	65.10 (18.76)	$t(382) = -3.23$, $p = 0.01^{**}$	0.87
Phase III	45.31 (22.10)	61.46 (19.13)	$t(382) = -3.21$, $p = 0.01^{**}$	0.75
Phase II-III	47.14 (19.19)	63.28 (18.90)	$t(766) = -4.55$, $p = 0.01^{**}$	0.80

$^*p < 0.05$, $^{**}p < 0.01$

II and III, and since each phase has a total of four examples of each expression there are 192 distinct values for both the pre- and post- tests for each phase. Review of these results indicate an increase in expression recognition for all six expressions for the entire group. Moreover, we see a statistically significant increase in expression recognition for afraid, sad, and surprised, all with $p < 0.05$. Additionally, if we consider the combination of all expressions for phase II and phase III individually and combined we have a statistically significant increase in expression recognition with $p < 0.01$ (See Table 5.4).

Individual Participant Analysis of Expression Recognition

To better understand the changes in expression recognition, we performed individual analysis for each participant (See Tables 5.5-5.7). When examining the individual changes several interesting findings appeared. All participants improved overall. In addition, most individuals had statistically significant improvement in one or more expression categories, and/or overall combined expression responses (See Tables 5.5-5.7). Second, a small group (N=3) improved in overall expression recognition by at least 15% (See Tables 5.5, 5.6).

When we considered expression recognition improvement, age, KBIT-composite, and CARS

Table 5.5: *Analysis of pre- and post- measures of expression recognition responses for participant 01-02 for all expressions individually and combined. Percentage of correct responses for each expression are given along with t-test results.*

Participant 01 Expression Recognition

Expression	**Pre-test**	**Post-test**	**T-test**
Afraid	25.00	50.00	$t(14) = -1.00$, $p = 0.33$
Angry	25.00	0.00	$t(14) = 1.52$, $p = 0.15$
Disgusted	25.00	62.50	$t(14) = -1.52$, $p = 0.15$
Happy	50.00	50.00	$t(14) = 0.00$, $p = 1.00$
Sad	12.50	50.00	$t(14) = -1.66$, $p = 0.12$
Surprised	50.00	62.50	$t(14) = -0.48$, $p = 0.64$
Phase II	41.67	45.83	$t(46) = -0.29$, $p = 0.78$
Phase III	20.83	45.83	$t(46) = -1.87$, $p = 0.07$
Phase II-III	31.25	45.83	$t(94) = -1.47$, $p = 0.15$

Participant 02 Expression Recognition

Expression	**Pre-test**	**Post-test**	**T-test**
Afraid	12.00	87.50	$t(14) = -4.24$, $p = 0.01^{**}$
Angry	25.00	75.00	$t(14) = -2.16$, $p = 0.05$
Disgusted	25.00	37.50	$t(14) = -0.51$, $p = 0.62$
Happy	50.00	87.50	$t(14) = -1.66$, $p = 0.12$
Sad	62.50	87.50	$t(14) = -1.13$, $p = 0.28$
Surprised	12.50	87.50	$t(14) = -4.24$, $p = 0.01^{**}$
Phase II	33.33	79.17	$t(46) = -3.53$, $p = 0.01^{**}$
Phase III	29.17	75.00	$t(46) = -3.50$, $p = 0.01^{**}$
Phase II-III	31.25	77.08	$t(94) = -5.02$, $p = 0.01^{**}$

$^{*}p < 0.05, ^{**}p < 0.01$

scores we also found that the previously mentioned three individuals shared common characteristics. This subgroup's expression recognition average percent correct improvements were statistically different $m_1 = 29.17$ (15.03) compared to the remaining participants $m_2 = 8.33$ (2.95); $t(6) = 3.17$, $p < 0.05$. They constituted the younger participants $m_1 = 8.78yrs$ ($1.82yrs$) compared to the other five participants $m_2 = 14.63yrs$ ($4.16yrs$). Likewise, comparisons of the averaged KBIT-composite scores for the first subgroup are $m_1 = 55.33$ (14.19) compared to $m_2 = 43.20$ (7.16). The CARS scores indicate no difference $m_1 = 33.00$ (4.36) compared to $m_2 = 33.30$ (4.41) (See Table 5.8). Given these results, we considered subgroups analysis. The first subgroup (SG1) contains the previously mentioned three individuals (participants 02, 03, and 04), while the remaining five individuals made up subgroup 2 (SG2). Section 5.1.3 examines subgroup differences.

Table 5.6: *Analysis of pre- and post- measures of expression responses for participant 03-05 for all expressions individually and combined. Percentage of correct responses for each expression are given along with t-test results.*

Participant 03 Expression Recognition

Expression	Pre-test	Post-test	T-test
Afraid	87.50	100.00	$t(14) = -1.00$, $p = 0.33$
Angry	87.50	87.50	$t(14) = 0.00$, $p = 1.00$
Disgusted	25.00	87.50	$t(14) = -3.03$, $p = 0.01^{**}$
Happy	100.00	100.00	$t(14) = 0.00$, $p = 1.00$
Sad	100.00	100.00	$t(14) = 0.00$, $p = 1.00$
Surprised	100.00	100.00	$t(14) = 0.00$, $p = 1.00$
Phase II	79.17	95.83	$t(46) = -1.77$, $p = 0.08$
Phase III	87.50	95.83	$t(46) = -1.03$, $p = 0.31$
Phase II-III	83.33	95.83	$t(94) = -2.03$, $p = 0.05^{*}$

Participant 04 Expression Recognition

Expression	Pre-test	Post-test	T-test
Afraid	25.00	75.00	$t(14) = -2.16$, $p = 0.05^{*}$
Angry	62.50	62.50	$t(14) = 0.00$, $p = 1.00$
Disgusted	62.50	62.50	$t(14) = 0.00$, $p = 1.00$
Happy	87.50	100.00	$t(14) = -1.00$, $p = 0.33$
Sad	50.00	50.00	$t(14) = 0.00$, $p = 1.00$
Surprised	12.50	63.50	$t(14) = -2.26$, $p = 0.05^{*}$
Phase II	45.83	70.83	$t(46) = -1.78$, $p = 0.08$
Phase III	54.17	66.68	$t(46) = -0.87$, $p = 0.39$
Phase II-III	50.00	68.75	$t(94) = -1.89$, $p = 0.06$

Participant 05 Expression Recognition

Expression	Pre-test	Post-test	T-test
Afraid	37.50	50.00	$t(14) = -0.48$, $p = 0.64$
Angry	37.50	37.50	$t(14) = 0.00$, $p = 1.00$
Disgusted	37.50	25.00	$t(14) = 0.51$, $p = 0.62$
Happy	100.00	100.00	$t(14) = 0.00$, $p = 1.00$
Sad	37.50	50.00	$t(14) = -0.48$, $p = 0.64$
Surprised	25.00	50.00	$t(14) = -1.00$, $p = 0.33$
Phase II	45.83	54.17	$t(46) = -0.57$, $p = 0.57$
Phase III	45.83	50.00	$t(46) = -0.28$, $p = 0.78$
Phase II-III	45.83	52.08	$t(94) = -0.61$, $p = 0.55$

$^{*}p < 0.05, ^{**}p < 0.01$

Table 5.7: *Analysis of pre- and post- measures expression responses for participant 06-08 for all expressions individually and combined. Percentage of correct responses for each expression are given along with t-test results.*

Participant 06 Expression Recognition

Expression	**Pre-test**	**Post-test**	**T-test**
Afraid	50.00	37.50	$t(14) = 0.48$, $p = 0.64$
Angry	62.50	75.00	$t(14) = -0.51$, $p = 0.62$
Disgusted	62.50	75.00	$t(14) = -0.51$, $p = 0.62$
Happy	87.50	75.00	$t(14) = 0.61$, $p = 0.55$
Sad	37.50	100.00	$t(14) = -3.42$, $p = 0.01^{**}$
Surprised	37.50	37.50	$t(14) = 0.00$, $p = 1.00$
Phase II	62.50	70.83	$t(46) = -0.60$, $p = 0.55$
Phase III	50.00	62.50	$t(46) = -0.86$, $p = 0.39$
Phase II-III	56.25	66.67	$t(94) = -1.04$, $p = 0.30$

Participant 07 Expression Recognition

Expression	**Pre-test**	**Post-test**	**T-test**
Afraid	37.50	12.50	$t(14) = 1.13$, $p = 0.28$
Angry	0.00	50.00	$t(14) = -2.65$, $p = 0.02^{*}$
Disgusted	87.50	37.50	$t(14) = 2.26$, $p = 0.04^{*}$
Happy	0.00	62.50	$t(14) = -3.42$, $p = 0.04^{*}$
Sad	0.00	37.50	$t(14) = -2.05$, $p = 0.06$
Surprised	12.50	12.50	$t(14) = 0.00$, $p = 1.00$
Phase II	25.00	37.50	$t(46) = -0.92$, $p = 0.36$
Phase III	20.08	33.33	$t(46) = -0.96$, $p = 0.34$
Phase II-III	22.92	35.42	$t(94) = -1.35$, $p = 0.18$

Participant 08 Expression Recognition

Expression	**Pre-test**	**Post-test**	**T-test**
Afraid	12.50	25.00	$t(14) = -0.61$, $p = 0.55$
Angry	100.00	87.50	$t(14) = 1.00$, $p = 0.33$
Disgusted	0.00	62.50	$t(14) = -3.42$, $p = 0.01^{**}$
Happy	100.00	87.50	$t(14) = 1.00$, $p = 0.33$
Sad	25.00	37.50	$t(14) = -0.51$, $p = 0.62$
Surprised	100.00	87.50	$t(14) = 1.00$, $p = 0.33$
Phase II	58.33	66.67	$t(46) = -0.58$, $p = 0.56$
Phase III	54.17	62.50	$t(46) = -0.58$, $p = 0.56$
Phase II-III	56.25	64.58	$t(94) = -0.83$, $p = 0.41$

$^{*}p < 0.05, ^{**}p < 0.01$

Table 5.8: *Analysis of subgroup characteristics. Average expression recognition improvement, age, KBIT, and CARS for subgroup 1 ($N_1 = 3$), and subgroup 2 ($N_2 = 5$). Expression recognition represents the average percent improvement in recognizing facial expressions between pre- and post- test.*

Subgroup Characteristics

Measure	**SG1**	**SG2**	**T-test**
Recognition improvement	29.17 (15.03)	8.33 (2.95)	$t(6) = 3.17$, $p = 0.02^*$
Age	8.78 (1.82)	14.63 (4.16)	$t(6) = -2.25$, $p = 0.07$
KBIT-composite score	55.33 (14.19)	43.20 (7.16)	$t(6) = 1.65$, $p = 0.15$
CARS score	33.00 (4.36)	33.30 (4.41)	$t(6) = -0.09$, $p = 0.92$

$^*p < 0.05,^{**}p < 0.01$

5.1.3 Gaze behavior and Expression Recognition Analysis of Subgroups

Given the results of individual participant's expression recognition scores, and analysis of group characteristics we have further divided the participants into two subgroups of size $N_1 = 3$, and $N_2 = 5$. Subgroup 1 consists of participants 02, 03, and 04, while subgroup 2 consists of participants 01, 05, 06, 07, and 08. Furthermore, given group gaze behavior results, the general trend of better expression recognition during phase II and the similarities to previous work [44] we have chosen to focus our subgroup analysis to phase II.

Gaze Behavior Analysis for Subgroups

Analysis of gaze behavior during phase II for SG1 and SG2 reveals a general tend in both groups increasing attention to facial features and the eyes region in particular. Additionally, a statistically significant change is seen for SG1 for face fixations $m_{pre} = 18.10$ (9.27), $m_{post} = 26.46$ (10.66), $t(22) = -2.05$; $p < 0.05$ (See Table 5.9).

Further analysis between pre- and post- measures indicates a statistically significant difference in fixations to the upper face region $t(30) = -2.79$; $p = 0.01$, with SG2 spending more time fixating on this region (See Table 5.10). It is interesting to note that gaze behavior is similar between groups, especially post-intervention, yet there are statistically significant differences in expression recognition (See Table 5.12). This may indicate that there are other non-gaze behavior factors that are more significant for expression recognition accuracy.

Table 5.9: *Analysis of pre- and post- measures of gaze behavior across subgroup 1* ($N_1 = 3$) *and subgroup 2* ($N_2 = 5$) *for phase II condition. Mean and (standard deviation) of the percentage of fixation time to a particular ROI are given, the mean absolute time of all face fixations in seconds for the image set, and the mean image display time in seconds for the image set are also presented. In addition, t-test results and Cohen's d value are presented.*

Subgroup 1 Phase II

Region	Pre-test	Post-test	T-test	Cohen's d
Upper Face	1.52 (0.23)	5.62 (8.30)	$t(22) = -1.68$, $p = 0.11$	0.66
Eyes	45.66 (21.75)	50.81 (9.24)	$t(22) = -0.75$, $p = 0.46$	031
Lower Face	50.50 (22.13)	40.08 (13.83)	$t(22) = 1.38$, $p = 0.18$	0.55
Other	2.31 (2.53)	3.49 (3.53)	$t(22) = -0.94$, $p = 0.36$	0.39
Face Fixations	18.10 (9.27)	26.46 (10.66)	$t(22) = -2.05$, $p = 0.05^*$	0.78
Face Fixation Time	59.23 (11.19)	65.41 (16.92)	$t(22) = -0.53$, $p = 0.62$	0.47
Image Display Time	336.40 (31.30)	258.72 (80.40)	$t(22) = 1.56$, $p = 0.19$	1.12

Subgroup 2 Phase II

Region	Pre-test	Post-test	T-test	Cohen's d
Upper Face	20.83 (23.82)	9.19 (15.23)	$t(38) = 1.84$, $p = 0.07$	0.57
Eyes	36.92 (21.94)	47.71 (20.17)	$t(38) = -1.62$, $p = 0.11$	0.50
Lower Face	39.91 (31.68)	38.74 (20.29)	$t(38) = 0.14$, $p = 0.89$	0.05
Other	2.33 (3.35)	4.35 (5.97)	$t(38) = -1.32$, $p = 0.19$	0.41
Face Fixations	20.65 (14.11)	29.41 (15.47)	$t(38) = -1.87$, $p = 0.07$	0.57
Face Fixation Time	69.34 (43.68)	75.65 (44.51)	$t(38) = -0.23$, $p = 0.83$	0.15
Image Display Time	362.15 (55.05)	279.92 (116.77)	$t(38) = 1.42$, $p = 0.19$	0.85

$^*p < 0.05, ^{**}p < 0.01$

Table 5.10: *Comparison of pre- and post- measures of gaze behavior between subgroup 1* ($N_1 = 3$) *and subgroup 2* ($N_2 = 5$) *for phase II condition. Mean and (standard deviation) of the percentage of fixation time to a particular ROI are given, the mean absolute time of all face fixations in seconds for the image set, and the mean image display time in seconds for the image set are also presented. In addition, t-test results and Cohen's d value are presented.*

Subgroup 1-2 Phase II Fixation Comparison

Region	Pre-test	Post-test
Upper Face	$t(30) = -2.79$, $p = 0.01^{**}$	$t(30) = -0.75$, $p = 0.46$
Eyes	$t(30) = 1.09$, $p = 0.28$	$t(30) = 0.50$, $p = 0.62$
Lower Face	$t(30) = 1.02$, $p = 0.32$	$t(30) = 0.20$, $p = 0.84$
Other	$t(30) = -0.02$, $p = 0.98$	$t(30) = -0.45$, $p = 0.65$
Face Fixations	$t(30) = -0.56$, $p = 0.58$	$t(30) = -0.58$, $p = 0.56$
Face Fixation Time	$t(30) = -0.38$, $p = 0.72$	$t(30) = -0.37$, $p = 0.72$
Image Display Time	$t(30) = -0.73$, $p = 0.49$	$t(30) = -0.27$, $p = 0.79$

$^*p < 0.05, ^{**}p < 0.01$

Expression Recognition Analysis for Subgroups

The most statistically significant differences between SG1 and SG2 are related to expression recognition scores. Investigation of subgroups shows an overall positive increase in expression recognition across all six expression categories for SG1 with statistically significance levels of of change, $p < 0.01$, for afraid and surprised (See Table 5.11). In addition, the overall results for phase II are statistically significant for the post-test vs. pre-test condition, $t(142) = -3.03$; $p = 0.01$ relative to the pre-test. SG2 does not follow a similar pattern, with lower average correct responses for afraid and anger in the post-test condition, though SG2 does see a statistically significant improvement in recognizing sadness, $t(78) = -3.12$; $p = 0.01$ (See Table 5.11).

Comparing SG1 and SG2 pre- and post- measures reveals statistically significant better recognition of sad in the pre-test, $t(62) = 4.27$; $p = 0.01$ for SG1 compared to SG2, but the overall combined recognition of expressions were not statistically different for the pre-test. Post-test analysis reveals a statistically significant increase in the ability to recognize afraid $t(62) = 4.68$; $p = 0.01$, happiness $t(62) = 2.18$; $p = 0.03$, and surprised $t(62) = 2.78$; $p = 0.01$ for SG1 compared to SG2. Additionally, the overall response rate was statistically significant for SG1 compared to SG2 across all expressions, $t(190) = 3.92$, $p = 0.01$ for the post-test (See Table 5.12).

Table 5.11: *Analysis of pre- and post- measures of expression recognition responses for subgroup 1 ($N_1 = 3$) and subgroup 2 ($N_2 = 5$) for all expressions individually and combined. Mean and (standard deviation) of the percentage of fixations to a particular ROI are given along with t-test results.*

Subgroup 1 Expression Recognition

Expression	Pre-test	Post-test	T-test	Cohen's d
Afraid	41.67 (28.87)	83.33 (14.43)	$t(46) = -3.23$, $p = 0.01^{**}$	1.36
Angry	50.00 (25.00)	75.00 (25.00)	$t(46) = -1.22$, $p = 0.23$	0.95
Disgusted	25.00 (25.00)	58.33 (38.19)	$t(46) = -0.29$, $p = 0.77$	0.98
Happy	91.67 (14.43)	100.00 (0.00)	$t(46) = -1.77$, $p = 0.08$	0.82
Sad	75.00 (25.00)	91.67 (14.43)	$t(46) = -0.66$, $p = 0.52$	0.82
Surprised	33.33 (57.74)	83.33 (14.43)	$t(46) = -3.23$, $p = 0.01^{**}$	1.07
Phase II	52.78 (23.69)	81.94 (12.73)	$t(142) = -3.03$, $p = 0.01^{**}$	1.25

Subgroup 2 Expression Recognition

Expression	Pre-test	Post-test	T-test	Cohen's d
Afraid	50.00 (25.00)	40.00 (22.36)	$t(78) = -0.23$, $p = 0.80$	0.44
Angry	55.00 (37.08)	50.00 (30.62)	$t(78) = -0.44$, $p = 0.66$	0.16
Disgusted	45.00 (32.00)	50.00 (17.68)	$t(78) = -0.89$, $p = 0.38$	0.20
Happy	65.00 (41.83)	75.00 (25.00)	$t(78) = -0.73$, $p = 0.47$	0.30
Sad	20.00 (20.92)	60.00 (28.50)	$t(78) = -3.12$, $p = 0.01^{*}$	1.26
Surprised	45.00 (37.08)	55.00 (27.39)	$t(78) = -0.44$, $p = 0.66$	0.32
Phase II	46.67 (14.85)	55.00 (13.94)	$t(238) = -1.29$, $p = 0.20$	0.58

$^{*}p < 0.05,^{**}p < 0.01$

Table 5.12: *Analysis of pre- and post- measures of expression recognition responses for subgroup 1 and 2 for all expressions individually and combined.*

Subgroup 1-2 Phase II Expression Recognition Comparison

Region	Pre-test	Post-test
Afraid	$t(62) = 0.73$, $p = 0.47$	$t(62) = 4.68$, $p = 0.01^{**}$
Angry	$t(62) = 1.03$, $p = 0.30$	$t(62) = 2.00$, $p = 0.05^{*}$
Disgusted	$t(62) = -0.39$, $p = 0.70$	$t(62) = 0.77$, $p = 0.44$
Happy	$t(62) = 1.00$, $p = 0.32$	$t(62) = 2.18$, $p = 0.03^{*}$
Sad	$t(62) = 4.27$, $p = 0.01^{*}$	$t(62) = 1.98$, $p = 0.05$
Surprised	$t(62) = -0.25$, $p = 0.80$	$t(62) = 2.78$, $p = 0.01^{**}$
Phase II	$t(190) = 0.46$, $p = 0.66$	$t(190) = 3.92$, $p = 0.01^{*}$

$^{*}p < 0.05,^{**}p < 0.01$

5.2 Intervention Game Play Data Analysis

In addition to pre- and post- measures, game play data was recorded for each participant for each session. Results are presented for all 10 participants for the remainder of the chapter. Furthermore, the two participants not included in the above gaze behavior and expression recognition analysis were grouped into SG2 for further subgroup analysis. Their inclusion into SG2 was made based on the participant's age, KBIT, and CARS scores. The following presents results for puzzle-construction-specific-interactions and for puzzle expression labelling for the entire group ($N_g = 10$) and for SG1 ($N_1 = 3$) and SG2 ($N_2 = 7$).

5.2.1 Group Puzzle Interaction Results

During each session, participant's play would include the construction of many puzzles over a time period of 10 to 15 minutes. For each session the number of puzzles completed, i.e. constructed correctly and labelled with the correct expression label, was recorded. In addition, the amount of time to complete each puzzle, the number of puzzle pieces used while attempting to construct the puzzle, the number of correct pieces used while constructing the puzzle, and the selection of expression labels was recorded.

While the total number of puzzle pieces used to construct the puzzle gives some indication of how well the participant was able to match the target puzzle with the available puzzle pieces, the total number of correct puzzle pieces used gives us insight into how well the participant was able to organize the pieces into the correct spatial relationships. The amount of time required to complete a puzzle relates to the overall ease or difficulty of scanning, matching, manipulating, organizing and labelling the puzzle. Lastly, the number of expression labels chosen for each puzzle provides insight into how well the participant recognizes a particular facial expression solely from information surrounding the eyes.

Figure 5-2 presents the averaged group results for puzzles completed, puzzle completion time, correct puzzle pieces used, and total number of puzzle pieces used. For all graphs, error-bars represent standard error (See Equation 5.1), that is the sample standard deviation divided by the square root of the sample size.

$$\frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}}{\sqrt{n}} \tag{5.1}$$

We see in figure 5-2 top left and right a trend of increasing the number of puzzles completed each session while decreasing the amount of time required to complete each puzzle. Furthermore, we see that the number of correct pieces placed and the total number of pieces used decrease steadily across sessions, figure 5-2 bottom left and right. These results correspond to experimenter observations that children became better at constructing puzzles. Many participants were able to quickly construct puzzles using only the correct pieces placed in the correct spacial relationship by the fifth session. Interestingly, at the group level we see little change in the average number of wrong expression label selections for each puzzle (See Figure 5-3).

5.2.2 Group Puzzle Expression Recognition Results

After correctly constructing a puzzle the participant is presented with four expression choices. Of these four one is correct and the other three are randomly chosen from the remaining five incorrect expressions. In addition, the association of a label to a button is programatically controlled and assigned randomly. This is done to ensure that the child is actually reading the label and not simply memorizing the labels location. Participants are able to make as many expression label selections as they wish, until the correct expression is chosen, each selection has a 25% chance of being correct. Participants seldom made more than three total expression selections for any one puzzle. Bellow we present averaged

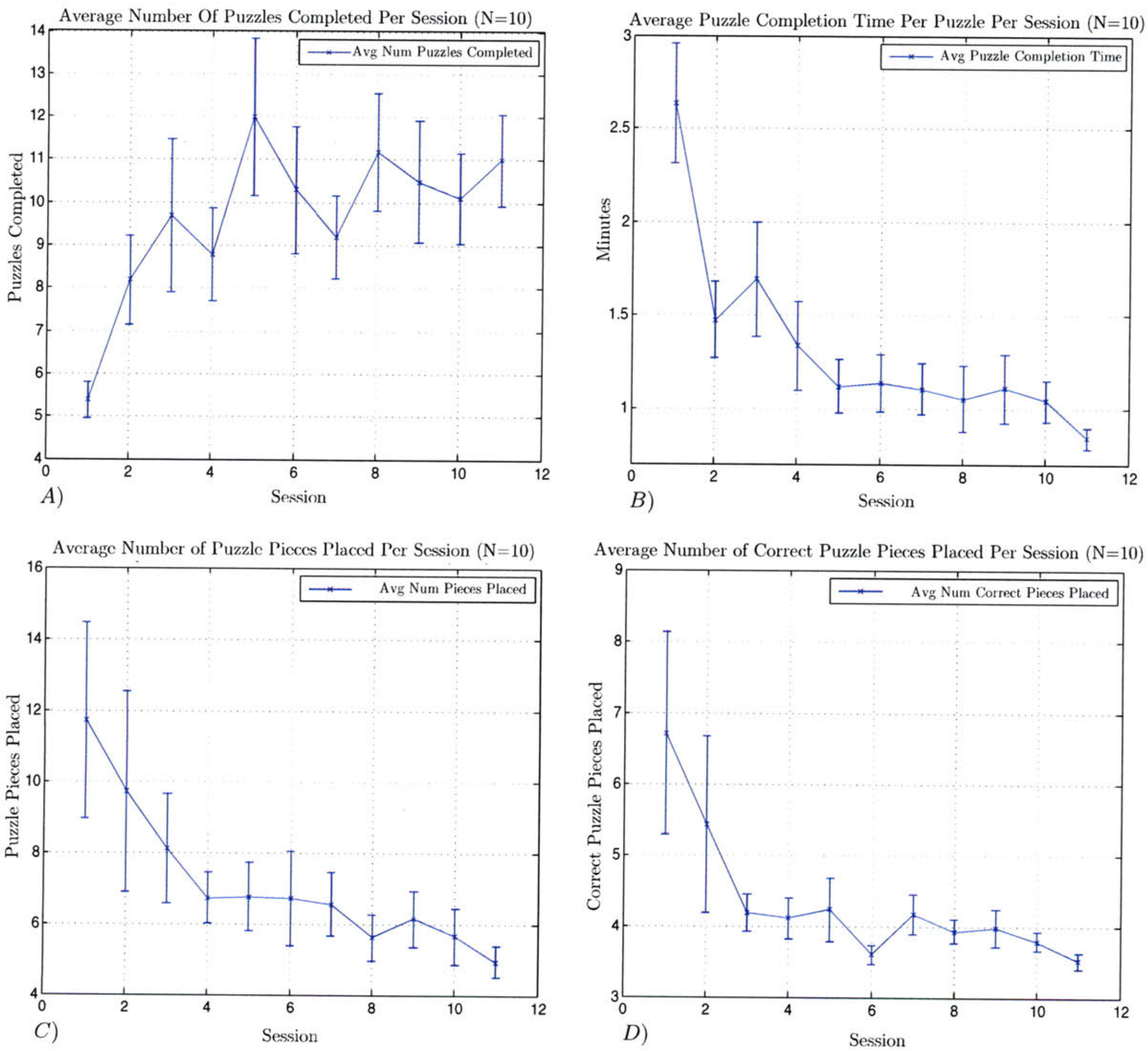

Figure 5-2: *Group (N=10) analysis of puzzle play per session. A) Average number of puzzles completed per session. B) Average amount of time to complete each puzzle per session. C) Average number of puzzle pieces, correct and incorrect, used during puzzle construction. D) Average number of correct puzzle pieces used during puzzle construction.*

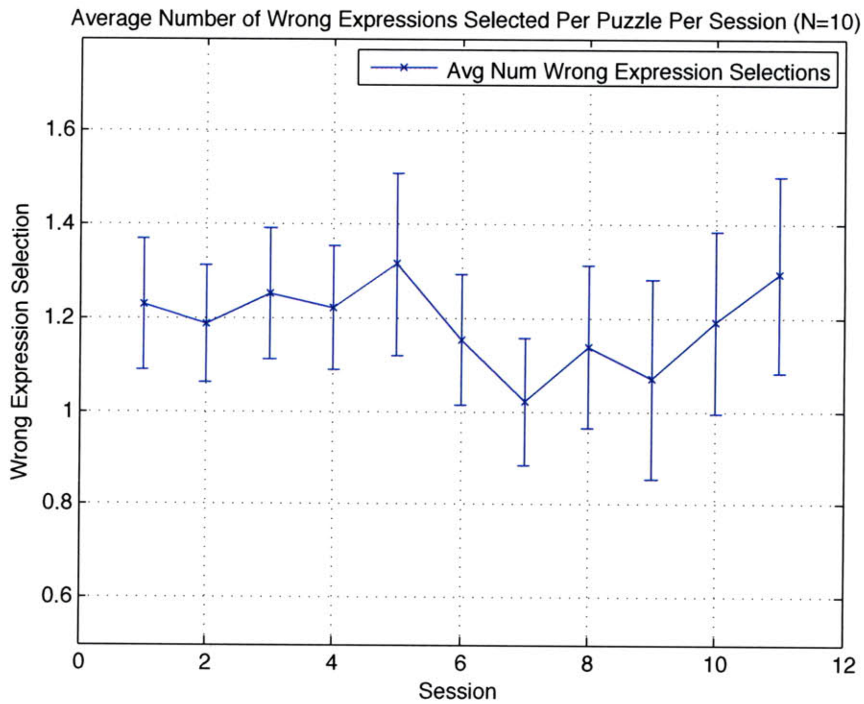

Figure 5-3: *Group (N=10) analysis of average number of wrong expression labels selected per puzzle.*

Table 5.13: *Group analysis of percentage of correct puzzle expression responses for the first three possible responses.*

Group Analysis of Correct Expression Responses

Expression	First Response (%)	Second Response (%)	Third Response (%)
Fear	31.52	60.87	83.70
Angry	47.40	69.36	83.82
Disgusted	57.46	72.93	84.53
Happy	57.87	78.09	89.33
Sad	46.78	66.08	84.80
Surprised	31.07	59.32	76.27

response for each expression for the first three possible selections.

We see the group was above chance at recognizing the presented expression on the first try, though only slightly for afraid and surprised (See Table 5.13). For those instances when a participant was incorrect on their first expression selection their second response was well above chance. For those that required a third choice their accuracy nears perfect. It should be noted that it is generally accepted that identifying facial expression from only the eyes is a difficult task, therefore the relatively low first response correctness is not surprising. Studies have shown that even NTs score low on such tasks [7]. It is also valuable to understand how participants are mislabelling expressions. Table 5.14 presents similar data related to mislabels for the first three possible responses. For each expression the most common mislabel selection is shown for each of the three possible responses.

Table 5.14: *Group analysis of most common incorrect puzzle expression incorrect responses for the first three possible responses.*

Group Analysis of Incorrect Expression Responses

Expression	**First Response**		**Second Response**		**Third Response**	
	Expression	**(%)**	**Expression**	**(%)**	**Expression**	**(%)**
Fear	Disgusted	32.07	Disgusted	11.41	Sad	4.89
Angry	afraid	18.50	Happy	10.98	Happy	7.51
Disgusted	Sad	14.92	Happy	9.94	Surprised	4.42
Happy	afraid	16.29	Sad	9.55	Surprised	5.06
Sad	Angry	23.39	Happy	11.70	Angry	4.68
Surprised	Disgusted	29.38	afraid	12.99	Happy	9.60

5.2.3 Subgroups Puzzle Interaction Results

Given the above differences in gaze behavior and expression recognition between SG1 and SG2 we investigated their puzzle interaction results separately. Results show that SG1 generally performed better than SG2, but that performance levels near each other as the intervention proceeded (See Figure 5-4). The most telling difference between groups is related to the average number of wrong expression labels selected per-puzzle (See Figure 5-5). Analysis reveals a statistically significant difference between SG1 and SG2, $t(108) = -3.32; p < 0.01$, with SG1 becoming better at correctly labelling expressions. These results fit with the expression recognition results from the pre- and post- test discussed above, as SG1 performed better overall than SG2 at recognizing expressions.

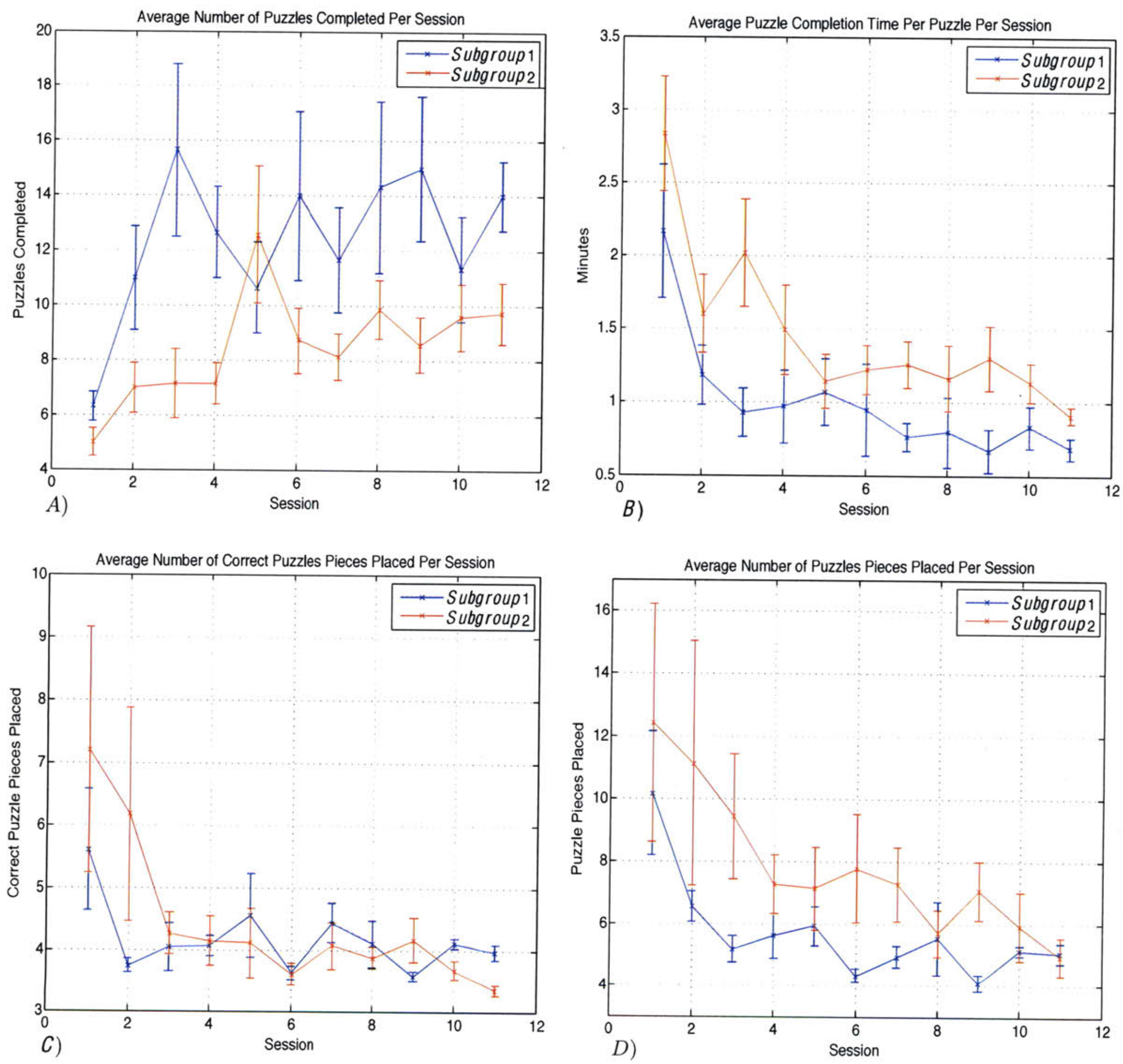

Figure 5-4: *Subgroup analysis of puzzle play per session, subgroup 1* ($N_1 = 3$), *subgroup 2* ($N_2 = 7$). *A) Average number of puzzles completed per session. B) Average amount of time to complete each puzzle per session. C) Average number of puzzle pieces, correct and incorrect, used during puzzle construction. D) Average number of correct puzzle pieces used during puzzle construction.*

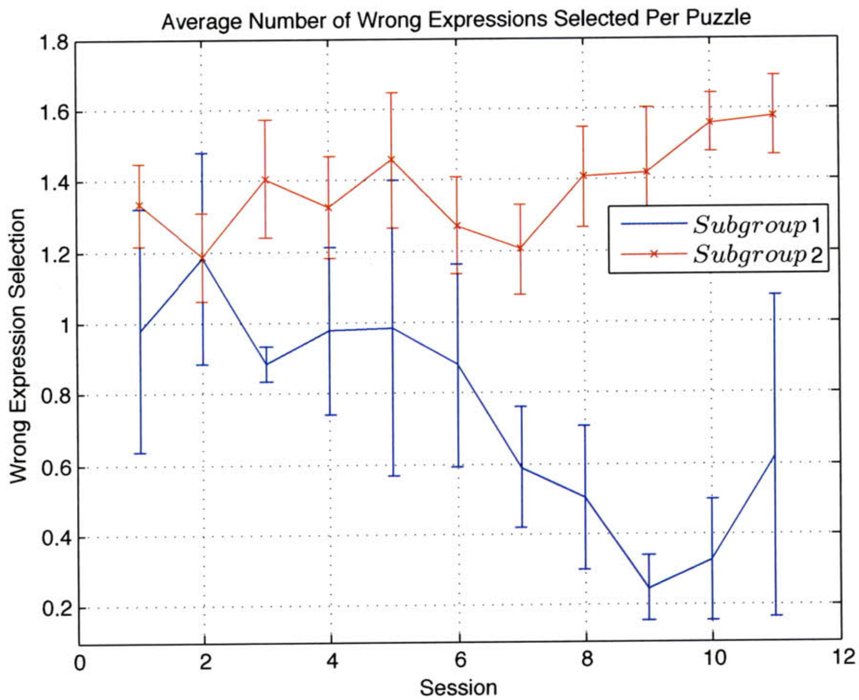

Figure 5-5: *Analysis of average number of wrong expression labels selected per puzzle for subgroups SG1* ($N_1 = 3$) *and SG2* ($N_2 = 7$)

Table 5.15: *Subgroup 1-2 comparison of correct puzzle expression responses for the first three possible responses for subgroup 1 (SG1) and subgroup 2 (SG2), N=3 and N=7 respectively.*

Subgroup 1-2 Analysis of Correct Expression Responses

Expression	**First Response** (%)		**Second Response** (%)		**Third Response** (%)	
	SG1	**SG2**	**SG1**	**SG2**	**SG1**	**SG2**
Fear	47.95	20.72	79.45	48.65	91.78	78.38
Angry	67.16	34.91	80.60	62.26	88.06	81.13
Disgusted	79.10	44.74	91.04	62.28	94.03	78.95
Happy	71.64	49.55	89.55	71.17	98.51	83.78
Sad	82.09	24.04	94.03	48.08	98.51	75.96
Surprised	37.68	26.85	71.01	51.85	85.51	70.37

5.2.4 Subgroups Puzzle Expression Recognition Results

When comparing the averaged percent correct for each expression between subgroups we again see that SG1 performs better at correctly labelling expressions for all three response instances. More importantly, SG1 correctly labels many of the expressions at a high accuracy for the first response (See Table 5.15). For each subgroup we also considered their incorrect responses (See Table 5.16).

5.2.5 Correlation Analysis

This section examines the relationships between KBIT and CARS scores with post-measures of gaze behavior of facial features in general and to each ROI in addition to the overall expression recognition scores (See Table 5.17). We have included the superscript marker † to indicate those relationships that near significant levels. Our reasons being that a larger sample size may have revealed such relationships to be significant.

Results show (See Table 5.17) positive correlation for attention to eyes and higher KBIT-verbal scores. Likewise, a statistically significant positive correlation was found for expression recognition and higer KBIT-verbal scores. Additionally, we found a statistically significant relationship between lower CARS scores and fixation time to facial features.

Table 5.16: *Subgroup 1 and 2 analysis of incorrect puzzle expression responses for the first three incorrect responses. Subgroup 1* ($N_1 = 3$)*, subgroup 2* ($N_2 = 7$).

Subgroup 1 Analysis of Incorrect Expression Responses

Expression	**First Response**		**Second Response**		**Third Response**	
	Expression	(%)	Expression	(%)	Expression	(%)
Fear	Disgusted	34.25	Disgusted	8.22	Surprised	4.11
Angry	afraid	16.42	Surprised	7.46	Happy	8.96
Disgusted	Sad	5.97	afraid	2.99	afraid	1.49
Happy	Surprised	11.94	Surprised	4.48	Sad	1.49
Sad	Disgusted	5.97	Disgusted	4.48	Angry	1.49
Surprised	Disgusted	34.78	afraid	13.04	Disgusted	7.25

Subgroup 2 Analysis of Incorrect Expression Responses

Expression	**First Response**		**Second Response**		**Third Response**	
	Expression	(%)	Expression	(%)	Expression	(%)
Fear	Disgusted	30.63	Happy	16.22	Sad	8.11
Angry	Happy	23.58	Happy	14.15	Surprised	9.43
Disgusted	Sad	20.18	Happy	14.04	Surprised	6.14
Happy	afraid	19.82	Sad	13.51	Surprised	8.11
Sad	Angry	37.50	Happy	18.27	Angry	6.73
Surprised	Disgusted	25.93	Happy	16.67	Happy	14.81

Table 5.17: *Correlation between KBIT and CARS scores with the percentage of fixation times to eyes, lower face, upper face from all fixations. In addition, the percent of time fixating facial features from all possible viewing time and expression recognition scores are considered.*

	% fixation time to upper face	% fixation time to eyes	% fixation time to lower face	% fixation time to facial features	% correct expression responses
KBIT-verbal	-0.15	0.41	-0.27	0.16	0.73^{*}
KBIT-non-verbal	$0.64^{\dagger}$	0.08	-0.54	-0.41	0.50
KBIT-composite	0.41	0.26	-0.55	-0.17	$0.69^{\dagger}$
CARS	0.17	$-0.67^{\dagger}$	0.54	-0.80^{*}	-0.20

$^{\dagger}p < 0.10, ^{*}p < 0.05, ^{**}p < 0.01$

5.3 Review of Qualitative Measure

In addition to the quantitative measures presented above a number of qualitative measures were recorded to help better understand the interaction and efficacy of the intervention. The following presents results from these measures.

5.3.1 Engagement and Mood

During each session observations and notes were taken from which the experimenter would rate the child's level of engagement in the activity. A five point Likert scale was used with one representing "uninterested" and five repsenting "very engaged." In addition, after each session the teacher was asked to assess the child's mood on a Likert scale with one representing "bad" and five representing "very good." Examining these factors we have a mean engagement of $m_e = 4.31$ (0.89) and a mean post game mood of $m_m = 4.22$ (0.72). Furthermore, correlation analysis reveals $r(222) = 0.59$, $p < 0.001$, indicating that high engagement resulted in positive post game mood.

5.3.2 Intervention Efficacy

At the end of the study teachers were given a survey to assess overall aspects of the intervention. The questions presented where:

- The child enjoyed playing *Frame It*
- The child became better at matching and constructing puzzles
- The child became better at recognizing facial expressions
- The child began making more eye contact after using *Frame It*

A four point Likert scale was used for all questions with one representing "not true" and four representing "almost always true." Table 5.18 presents the averaged response for the

Table 5.18: *Intervention efficacy survey results for all participants (N=10). Values reported are mean and standard deviation*

Question	Mean Response
Child enjoyed playing *Frame It*	2.70 (1.16)
Child became better at matching and constructing puzzles	3.30 (1.06)
Child became better at recognizing facial expressions	2.50 (1.08)
Child began making more eye contact after using *Frame It*	2.00 (1.05)

Table 5.19: *Comparison of subgroups 1 and 2 intervention efficacy survey results* ($N_1 = 4$) *and* ($N_2 = 6$). *Values reported are mean and standard deviation*

Question	Mean Response SG1	Mean Response SG2
Child enjoyed playing *Frame It*	4.00 (0.00)	1.83 (0.41)
Child became better at matching and constructing puzzles	4.00 (0.00)	2.83 (1.17)
Child became better at recognizing facial expressions	3.25 (0.50)	2.00 (1.10)
Child began making more eye contact after using *Frame It*	2.75 (1.26)	1.50 (0.55)

group. In addition, we considered subgroups based on teacher responses to the survey. Grouping was performed using the response to the question "the child enjoyed playing *Frame It*." Based on these responses we divided all participants into two subgroups, those that responded with the maximum score, four, and those that responded with a three or less.

Results indicate that subgroup 1 enjoyed the game more than subgroup 2, and appears to have become better at both constructing puzzles and recognizing expressions (See Table 5.19). Analysis of survey results indicate that subgroup 1 scored significantly better across all categories than subgroup 2, $t(38) = 5.00$, $p < 0.001$.

Interestingly, even though these subgroups were formed using different criteria than early subgroups, the composition is nearly identical to the subgroups formed for the gaze behavior and expression recognition analysis. The difference in subgroups relates to the two participants not included in that initial gaze behavior analysis, one of these individuals was assigned to the subgroup 1 and the other to subgroup 2.

Chapter 6

Discussion and Conclusion

6.1 Discussion

The hypothesis proposed in this work is that engagement in play-centered activities focused around the construction and labelling of people's expressive eyes will increase attention to others face and eyes region and will improve expression recognition ability. To explore this hypothesis we developed a new technology, *Frame It*, and developed a set of experiments to test the hypothesis. The following discusses the conclusions of this study.

6.1.1 Frame It

Central to our investigation was the development of an accessible, engaging and fun system, *Frame It*, which could be used to study the ability to influence gaze behavior and expression recognition through play-centered activities. Critical to the design of *Frame It* were the methodologies of user-centered design and the considerations used to guide its development. The inclusion of children, teachers, therapists, psychologists and educational technology researchers in the design process helped us meet the challenges of designing a system for a group of children with many challenging conditions.

In addition to the usability test carried out and reported in chapter 3, the subsequent *Eyes Up* intervention further validated the system. Children in the study intuitively knew how to use the system and it was accessible to even the most challenged participants; those with little or no spoken language, low-compliance, difficulty focusing on tasks, poor motor skills, and cognitive challenges.

Post intervention surveys given to the participants' teachers indicate that several of the children thoroughly enjoyed playing with *Frame It.* Comments from teachers included "he loved doing the study!" and "...he always looked forward to attending the study." Four of the children would literally run into the study room saying phrases from the game and immediately start constructing puzzles.

One of the more challenged child's teacher commented, "he loves playing *Frame It.*" During several sessions with this same child the teacher commented, "I cannot believe how much better he is doing, I really didn't think he would be able to do this." It should be noted that these children represented some of the youngest participants, $m_1 = 8.78$ (1.82) vs $m_2 = 14.63$ (4.16).

While we feel that we accomplished one of our goals, developing an accessible and engaging technology for young children, our final design may not have been as engaging for older participants. This could be one reason that we have seen distinct differences, both quantitatively and qualitatively, between participants.

6.1.2 Gaze Behavior

Substantial therapy is directed towards influencing gaze behavior in persons diagnosed with ASD. A common phrase heard in therapy and school settings is "eyes up." This is both because of the importance of seeing other peoples social-emotional cues and for establishing expected gaze patterns to facilitate social interaction. Previous works have explored the social gaze behavior of people with Aspergers or with Autism diagnoses, who had normal or above normal scores on intelligence test, age appropriate speech skills and relatively few behavioral challenges, using still images and films [44, 33, 41]. This work continues this investigation with younger and more challenged children; those with little or no spoken language, low-compliance, difficulty focusing on tasks, poor motor skills, and cognitive challenges. Furthermore, this work seeks to influence gaze behavior and expression recognition abilities.

An initial finding we discovered with the gaze behavior of our participants, compared to previous works with HFA, was that they simply avoid looking at the face entirely. During pre-trial testing of the eye-tracker and data collection methods we discovered that even though the participants would face the monitor and images they would avert their eyes such that their focus would be outside of the image region or even so averted that the eye-tracker would not detect their gaze. This behavior continued with our study participants. Not only was this behavior captured in our eye-tracking and observational data, but surveys given to teachers noted that most of the participants typically avoided looking at other peoples eyes and face as much as possible (See Appendix B.1). Compared to other works we had substantially less gaze behavior directed to the face.

While other works have shown differences in gaze behavior of the face, i.e. HFA directing more attention to the lower face and mouth region than NTs, they did not show a significant difference in actual viewing of the face. That is, both groups looked at the face approximately the same amount of time–just in different ways. Our work does not compare ASD and NT groups, but we feel the results indicate a possible difference in gaze behavior between those with more challenging conditions, and HFA and NT individuals.

Group Dynamics of Gaze Behavior

Each phase of the pre- and post- testing has distinctly different characteristics and results. While phase I entails undirected viewing of images phase II and III direct the participant to determine the person's facial expression. Furthermore, although phase III presents the same images as phase II they have been altered to highlight the eyes region and obscure the rest of the face.

Results from the pre- and post- measures of gaze behavior indicate that the the intervention was associated with an improvement in participants' gaze behavior. Across all three phases of testing we see increases in attention to the face and eyes. Though not statistically significant, $p = 0.09$, it is interesting that during phase I participants had a greater number of fixations to the eyes, while during phase II and III, when directed to determine the person's facial expression, the participants increased their fixations to the lower face region. Since the eyes and mouth regions are the most dynamic in facial expressions it is reasonable that more attention would be directed to these areas when trying to determine a persons facial expression, and may account for this difference.

It is only for phase II that we see significant changes in gaze behavior for the entire group. Participants spent statistically significantly more time fixating on the face, $m_{pre} = 19.70\%$ (12.41%) vs $m_{post} = 28.31\%$ (13.75%), $p < 0.05$. Although phase II and III are the same with the exception of the image modification we do not see a similar result. This may be a result of our small sample size or because of the occlusion mask applied to the image. Perhaps the mask made the image less interesting to view because of the reduced global information. It is also possible that it made expression recognition more difficult, which may have frustrated some children and made them less interested in doing the task.

Subgroup Dynamics of Gaze Behavior

Similar to the analysis of the entire group both subgroups exhibited positive changes in gaze behavior, with more attention given to the face region and a substantial increase in attention to the eyes region. Only subgroup 1 had a statistically significant increase in fixations of the face, $p < 0.05$. Subgroup 2 approached significance, $p = 0.07$, for facial fixations with a moderate effect size, $d = 0.57$. These results suggest that given a larger sample size for subgroup 2 we may have seen statistically significant change for this measure too. Subgroup 2 also had a substantial change in fixations to the upper face region, $m_{pre} = 20.83\%$ (23.82%) vs $m_{post} = 9.19\%$ (15.23%); $p = 0.10$, with $d = 0.57$. All of these changes were accompanied with increases in expression recognition.

Interestingly, it is only in the pre-test that we see statistically significant differences between the two subgroups for phase II gaze behavior. Subgroup 1 had statistically significant fewer fixations on the upper face region than did subgroup 2, $m_{1-pre} = 1.52\%$ (0.23%), $m_{2-pre} = 20.83\%$ (23.82%); $p < 0.01$. For subgroup 1, they fixated to the lower face region the most in the pre-test, while for subgroup 2 they fixated to the upper face region. Given past findings showing that HFA tend to fixate the lower face (mouth) more, it is interesting to see this substantial difference between the groups. Such differences were not seen in the post-test, and in general the two groups have similar post-test results. This may indicate that the intervention actually had a substantial influence on subgroup 2, making them direct more attention to the eyes region and away from the upper face. These results may also indicate that there are different eye avoidance strategies used by people diagnosed with ASD, one to fixate on the mouth and another to fixate on the forehead.

6.1.3 Expression Recognition

Understanding other's social-emotional cues is of great importance and taken for granted by most individuals. Most people are experts at understanding others social signals without even realizing it, it is those that struggle with such tasks that make us understand the importance and difficulty of recognizing these subtle cues. We believe that by increasing attention to faces and eyes in those diagnosed with ASD we can improve their ability to recognize facial expressions. The aforementioned work on gaze behavior, using the *Frame It* system, shows that we have influenced gaze behavior towards viewing the face and eyes more. This has been accompanied by a substantial increase in expression recognition.

During phase II and III of the pre-test, participants were presented with images of peoples' entire face depicting different expressions and asked to choose one of three possible expressions for the person. During the subsequent intervention participants constructed puzzles of peoples eyes and then labelled the constructed puzzle with an expression label chosen from a set of four possible expressions. All participants took part in 11 sessions using *Frame It* before taking the post-test, which were identical to the pre-test.

Across all participants there was an increase in overall expression recognition between pre- and post- tests. These results are very encouraging and indicate a practical usefulness of the designed system and methodologies. It also indicates that children were able to generalize characteristics of expressions between the intervention and test images, as testing and intervention images were substantially different. Testing images were grey scale, closely cropped full faces at an image size of 384×570 pixels. Images used for the intervention were high resolution, full color of the eyes region only, and of size 1440×540 pixels, slightly larger than real-life size. Despite these substantial differences, participants were able to relate expressions learned over the intervention to test images.

Group Dynamics of Expression Recognition

Post intervention analysis of group expression responses shows improvements in all categories with statistically significant improvements in three of the six expression categories: sad, surprised, and afraid. In addition, their combined scores across all expressions for phase II and III both improved a statistically significant amount, $p < 0.001$. These changes are very encouraging and led us to investigate individual participants changes, which resulted in the discovery of two distinct subgroups.

Subgroup Dynamics of Expression Recognition

Those participants that had one or more areas of significant improvement, and at least a 15% overall improvement in recognition were grouped together, $N_1 = 3$, while all others formed the second group, $N_2 = 5$. Further analysis revealed that subgroup 1 was younger and had higher intelligence scores, $m_{age} = 8.78yrs$ and $m_{kbit} = 58.44$, than their peers in subgroup 2, $m_{age} = 14.63yrs$ and $m_{kbit} = 47.07$, respectively. Subgroup 1 performed better, with statistical significance, at recognizing fear, anger, happiness and surprise than subgroup 2. In addition, the overall response correctness for all expressions is significantly better for subgroup 1, $t(190) = 3.92; p < 0.01$.

For both gaze behavior and expression recognition abilities we found subgroup 1 performing markedly better than subgroup 2. As discussed above, these subgroups were different with respect to age, IQ, and expression recognition improvement. Subgroup 1 was younger, had higher IQ scores, and larger expression recognition improvements. Interestingly, the three members of subgroup 1 also had the highest levels of engagement and fun as reported by teacher surveys and experimenter observations. While age and IQ scores were not statistically different, both engagement levels and expression recognition improvements were different with statistical significance.

6.1.4 Puzzle Play

The *Frame It* system used for the intervention not only allowed for the study of gaze behavior and expression recognition, but recorded play data that was able to reveal patterns in participants' play. Analysis of game data showed that the entire group steadily became better at constructing and labelling puzzles. These results further validate the usability of the system and the methodologies used in its design.

To gain a better understanding of the differences between participants we conducted analysis on the two subgroups mentioned previously. Results show that subgroup 1 performed better in earlier stages of the intervention, but that performance levels near each other as the intervention proceeded. The most telling difference between the two groups is related to the average number of wrong expression labels selected per-puzzle. Analysis revealed a statistically significant difference between the groups, $t(108) = -3.32$; $p < 0.01$, with subgroup 1 becoming better at correctly labelling expressions.

Furthermore, these results correspond with the pre- and post- tests results for expression recognition, which show subgroup 1 becoming better than subgroup 2 at recognizing facial expressions.

6.2 Conclusion

This work was guided by an early intuition that play can be a powerful mediator of behavior change and learning. During early development of *Frame It* we worried that the eye images would be too stimulating and children would not want to interact with them. Subsequent testing proved this wrong, no child found the puzzles disturbing and in general children enjoyed playing the puzzle games. Furthermore, the system proved to be engaging and maintained interest over many sessions. This provided us with a means of exploring play-centered interactions with the goal of changing gaze behavior and expression recognition abilities. The pre- and post- tests, along with the intervention, provided a testable framework to detect change.

Overall, results indicate that children that found the game engaging, were younger, and had higher IQ scores performed better across all measures and found the system more enjoyable. Since we designed the system to be for younger children these finding are encouraging, but challenges us to extend these benefits to older children and those with more cognitive impairments.

Given the relative shortness of the intervention, 11 sessions over 5 weeks, the changes seen are encouraging. When the intervention started it became evident that we might not see any positive results, given the many challenges our participants were faced with. The fact that we have seen changes in gaze behavior and expression recognition is rewarding to everyone involved. We feel that the engagement and fun of playing the game had a significant influence in facilitating positive gains.

6.3 Future Work

Although this work has resulted in measurable benefits and accomplished its initial goals it has also revealed further questions and considerations. In part, we feel having more participants would allow us to better understand aspects of our findings. The small sample size makes us cautious of our findings and highlights our need to pursue more work with this population. Furthermore, the inclusion of a control group would better differentiate certain changes that were discovered. In addition, the study was relativity brief and may account for some participants not making more significant gains. Most importantly, we have not been able to perform assessment of real-world generalization of gains seen in experimental tests.

At the system level we are interested in making the *Frame It* platform more configurable and easier for users to customize. This would allow users to make very personalized games, which we feel would be of benefit. In addition, we would like to explore the use of dynamic media. Instead of constructing puzzles of still images, users could construct puzzles of social situations which involve a specific order of social events. The resulting puzzle would form a short film that would play out the constructed scenario. This could allow a child to explore different combinations of social interactions and the resulting consequences of those actions.

Particular questions that have arisen as a result of this study are:

- Is a tangible interface more engaging than a non-tangible interface

- Is engagement dependent on age or IQ
- Are there differences between "low" functioning ASD gaze behavior and that of HFA and NT individuals
- Would images of people in the child's life work better than images of strangers for puzzles
- Would dynamic media types work better than static images
- How does *Frame It* compare to similar ASD technologies
- How much do we need to look at the face and eyes to tell how someone feels
- Do the changes in gaze behavior translate to real-world social interactions

Further studies with regular physical puzzles, digital only puzzles, tangible-digital puzzles, and dynamic media puzzles that include familiar and unfamiliar faces could be structured to answer many of the above questions. A larger participant group that includes HFA and NT children with corresponding control groups would allow the investigation of group specific differences in gaze behavior, expression recognition and game play. In addition, the analysis of group specific time required to recognize facial expressions could be carried out. With the use of wearable eye-tracking equipment test of real-world gaze behavior could be carried out to determine if gains seen in testing are generalized to real-world situations.

Bibliography

[1] Tobii eye tracking technologies. `http://www.tobii.com`.

[2] R. Adolphs, L. Sears, and J. Piven. Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience*, 13(2):232–240, 2001.

[3] S. Baron-Cohen. *Mindblindness: An essay on autism and theory of mind.* The MIT Press, 1997.

[4] S. Baron-Cohen. Mind reading: The interactive guide to emotions (version 1.0/DVD ROM). *Human Emotions, Sheffield*, 2002.

[5] S. Baron-Cohen, O. Golan, E. Chapman, and Y. Granader. Transported to a world of emotion. *Psychologist*, 20(2):76, 2007.

[6] S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, and I. Plumb. The Reading the Mind in the Eyes Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(02):241–251, 2001.

[7] S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, and I. Plumb. The Reading the Mind in the Eyes Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(02):241–251, 2001.

[8] S. Baron-Cohen, S. Wheelwright, and T. Jolliffe. Is there a" language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, 4(3):311–331, 1997.

[9] S. Baron-Cohen, S. Wheelwright, A. Spong, V. Scahill, and J. Lawson. Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of Developmental and Learning Disorders*, 5(1):47–78, 2001.

[10] N. Bauminger, E. Gal, D. Goren-Bar, J. Kupersmitt, F. Pianesi, O. Stock, P. Weiss, R. Yifat, and M. Zancanaro. Enhancing social communication in high-functioning children with autism through a co-located interface. In *Proc 6th Intl Workshop Social Intelligence Design*, pages 15–23, 2007.

[11] M. Belmonte. A New Way to Study and to Treat Autism Spectrum Conditions: Video Games for Ecologically Valid Measurement and Therapy. 2008.

[12] K. Blocher and R. Picard. Affective social quest: emotion recognition therapy for autistic children. *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, pages 133–140.

[13] J. Bohn. The smart jigsaw puzzle assistant: Using RFID technology for building augmented real-world games. In *Workshop on Gaming Applications in Pervasive Computing Environments at Pervasive*, volume 2004. Citeseer, 2004.

[14] G. Celani, M. Battacchi, and L. Arcidiacono. The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders*, 29(1):57–66, 1999.

[15] J. Cohen. *Statistical power analysis for the behavioral sciences.* Lawrence Erlbaum, 1988.

[16] K. Dautenhahn. Design spaces and niche spaces of believable social robots. In *Proceedings of the International Workshop on Robots and Human Interactive Communication*, pages 192–197. Citeseer, 2002.

[17] K. Dautenhahn and A. Billard. Games children with autism can play with robota, a humanoid robotic doll. *Universal Access and Assistive Technology*, pages 179–190, 2002.

[18] P. Dietz and D. Leigh. DiamondTouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, page 226. ACM, 2001.

[19] A. Druin. The role of children in the design of new technology. *Behaviour and Information Technology*, 21(1):1–25, 2002.

[20] M. Eckhardt and R. Picard. A more effective way to label affective expressions. *Proceedings of Affective Computing and Intelligent Interaction ACII-2009. Amsterdam*, 2009.

[21] P. Ekman. The argument and evidence about universals in facial expressions of emotion. *Handbook of social psychophysiology*, 58:342–353, 1989.

[22] P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3):169–200, 1992.

[23] P. Ekman and W. Friesen. Pictures of facial affect, 1976.

[24] M. Ganz. The lifetime distribution of the incremental societal costs of autism. *Archives of Pediatrics and Adolescent Medicine*, 161(4):343, 2007.

[25] T. Grandin. Teaching tips for children and adults with autism. *Autism*, 2:5.

[26] K. Hendrixa, R. van Herka, J. Verhaegha, and P. Markopoulosb. Increasing Childrens Social Competence Through Games, an Exploratory Study. 2009.

[27] L. Hildman, P. Friedberg, and P. Wright. Kaufman Brief Intelligence Test. *Journal of Psychoeducational Assessment*, 11(1):98, 1993.

[28] M. Hopkins and F. Biasini. Facesay.

[29] W. Jones, K. Carr, and A. Klin. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of general psychiatry*, 65(8):946, 2008.

[30] R. Joseph and J. Tanaka. Holistic and part-based face recognition in children with autism. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 44(4):529, 2003.

[31] R. Kaliouby and P. Robinson. The emotional hearing aid: an assistive tool for children with Asperger Syndrome. *Universal Access in the Information Society*, 4(2):121–134, 2005.

[32] J. Kientz, G. Hayes, T. Westeyn, T. Starner, and G. Abowd. Pervasive computing and autism: Assisting caregivers of children with special needs. *IEEE Pervasive Computing*, pages 28–35, 2007.

[33] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism, 2002.

[34] M. Madsen, M. Eckhardt, M. Hoque, M. Goodwin, and R. Picard. Lessons from participatory design with adolescents on the autism spectrum. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 3835–3840. ACM New York, NY, USA, 2009.

[35] J. Manjiviona and M. Prior. Comparison of Asperger syndrome and high-functioning autistic children on a test of motor impairment. *Journal of Autism and Developmental Disorders*, 25(1):23–39, 1995.

[36] R. MAYER-JOHNSON. Boardmaker: the picture communication symbols libraries on disk. *Solana Beach: Mayer Johnson Co*, 2004.

[37] R. Morris, C. Kirschbaum, and R. Picard. Broadening Accessibility Through Special Interests: A New Approach for Software Customization. *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, 2010.

[38] J. Movellan, M. Eckhardt, M. Virnes, and A. Rodriguez. Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 307–308. ACM New York, NY, USA, 2009.

[39] J. Movellan, F. Tanaka, I. Fasel, C. Taylor, P. Ruvolo, and M. Eckhardt. The RUBI project: a progress report. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 333–339. ACM New York, NY, USA, 2007.

[40] S. Myers, C. Johnson, et al. Management of children with autism spectrum disorders. *Pediatrics*, 120(5):1162, 2007.

[41] C. Norbury, J. Brock, L. Cragg, S. Einav, H. Griffiths, and K. Nation. Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry*, 50(7):834–842, 2009.

[42] W. H. Organization. International statistical classification of diseases and related health problems, 10th revision. 1992.

[43] A. Paiva, R. Prada, R. Chaves, M. Vala, A. Bullock, G. Andersson, and K. Höök. Towards tangibility in gameplay: building a tangible affective interface for a computer game. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 60–67. ACM New York, NY, USA, 2003.

[44] K. Pelphrey, N. Sasson, J. Reznick, G. Paul, B. Goldman, and J. Piven. Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, 32(4):249–261, 2002.

[45] A. Piper, E. O'Brien, M. Morris, and T. Winograd. SIDES: a cooperative tabletop computer game for social skills development. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 1–10. ACM New York, NY, USA, 2006.

[46] D. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and Brain sciences*, 1(04):515–526, 2010.

[47] B. Prizant and A. Wetherby. The SCERTS Model. *Infants and Young Children*, 16(4):296–316.

[48] J. Read. The ABC of CCI. *Interfaces*, 62:8–9, 2005.

[49] C. Rice. Prevalence of autism spectrum disordersautism and developmental disabilities monitoring network, six sites, United States, 2000. *MMWR Surveill Summ*, 56:1–11, 2007.

[50] B. Robins, K. Dautenhahn, R. Te Boekhorst, and A. Billard. Robotic Assistants in Therapy and Education of Children with Autism: Can a Small Humanoid Robot Help Encourage Social Interaction Skills? 2008.

[51] S. Rogers. Interventions that facilitate socialization in children with autism. *Journal of Autism and Developmental Disorders*, 30(5):399–409, 2000.

[52] N. Sasson, L. Turner-Brown, T. Holtzclaw, K. Lam, and J. Bodfish. Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research*, 1(1), 2008.

[53] B. Scassellati. How social robots will help us to diagnose, treat, and understand autism. In *Proc. of the 12th International Symposium of Robotics Research (ISSR05)*. Springer, 2005.

[54] E. Schopler, R. Reichler, R. DeVellis, and K. Daly. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders*, 10(1):91–103, 1980.

[55] F. Shic, K. Chawarska, J. Bradshaw, and B. Scassellati. Autism, eye-tracking, entropy. In *7th IEEE International Conference on Development and Learning, 2008. ICDL 2008*, pages 73–78, 2008.

[56] M. Spezio, R. Adolphs, R. Hurley, and J. Piven. Analysis of face gaze in autism using. *Neuropsychologia*, 45(1):144–151, 2007.

[57] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science: A Multidisciplinary Journal*, 12(2):257–285, 1988.

[58] F. Volkmar, S. Sparrow, R. Rende, and D. Cohen. Facial perception in autism. *Journal of Child Psychology and Psychiatry*, 30(4):591–598, 1989.

[59] T. L. Westeyn, T. E. Starner, G. D. Abowd, R. I. Arriaga, and P. W. Presti. An Initial Investigation Using Augmented Toys and Statistical Models to Automatically Categorize Object Play Behaviors. *(Abstract) International Meeting for Autism Reserach (IMFAR)*, May 7-9 2009.

Appendix A

Committee on Use of Humans as Experimental Subjects forms

MIT Committee On the Use of Humans as Experimental Subjects

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
77 Massachusetts Avenue
Cambridge, Massachusetts 02139
Building E 25-143B
(617) 253-6787

To:	Rosalind Picard E15-448
From:	Leigh Firn, Chair COUHES
Date:	03/24/2009
Committee Action:	**Approval**
COUHES Protocol #:	0902003125
Study Title:	Frame It: Social Signals
Expiration Date:	03/18/2010

The above-referenced protocol has been APPROVED following Full Board Review by the Committee on the Use of Humans as Experimental Subjects (COUHES).

If the research involves collaboration with another institution then the research cannot commence until COUHES receives written notification of approval from the collaborating institution's IRB.

It is the Principal Investigator's responsibility to obtain review and continued approval before the expiration date. Please allow sufficient time for continued approval. You may not continue any research activity beyond the expiration date without COUHES approval. Failure to receive approval for continuation before the expiration date will result in the automatic suspension of the approval of this protocol. Information collected following suspension is unapproved research and cannot be reported or published as research data. If you do not wish continued approval, please notify the Committee of the study termination.

Adverse Events: Any serious or unexpected adverse event must be reported to COUHES within 48 hours. All other adverse events should be reported in writing within 10 working days.

Amendments: Any changes to the protocol that impact human subjects, including changes in experimental design, equipment, personnel or funding, must be approved by COUHES before they can be initiated.

Prospecitve new study personnel must, where applicable, complete training in human subjects research and in the HIPAA Privacy Rule before participating in the study.

COUHES should be notified when your study is completed. You must maintain a research file for at least 3 years after completion of the study. This file should include all correspondence with COUHES, original signed consent forms, and study data.

Figure A-1: *MIT COUHES Approval*

2/26/09 D

MIT IRB APPROVAL MAR 2 4 2009 COUHES

MIT IRB APPROVAL EXPIRES SEP 1 8 2010 COUHES

Children's Assent Form

Social Signals: Emotional expression recognition
of face and eye region for children with an Autism Spectrum Disorder

My name is [ ]. I am doing a study to determine if an interactive puzzle game, Frame It, can be used to test peoples ability to recognize emotional expressions of the face and eye region. Additionally, I will be studying whether *Frame It* can be used to teach people to recognize facial expressions and reduce the stress that many autistics feel when making eye contact with others.

If you agree to this study, here is what will happen: I will ask you to wear a special *ICalm* wristband that records when you are excited. Once you have the wrist band on you play a game that requires you to match a facial expression that you see on the computer monitor. You will have to select the correct physical puzzle pieces and put them together so that they match the facial expression on the computer monitor. You will be asked to do this for several different faces. Over a two week period you will be asked to play this game once a day for approximately 15 minutes.

We will show you how to move the wristband around so it is not uncomfortable and we will show you how to play *Frame It*. At anytime if the wristband is uncomfortable or you otherwise do not want to wear it you can take it off. The wristband should not ever hurt. If they do, please take them off and show you wrist to your teacher or me. Also, if for any reason you do not want to play *Frame It* you can stop.

Your part in this study is confidential. No one else will know if you were in this study and no one else can find out how you felt during the study. We will keep all the records for this study in locked files and only people involved in the study will be able to access them.

You should talk this over with your parents before you decide whether or not to be in the study. I will also ask your parents to give their permission for you to be in this study. But even if your parents say "yes," you can still decide not to do this.

You may or may not benefit personally from participating in this study. It is possible, but not guaranteed, the use of the puzzle game *Frame It* will help you become more comfortable with eye contact and to better recognize facial expressions.

Signing this paper means that you have read this form or had it read to you and that you want to be in the study. If you don't want to be in the study, don't sign the paper. Remember, being in the study is up to you. No one will be mad if you don't sign this paper or even if you change your mind later.

Signature of participant: ____________________ Date: ____________

Signature of witness: ____________________ Date: ____________

Signature of investigator: ____________________ Date: ____________

Figure A-2: *Childrens Consent Form.*

REVISED

MAR 24 2009 COUHES

EXPIRES SEP 18 2010 COUHES

PARENTAL CONSENT LETTER

Frame It: Emotional expression recognition
of face and eye region for children with Autism Spectrum Disorder

Date:

Dear Parent/Guardian,

I am writing to ask your consent for your child's participation in a research project at the Groden Center, Inc. in collaboration with the Massachusetts Institute of Technology Media Lab. The Media Lab has developed an interactive, tangible-digital puzzle game, *Frame It*, that is intended for use as a testing and teaching tool for recognizing facial expressions. The proposed study will also incorporate the Media Lab technology *iCalm*, a wrist-worn sensor for measuring autonomic arousal (for example, skin conductance and heart rate change when one is frustrated, stressed, or thinking about something particularly meaningful). The purpose of the research is to establish if the *Frame It* game can be used as a reliable testing tool to determine if a child has difficulty recognizing emotional expressions of the face and eye region. Additionally, we will be testing if *Frame It* can be used as a a means to teach children to recognize facial expressions and reduce their anxiety with direct eye contact. You have the right to be present at any and all sessions. You should read the information below, and ask questions about anything you do not understand before deciding to let your child participate in this research project.

PURPOSE OF THE STUDY

Difficulty recognizing facial expressions and anxiety with direct eye contact are common among people with Autism Spectrum Disorder. The MIT Media Lab has developed an interactive, tangible-digital puzzle game, *Frame It*, for the purpose of testing and teaching children to recognize emotional expression of the face and eye region. We wish to evaluate this technology to help children with and without autism to learn important social signals.

PROCEDURES

If you volunteer your child to participate in this study, here is what will happen:

1. We will ask your child to wear a small sensor, *iCalm*, sewed into a wristband during a 15 minute regularly scheduled testing and teaching sessions at the Groden Center over a two week period. The sensor will record physiological signals. In a previous study done at the Groden Center children had no problems wearing these sensors. It is possible that the wearable sensor could be uncomfortable, for this reason we will ask your child's teacher(s) to make sure your child only wears the sensor if your child finds it comfortable. We will ask your child's teacher(s) never to pressure him or her to wear the sensor. We will be happy to let you try the sensors before we put them on your child so you have a better understanding of how it feels and how it should be worn.
2. Once wearing the *iCalm* sensor, or not if they choose, your child will play an emotional expression matching game that requires them to put together a physical puzzle to match a facial expression that is shown on the game monitor. The puzzle pieces will be of human eyes. As the your child puts the puzzle together the computer monitor will provide them with feedback about the correctness of their puzzle, helping to guide them to the correct solution.

POTENTIAL RISKS AND DISCOMFORTS

Figure A-3: *Parental Consent Form.*

Appendix B

Study Survey Results

Table B.1: *Teacher survey results for student gaze behavior related to viewing others eyes and face, 0 = not child's typical gaze behavior, 1 = child's typical gaze behavior*

Participant	Normal Gaze	Strong Avoidance	Angled Gaze	Brief Engagement*
P01	0	0	0	1
P02	0	0	0	1
P03	0	0	1	0
P04	0	0	0	1
P05	0	0	0	1
P06	0	0	0	1
P07	0	1	0	0
P08	0	1	0	0
P09	0	0	0	1
P10	1	0	0	0

*Child looks at eyes/face very quickly then avoids looking at eyes/face.