Developing an Interactive Social-Emotional Toolkit for Autism Spectrum Disorders

By

Miriam A. Madsen

S.B., C.S. M.I.T., 2009

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

August 2010

[September 2010]

Author_____

Department of Electrical Engineering and Computer Science

August 20, 2010

Certified by_____

Rosalind Picard

Thesis Supervisor

Certified by___

Rana el Kaliouby

Thesis Co-Supervisor

Accepted by_____

Dr. Christopher J. Terman

Chairman, Department Committee on Graduate Theses

Abstract

A development process consisting of participatory design and iterative implementation was carried out to create a framework for interactive emotion-learning, the Interactive Social-Emotional Toolkit (iSET). iSET is a novel intervention consisting of live video recording and annotation software for use during social interactions as well as video review and discussion components. It is suitable for persons diagnosed with Autism Spectrum Disorders (ASDs) including Autistic Disorder (AD) and Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS), the target groups for the intervention, as well as for persons with Asperger's Syndrome (AS). The iSET intervention was tested with a group of AD/PDD-NOS participants (n=20) with mean age 22.7 ± 8.55 years; these students were divided into an experimental group to test the iSET paradigm (n=10) and a control group following Golan and Baron-Cohen's Mind Reading DVD intervention approach (n=10). An age- and sex-matched group of neurotypical participants (n=20) were also tested with the pretest measures. Preliminary results show an increasing ability to use the iSET materials and to capture videos that neurotypical teachers considered "good examples" of emotions considered in the intervention.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 0555411 and the Things That Think Consortium. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The hardware was generously provided by Samsung and Logitech, and the tracker used as a front-end for the face analysis software was provided by Google.

The author would like to thank the participants, teachers, and staff of the Groden Center, particularly Betsy Stratis, Eliza Lane, and Meghan Scrimgeour, with extra thanks to Daniella Aube. She would also like to thank Matthew Goodwin, Rana el Kaliouby, and Rosalind Picard for their incredible support during this project.

Contents

1 Introduction

1.1 Autism Spectrum Conditions

Persons diagnosed with Autism Spectrum Disorders (ASDs) can face a variety of challenges during everyday tasks and activities. (The rest of this thesis will use the alternative language Autism Spectrum Conditions [ASCs] instead of Autism Spectrum Disorders [ASDs] to highlight the range of abilities and strengths associated with this population, including enhanced perception, increased systematizing skills, and sensory hypersensitivity [1, 2, 3, 4].) Some of the issues associated with ASCs are limited to non-social behavior such as stereotyped movements or restricted interests [5]. However, ASCs are also characterized by deficits directly related to social communication and emotion-reading. These deficits include: 1) issues in understanding the nonverbal cues and mental states of others [6, 7, 8, 9]; 2) reduced duration of facial feature fixation time (especially on the eyes) and atypical processing of eye gaze [10, 11, 12]; 3) difficulty in expression and comprehension of one's own feelings [13]; and 4) tendency toward monologuing and non-reciprocating conversation [14]. Persons on the spectrum have also reported issues with "mind reading" abilities, known as "Theory of Mind," which include understanding the beliefs, perspectives, and behavior of others [15].

For persons on the spectrum, many of whom have moderate to severe emotion-reading deficits, daily life and social interactions can be a continuous challenge. Like someone visiting a foreign country who does not speak the native language, a person diagnosed with an ASC such as Asperger's Syndrome (AS), Autistic Disorder (AD), or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) may struggle to communicate with others, impeded by a lack of a common interaction paradigm. As they misinterpret nonverbal communication, fail to issue socially appropriate pleasantries, and miss important emotional cues, they may struggle to make friends, create meaningful emotional connections, and successfully navigate important settings such as the workplace.

There are a variety of methods to help people with social/communicative impairments. Many of these methods rely on predigested images, audio, and video packages designed by persons without any personal mental history of emotion-recognition impairment. This process can be like learning a foreign language in a classroom, using a textbook written entirely in the language of study; while some learning may occur, using that learning in daily interactions can be a difficult task, and many subtle points may be lost without an appropriate translation between the learning setting and a real-world practice setting. People who learn to read emotions from video clip lessons that claim a large smile means "happy" and a deep frown means "sad" may be left adrift in a world full of spontaneous expressions of interest, disgust, surprise and confusion. Real-time recognition of complex natural emotions produced during face-to-face conversations are nontrivial for most persons on the autism spectrum. One young adult user of the Mind Reading DVD intervention - a cutting-edge therapeutic approach that includes many different games and lessons illustrating a variety of emotions - told us that he became an expert user, acing all the quizzes on the DVD; however, he experienced almost none of this success in real-life emotion recognition. He needed technology that helped him in real life interactions [16].

The Interactive Social-Emotional Toolkit (iSET) intervention provides a real-world practice setting as an integral part of the approach; rather than rote classroom practice, the iSET intervention teaches the ability to learn emotion in a context that involves all the complexity of real-world interactions. By allowing students to record, annotate, and review video of their peers and teachers, students can learn from videos of familiar people making expressions relevant to the students' own experiences. We hope to show that our mobile, *in situ* intervention represents a significant improvement over existing approaches to teaching emotion to people on the spectrum.

1.2 Autism and Technology

The Affective Computing Group at the MIT Media Lab works to improve the lives of persons on the spectrum through novel technology development. The group is leading a new area of research developing techniques to understand and address challenges faced by people on the spectrum and strives to integrate direct intellectual contributions from the autism community. Technologies are designed in conjunction with this community and their input is frequently sought during development of these devices. The iSET project applies these participatory design goals to the area of learning emotion reading.

iSET aims to make a severalfold contribution in this area. First, the project works to create a technology "partner" for persons on the spectrum. By creating a social "guide" that is repeatable, consistent, and calming (by reducing chaos and uncertainty in social interaction) - and perhaps even fun! - iSET will help its users learn to navigate their everyday difficulties in social interaction in a way that is meaningful and expressive rather than passive. Second,

the iSET project strives to create a meaningful assessment of social interaction ability. Bringing technology into the picture allows for creation of more objective measures. Rather than subjective evaluation by an occasional viewer of how much eye contact is made over time or how atypical a persons's expressions are, technology can help its users develop an objective and more meaningful measure of how social interactions are taking place, what factors affect social behavior, and how interactions can be better navigated. iSET seeks to quantify, simplify, and clarify emotion-reading social issues.

The solution of technology to a difficult problem, such as teaching emotion, is a new capability; in the past, computers equipped with onboard cameras were prohibitively expensive, meaning that subjective human-based assessment of social capabilities was the only feasible solution. Because of the current ubiquity of this type of technology, however, the iSET approach shows promise as a way to teach emotion-reading in the natural environment of a person on the spectrum.

1.3 Aims and Goals of Thesis

This thesis will cover the participatory design, implementation, and clinical application that comprise the Interactive Social-Emotional Toolkit (iSET) intervention, conducted collaboratively between the Affective Computing Group of the Media Lab and the Groden Center of Providence, Rhode Island. The iSET intervention was developed in order to show the possible benefits to users on the spectrum of recording and viewing naturalistic video in an open-ended intervention. The iSET approach was contrasted in this study with the Mind Reading DVD approach [16]. The Mind Reading DVD, an educational tool illustrating the expression of emotion through videos of actors expressing emotions, was tested with users by Golan and Baron-Cohen and acted as our standard for emotion-teaching tools [20].

1.4 Outline

This thesis will be arranged as follows:

- Section 2: *Related work.* This section will describe related tools in emotion communication (technology-mediated and otherwise), as well as other tools (which can be used for communication of emotion as well as other purposes) that were included as components of the iSET intervention.
- Section 3: *Technology Development and Participatory Design.* This section will introduce the iSET intervention, including each of its software/web components (iSET Interface and VidL), the Emotion Hunt game used during the intervention, and the iterations on each component of the intervention that were part of the design process.
- Section 4: *Experimental Design.* This section will cover the hypotheses, participants, instructions, and procedure of the iSET intervention.
- Section 5: *Construction and Administration of the VidL Pretest.* This section will discuss the creation of the VidL pretest, a set of naturalistic and acted videos illustrating emotion expression.
- Section 6: *Pretest Results and Analysis.* This section will cover the goals and results of each pretest used in the intervention, as well as correlations between the pretest results and possible clinical relevance.
- Section 7: *Preliminary Intervention Data.* This section will analyze results collected so far in the intervention.

2 Related Work

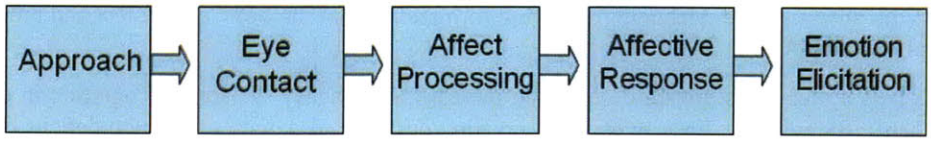
2.1 Introduction

The Interactive Social-Emotional Toolkit (iSET) intervention is designed as a technology-based paradigm that will allow users to understand emotions through recording, annotating, and viewing naturalistic clips of nearby persons expressing emotions.

In order to understand the potential impact of iSET, the field of existing approaches for social/pragmatic interventions must be more closely examined. Many approaches that exist to address the needs of people on the spectrum have usually been developed with a particular type of participant in mind; furthermore, the type of technology can vary for each intended participant group.

Interventions to teach emotion-reading usually address some, although not all, components of interpersonal communication. Figure 1 shows a diagram of the important nonverbal components of social interactions.

Figure 1: Nonverbal Components of Interpersonal Interactions. Each component shown here is crucial to success in everyday social interactions; leaving out any element, such as eye contact or affective response, can result in a negative social results and suboptimal communication.



These interventions can vary greatly on the question of whether technology is a major component of the interactions. Because of the relative recency of inexpensive and easily accessible technology, many current structures for teaching social skills do not include it. However, technology is increasingly included as a major component because it is fun and repeatable and may also be more scalable than existing approaches.

One variation across emotion-teaching interventions is degree of structure involved in the approach. These interventions can usually be administered in a variety of ways; however, depending upon the approach's goals, directions for use may specify varying degrees of structure to be carried out by the caretaker of the user, the person administering the technology, or the user themself. For instance, in the application of one technology that will be discussed further, the Mind Reading DVD, participants were asked to spend a certain amount of time in each area of the software (so, for instance, they would avoid spending too much time in the Rewards area).

Another differentiator between emotion-teaching approaches is the degree of generalizability that can be applied. Emotion-reading teaching tools must take into account known difficulties in terms of generalization for these users. Baron-Cohen has found that people on the spectrum may instead have relative strengths in the area of systematizing [19]; interventions that address emotion-reading deficits must therefore be evaluated partially on the opportunities they give to develop systems pertaining to the information presented (rather than heuristics more suitable to generalizing strengths).

One important differentiation in the design of emotion-teaching interventions is the age group of the intended participant. Depending on the age and developmental stage of the participant, emotion-reading educational approaches may be designed differently and have different end goals. With young participants (and particularly with very young participants) there is a significant focus on "intervention" with a goal of training the user in a standard social paradigm, so that approximately normal social behavior will eventually take place.

In contrast, older users may have different goals. For adults, it may be more helpful to consider coping strategies for everyday interactions. Furthermore, both the chronological age and the mental age of the participants must be taken into account. Persons who are more independent or socially involved will have different needs than those who are in a caretaking setting for most of each day (such as the persons in our study). One facet of the iSET design was to create a technology that would be easy to use and relevant for people on the spectrum with a variety of ages and mental abilities.

Finally, approaches to help people on the spectrum learn to read emotion should incorporate information that occurs across the timespan of interpersonal interactions. Rather than addressing face-reading in a vaccuum, tools to teach this skill should assist with the other components of these interactions, such as approach, eye contact, and emotion elicitation. (See Figure 1.) iSET aims to address all of these elements in a single learning process.

2.2 Existing Affect-Teaching Interventions

A number of tools exist at the present time to facilitate communication and understanding of emotion. Some examples are low-tech and have been incorporated across a variety of media; some are higher-tech and must be administered in a particular fashion for effective use; and some may be used in a variety of contexts, resulting from a fusion of lower-tech and higher-tech media.

- The Mind Reading DVD was designed and implemented at the University of Cambridge in a multi-year project designed to collect a large and varied collection of emotion expressions in images and video, as well as through sound clips, written stories, and games. The DVD includes 24 basic categories encompassing 412 emotions total, each of which has many included video clips, images, and stories. The program can be used at multiple levels of complexity. A study of its efficacy was carried out by Golan and Baron-Cohen and replicated by LaCava et al. [20, 21]. (See Section 2.3 for more information.) The DVD can be purchased from Jessica Kingsley Publishers [16].

- The SelfCam paradigm, which was the start of the iSET project and was created in the MIT Media Lab's Affective Computing Group in 2006, consisted of a wearable camera trained on the expressions of its wearer [22, 23]. This data could be read and analyzed in real time by the iSET Interface (see Section 3.2); the information gathered would be communicated back to the wearer *in situ*, allowing for dynamic reading of affect in interpersonal situations with an assist from technology. More information about the SelfCam project and previous iterations of iSET is in Section 3.5.

- Let's Face It is a piece of software designed to increase the efficacy of facial recognition in children on the spectrum [26]. By enlisting "face tutors" and creating customized games suiting the deficits of a pilot group of children affected with face-reading deficits, Let's Face It aimed to produce an intervention suitable for use in early childhood that would improve the ability of people on the spectrum to improve face and emotion recognition.

- FrameIt was a game developed by Eckhart and Picard at the MIT Media Lab in 2008 [24]. It is a "tangible-digital" game which includes both physical puzzle pieces (showing parts of the eye area of the face) and a screen which correctly selects and renders the pieces used through RFID tag connection.

- The PECS (Picture Expression Communication System), which contains a variety of standardized images that can be combined with words on cards and used across a variety of media, is perhaps the best known; it is frequently used in combination with the Boardmaker system [25]. This system can be used to communicate simple concepts word-by-word without use of verbal languages and in a simple, direct way for users who may have deficits in cognition or speaking. More recent systems usually incorporate a greater degree of technological sophistication. Direct descendants of the Boardmaker/PECS paradigm can be seen in devices used with persons who have communication difficulties.

- Social Stories are a series of text-based teaching activities oriented around the activities that a child might face in everyday life, e.g., strong feelings about where to stand in a classroom line [27]. By adapting a storytelling format to help people on the spectrum think through interpersonal interactions in their daily lives, Social Stories seeks to help people on the spectrum navigate social occasions with less stress.

- TeachTown is an environment simulating a real-world town suitable for use with children on the autism spectrum [28]. The primary goal is to allow players to simulate many everyday reactions in a safe virtual environment. The lifelike animations shown can be varied in naturalism from very abstract to very realistic (e.g., from geometric shapes approximating eye gaze to pictures of actual people in a joint attention exercise.)

- The Transporters is a software application designed to teach emotion-reading using lifelike images of animated trains [29]. Similar to TeachTown, lifelike animations work to help persons on the spectrum, especially children, imitate and practice the skills needed for everyday interactions.

- FaceSay is an innovative new software package for teaching emotion reading and social skills [56]. It incorporates realistic-looking faces and a variety of social scenarios to teach the skills to navigate interpersonal situations. FaceSay is unusual in that it has results indicating broad generalization, i.e., fewer negative interactions, more social behavior, and more person-to-person approach in a playground setting.

Section 2.4 will cover the shortcomings of these interventions.

2.3 Mind Reading DVD Intervention

The Mind Reading DVD is a tool developed by Baron-Cohen et al. in order to facilitate learning of emotions and affective communication, particularly by people on the spectrum. By providing a form of interactive multimedia appropriate for several different levels of developmental ability, the DVD offers many audio, visual (both image- and video-based), and textual examples of a wide variety of emotions [16]. The Mind Reading DVD is also one of the few existing interventions to include emotions beyond the basic ones, including more complex cognitive states; this aspect made it a good comparison choice for the iSET intervention. The examples are presented dynamically (i.e. with facial gestures in video, rather than static images, illustrating an emotion), allowing it to serve as a control for other video-based interventions.

Baron-Cohen and Golan created an intervention to test the effects of consistent use with the Mind Reading DVD in both group and non-group situations. (See Table 1 for information about the intervention structure.) During their intervention, they focused on practice with persons diagnosed with AS and high-functioning autism (HFA). They included two groups of AS/HFA participants (one, the control group, participated in before and after assessments but did not use the DVD) as well as a matched group of neurotypical participants with no diagnosis or family history of ASCs.

The first experiment took place at home for each participant. After the initial assessment session, participants in the experimental group were asked to use the DVD for 10 hours total over a period of 10-15 weeks. In the second experiment, a similiar setup took place but the Mind Reading DVD ASC group (as well as the control ASC group) took part in a social skills workshop concurrently.

Table 1: Format of Golan/Baron-Cohen Mind Reading DVD Intervention. Each of the two experiments included two AS/HFA and one NT participant group; the AS/HFA groups experienced either the experimental or the control setting, including a social skills group (depending on the experiment), and NT groups were tested but not otherwise involved in the intervention.

Experiment	Participants	Pretest	Posttest	Time with DVD	DVD at home	Social Skills Group
1	AS/HFA Experimental (n=19)	yes	yes	10 hours	yes	no
1	AS/HFA Control (n=22)	yes	yes	-	no	no
1	NT Control (n=24)	yes	no	-	no	no
2	AS/HFA Experimental (n=13)	yes	yes	10 hours	yes	yes
2	AS/HFA Control (n=13)	yes	yes	-	no	yes
2	NT Control (n=13)	yes	no	-	no	no

Before and after both experiments of the intervention, participants were asked to test their acquisition of emotion-reading knowledge across two modalities, audio and visual, and across three types of generalization: *close generalization* (video shown in Mind Reading, presented in a different format with different answers; *feature-based distant generalization* (similar video/audio not shown in Mind Reading), and *holistic distant generalization* (scenes from feature films).

In both experiments, the AS/HFA groups who used the Mind Reading DVD improved from baseline in their abilities to recognize the Mind Reading DVD clips; however, they did not improve on either feature-based or holistic generalization tasks, suggesting that the time spent using the DVD was not directly useful to the participants in everyday face recognition tasks.

2.4 Shortcomings of Existing Affect-Teaching Interventions

Existing behavioral interventions to teach emotion-reading usually had some problematic characteristics that are addressed in the design of the iSET intervention. (See Table 2 for comparison criteria across intervention technologies.) The main shortcomings common to these interventions are: a) lack of relevance to conversational partners that the student will encounter in day-to-day life; b) lack of relatedness to situations that the student will encounter in day-to-day life, and c) stimuli that are significantly different than the stimuli presented in *in situ* emotion-reading interactions. These issues will be covered individually below.

Most interventions that teach emotion-reading skills to persons on the spectrum rely on hypothetical conversational partners so that the interventions can portray the interactions from a third person perspective. For instance, in the

BoardMaker kit, the user as well as any conversational partners are represented graphically by minimally expressive face images; in the Mind Reading DVD, users are asked to empathize with actors portraying various characters used throughout the program. The objects of emotion-reading in these interventions do not include nearby persons who would be included in real-life interactions.

Furthermore, the situations discussed in the interventions currently used to teach emotions cannot be customized in most cases to include particular subjects with which the user would be familiar. For instance, in the Mind Reading DVD, many interactions concern types of relationships that the user may not have had or many not be able to relate to (such as a boyfriend/girlfriend); similarly, while Boardmaker may be able to handle generic topics such as music or baseball while modeling day-to-day conversations, it would struggle to effectively include the specific topics of high interest to our participants, such as particular sports teams or bands. Relevance to personal interests and daily life to users is a critical part of a successful intervention, since relevance can significantly boost motivation and interest; this factor is particularly important to consider when working with persons on the spectrum, who may have had significant emotional issues with social interactions in the past.

Few of these interventions in any way address the critical need for eye contact, except through animated exercises; without a realistic practice setting and application for this skill, the success of an intervention is likely to be impaired. Eye contact is a valuable and often-required skill for everyday social interactions, and leaving out a direct application is problematic.

Finally, many of the stimuli presented in these artificially created interactions or situations lack many of the salient aspects of *in situ* interpersonal interactions. In the Mind Reading DVD, images and video are presented by actors with clear facial expressions wearing non-distracting clothing on white backgrounds; each actor only presents one emotion at a time in a given image or video.

Table 2: Comparison of iSET with Existing Affect-Teaching Interventions. The iSET intervention includes significantly more practice of skills crucial to everyday reaction, including social approach and eye contact. Many other interventions do not show generalizability to real-life situations, such as through *in situ* significant observations or testing using stimuli that mimic everyday interpersonal interactions, e.g. naturalistic video. The iSET intervention was developed in participatory sessions with people on the spectrum across a variety of ages; many other interventions only show evidence of design and testing with children.

	iSET	*FaceSay*	*Mind Reading*	*Transporters*	*TeachTown*	*Let's Face It*	*Social Stories*
Is Engaging	**yes**	yes	yes	yes	yes	yes	yes
Is Designed for Children	**yes**	yes	yes	yes	yes	yes	yes
Is Designed for Adults	**yes**	no	no	no	no	no	no
Includes Basic Emotions	**yes**	yes	yes	yes	yes	yes	yes
Includes Complex Emotions/Cognitive States	**yes**	no	no	no	no	no	no
Includes Static Presentations of Emotions	**yes**	yes	yes	yes	yes	yes	yes
Includes Dynamic Presentations of Emotions	**yes**	yes	yes	yes	yes	yes	yes
Requires Social Approach	**yes**	no	no	no	no	no	no
Requires/Encourages Eye Contact	**yes**	no	no	no	no	no	no
Requires Emotion Elicitation	**yes**	no	no	no	no	no	no
Includes Socializing *in situ*	**yes**	no	no	no	no	no	no
Contextualizes everyday social interactions	**yes**	no	no	no	no	no	no
Has evidence of real-life generalizability	**?**	yes	no	no	no	no	no

2.5 Conclusion

Current technologies for addressing emotion-learning in autism have been improving; however, many of them still assume significant generalization abilities on the part of their users, which is a known weakness for this population. By addressing situations that may not be familiar to less-socialized target populations, these approaches may not be well suited for their audiences. Existing interventions provided the inspiration behind the iSET Project; iSET addresses the ecological validity of these interventions and strives to create a better learning situation for emotion-reading skills that can be generalized to real-life interaction.

3 Technology Development and Participatory Design

3.1 Introduction

iSET, the Interactive Social-Emotional Toolkit, was developed through many rounds of collaborative iteration with users on the spectrum and their caregivers in order to create a tool that would facilitate the needs of users by improving their abilities to express, read, and understand facial affect and emotion cues. The overall structure of the experimental group's procedure in the iSET intervention consists of two main components:

1. *Recording Sessions.* In a recording session, the user first watched example video clips approved by neurotypical labelers as demonstrating a particular emotion (see Section 5.3). These clips (as well as the clips in the viewing session) were watched in the VidL interface described in Section 3.4, and only one possible answer - the right one - was offered for each clip. Next, the participant worked with a caregiver to record video of the faces of nearby people, sometimes including the caregiver. During this time, the participant had an opportunity to annotate the faces that they are are recording with emotion labels during the real-time recording process. The software used for this component is the iSET Interface, a software component custom-designed for iSET at the Media Lab; this software is run on the Samsung Ultramobile computer platform, which is well-suited to the project because of its easily holdable size, large screen, and rear-facing camera ideal for recording faces of conversational partners (see Section 3.2). During a Recording Session, participants played the Emotion Hunt game in order to collect examples of emotions; this game is described in Section 3.3.

2. *Viewing Sessions.* In a viewing session, participants watch three types of clips through the VidL interface (see Section 3.4).

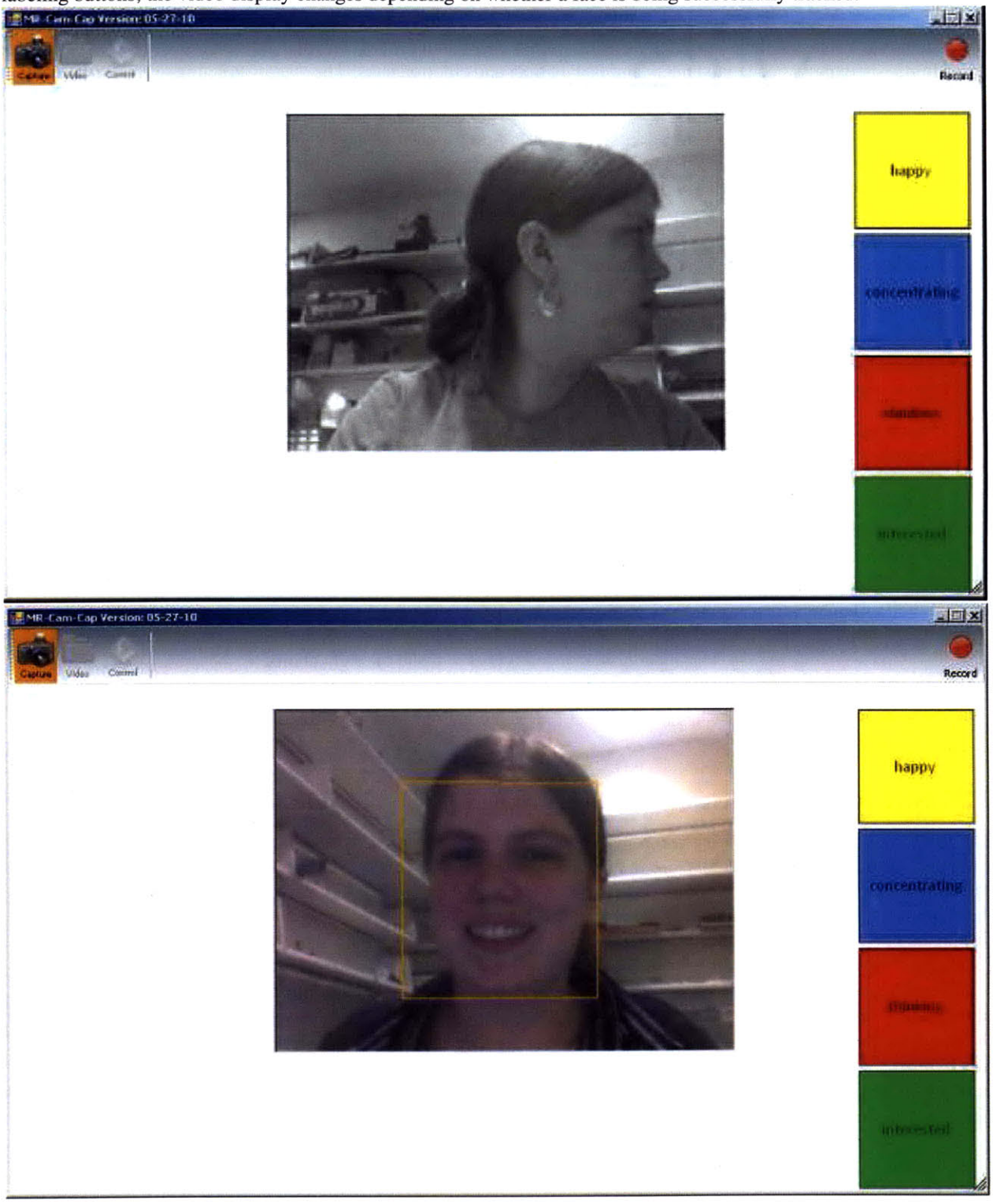
    - First, participants watched example clips approved by neurotypical labelers as representative of the expressed emotion. In the examples viewing, which is identical to the examples viewing during the recording session, only one option (the correct one) was given for each clip in order to provide an errorless learning environment conducive to motivated learning and engagement.
    - Participants then watched preapproved clips; these came from the same corpus of pre-recorded and NT-approved clips as the examples clips, except that starting in Week 05, these clips are shown with more than one possible choice. Participants were incentivized to make the right choice by the showing of customized Reinforcing Images.
    - Finally, participants watched clips that they recorded of peers or nearby strangers during the most recent recording session.

3.2 iSET Interface Development and Samsung Ultramobile Hardware Use

The iSET Interface allows for easy recording and annotation of video. The key components of the iSET Interface, which will be described individually, are a) the video display in the center of the window, in which real-time video is shown; b) the annotation buttons to the right of the video; c) the annotation notification area below the video; and d) the start/stop video recording button (see Figure 2). There is also a control panel in which settings which affect the display and processing can be altered. The iSET Interface is based on the FaceSense software created by el Kaliouby [31].

The video display area shows video captured in real time from the camera on the reverse side of the Samsung ultramobile computer; thus, the video shown is what would be seen by the device user in the absence of the device if the user were looking forward. This format allows for a natural transition between device use and normal conversation, highlighting the naturalistic aspects of the software/hardware arrangement. The video display is in black-and-white until a face is recognized on the screen by Google Tracker, a face-tracking algorithm; when a face is recognized, the face is outlined in an orange box and the video display changes to full color until no face is recognized again.

Figure 2: The iSET Interface Screen for Recording Sessions with emotion buttons labeled *Happy, Concentrating, Thinking,* and *Interested*. The display allows for toggling of the recording functionality and shows the emotion-labeling buttons; the video display changes depending on whether a face is being successfully tracked.



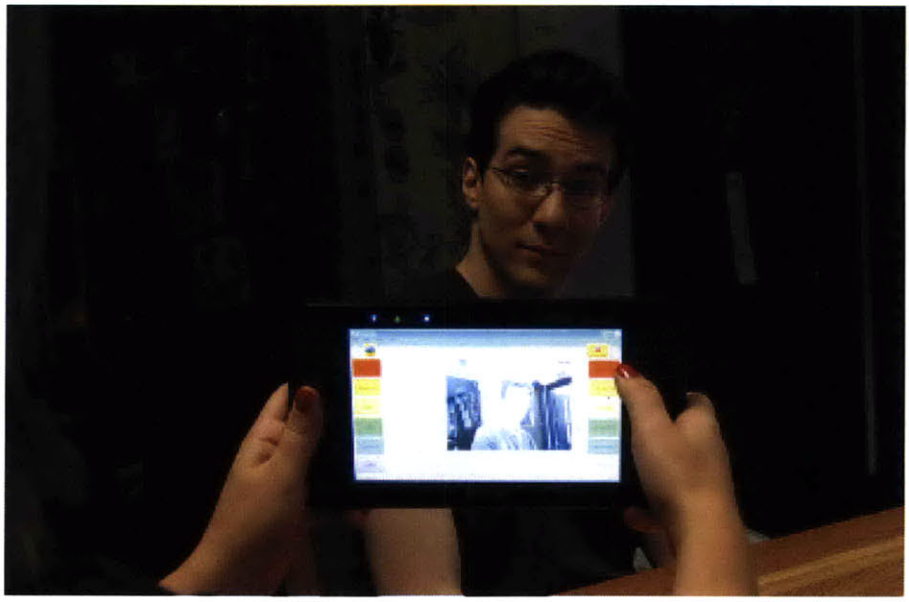
The annotation buttons to the right of the screen are fully customizable. During the iSET intervention, they consisted of one, two, or four of the following emotions (see Section 4.5 for more information): *Happy*, *Concentrating*, *Thinking*, *Interested*. Each of these had a characteristic color (yellow, blue, red, and green, respectively); these colors were associated with their respective emotion throughout all interfaces and reference material in the intervention.

Note that in this section and all subsequent ones, the word "emotions" is used as a shorthand for "emotions and cognitive states"; of the four final "emotions" used, three are more traditionally regarded as cognitive states (*Concentrating, Interested*, and *Thinking*) and only one is usually regarded as an "emotion" *(Happy)*. The goal of the iSET interaction was to improve processing by our participants with regard to affective communication conveyed on a daily basis, and as discussed in Section 5.2, of the set of 6 basic emotions and 6 cognitive states initially recorded, the four affective expressions seen most often were a combination of these two types of information. The iSET project discusses these as "emotions" both for simplicity and because understanding of these states frequently has emotional components and consequences.

3.3 The Emotion-Hunt Game

The emotion hunt game was the primary activity carried out during recording sessions. After viewing example videos for 5-10 minutes in VidL, partipants would go on to participate in 15-20 minutes of recording and labeling emotions with the iSET Interface as part of this game. See Figure 3 for an illustration.

Figure 3: An Illustration of the Emotion Hunt Game. During the Emotion Hunt game, the user of the Samsung/iSET Interface holds the computer with the camera facing their conversational partner. During the course of conversation, video of the partner appears on the screen and is recorded; the user can press labeling buttons to indicate the viewed emotion at any time.



During the Emotion Hunt game, participants follow this series of steps:

- Hold the Samsung with the screen facing the participant and the back-facing camera facing the converstional partner.
- Activate video recording in the iSET Interface.
- Elicit emotional expression from the conversational partner.
- Label expressions with emotions available in the iSET Interface.

In some cases, gross motor skills were not strong enough to allow holding of the Samsung computer, or fine motor skills were insufficient to allow pushing buttons while holding the device; in these cases, the caregiver(s) who participated in the sessions with the student would assist the student by holding the Samsung computer while the student pushed the labels on the touchscreen.

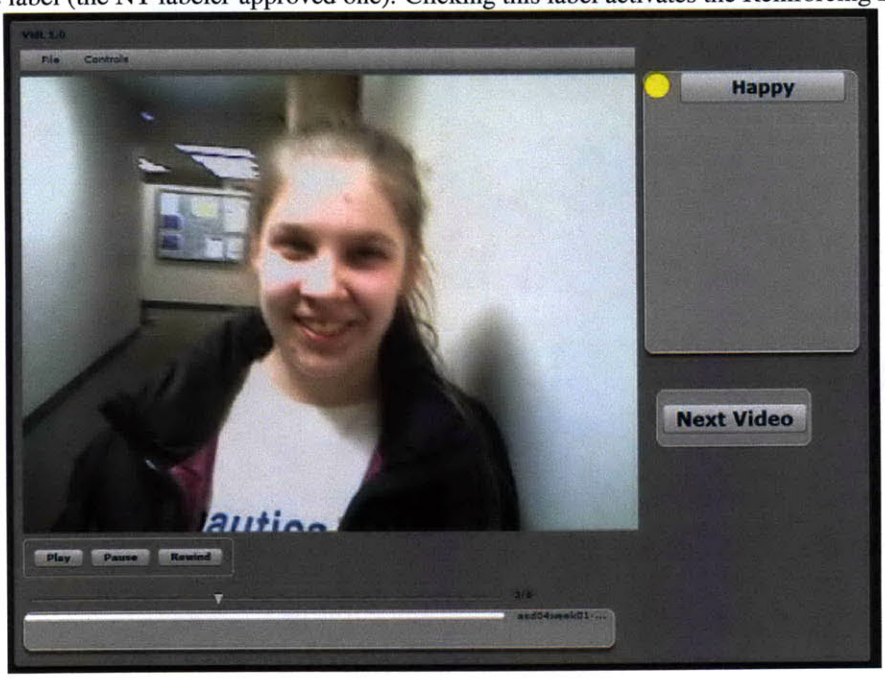
3.4 VidL Web Application Development

The VidL web application was developed by Eckhardt as way to easily label video and images on a web interface accessible from anywhere [42]. The VidL application consists of a browser window showing a video with annotation buttons and a "Next Video" button on the right, video play controls on the bottom left, and a video play timeline at the bottom right. The user clicks the "Next Video" button to trigger the next video, which begins playing automatically. The video can be annotated during or after playback with the labeling buttons on the right. After it plays, the user can click the "play" button at the bottom of the window to replay the video, and more labels can be added before or after replay. As each annotation button is clicked, the selected label is shown in large text at the bottom of the window to confirm the user's choice.

VidL has fully customizable buttons and several different types of playback. Three major interfaces were constructed for use during iSET by Eckhardt and Madsen.

- In the Example Videos interface (see Figure 4), one choice was presented for each video throughout the intervention; whenever that choice was clicked, the Reinforcing Image for that student on that day (see Figure 6) was displayed.

Figure 4: Screenshot of the Example Videos interface in VidL. In the Example Videos interface, videos can be labeled with one possible label (the NT-labeler-approved one). Clicking this label activates the Reinforcing Image.



- In the Preapproved Videos interface (see Figure 5), students were able to click one label during Period 1 and two labels in Periods 2 and 3; if they clicked the correct label, the Reinforcing Image was shown (see Figure 6). If an incorrect label is picked by the labeler (which is possible in Periods 2 and 3) then the "Try Again!" screen is shown (see Figure 7). The correct label must be clicked in order to proceed, but if the incorrect label was chosen first, the "Correct!" screen is shown without the Reinforcing Image. (See Figure 8.)

Figure 5: Screenshot of the Preapproved Videos interface in VidL.

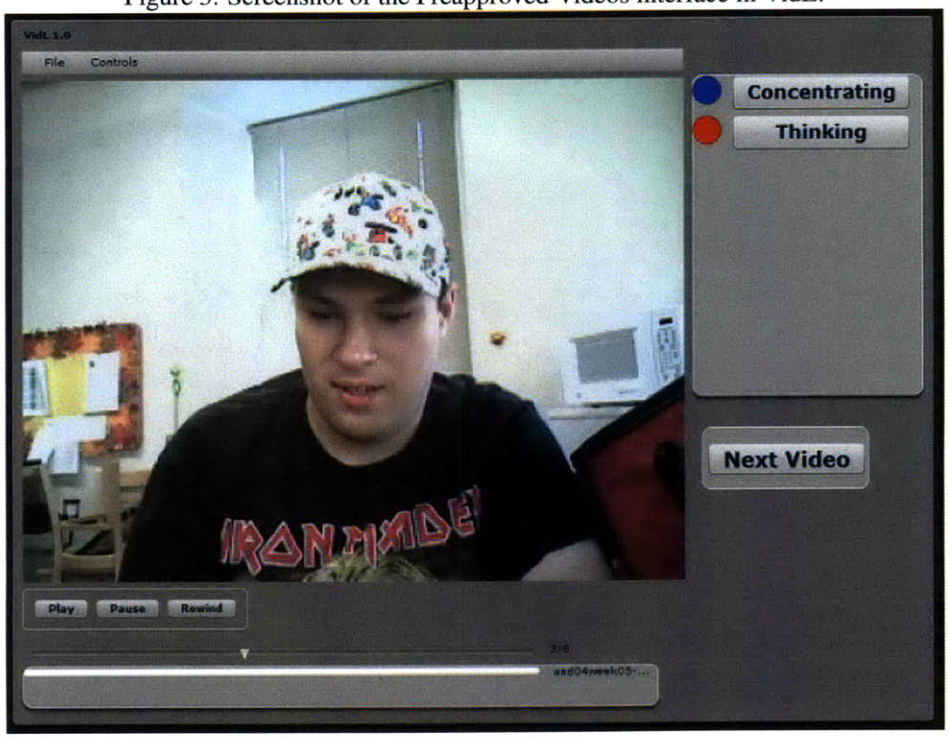


Figure 6: Example of Reinforcing Image and "Correct!" screen that can appear during Example Videos and Preapproved Videos Interfaces. Each student has customized Reinforcing Images which are changed each day; these images can appear in a variety of fashions, including multiple copies of the image showing at once, one copy of the image bouncing, and one copy of the image rapidly growing bigger and then smaller.

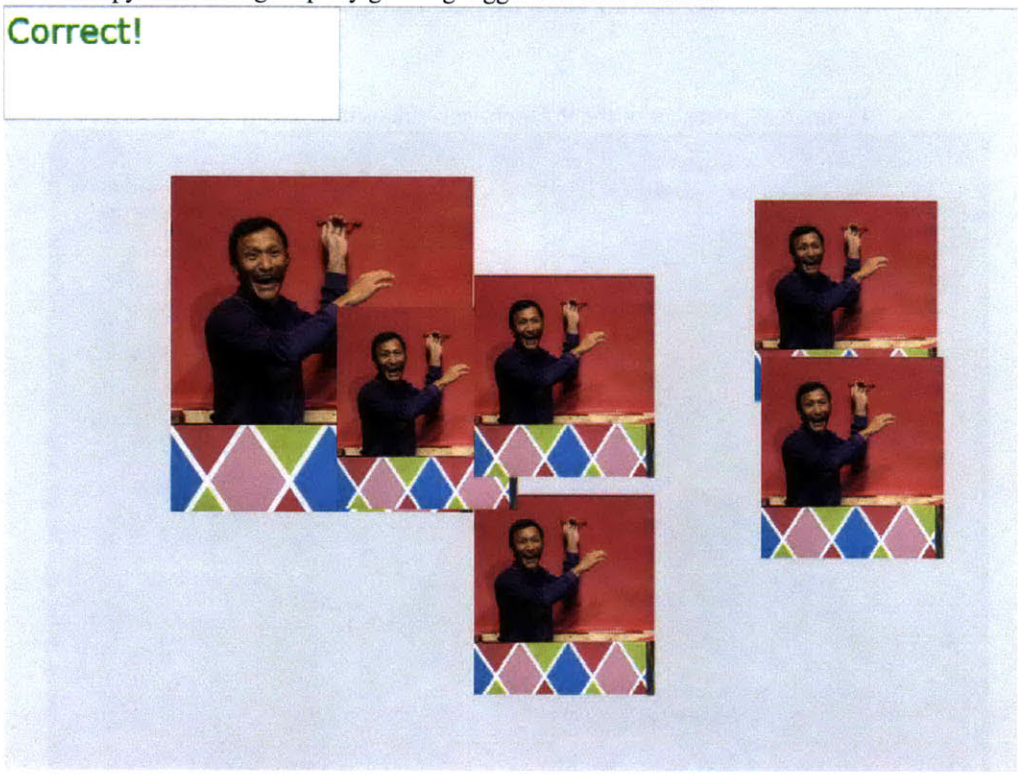


Figure 7: The "Try Again!" screen shown after a wrong answer. This screen shows for a few seconds and then automatically disappears.

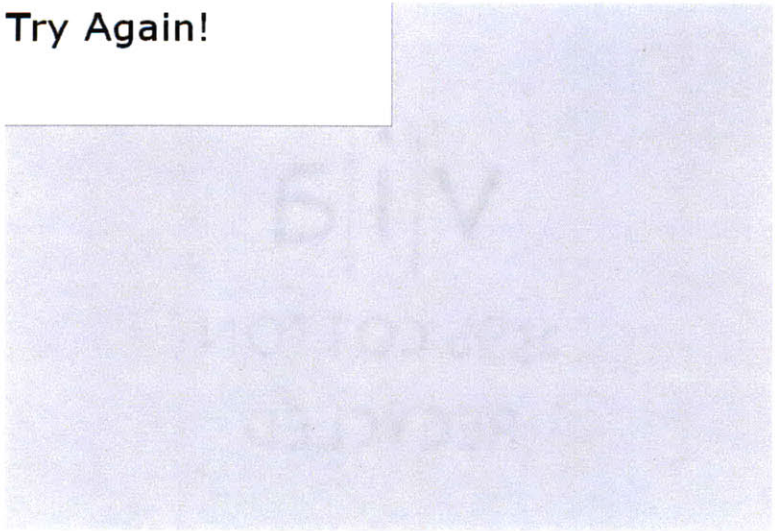
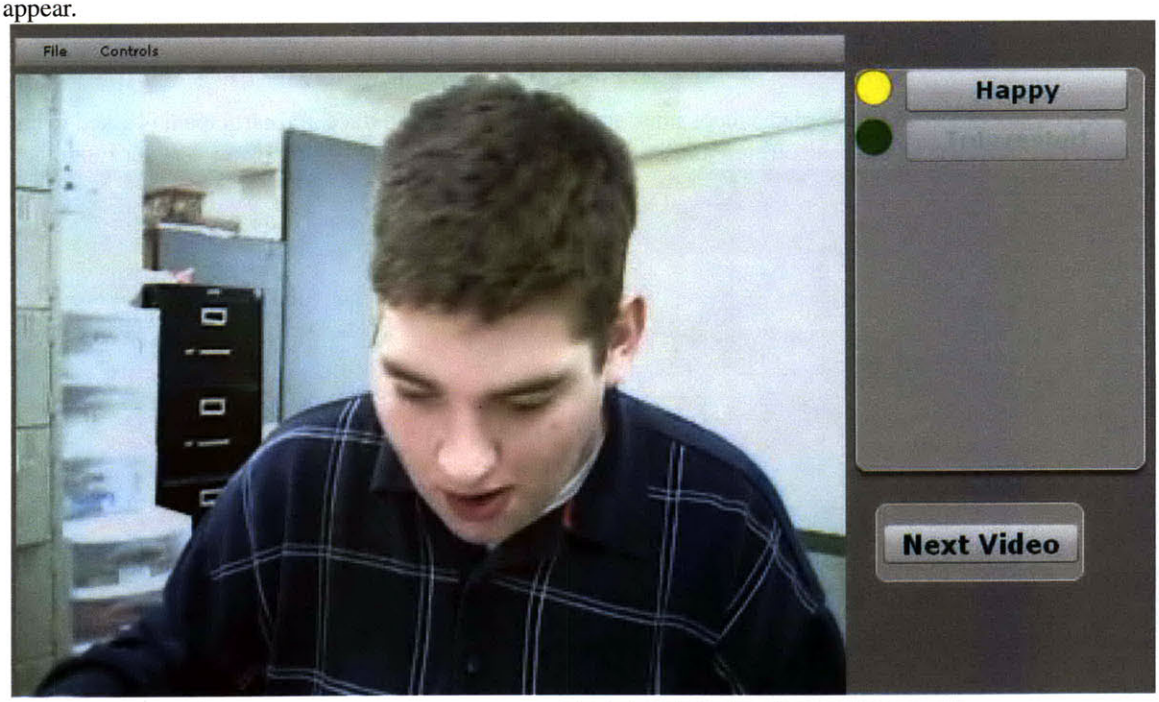
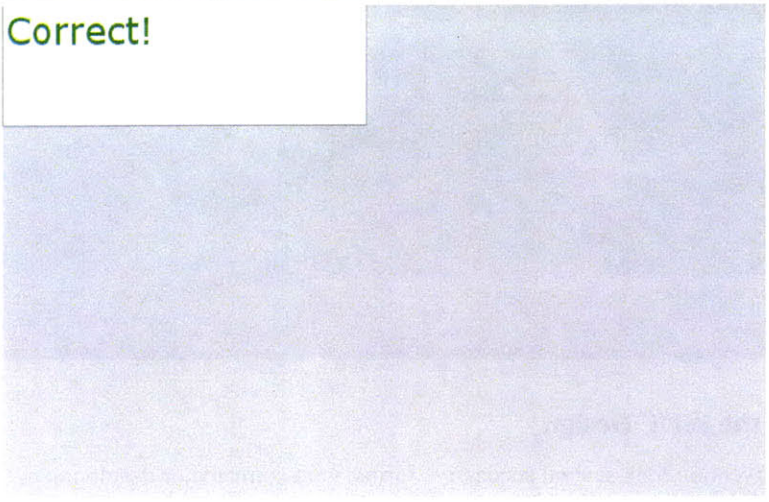


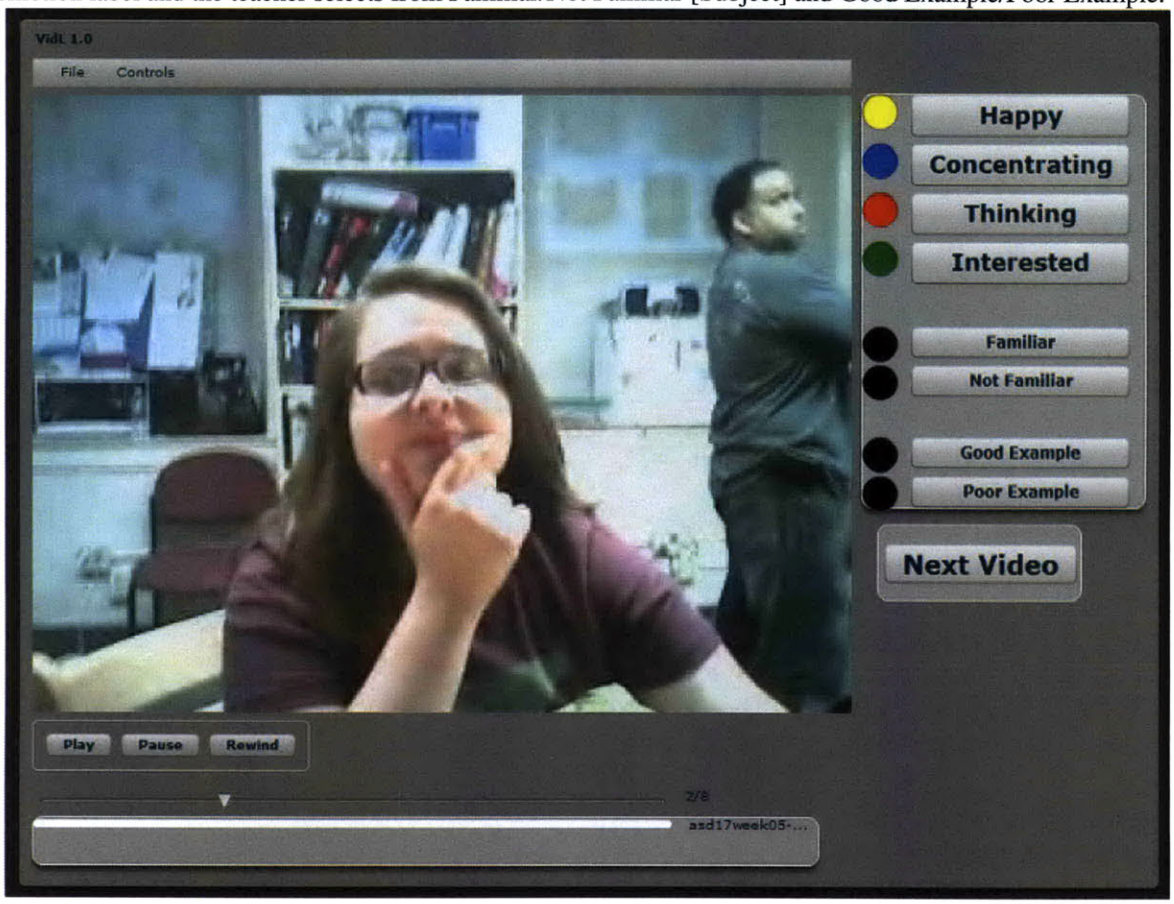
Figure 8: Choosing a Wrong Answer after a Right Answer in Periods 2 and 3. If a wrong answer is chosen after a right one, the wrongly chosen answer is grayed out; when the "Correct!" screen is shown, the Reinforcing Image does not appear.





- In the Student-Recorded Videos interface (see Figure 9), each video was shown with four emotion buttons to be clicked by the students and two other choices (Familiar/Not Familiar [Subject] and Good Example/Poor Example) to be made by the teacher. No reinforcing screen is presented during the Student-Recorded Video section.

Figure 9: Screenshot of Student-Recorded Videos interface in VidL. In this interface, the participant is asked to choose an emotion label and the teacher selects from Familiar/Not Familiar [Subject] and Good Example/Poor Example.



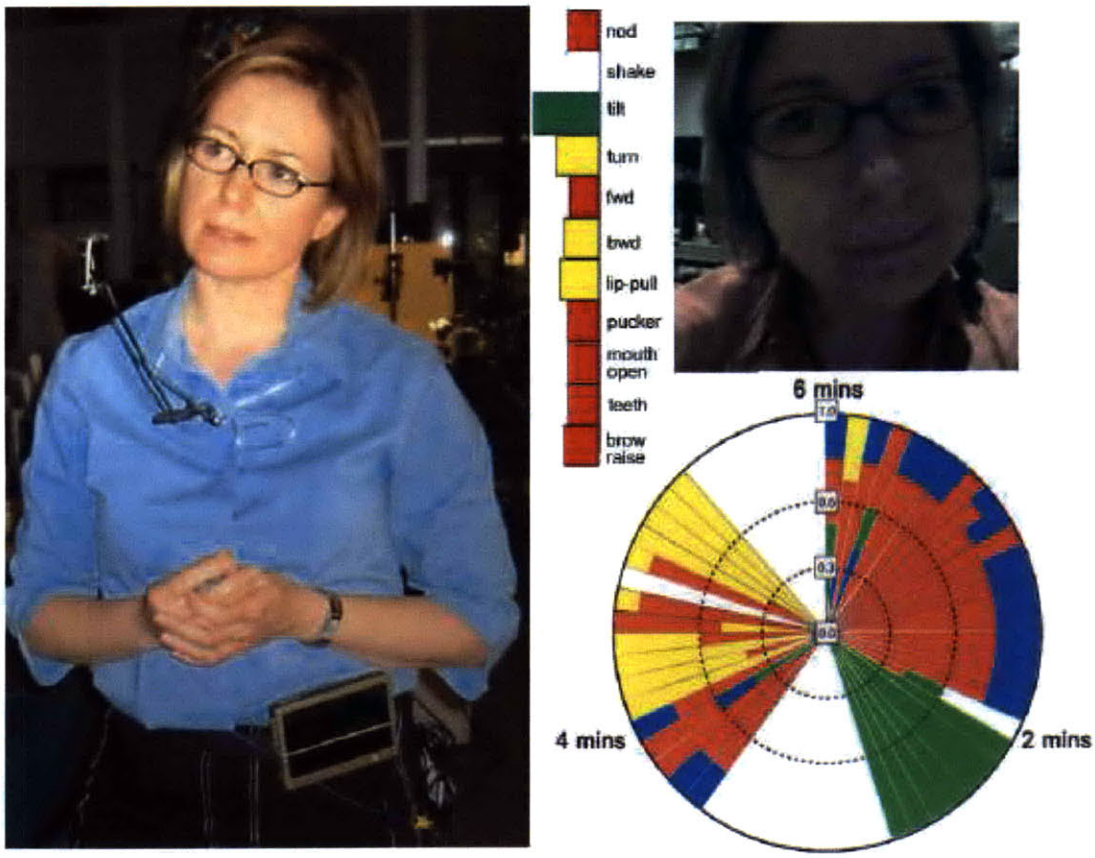
3.5 Iterations on the iSET Design

Design of the iSET intervention took several successive forms, with some major development (such as from SelfCam to Emotion Bubbles and on to the current iSET Interface) and some minor ones (such as color, button size, and layout tweaking). Some of the factors taken into account in relation to use with a population on the spectrum have been reviewed at the ASSETS conference [43]; the design process and interactions with students were outlined at the CHI conference [44].

The underlying software used for iSET was FaceSense, a face reading algorithm developed by el Kaliouby [31]. FaceSense was used to automatically assess emotions shown in facial expressions. By tracking 22 feature points on the face and using Ekman's FAC (Facial Action Coding) system, the software automatically detected a range of "gestures" (such as *head-tilt* and *lip-pucker*) and translated those into a range of congitive states, which were *Agreeing*, *Concentrating*, *Disagreeing*, *Interested*, *Thinking*, and Unsure [55].

In the original iteration of the iSET Interface, the iSET project used information from the FaceSense software to create a wearable technology that would allow the user to record and process their own emotional state, as conveyed through facial expression, in real time. This approach, called SelfCam, was self-oriented for its users, allowing them to focus on their own expressions as opposed to directly engaging with affective expression by conversational partners [22]. (See Figure 10 for an illustration.)

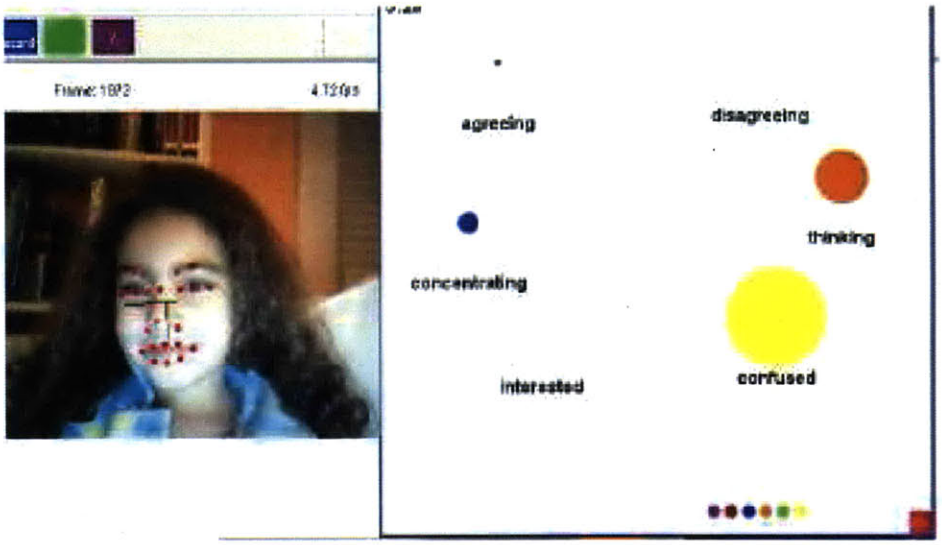
Figure 10: The Selfcam Apparatus and Output. The SelfCam is a portable device that focuses on the wearer's face; it produces a display that can signal cognitive states and facial gestures.



While the SelfCam pilot focused on more-communicative individuals - many of whom had Asperger's Syndrome - it became clear that this population might be able to benefit from a broad range of approaches, some of which may not be suitable for individuals with a lower social comfort level. Thus, in order to serve the most people on the spectrum, it made sense to design a system for use by persons with AD/PDD-NOS and then adapt as necessary for persons with AS, rather than doing the more substantial adaption/reconceptualization in the opposite direction. It was also determined that understanding of one's own emotional presentation, the focus of the SelfCam project, was helpful for the learning and self-reflection process. However, that paradigm was insufficient for conversation-oriented learning of interpersonal affective cues.

A subsequent version of the iSET Interface for use at the Groden Center included a component created by the author called Emotion Bubbles (see Figure 11.) In the Emotion Bubbles Interface, six emotions were chosen for automatic recognition by the FaceSense backend, and each was assigned a distinctive color. The three negative emotions assigned "warm" colors (indicating that the viewer of the Emotion Bubbles may want to change conversational style) were *Disagreeing* (red), *Thinking* (orange), and *Confused* (yellow); the positive emotions assigned "cool" colors (indicating current conversational success) were *Interested* (green), *Concentrating* (blue), and *Agreeing* (purple).

Figure 11: Emotion Bubbles Interface. The Emotion Bubbles reflected the cognitive states that were interpreted by the FaceSense Software: Agreeing, Concentrating, Interested, Disagreeing, Thinking, and Confused. These states were displayed in characteristic colors; the circles that represented them could be touched and dragged around the screen.



The Emotion Bubbles interface relied on the original FaceSense emotion-sensing algorithms [31]. However, although the Emotion Bubbles were engaging and interesting for the students to view, they did not involve any action other than directing the camera on the part of the students, and therefore did not require ongoing intellectual engagement. Furthermore, since the emotion-recognition algorithms were not consistently correct, the teachers sometimes orally corrected the "perception" of the software and undermined its claim to provide helpful information during interpersonal interactions.

The iSET project team determined through a series of participatory design interactions that, in addition to providing sometimes-incorrect information about emotion conveyed through the face, the FaceSense software back-end did not promote maximum engagement between the participant on the spectrum and the learning system.

The current iSET Interface was therefore adopted; it removed all of the automatic emotion-recognition components but retained the Google-Tracker-supported face tracking algorithm, allowing users to easily (through changes to and from a black-and-white display) understand whether or not a face was being tracked. The addition of large, colorful labeling buttons, added in a pattern of increasing complexity, encouraged engagement with the system and allowed for ongoing review of emotions between weeks of the intervention.

3.6 Conclusion

The Interactive Social-Emotional Toolkit is a multi-component system designed to facilitate the recording and annotation of affective information. The iSET system was refined and improved through many rounds of feedback from ASC participants and their caregivers, resulting in components facilitating both naturalistic interpersonal interaction and a reflection/discussion period of recorded observations in order to support the learning process.

4 Experimental Design

4.1 Introduction

The goal of the intervention was to determine the efficacy of iSET as a tool for learning emotions. The iSET intervention was designed to be as consistent with the Mind Reading DVD intervention as possible, with the major differences derived from recording and review of live videos instead of viewing of prerecorded videos of acted emotion [20]. Consistency between the two interventions assisted with later comparison and increased ability to compare the two interventions on the basis of naturalistic vs. acted video and on the basis of on-computer vs. *in situ* affect-reading interactions.

4.2 Hypotheses

The primary hypotheses for the iSET intervention were as follows:

1. Persons on the spectrum will benefit from active involvement with their own emotion-learning intervention, including with the design process; because of a lack of usability studies with persons on the spectrum in the literature, direct contact with target users during the design process will produce a more useful final intervention for emotion learning.

2. Naturalistic videos, recorded and viewed *in situ* and at a later time by the user, will carry more benefit in terms of terms of teaching the user to recognize emotions and affect in an environment similar to ones in which he/she will usually be asked to do so.

3. Teacher feedback and cooperative analysis of emotion videos is a stronger teaching tool for understanding of these videos than non-social use of emotion-teaching software.

4.3 Participants

Two groups of participants took part in the intervention - a control group of neurotypical participants (n=20) and a group of participants on the spectrum with PDD-NOS or Autistic Disorder (n=20), recruited from the Groden Center's three sites. The ASC and NT participants were matched on gender and age (with the NT participants being $\pm$ 6 months older than than the ASC participants.)

The ASC group was further divided into two subgroups; one group (n=10) took part in a version of the original Mind Reading DVD intervention, and the other group (n=10) took part in the iSET intervention. The two ASC groups were matched on the basis of gender, age, and scores on the KBIT-2 intelligence test (see Section 4.4, Instruments.)

4.4 Instruments

Several tests were used with the ASC participants only (for instance, autism-specific tests), and some were used with the matched NTs to compare the two groups. Some tests were used as both pretests and posttests with the ASC group, in order to gauge anticipated change over the course of the intervention, and some were used only during the pretest phase to assess factors not expected to change over the course of the intervention. None of the tests were used as posttests in the NT group. Table 3 shows the breakdown of standardized measure administration.

Tests used during the intervention included:

- Autism Diagnosis Observation Schedule (ADOS) [48]: a test for diagnosing and assessing autism, consisting of probes for social, communicative, and restricted & reptitive play and behavior. The ADOS is the gold-standard test for diagnosing autism from toddlerhood to adulthood. Evaluation takes place across a variety of social and communicative situations, including both activities and conversational settings.

- The Benton Facial Recognition Test[51]: the short form of this test (27 items) was used to test ability to compare matching versus not-matching photos of faces, as well as to generalize facial recognition between multiple photographs of the same person. The test can be used for ages 6-74 years, and it is designed to be administered quickly with minimal verbal ability required.

- The Kaufman Brief Intelligence Scale (KBIT-2) [45]: a short assessment of intelligence similar to IQ testing, but covering a more limited scope of detail. There are two subscales for verbal and nonverbal scores as well as a composite scale. (Note that the KBIT-2 was used for brevity; however, Dawson has presented evidence that the Raven' Progressive Matrices test may be a better measure of the unique intelligences seen in autism [53].)
- The Reading the Mind in the Eyes Test [49]: this test created by Dr. Simon Baron-Cohen includes one practice question and 36 test questions, each showing a picture of a pair of eyes and suggesting four choices that could describe what is being thought or felt. The test-taker is instructed to choose the most suitable option. (This test was administered to teachers and neurotypical labelers evaluating the video corpus only; the purpose was to ensure an above-average level of facial-expression-based emotion recognition in persons crucial to the success of the intervention.)
- The Social Responsiveness Scale (SRS)[50]: a 65-question list of traits applicable to social interactions. Five subscale scores are generated: Social Awareness, Social Cognition, Social Communication, Social Motivation, and Autistic Mannerisms. T-Scores are generated for each category as well as the overall test and may be used to evaluate particular deficits or overall impairment in the area of social interaction.
- The Toronto Alexithymia Scale (TAS-20)[46]: a 20-question inventory of traits related to alexithymia, the ability to express self-awareness and personal emotion awareness. Three subscale scores are compiled in addition to the overall score: Difficulty Describing Feelings, Difficulty Identifying Feelings, and Externally Oriented Thinking. The TAS is intended to identify and quantify character traits related to minimization of internal emotional state and tendency to focus emotions in an external direction. The TAS-20 was administered by teachers and also taken directly by participants; whenever possible, the participants score was used in analysis.
- Customized VidL pretest containing videos of a study subject and his/her -self, peers, and teachers, as well as some unfamiliar adults (other adults at the Groden Center.) See Section 5 for more information.

Table 3: Tests Administered to ASC Participants, NT Participants, and Teachers. NT participants did not participate in any post-tests, and teachers only participated in two tests for reasons for time. Major baseline measures such as the KBIT-2 and autism-specific measures such as the ADOS were administered to the ASC group only.

Test	ASC (pretest)	ASC (posttest)	Teachers	NT
Autism Diagnosis Observation Schedule (ADOS)[48]	yes	no	no	no
Benton Face Recognition Test [51]	yes	yes	no	yes
KBIT-2 [45]	yes	no	no	no
Reading the Mind in the Eyes test [49]	no	no	yes	no
Social Responsiveness Scale (SRS) [50]	yes	yes	no	yes
Toronto Alexithymia Scale (TAS-20) [46]	yes	yes	no	yes
VidL Naturalistic Videos Test	yes	yes	yes	yes

The pretests were not be readministered to NT participants; since they did not participate in the intervention, there is no need for gauging their progress during the time of the intervention.

Since both groups of ASC participants are engaging in full emotion-teaching interventions, it was necessary to gauge the relative progress of groups and individuals on these measures relative to the beginning of the intervention.

Post-tests were be administered to ASC participants starting the week of 8/22/10.

4.5 Procedure

The iSET intervention was 12 weeks long and consisted of three four-week periods of increasing complexity, with three to four sessions per week for each participant. Participants were accompanied by a teacher during each session and were prompted throughout, as per the instructions in Appendix 9.4.3 (recording sessions) and Appendix 9.4.6 (viewing sessions).

For all participants, one emotion was looked at per session in the first period; two emotions were looked at per session in the second period; and the four emotions of the study were looked at in the third period. (See Table 4 for a full schedule.) The emotions were chosen for each week as shown in Table 4. (The emotions for Period 1 were ordered from easiest to hardest by the Groden Center teachers; the emotions in Period 2 were selected in random pairs so that each was chosen twice during that time.)

Table 4: Emotion presention for iSET and MR participants during the iSET intervention. During Period 1, the same emotion was presented during each of the 3-4 sessions in a given week, and the emotion changed from week to week. In Period 2, a particular pair was selected for each week, and in Period 3, all emotions were looked at during each week and session.

Period	Week #s	Emotions/week	Week 1	Week 2	Week 3	Week 4
1	1-4	1	Happy	Thinking	Concentrating	Interested
2	5-8	2	Concentrating/Thinking	Thinking/Interested	Happy/Concentrating	Happy/Interested
3	9-12	4	All emotions	All emotions	All emotions	All emotions

Participants in the iSET group versus the MR group experienced different procedures.

4.5.1 iSET Group Procedure

During each week, each student experienced alternating recording and viewing sessions. In the 3-4 sessions per week, students always had at least a recording session and a viewing session, followed by another recording session; if scheduling permitted, another viewing session took place as well

Recording Sessions consisted of:

- 5-10 minutes of labeling videos from the corpus, called "Example Videos", through the VidL interface (see Section 3.4). Each Example video had one label offered; this label was the label approved for the clip by the neurotypical labelers.
- 15-20 minutes of recording video through the Samsung hardware/iSET Interface (see Section 3.2) In Period 1, the iSET Interface had the emotion for that week available for live annotation; in Period 2, both emotions being focused on that week were available for live annotation; and in Period 3, all emotions were available for labeling.

Viewing sessions consisted of:

- 5-10 minutes of labeling videos from the corpus, called "Example Videos", through the VidL interface (see Section 3.4). Each Example video had one label offered; this label was the label approved for the clip by the neurotypical labelers.
- 5-10 minutes of labeling videos from the corpus, called "Preapproved Videos", through the VidL interface (see Section 3.4). Each Preapproved video had 1 or two labels offered. In Period 1, one label (the correct one) was offered; in Period 2, both labels covered that week were offered; and in Period 3, two of the four labels in the intervention were offered (one was the correct one and one was selected randomly from the set of the other three.)
- 10-15 minutes of video clips recorded during the Recording Session. Videos shown during the the Viewing Session were the recorded clips (four seconds before and after a label), except for the clips in which the tracker in the iSET Interface did not successfully find a face at the time when the labeling button was pressed in the iSET Interface (see Section 3.2).

During both Recording and Viewing sessions, students received Reinforcing Images with different visual effects (e.g. bouncing, changing size, and appearing many times) whenever a right answer was chosen. These images were chosen by students before and during the intervention and switched on a daily basis (with the same group of 10 images used each week for each students.

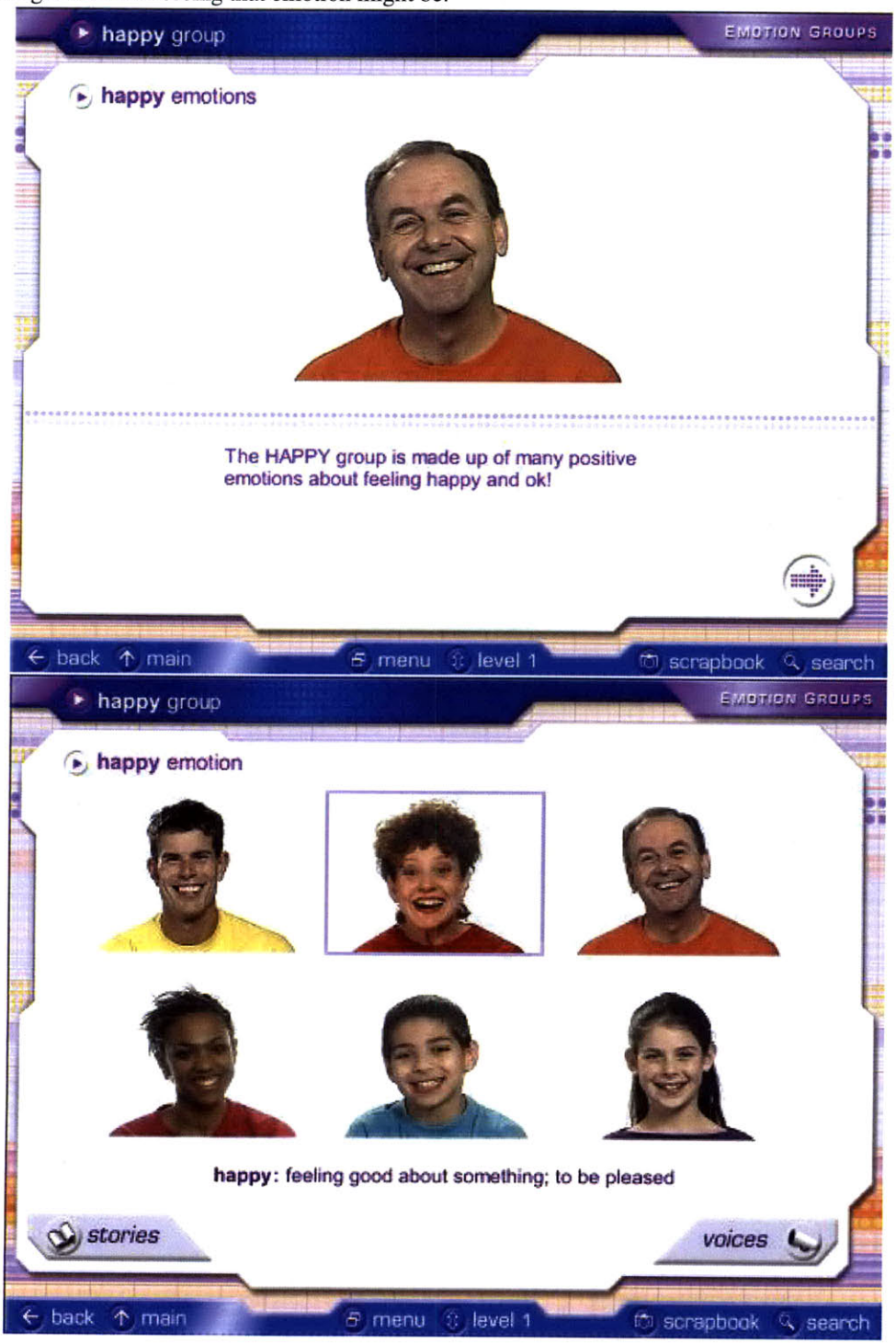
4.5.2 Mind Reading Group Procedure

The Mind Reading DVD procedure was adapted from the procedure followed by Golan and Baron-Cohen [20].

In each session, students were asked to visit three components of the Mind Reading DVD for each emotion being covered that week. These sections were: Learning Center/Emotion Groups, Learning Center/Lessons, and Rewards.

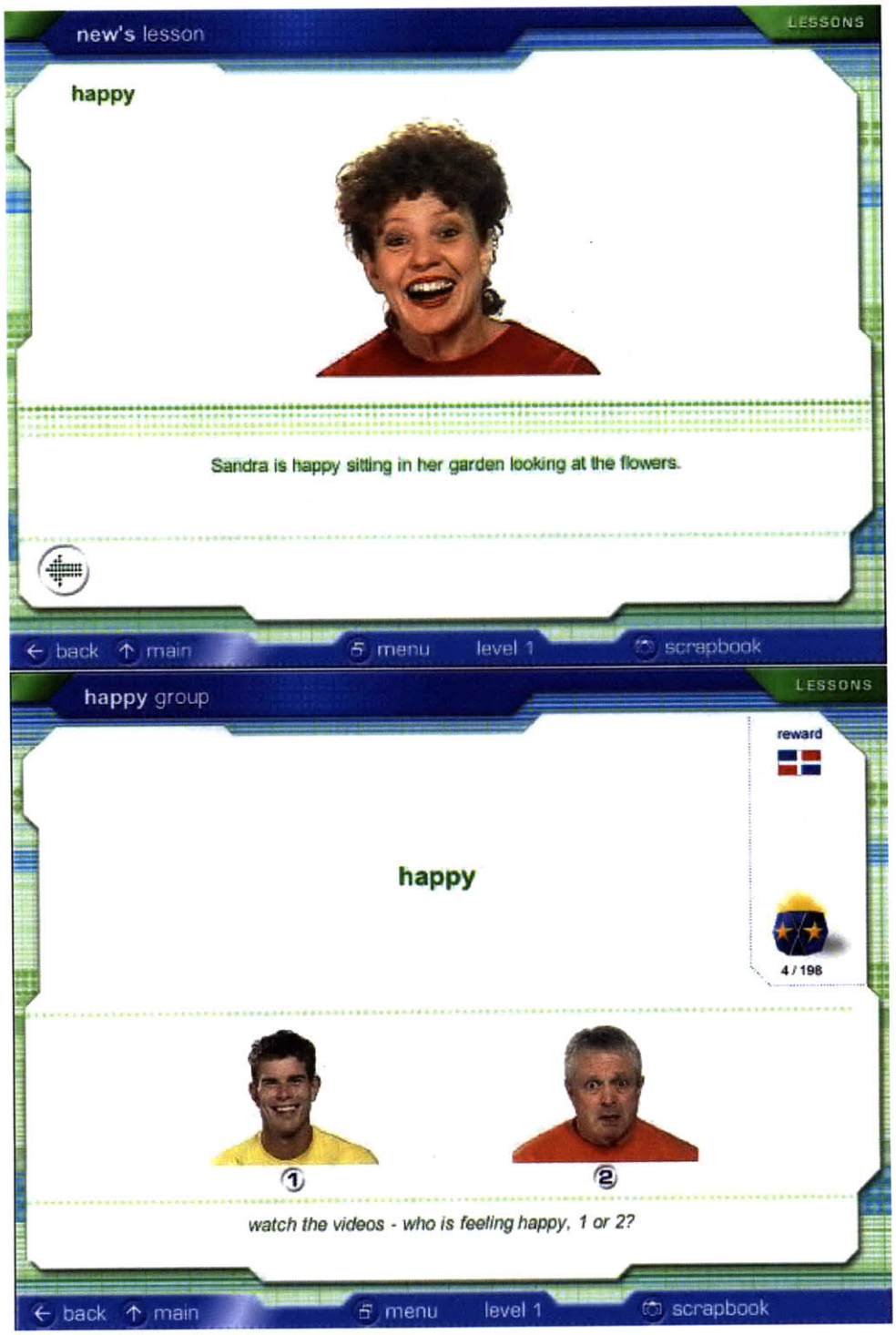
- In the Emotion Groups section (part of the Learning Center), students watched a series of videos of actors illustrating that week's emotions. For instance, in Week 04 (Happy/Interested), students would watch videos from both the Happy and Interested emotion sets, found in the Happy and Interested groups, respectively. The students would then listen to their teachers read short stories from the software describing situations in which the emotion may be relevant. See Figure 12 for illustrations of the screens participants encountered.

Figure 12: Mind Reading DVD: Emotion Groups Area. In the Emotion Groups Area, students are shown videos illustrating an emotion and statements about an emotion that illustrate how the emotion might look in social interactions and what the significance of seeing that emotion might be.



- In Lessons (part of the Learning Center), students would re-watch those same videos; in this case, the videos would be preceded by more context about the overall group, as well as a series of written stories read by teachers at the end of the video showing. Some short quizzes (two questions about a presented video, each question having two choices) are also presented. See Figure 13 for an illustration of the Lessons Area.

Figure 13: Mind Reading DVD: Lessons Area. In the Lessons area of the Mind Reading DVD, students are shown a series of videos for practice and then asked to choose between representative and non-representative videos for a given emotion.



- In the Rewards Center, students are able to see images that they earned in the Lessons area.

4.6 Conclusion

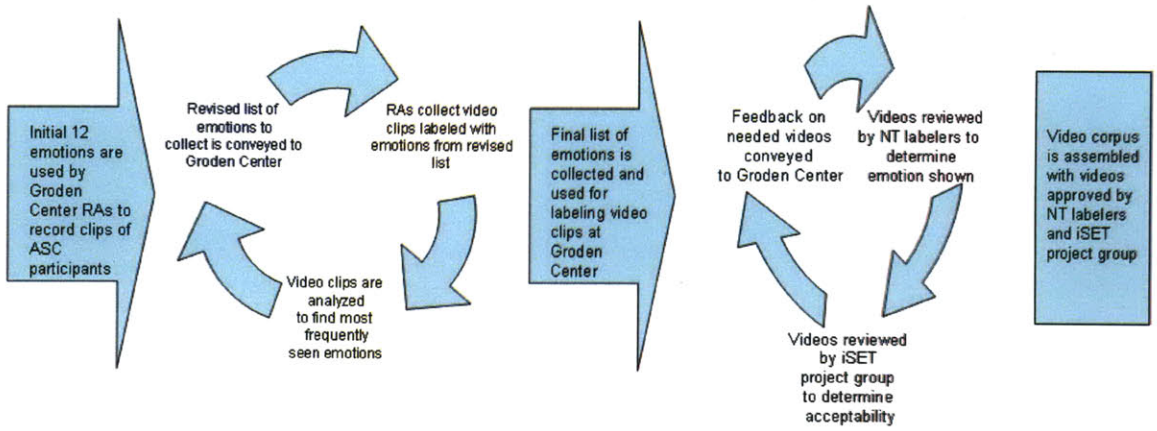
Two procedures were developed for the ASC participant groups. The procedures were designed to be as similar as possible across the two software interfaces, reflecting a desire to see the impact of adding naturalistic video and increased feedback and social interaction and controlling for number of emotions presented, audio input or lack thereof, and speed and complexity with which the two participant groups proceeded.

5 Construction and Administration of the VidL Pretest

5.1 Introduction

The VidL pretest was created as a tool for assessing improvement in emotion recognition for the participants in the Interactive Social-Emotional Toolkit (iSET) intervention. Each student was to see videos of themselves, their peers, and adults both familiar and unfamiliar as part of the test. Each representative of these groups would be shown in videos expressing each of the emotions (*Concentrating, Happy, Interested,* and *Thinking*) in clips that had achieved high inter-rater reliability. In order to choose these emotions originally and to collect the data for the pretests/posttests and the Examples/Preapproved Videos used during the intervention, an extended data collection process was carried out between July 2009 and April 2010. A summary of the full process is in Figure 14.

Figure 14: Summary of the Pretest Data Collection Process. Each video clip collected for the corpus was initially labeled with an emotion from a set selected by identifying which emotions were seen most often in the classroom; the clip was then approved by both neurotypical labelers and the iSET project group.



5.2 Selection of Emotions for iSET Intervention

The data for the pretest was collected between July 2009 and April 2010 at all three Groden Center sites. Recordings done with VidL focused on natural elicitation of emotions; the research assistants carrying out the records were requested to work with interactions that were already taking place, or to try to have conversations in which the desired emotions would naturally take place. Emotions to be labeled during the intervention were chosen in order to populate the iSET Interface and allow for easy *in situ* labeling. The process of choosing emotions for the intervention was carried out via the following set of steps:

- We initially started with 12 emotions available for labeling in the iSET Interface. This set was assembled from Baron-Cohen's basic emotions, *Angry, Disgusted, Afraid, Happy, Sad,* and *Surprised*; and the list of cognitive states recognizable by el Kaliouby's FaceSense software, *Agreeing, Concentrating, Disagreeing, Interested, Thinking*, and *Unsure* [54, 31]. Research assistants working in the Groden Center context were asked to record those emotions on the Samsung/iSET Interface system whenever the emotions were observed in the classroom.

- From that set of emotions, the following emotions were seen most often (from most to least frequently): *Happy, Thinking, Concentrating, Interested, Sad, Unsure, Agreeing.* In order to make the learning that the students did during the intervention most relevant to their everyday experiences, it was decided that these emotions would tentatively be the ones taught through the iSET intervention.

- Teachers were asked to record with only that set of emotions. It became clear that *Sad*, although it had appeared during the initial free-recording process many times, did not naturally occur on a reliable basis during day-to-day classroom activity, and the teachers were not comfortable eliciting it, so it was removed from the list of emotions.

- *Agreeing* was removed from the set of emotions once it was determined that there was no appropriate match to it in our control intervention, the Mind Reading DVD; the closest equivalent was *Sure*, which was inconsistent visually with the usual depictions of *Agreeing* recorded with the iSET Interface.
- Finally, *Unsure* was removed from the set because the teachers had an extremely hard time recording clips that could be affirmatively reviewed by the NT labelers; after several months of only trying to elicit *Unsure* clips, this emotion was eventually discarded as one that would be less valuable to teach because there was so much inconsistency in its elicitation and interpretation.

The final set of emotions recorded in the iSET intervention was therefore *Concentrating*, *Happy*, *Interested*, and *Thinking*.

5.3 Labeling Process for Neurotypical Labelers

Neurotypical labelers (who were a distinct group from the neurotypical age-and-gender-matched participants) were included in the study if they had scored a 26 or above in the Reading the Mind in the Eyes online test and had no history of ASC diagnosis [49]. These labelers were shown many of the clips recorded by teachers at the Groden Center during the clip collection process.

During the pre-intervention labeling process, the NT labelers were asked to watch the videos recorded by teachers during the intervention. For each video, they were shown the teacher-selected label for that clip along with three foils from the set of twelve initial iSET labels, as described in 5.2. (The foils list is given in Appendix 9.1.3; the foils for each labeler contained two similar foils and one "opposite" foil from the 12-label set, e.g. for *Happy*, the similar foils were *Agreeing* and *Interested* and the "opposite" foil was *Sad*.)

Each video was reviewed by at least 8 neurotypical labelers. A video was selected for use in the approved-video corpus if at least 75% of the NT labelers who viewed it agreed on a common label. The common label selected by NT labelers did not have to be consistent with the label originally selected by the teachers for that clip. Potential Example/Preapproved Videos were also evaluated by members of the iSET project research group; if any of 4 members voted a video clip as "not usable" for reasons of blurriness, unclear subject, etc., it was not added to the corpus.

5.4 Collection of Naturalistic Video Data

A total of 4490 eight-second clips were collected by teachers and RAs to prepare for the pretest. A subset of these videos (3655) were labeled by neurotypical labelers as described in Section 5.3.

Each participant had 8 videos of themselves (two for each emotion) in the pretest, for a total of 160 videos of students; in addition, 7 staff members formed the group of familiar/unfamiliar teachers, and each had 8 videos (two for each emotion) for a total of 216 videos appearing across all pretests. Each student's pretest consisted of 40 naturalistic videos (4 emotions * 2 examples per emotion for each of Self, Peer 1, Peer 2, Familiar Adult, and Unfamiliar Adult) along with 20 videos of Mind Reading DVD videos (5 examples of each emotion); there were also 4 training videos presented at the beginning of each pretest, which were shown with only the correct choice available. Thus, each pretest consisted of 64 videos. (See Table 5 for a summary of the contents of each pretest.)

Table 5: Contents of pretest for each ASC participant/NT counterpart. Each participant had a 64-video pretest consising of 4 test videos, 20 acted videos, and 40 naturalistic videos, broken down into particular video subjects and emotions.

Question Type	Video Type	Subject	Number of videos	Breakdown of videos	Position in test
Training	Acted	Mind Reading Actors (assortment)	4	1 of each: Happy, Thinking, Concentrating, Interested	Beginning
Test	Naturalistic	Self	8	2 of each: Happy, Thinking, Concentrating, Interested	Order randomized
Test	Naturalistic	Peer 1	8	2 of each: Happy, Thinking, Concentrating, Interested	Order randomized
Test	Naturalistic	Peer 2	8	2 of each: Happy, Thinking, Concentrating, Interested	Order randomized
Test	Naturalistic	Familiar Adult	8	2 of each: Happy, Thinking, Concentrating, Interested	Order randomized
Test	Naturalistic	Unfamiliar Adult	8	2 of each: Happy, Thinking, Concentrating, Interested	Order randomized
Test	Acted	Mind Reading Actors (assortment)	20	5 of each: Happy, Thinking, Concentrating, Interested	Order randomized

Pretests were administered to each of the 20 participants. Almost all participants had a primary and secondary teacher participating, with some overlap between teachers, resulting in a total of 28 staff members taking their students' pretests in order to familiarize them with the intervention. Three tertiary and three quaternary pretests were also administered when students needed extra staff members. Finally, each of the students also had a neurotypical (NT) counterpart take their pretest. These NT counterparts were matched to the ASC students on the basis of age ($\pm$ six months from the ASC participant's age) and gender. The neurotypical participants were further required to have not been diagnosed with any form of autism.

After the 216 videos included in the pretest were removed from consideration during the intervention, the remaining 477 were used as "example" videos (89 for Happy, 167 for Concentrating, 169 for Thinking, and 52 for Interested).

5.5 Analysis of Pretest Data Collection

The videos collected for the pretest yielded some surprising information about the likelihood of labelers to agree on videos of NT vs. ASC subjects and also indicated interesting differences in agreement between labelers depending on emotion shown. Labelers were shown images of both ASC subjects (students participating in the intervention) and NT subjects (teachers assisting with the intervention) to construct the pretest corpus. As a percentage of videos approved by NT labelers out of total videos collected, on average across the set of participants, the groups were fairly similar: 36.8% $\pm$ 6.25% for the neurotypical video subjects and 34.4% $\pm$ 8.30% for the ASC video subjects. This similarity between emotion recognition/agreement rates for NT vs. ASC video clip subjects suggests that more quantitative research should be conducted in the area of flat affect, and that emotions expressed by persons not on the spectrum may not be as clear as previously thought.

In the measure of "convergence" (total videos approved for an emotion divided by total number of labels of that emotion applied to all videos by all labelers), there were clear differences between emotions suggested. From most to least, convergence rates were 30.4% for *Happy*, 17.7% for *Concentrating*, 13.3% for *Interested*, and 12.6% for *Thinking*. Overall, convergence rates were 18.5% $\pm$ 7.16%. (See Appendix 9.1.4.)

5.6 Conclusion

An extended data collection process was used to create a full corpus of video with high inter-rater agreements suitable for using as a pretetest/posttest measure. Some emotions, such as *Disgusted, Disagreeing*, and *Surprised,* were recorded very rarely by teachers in a free-recording situation; conversely, some, such as *Happy, Interested,* and *Thinking*, were observed and recorded frequently by the teachers. Emotions were removed from consideration after their initial tentative inclusion for a variety of reasons, such as inappropriateness of elicitation (*Sad*), difficuilty in achieving good inter-rater reliability (*Unsure)*, and lack of concordance with the Mind Reading DVD (*Agreeing*).

The corpus of representing videos showed many examples of the final selected group of emotions across a variety of ages, settings, activities, and subjects. Videos of NT and ASC participants (teachers and students, respectively) were approved at similar agreement rates by the NT labelers evaluating their videos, and *Happy* was almost twice as easy to converge on for NT videos compared to the other possible emotions.

6 Pretest Results And Analysis

6.1 Administration of Pretests

Three tests were given to both the NT and ASC groups: the TAS-20, a 20-item test of alexithymia (lack of internal emotion understanding) with three subscales for Difficulty Describing Feelings, Difficulty Identifying Feelings, and Externally Oriented Thinking; the Benton Face Recognition Test, a 13-item test of face recognition ability; and the Social Responsiveness Scale, a 65-item test with five subscales measuring Social Awareness, Social Cognition, Social Communication, Social Motivation, and Autistic Mannerisms. The NT participants took these tests only once, as they did not take part directly in the intervention; the ASC students took each of these tests (along with the VidL naturalistic-videos test covered in the next section) before the intervention, regardless of control/experimental group status, and will take them immediately afterwards as well.

In addition, the following two tests were given exclusively to ASC participants, and were not readministered after the intervention: the Kaufman Brief Intelligence Test (KBIT-2), a short test of intelligence quotient, and the Adult Diagnostic Observation Schedule (ADOS), a test of many traits relevant to ASCs [45, 48].

At the time of writing (8/6/10), the VidL naturalistic video pretest had been administered to 18 NT participants and 20 VidL participants.

6.2 Results of Pretests

The ASC and NT participant groups were matched on age and gender: each NT participant was within 6 months in age of their ASC counterpart and was of the same gender. NT participants had not been diagnosed with an ASC. Means and standard deviations for each pretest administered are summarized in Table 6, including subscale scores for each test that varied the most between NT and ASC participants. See Figure 15 for a comparison of overall scores.

Figure 15: Overall Scores Comparison for the Benton Facial Recognition Test, the Social Responsiveness Scale (SRS), and the Toronto Alexithymia Scale (TAS-20). ASC participants were more impaired than NT participants at face-reading (Benton), had weaker social responsiveness skills (SRS), and were more alexithymic (TAS-20). These three tests were given to all participants and evaluated on similar scales (unlike the VidL Pretest).

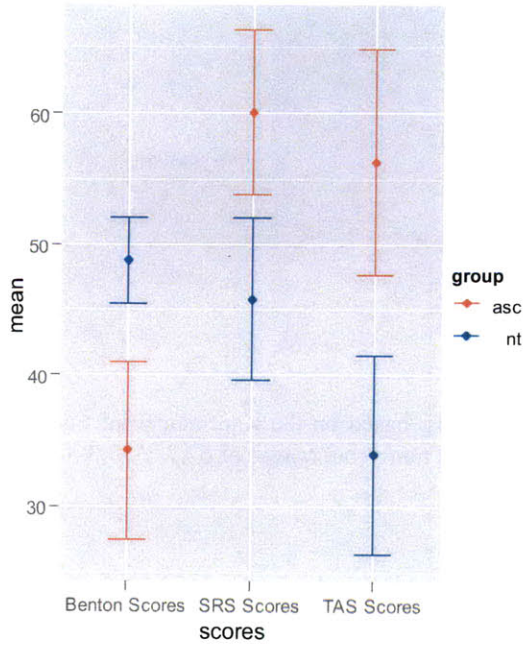
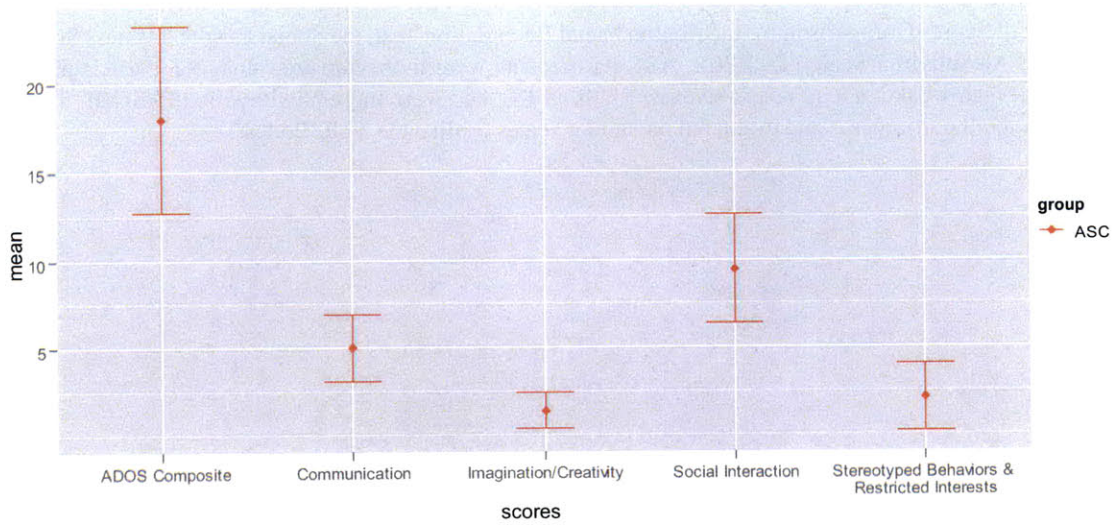


Table 6: Summary of Pretest Scores for ASC and NT Participant Groups. Overall, the ASC participants were much more likely to score as autistic, alexithymic, and face-reading impaired.

Test name	NT mean/sd	ASC mean/sd	Cutoff Point(s)
Adult Diagnostic Observation Schedule (ADOS)	-	17.8 ± 5.21	-
ADOS subscale: Social + Communication	-	14.4 ± 4.13	$\geq$7=ASC,$\geq$12= autism
Benton Face Recognition Test	48.4 ± 3.75	34.6 ± 6.29	$\leq$40=impaired face recognition
Kaufman Brief Intelligence Test (KBIT-2)	-	54.9 ± 12.3	-
KBIT-2 subscale: nonverbal	-	62.7 ± 17.3	-
KBIT-2 subscale: verbal	-	54.5 ± 11.8	-
Social Reponsiveness Scale (SRS) T-Scores	46.1 ± 6.51	60.1 ± 6.31	-
SRS subscale: Social Cognition	43.7 ± 6.93	62.4 ± 5.41	-
Toronto Alexithymia Scale (TAS-20)	35.6 ± 9.39	56.2 ± 8.17	$\geq$51=alexithymia, $\geq$41=probable alexithymia
TAS-20 subscale: Externally Oriented Thinking	15.0 ± 5.12	24.0 ± 3.90	-
VidL Naturalistic Videos Pretest	$70.1\% \pm 8.84\%$	$42.0\% \pm 13.3\%$	-

The ADOS subscale scores for our ASC participants were: Communication, 5.0 ± 1.8; Social Interaction 9.4 ± 3.0; Imagination, 1.5 ± 1.1; Stereotyped Behaviors and Restricted Interests, 2.0 ± 1.9. Overall, ADOS scores had a mean of 17.8 ± 5.21. The average for a combined Social + Communication score was 14.4 ± 4.13; the cutoff for "autism" on this subscale is 12, and the cutoff for an ASC is 7. Overall, 17 participants scored in the "autism" range, 2 scored in the ASC range, and one scored in the "no autism" range. (See Figure 16 and Appendix 9.3.1.)

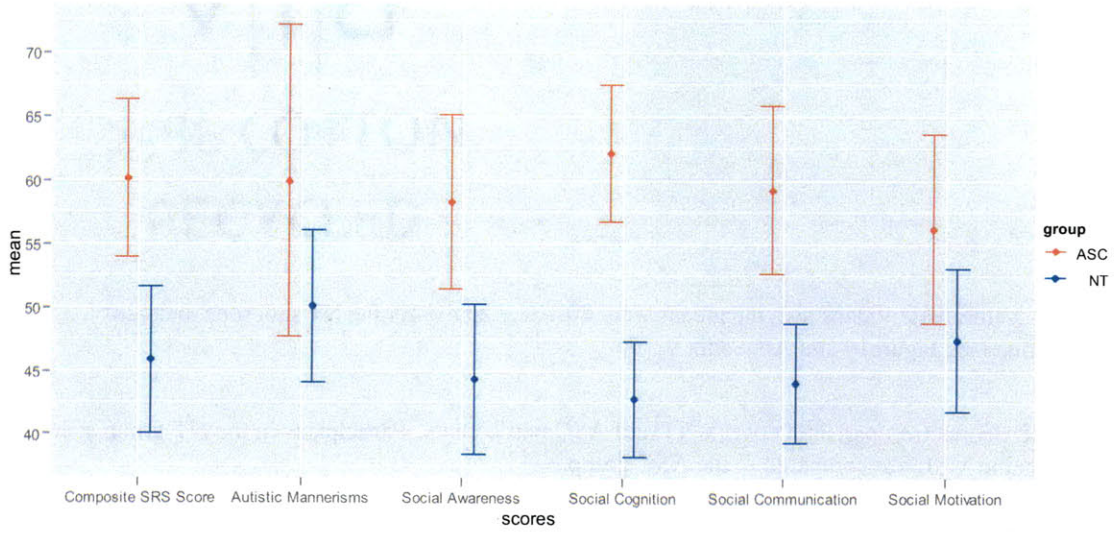
Figure 16: Adult Diagnostic Observation Schedule (ADOS) Composite and Subscale Scores for ASC Participants.



The IQ ranges for our ASC participants, based on the Kaufman Brief Intelligence Test (KBIT-2), were 54.9 ± 12.3, with verbal ranges of 54.5 ± 11.8 and nonverbal ranges of 62.7 ± 17.3. (See Appendix 9.3.4.)

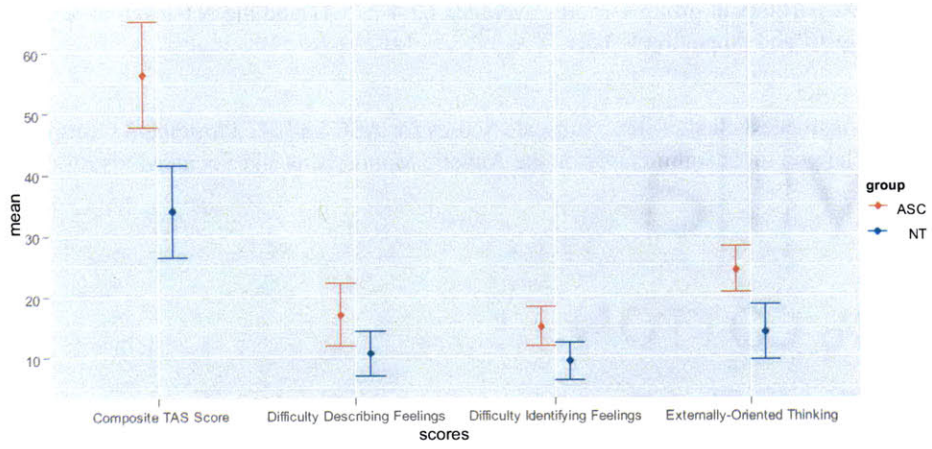
When comparing our participants using the Social Responsiveness Scale, we saw a significant difference between the groups, as expected, based on the SRS questions relating to many typical autistic traits. The overall T scores for the ASC population (n=20) had a mean of 60.1 $\pm$ 6.31, while the overall T scores for the matched NT population (n=20) had a mean of 46.1 $\pm$ 6.51. The difference between the two groups was most pronounced on the Social Cognition subscale, where the ASC participant group T-scores averaged 62.4 $\pm$ 5.41 and the NT group T-scores had a mean of 43.7 $\pm$ 6.93. (See Figure 17 and Appendix 9.3.5.)

Figure 17: Social Responsiveness Scale (SRS) Subscale Scores for ASC and NT Participant Groups. Subscale scores for the two groups overlapped most significantly on the Autistic Mannerisms and Social Motivation subscales.



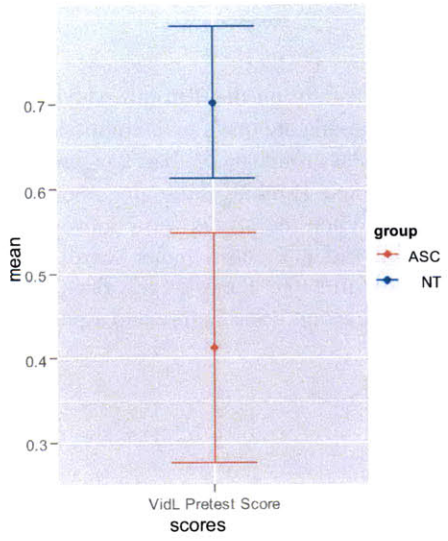
Similarly, participant groups varied significantly on the Toronto Alexithymia Scale (TAS-20), a test measuring alexithymia, the ability to understand and communicate one's own emotional state. The ASC group (n=20) scored an average of 56.2 $\pm$ 8.17, while the NT participant group (n=20) had a mean of 35.6 $\pm$ 9.39. The subscale differences were most pronounced in the Externally-Oriented Thinking subscale, where the NT subgroup averaged 15.0 $\pm$ 5.12 and the ASC subgroup scored 24.0 $\pm$ 3.90. When the scores were converted to alexithymia subscores - a score of 41-51 on the test is rated as probable alexithymia (1), while higher scores indicate a high likelihood of alexithymia (2), and lower scores indicate a lack of alexithymia (0) - the average values were 1.05 $\pm$ 0.740 in the ASC participant group and 0.158 $\pm$ 0.488 in the NT participant group. (See Figure and Appendix 9.3.6.)

Figure 18: Toronto Alexithymia Scale (TAS-20) Composite and Subscale Results for ASC and NT Participant Groups. A higher score indicates a higher likelihood of alexithymia; the ASC group scored as significantly more alexithymic on each subscale measure and in the composite measure.



In the VidL Naturalistic Videos test, the means were 70.1% $\pm$ 8.84% for the NT subgroup and 42.0% $\pm$ 13.3% for the ASC subgroup (see Figure19 and Appendix 9.3.7).

Figure 19: VidL Scores Comparison between NT and ASC participants. Participants in the NT group scored significantly higher on the VidL test compared to the ASC group.

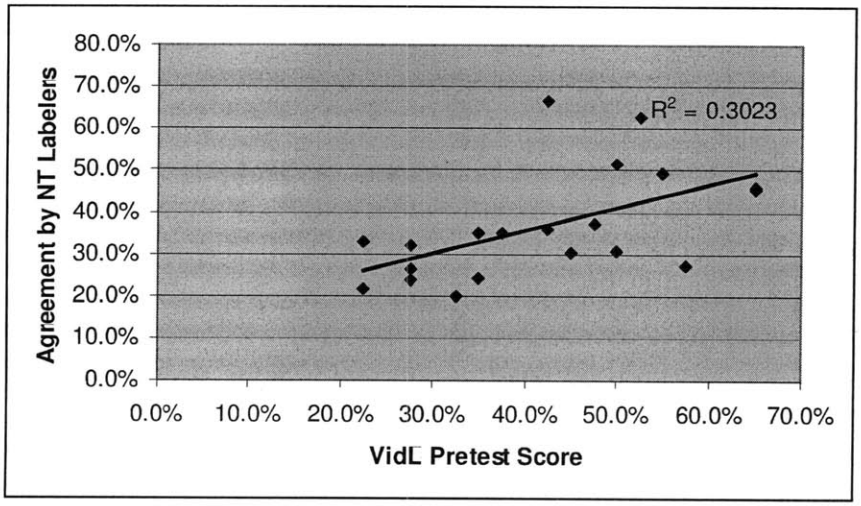


The final standardized test administered to both groups was the Benton Face Recogition Test. In this test, participants were asked to look at a series of pictures and match identical pictures as well as pictures from different angles of the same person. The neurotypical group (n=20) scores on this test had a mean of 48.4 $\pm$ 3.75, while the ASC group (n=20) had scores averaging 34.6 $\pm$ 6.29. When the cutoff for "face recognition impairment" (41 and above is classified as "not impaired", 0; 40 or below is "impaired", 1) was analyzed, the average was 0.750 $\pm$ 0.433 in the ASC group and 0.0500 $\pm$ 0.218 in the NT group. (See Appendix 9.3.2.)

6.3 Analysis of Pretest Results and Correlations

One piece of data that was calculated during pretest data collection and labeling by the NT labelers was "agreement", i.e., how often NT subjects agreed on a video. For instance, a 70% "agreement" rate would imply the NT labelers were able to come to a consensus - with $\geq$75% of NT viewers in agreement - for 70% of the videos of that participant that were viewed by the NT labelers. This variable was found to correlate relatively well with the VidL naturalistic video score of the ASC participant ($R^2=0.3023$, $p<0.05$) . This finding implies that participants who are better able to express emotion are also better at successfully reading emotions of persons around them. (See Figure 20 for a plot of these results.)

Figure 20: Correlation between Agreement and VidL Pretest Score. The relatively strong correlation betwene these variables indicates that participants who expressed emotion in more standard and clear ways (with a higher level of convergence by NT labelers) were also able to interpret emotion more accurately in naturalistic videos.



Furthermore, the VidL pretest score was positively correlated with the Benton Pretest Score ($R^2=0.1578$, $p<0.1$; see Figure 21) and the KBIT-2 Verbal Subscale ($R^2=0.263$, $p<0.05$; see Figure 22) and was negatively correlated with the Stereotyped Behaviors & Restricted Interested Subscale of the ADOS Test ($R^2=0.1492$, $p<0.1$; see Figure 23). These findings imply that a higher verbal IQ is correlated with ability to successfully read emotions, and that increased expression of stereotyped behaviors (such as hand-flapping) is poorly correlated with this emotion ability, perhaps by detracting from interpersonal interactions. The Benton Pretest Score correlation with the VidL pretest score is to be expected, given the similar content of these two tests.

Figure 21: Benton Pretest Score Correlation with VidL Naturalistic Pretest Score. The correlation (R^2=0.1579, $p<0.1$) indicates a relatively strong correlatation between these two variables, a finding which is predictable because of the similiarities of these two tests in terms of presentation and processing of realistic faces.

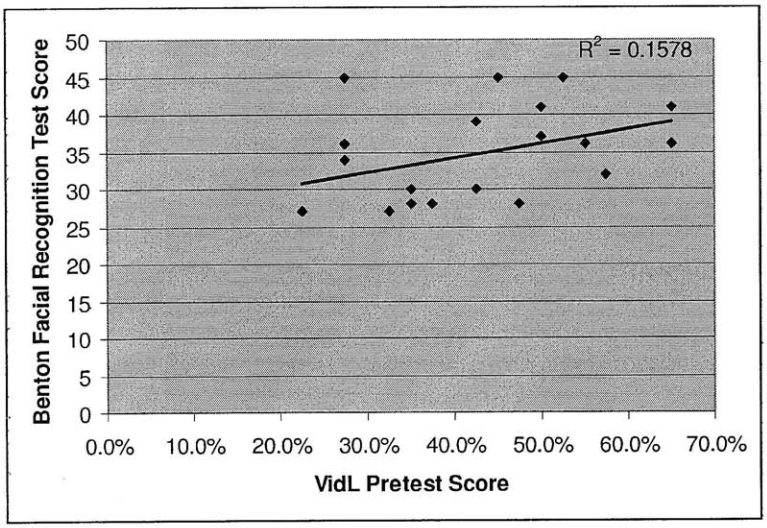


Figure 22: KBIT-2 Verbal Subscale Correlation with VidL Naturalistic Pretest Score. The correlation between the verbal subscale score and the VidL test score (R^2=0.263, $p<0.05$) suggests that increased ability to discuss affective communication appears in tandem with increased ability to recognize affective expression in naturalistic video.

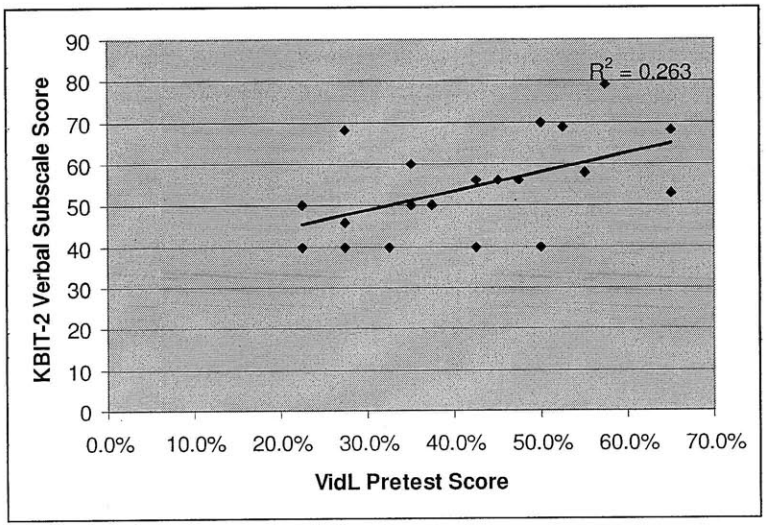
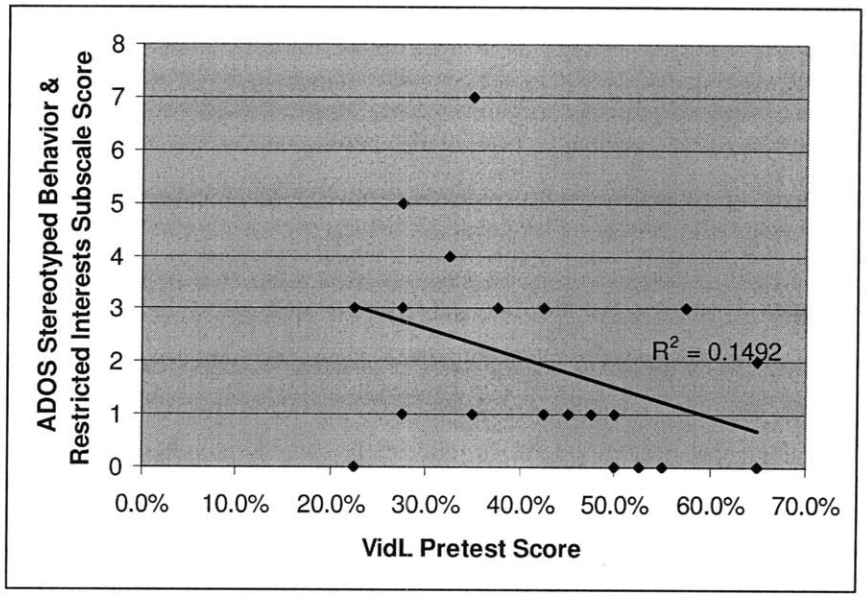


Figure 23: ADOS Stereotyped Movements & Restricted Interests Negative Correlation with VidL Naturalistic Pretest Score. The negative correlation between these variables (R^2=0.1492, p<0.1) implies a correlation between decreased ability to recognize emotion expression and increased tendency toward movements and behaviors particularly characteristic to autism, such as stereotyped behaviors (e.g. hand-flapping) and restricted areas of interest.



6.4 Conclusion

The ASC participants in the study had a stronger display of autistic traits compared to the NT counterparts and scored significantly lower on the VidL naturalistic video pretest as well as other standardized pretests. The VidL Naturalistic Videos pretest score was positively correlated with ability of NT labelers to to converge on a label for the videos collected by a participant; Benton Facial Recognition Score; and the KBIT-2 verbal subscale score. The VidL score was negatively correlated with the ADOS subscale score on Stereotyped Movements & Restricted Interests; this correlation indicates that persons who expressed more "autistic" behavioral characteristics, such as stereotyped behavior and limited interests, may experience more difficulty with successful recognition of emotion in naturalistic video.

7 Preliminary Intervention Data

7.1 Introduction

At the time of writing (8/6/10), progress is as follows:

- 5 ASC participants (04, 18, 19, 20, 24) are completing Period 1 of the intervention.
- 13 ASC students (07, 08, 10, 11, 12, 13, 14, 15, 16, 17, 21, 22, 23) are completing Period 2 of the intervention.
- 2 ASC students (05, 06) are completing Period 3 of the intervention.

Data collection is in progress and completion of the intervention is expected for most participants by the end of October 2010. Post-testing will begin on 8/22/10.

7.2 Increased Competence with Samsung Computer Use

During the intervention, all clips that students recorded by pushing an annotation button on the labeling interface were recorded as intended labels; however, only the clips in which the face-tracker was on at the time of labeling were shown to the student during the viewing component of the session.

In order to determine competence with the iSET Interface labeling system, the percentages of labeled clips in which the tracker was on during labeling ("tracked clips") divided by total clips labeled by that student during recorded ("recorded clips") were examined. The number of tracked clips as a percentage of recorded clips spiked in the first week, coinciding with the initial use by the participants and the most direct guidance of the participants by teachers and research assistants, as expected. Since that time, many students have been holding the system independently, and as they and the teachers become more familiar with the system, the overall percentage of tracked clips has significantly increased, indicating greater familiarity with the system and more productive recording sessions (see Table 7).

Table 7: Increased Tracking Percentage During Intervention. A high percentage of videos were tracked in the first week, when significant guidance was offered by RAs; a much lower percentage of clips were tracked in the second week when student were first using the system on their own, and in weeks 2-8 percentage of tracked videos steadily increased.

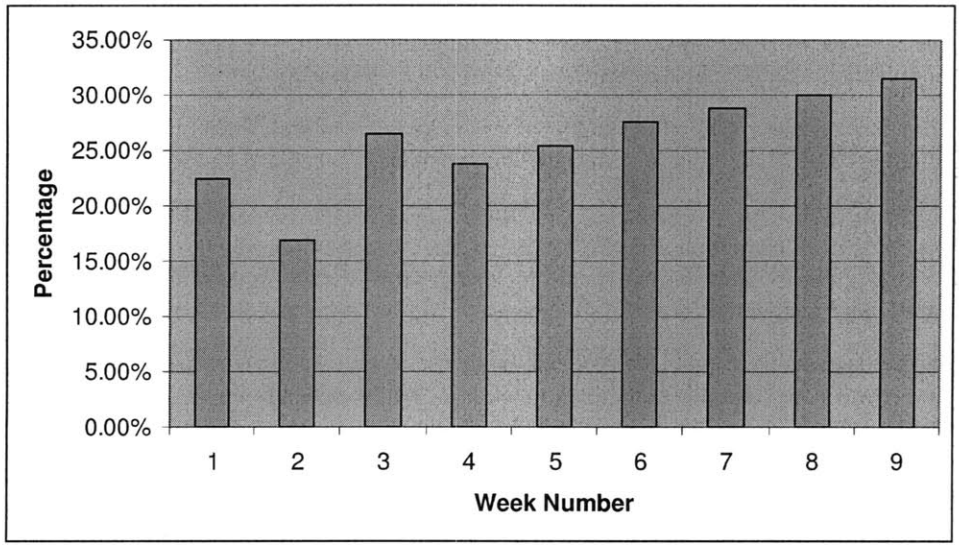
Week #	Percentage Tracked
1	76.9%
2	57.3%
3	70.2%
4	74.4%
5	72.9%
6	73.9%
7	86.7%
8	93.3%

One priority for this study was to gather group information showing the overall efficacy of the intervention with the students while also reviewing individual students to see how the study might work better for them or how their suitability for this type of intervention might be predicted by pretest scores. For example, some students did not increase significantly in their tracking percentage, e.g. student 11, whose performance after the first week (in weeks 2-6, the time period for which this participant's data was available) had a mean of 65.1% $\pm$ 10.2% tracked clips. In contrast, student 14 had a marked increase in clips which were tracked; 69.0% of clips were tracked in the second week and performance increased to 93.3% clips in week 8 (the most recent week for which both recording sessions took place for this student.) The average tracking percentage across all participants in weeks 2-8 was 77.2% $\pm$ 9.53%.

7.3 Increased Labeling of Good Examples

During the course of the intervention, students had significant success in collecting good examples as a percentage of overall clips in the first week, which then dropped sharply in the second week as teachers changed to a more hands-off approach. The percentage of good examples then rose fairly steadily during the ensuing weeks. (This data can be seen in Figure 24 and Appendix 9.4.12.)

Figure 24: Percentage of Videos Labeled by Teachers as "Good Examples" Per Week. A monotonically increasing percentage of videos labeled as "Good Examples" was seen in weeks 4-9; weeks 1-3 may indicate extra help and guidance from teachers in the first week (as was requested by the iSET project team), significantly less guidance in the second week (leading to less overall success), and some extra assistance in the third week to help with motivation to continue in the intervention after a less-successful second week.



7.4 Qualitative Observations of Students

In both the iSET and Mind Reading DVD components of the intervention, many of the qualitative observations made indicate a high degree of engagement by students. One major factor that indicated an advantage to the iSET approach was that many students had particular interest in working directly with - and, if possible - recording particular teachers or aides. Many qualitative observations were recorded in which students were particularly enthusiastic to record and then play back video of a particular peer or aide, and seeing these videos seemed to create a feedback loop in which students were then more enthusiastic about engaging in an affective-oriented mental state with these persons.

An intern for the MIT Media Lab made many qualitative observations of Mind Reading DVD and iSET group participants. Some insights gleaned through her work include:

- Mind Reading DVD interactions were sometimes easy to make personal: one MR participant in Week 1 was described as "...easily relating to 'happy' stories on a personal level."
- Students in the Mind Reading DVD were not always fully engaged by video of strangers: a participant in Week 2 of the intervention would "...[try] to answer [the quiz] before viewing both videos."
- Students remembered their iSET recording sessions as though they were social interactions; one participant in Week 1 "...reflected back on trying to make a particular person look and feel happy during the previous recording session."
- Students in the iSET intervention successfully navigated the inherent barriers to an *in situ* recording situation: one participant "...received several 'No' answers when asking people if they would like to be filmed; however, [he] was absolutely fine with that."
- iSET group participants were enthusiastic about the process of interacting with their peers and caregivers during the process. It was noted that one participant in Week 2 "really enjoys filming those [people] that she knows", and another participant was described in Week 2 as "...having a good time choosing subjects ... was laughing and enjoying themselves ... was at first tentative to ask if they could film others but then gained some more confidence in approaching others."

7.5 Conclusion

The iSET intervention shows a promising improvement in familiarity with the system by the participants. Each participant also had an increasing overall trend of labeling of "good examples" by teachers over the course of the reaction.

8 Conclusion

8.1 Design of the iSET Intervention

Participatory design was a major component of the iSET intervention. Starting with the SelfCam design, feedback was actively solicited by the iSET design team on a continuous basis. The major principles absorbed during the overall design process and intervention were:

- *Make the system as engaging as possible.* Our participants were interested in the initial FaceSense software, which contained many types of graphs, but they were unable to use it effectively. A later design iteration, Emotion Bubbles, contained interesting information for participants to look at while they focused the Samsung camera on a conversation partner's face, but did not allow sufficient simple interaction with the display to engage some participants. By switching to a paradigm requiring frequent and interesting user input, the iSET Interface significantly improved the experience and engagement levels of participants.
- *Make the system as easy to use as possible.* By creating simple step-by-step directions for Recording, Viewing, and Mind Reading DVD Sessions, we significantly reduced the number of issues reported by Groden Center staff members using the intervention. Furthermore, creating large, easy-to-use buttons and situating them on the same side of the viewing window allowed for logistically simpler use by participants.
- *Allow the real-life use scenarios and goals of persons on the spectrum to drive design decisions.* A core consideration of the design of the iSET intervention was that persons on the spectrum should have a significant say in the design and use of the system. Having many pilot sessions at the Groden Center and presenting at conferences that included people on the spectrum among their attendees (such as IMFAR, Autreat, and ASSETS) allowed for significant feedback on how the system could best look and feel to serve its users.

8.2 Pre-Intervention Data Collection

The major contributions of the data collection and analysis performed prior to the start of the iSET intervention are severalfold. Information learned can be grouped into several categories: Design Considerations, Data Collection Findings, and Information Learned During the iSET Intervention.

8.2.1 Design Considerations

The iSET Interface encountered a number of issues during the course of the participatory design sessions, including but not limited to: how to describe the emotions in the intervention; how to create an interface suitable for participants with all levels of cognitive levels, motor deficits, and vision impairments; and how to choose and adapt a piece of hardware suitable for participants with behavior issues, unpredictable behavior, and motor difficulties.

Ultimately, the guiding principles that answered many of these considerations were *simplicity* and *consistency*. For instance, the same emotion descriptions and colors were used in the iSET Interface, in VidL, and in the descriptions given to NT labelers, teachers, and both ASC and NT participants. This question also came into play when considering whether pictures, text, or both should be presented on the iSET interface along with the characteristic colors for each emotion. Many types of images were considered for the labeling buttons; ultimately, each possibility was discarded either for being too general (as with typical Boardmaker images used for those emotions/cognitive states) or too specific, interfering with the systematizing abilities of users on the spectrum. Intergrating caregivers significantly in the overall protocol allowed for the best combination of information transfer and situational adaptiveness for the iSET Interface.

8.2.2 Data Collection Findings

We had several dedicated research assistants who worked to collect video of 20 students and 25 teachers (later reduced to 7 teachers, the minimum necessary to cover the requirements for video of familiar- and unfamiliar-adult clips for all participants.) We also were able to use the relatively efficient VidL system to outsource labeling remotely to labelers, who could then label on their own time and have their labels transferred instantly to lab databases instead of coming into the lab. However, it still took us many months (from July 2009 until April 2010) to collect the eight NT-labeler-approved clips of each participant and teacher required. While the naturalistic videos are a valuable and integral part of the iSET intervention, it may be worth exploring strategies for more streamlined collection and labeling.

8.2.3 Information Learned During the iSET Intervention

During the iSET intervention, the ability to collect data was initially hampered by occasional software and hardware glitches pertaining to the equipment used for the intervention. For instance, a custom-made removable cover intended to remove the distraction of the Samsung Ultramobile computer keyboard inadvertently ended up triggering software unrelated to the intervention and had to be reattached permanently in a more limited capacity. Relatedly, some software needed during the intervention was simplified on the back-end early on so that data could be parsed more easily.

After RAs and teachers were able to familiarize themselves with the improved software, issues mostly arose from the scheduling and behavioral issues of the participants on the spectrum. Some teachers were reluctant to participate in the intervention since they did not feel it was part of their job description, and RAs had trouble scheduling some participants due to the participants' job commitments. Although there was a high degree of adherence to the study protocol overall, assisted by daily logs and fidelity checks, the scheduling issues highlighted the importance of having dedicated caregivers who are committed to the intervention in order to ensure its success.

8.3 Intended Analysis for Final iSET Results

Analysis of the iSET results will have several major goals. First, the final analysis should identify characteristics of people on the spectrum who may be helped most by the iSET intervention and its successors. In that vein, participants who register the most improvement in emotion-reading standardized tests, as well as the customized VidL test, will be assessed to determine which measures may predict success at learning in this type of intervention. By using clustering algorithms, participants can be grouped into persons who experienced the most and the least success in each type of intervention.

Each ASC participant looked at videos of persons with varying levels of familiarity to them in the pretest; by analyzing ability to correctly label these videos for each level of familiarity, information can be gained about how pretest results may predict generalization ability.

8.4 Considerations For Future Work

Many considerations for future work came to light in the overall design of the iSET project. For instance, the current project virtually ignored the role of speaking and audio distractions during social interactions. Future work should include these aspects of social interactions, perhaps both combined with and separate from the visual channels of these interactions. The FaceSense software was used to analyze videos labeled "good examples" during the intervention to determine the concordance of FaceSense's labeling with that of student participants and teachers in the intervention. However, automated emotion-reading and more sophisticated face-tracking and -reading tools could be used regularly during the intervention, with the feedback from these tools returned daily to participants and teachers.

In order to facilitate more natural social interactions, it may be beneficial to return to the wearable-technology model first tried with the SelfCam; the Samsungs were used in the final iSET design to provide a more flexible interface for labeling, but holding the Samsung computer can obscure the view of the other person's face in ways that may not be conducive to natural conversation and learning of emotion elicitation skills.

However, the converse of the Samsung computer blocking eye contact was that the device could be used as a "social buffer"; many participants on the spectrum in this intervention seemed to use it in exactly that way, and were able to negotiate conversational settings by increasing or decreasing eye contact to ensure maximal personal contact by using the Samsung as an intermittent eye-contact blocker. The use of similar tools as "social buffers" to ease interpersonal interactions and create comfortable social environments for persons on the spectrum may be worth investigation furhter.

Finally, the critical goal for iSET, as for many emotion-teaching interventions and tools, is to allow both close and distant generalization of skills learned. Participants should come out of the learning event with increased knowledge and comfort with interpersonal interactions and with the skills to interpret video and live interactions involving expressions, situations, and persons who were not directly part of the intervention. A primary motivation for the iSET project was the failure of the Mind Reading DVD intervention to succeed in improving generalization skills. Posttest will hopefully show an increase in these skills for the iSET project, and future work in this area with naturalistic videos may show promise in extending the inherent systematizing skills of people on the spectrum to the area of emotion-reading.

9 Appendices

9.1 Pretest Data Collection

9.1.1 Instructions for NT labelers

Hello! Thank you for agreeing to label videos with our online video interface.

Read these instructions in their entirety before you start working. Email me with questions.

Here is the link to the interface: [VidL labeling link] Please use your first and last name to log in.

For each labeling session: after logging in, go to the Mode menu and select "Single Label". (NOTE: this is very important! If you postpone this until partway through a labeling session, ALL YOUR LABELS WILL BE LOST.)

Select "Quick Menu" from the Controls menu and move the box that appears to the side of the screen, so that it does not obscure the black box. Press "Next Video" to access the first video.

Tag each video with one of the labels on the right side of the screen while it is playing or after it finishes. You can view a video more than once if necessary by clicking "Play" after it finishes. Press "Next Video" to see the next video in your queue after you've added a label.

Be sure to record each working session online at: [online timesheet]. You may only count time during which you were actively watching/labeling videos (not just when you were logged in.) Your time will also be tracked via the VidL system.

Please read the "emotion descriptions" at the bottom of this email and watch the linked videos now and anytime you feel that you need them as you are labeling.

Have fun! Your help is much appreciated. Contact us as soon as possible if you need help or have questions.

9.1.2 Emotion Descriptions for NT Labelers

Emotion Descriptions:

"Thinking" is used to mean that someone is considering something internally and withdrawing from outside stimuli in order to thinking about this thing. In order to think, someone might avert their eyes, sit back or tilt their head, and/or purse their lips.

http://affect.media.mit.edu/Rgrads/madsen/thinking1.mov

http://affect.media.mit.edu/Rgrads/madsen/thinking2.mov

"Concentrating" is used to mean that someone is outwardly engaged with some particular stimulus and ignoring distractions; in order to concentrate, someone might lean forward and/or look directly at something in front of them, and narrow the eyes slightly.

http://affect.media.mit.edu/Rgrads/madsen/concentrating1.mov

http://affect.media.mit.edu/Rgrads/madsen/concentrating2.mov

"Interested" is different from these; in order to express interest, someone might lean forward, look directly forward or at a number of interesting things, and/or widen their eyes and lift their eyebrows.

http://affect.media.mit.edu/Rgrads/madsen/interested1.mov

http://affect.media.mit.edu/Rgrads/madsen/interested2.mov

9.1.3 Foils for Neurotypical Labelers

Emotion	Similar Foils	Different Foil
Afraid	Angry, Sad	Happy
Agreeing	Happy, Interested	Disagreeing
Angry	Disagreeing, Disgusted	Happy
Concentrating	Interested, Thinking	Surprised
Disagreeing	Angry, Sad	Agreeing
Disgusted	Angry, Disagreeing	Happy
Happy	Agreeing, Interested	Sad
Interested	Agreeing, Happy	Sad
Sad	Afraid, Disagreeing	Happy
Surprised	Happy, Interested	Concentrating
Thinking	Concentrating, Unsure	Agreeing
Unsure	Concentrating, Thinking	Agreeing

9.1.4 Statistics on Pretest Data Collection

Videos Collected Per Intervention Participant for ASC Participants

Participant ID	Videos Approved By NT Labelers	Videos Collected	Percentage Approved
4	47	182	25.8%
5	21	80	26.3%
6	36	90	40.0%
7	31	107	29.0%
8	56	104	53.8%
10	10	26	38.5%
11	27	63	42.9%
12	27	88	30.7%
13	41	126	32.5%
14	37	136	27.2%
15	25	62	40.3%
16	33	128	25.8%
17	46	137	33.6%
18	26	96	27.1%
19	47	193	24.4%
20	30	83	36.1%
21	19	55	34.5%
22	24	59	40.7%
23	33	118	28.0%
24	51	99	51.5%

Videos Collected Per Intervention Participant for NT Participants (Teachers)

Participant (Teacher) ID	Videos Approved By NT Labelers	Videos Collected	% of Videos Approved
26	32	63	50.8%
59	32	84	38.1%
72	44	121	36.4%
77	45	153	29.4%
80	27	76	35.5%
89	57	172	33.1%
102	34	99	34.3%

Convergence By Emotion For Videos of All Participants

Emotion Emotion	Videos Approved	Labels Applied	Percentage Approved
concentrating	299	1690	17.7%
happy	218	716	30.4%
interested	180	1354	13.3%
thinking	282	2236	12.6%

9.2 Pretest Administration

9.2.1 Pretest Instructions

Please have the descriptions from [emotion description link] *[see Section 9.4.2]* printed out (in color) and available at all times. The subject should read them before beginning labeling and be able to refer to the instructions throughout.

Login at [VidL labeling link]

Participant XX (where XX is between 04-24) will have four associated pretests with the following login information.

For participant XX's own pretest, use:

asdXX

pretest

For participant XX's NT counterpart, use:

ntXX

pretest

For participant XX's primary teacher, use:

primaryXX

pretest

For participant XX's secondary teacher, if needed, use:

secondaryXX

pretest

Play the first movie by choosing "Quick Menu" from the "Controls" menu and then clicking "Next Video." The first video will begin to play.

Choose *exactly one* emotion label (by clicking on it) per video and click "Next Video" to save your video and move to the next video when you're done. You can watch a video several times before labeling it by clicking "Play" after the video is done playing.

9.2.2 Pretests Administered

Test Name	ASC (pretest)	ASC (posttest)	teachers	NT subjects
Autism Diagnosis Observation Schedule (ADOS)[48]	X	-	-	-
Benton Face Recognition Test [51]	X	X	-	X
Reading the Mind in the Eyes test [49]	-	-	X	-
KBIT-2 [45]	X	-	-	-
Social Responsiveness Scale (SRS)[50]	X	X	-	X
Toronto Alexithymia Scale (TAS-20) [46]	X	X	-	X
VidL Naturalistic Videos Test	X	X	X	X

9.3 Pretest Results

9.3.1 Autism Diagnosis Observation Schedule (ADOS) Results

ASC ID	Diagnosis	Communication	Social Interaction	Communication + Social Interaction	Imagination/Creativity	Stereotyped Behaviors & Restricted Interests
4	autism	7	11	18	1	1
5	autism	6	11	17	3	1
6	autism	4	7	11	1	0
7	autism	6	10	16	1	1
8	autism spectrum	4	3	7	1	0
10	autism	7	12	19	0	0
11	no autism	1	5	6	0	0
12	autism spectrum	4	5	9	2	3
13	autism	7	6	13	2	3
14	autism	4	7	11	1	2
15	autism	4	10	14	1	4
16	autism	5	9	14	0	3
17	autism	3	10	13	0	5
18	autism	8	13	21	2	1
19	autism	7	13	20	2	7
20	autism	4	12	16	2	1
21	autism	3	12	15	2	3
22	autism	7	11	18	4	1
23	autism	5	8	13	2	3
24	autism	4	13	17	2	0

9.3.2 Benton Face Recognition Test Results

ASC ID	ASC score	ASC face recognition impairment	NT score	NT face recognition impairment
4	28	yes	45	no
5	39	yes	39	yes
6	36	yes	50	no
7	45	no	47	no
8	45	no	45	no
10	27	yes	45	no
11	37	yes	43	no
12	27	yes	50	no
13	28	yes	45	no
14	41	no	49	no
15	27	yes	49	no
16	32	yes	49	no
17	36	yes	54	no
18	30	yes	52	no
19	28	yes	50	no
20	41	no	47	no
21	30	yes	52	no
22	34	yes	52	no
23	45	no	54	no
24	36	yes	50	no

9.3.3 Reading the Mind in the Eyes Test Results

Teacher ID	Corresponding Student ID(s)	Reading the Mind in the Eyes score
26	5	23
28	5	26
29	5	33
31	6	24
46	13	24
50	8, 14	23
52	13	29
56	8, 10, 14	28
66	17, 19	23
69	19	25
77	21, 24	28
85	18	26
86	15	30
91	20	31
93	16	31
96	20	23
97	12	24
102	15	28
105	21, 24	28
107	4	24
108	4	23
109	6	26
110	17	28
113	23	25
114	22	26
115	18	27
117	12	26
118	22	26
121	7, 11	23
122	7, 10, 11	23
124	6	31
125	23	28

9.3.4 Kaufman Brief Intelligence Test (KBIT-2) Results

ASC participant ID	Composite	Verbal	Nonverbal
4	62	56	77
5	49	56	52
6	55	58	62
7	60	56	74
8	69	69	78
10	59	50	78
11	40	40	48
12	40	40	40
13	40	50	40
14	73	68	85
15	40	40	40
16	76	79	79
17	68	68	70
18	48	50	50
19	53	60	57
20	69	70	76
21	40	40	42
22	64	40	98
23	45	46	55
24	47	53	52

9.3.5 Social Responsiveness Scale (SRS) Results

SRS T-Scores Scores for ASC Participants

ASC ID	Overall SRS Score	Social Awareness	Social Cognition	Social Communication	Social Motivation	Autistic Mannerisms
4	73	72	73	72	69	70
5	53	55	61	50	52	49
6	67	63	70	60	63	79
7	59	60	67	60	47	58
8	49	52	59	60	43	21
10	60	62	59	60	59	56
11	54	57	58	54	54	53
12	59	62	72	56	45	62
13	55	55	55	52	50	64
14	58	62	62	54	56	59
15	60	60	61	59	50	63
16	66	65	62	63	64	71
17	60	45	59	56	63	69
18	56	52	61	52	53	62
19	62	60	70	57	61	61
20	62	55	54	62	63	65
21	54	52	59	55	51	47
22	69	70	62	70	62	72
23	70	52	65	72	63	74
24	55	55	59	51	58	51

SRS T-Scores Scores for NT Participants

NT ID	Overall SRS Score	Social Awareness	Social Cognition	Social Communication	Social Motivation	Autistic Mannerisms
4	48	50	45	47	54	54
5	38	40	37	38	43	41
6	59	51	66	55	63	53
7	40	34	40	42	38	42
8	46	46	39	44	46	54
10	47	43	39	44	51	48
11	58	52	53	49	57	53
12	47	50	44	43	46	56
13	37	33	39	39	42	40
14	39	36	39	38	43	44
15	56	50	53	43	50	62
16	42	43	42	41	43	46
17	40	45	38	41	41	43
18	53	52	42	56	56	55
19	44	40	42	41	48	55
20	45	48	45	43	40	52
21	45	40	41	45	48	54
22	51	45	48	53	53	52
23	46	48	42	45	52	46
24	41	43	39	39	46	45

9.3.6 Toronto Alexithymia Scale (TAS-20) Results

TAS-20 Scores for ASC Participants

ASC ID	Alexithymia	Score, Overall	Score, Subscale 1	Score, Subscale 2	Score, Subscale 3
4	1	54	12	14	28
5	1	57	18	14	25
6	1	54	22	15	17
7	0	50	10	14	26
8	1	57	18	20	19
10	0	51	16	13	22
11	0	38	10	7	21
12	2	71	27	20	24
13	1	53	13	13	27
14	1	56	23	15	18
15	2	64	22	18	24
16	3	66	22	19	25
17	0	46	17	12	17
18	1	55	16	15	24
19	2	65	21	18	26
20	1	53	12	14	27
21	0	47	11	12	24
22	1	54	14	15	25
23	2	66	22	14	30
24	2	66	29	17	30

TAS-20 Scores for NT Participants

ASC Match ID	Alexithymia	Score, Overall	Score, Subscale 1	Score, Subscale 2	Score, Subscale 3
4	0	39	10	8	21
5	0	40	8	9	23
6	2	61	20	18	23
7	1	55	21	15	19
8	0	37	15	7	15
10	0	33	9	12	12
11	0	28	9	9	10
12	0	29	10	9	10
13	0	37	9	7	21
14	0	22	5	12	5
15	0	36	15	12	9
16	0	30	10	5	15
17	0	36	9	9	18
18	0	45	9	15	21
19	0	26	7	6	13
20	0	36	15	11	10
21	0	32	9	10	13
22	0	28	8	6	14
23	0	32	11	9	12
24	0	30	10	5	15

9.3.7 VidL Naturalistic Videos Test Results

VidL Naturalistic Videos Test Results Per Participant

ASC ID	ASC Participant Score	NT Participant Score
4	47.5%	75%
5	42.5%	-
6	55%	-
7	45%	62.5%
8	52.5%	67.5%
10	22.5%	82.5%
11	50%	50%
12	22.5%	70%
13	37.5%	57.5%
14	65%	70%
15	32.5%	65%
16	57.5%	60%
17	27.5%	80%
18	35%	80%
19	35%	67.5%
20	50%	75%
21	42.5%	72.5%
22	27.5%	72.5%
23	27.5%	82.5%
24	65%	72.5%

VidL Naturalistic Videos Test Results, Sorted By Relationship, for ASC (n=20) and NT (n=18) participants

Labeler type	Relation to ASC participant	ASC or NT clip subject	% Correct
ASC	self	ASC	35.6%
ASC	peer	ASC	40.9%
ASC	familiar teacher	NT	45.6%
ASC	unfamiliar teacher	NT	46.9%
ASC	**OVERALL**	-	**42.0%**
NT	self	ASC	67.4%
NT	peer	ASC	68.3%
NT	familiar teacher	NT	71.5%
NT	unfamiliar teacher	NT	75.7%
NT	**OVERALL**	-	**70.2%**

9.4 Intervention

9.4.1 Intervention Schedule

Period	Week #s	Emotions/week	First Week	Second Week	Third Week	Fourth Week
1	1-4	1	Happy	Thinking	Concentrating	Interested
2	5-8	2	Concentrating/Thinking	Thinking/Interested	Happy/Concentrating	Happy/Interested
3	9-12	4	All emotions	All emotions	All emotions	All emotions

9.4.2 Emotion Descriptions for ASC and NT Participants

Color	Emotion	Description
Yellow	HAPPY	Feeling good about something; to be pleased. *Expressing a positive emotion - may be smiling or laughing, with corners of the mouth turned upwards.*
Blue	CONCENTRATING	Focusing all your attention on a particular thing and not letting yourself be distracted. *Outwardly engaged with some particular stimulus and ignoring distractions; someone concentrating might lean forward and/or look directly at something in front of them, and narrow the eyes slightly.*
Red	THINKING	To work something out in your mind. *Considering something internally and withdrawing from outside stimuli in order to think about this thing. Someone thinking might avert their eyes, sit back or tilt their head, and/or purse their lips.*
Green	INTERESTED	Wanting to know about something. *Someone interested might lean forward, look directly forward or at a number of interesting things, and/or widen their eyes and lift their eyebrows.*

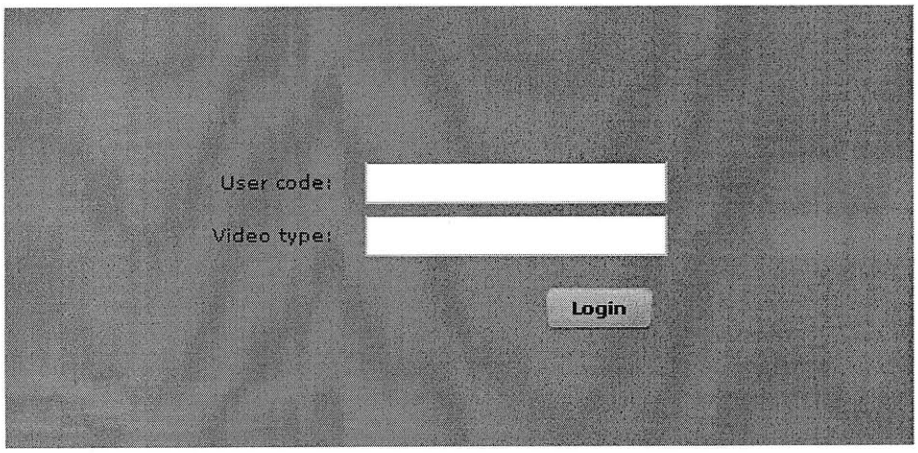
9.4.3 Recording Session Instructions (example for Week 01, Happy)

iSET Intervention – Recording Session Instructions – Week 01 (Happy)

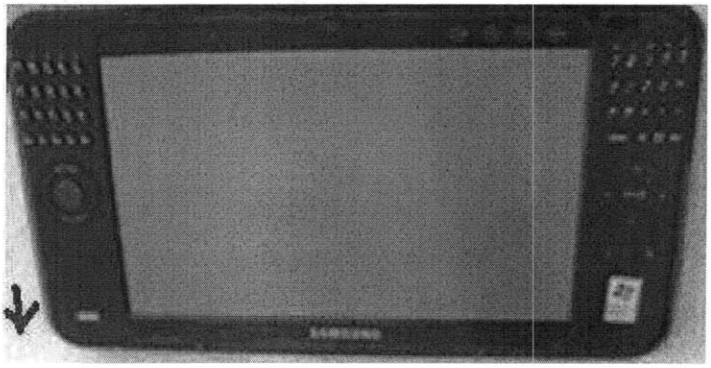
Recording Session Instructions (2 times weekly, 1st and 3rd sessions: do Examples for 5-10 minutes and Recording for 15-20 minutes)

- Please defer or close any pop-ups that appear on the Samsungs or laptops regarding VirusScan, automatic updates, etc, by clicking "Defer" or the X in the top-right corner of the window.
- Sign out the laptop and the Samsung at the location where they are kept. In the binder where you sign them out, make sure to mention time of sign-out and the staff who is signing out the equipment.
- Plug the laptop into the power cord and the ethernet cable (blue cable).

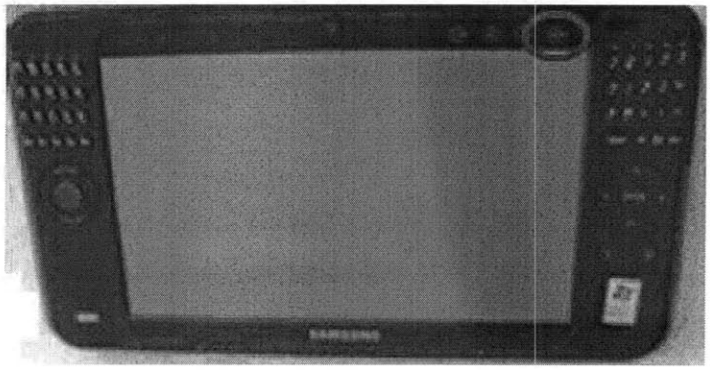
- Fill out log sheet in the binder regarding the times at which the session starts.
- Have the emotion descriptions and color associations available for use by student and teacher.
- On the velcroed list, remove any emotions that are not being worked on that week. The teacher should always have access to the full list of emotions.
- Review the scheduled emotion(s) with the student, discussing what the emotion looks like and when it occurs.
- During use of VidL, all mouse clicking should be performed by teachers. Students will verbally or gesturally indicate which emotion the teacher should click.
- On the laptop, open Firefox by double clicking the icon on the desktop.
- 

- The homepage should be the VidL log in screen. If not, go to http://tinyurl.com/VidL-iSET. This is also bookmarked in Firefox.
- Put Firefox into full-screen mode by selecting the "view" menu and then selecting "full-screen".
- 

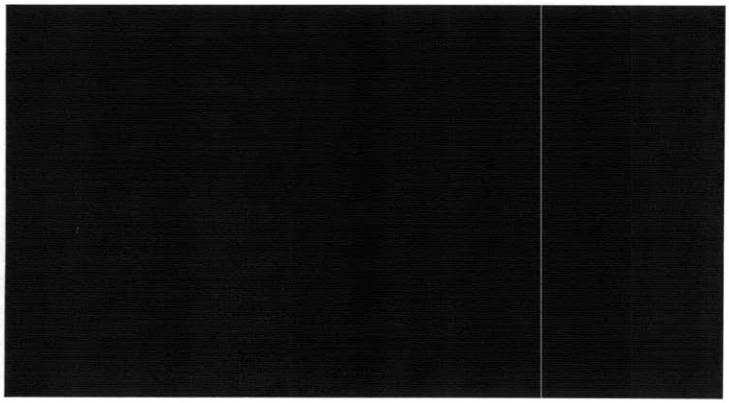
- Viewing Examples
  - Login as:
    - *User code:* asdXX
    - *Video type:* weekYY-examples
  - ...where XX is the student's user code (04-24) and YY is the current week of the intervention (01- 12). All codes are listed in binders and in "Participant IDs" folder on the computer desktop. Click "next video." Label the video by clicking the emotion button, then click "next video" again. Label example videos for 5-10 minutes.
  - Note on the log sheet what time the VidL part of the session ends and what time the Samsung part of the session starts.
- Recording on the Samsung
- Turn on Samsung. The power button is located on the lower left side of the Samsung. Slide the switch down and release to power on the laptop.

- 

  - If needed, adjust the brightness of the screen by pushing the menu button at the top right corner of the screen. On the second listed item, push the up arrow (button furthest to the right) repeatedly until the screen is sufficiently lit.

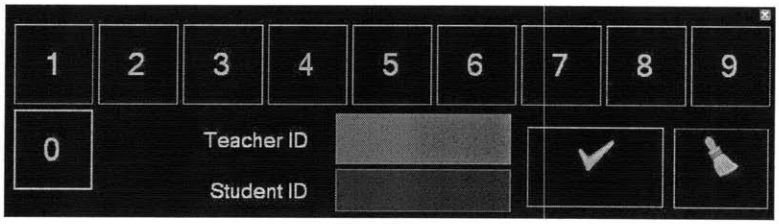
- 

- Slide out stylus from the bottom of the Samsung.

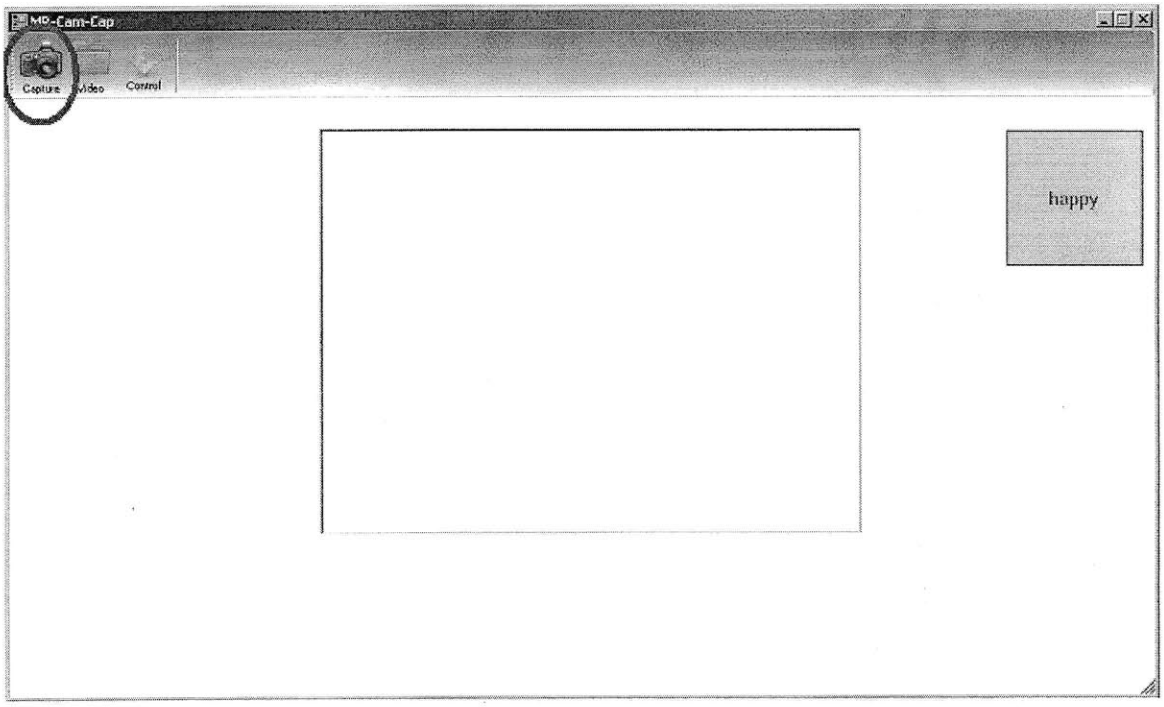
- 

- Double click the icon labeled "click me" in the middle of the screen.

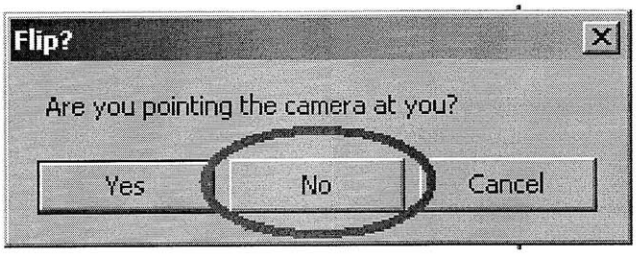
- Log in with the teacher's ID code and student ID code, provided in your binder of instructions. Click the green checkmark when finished.

- 

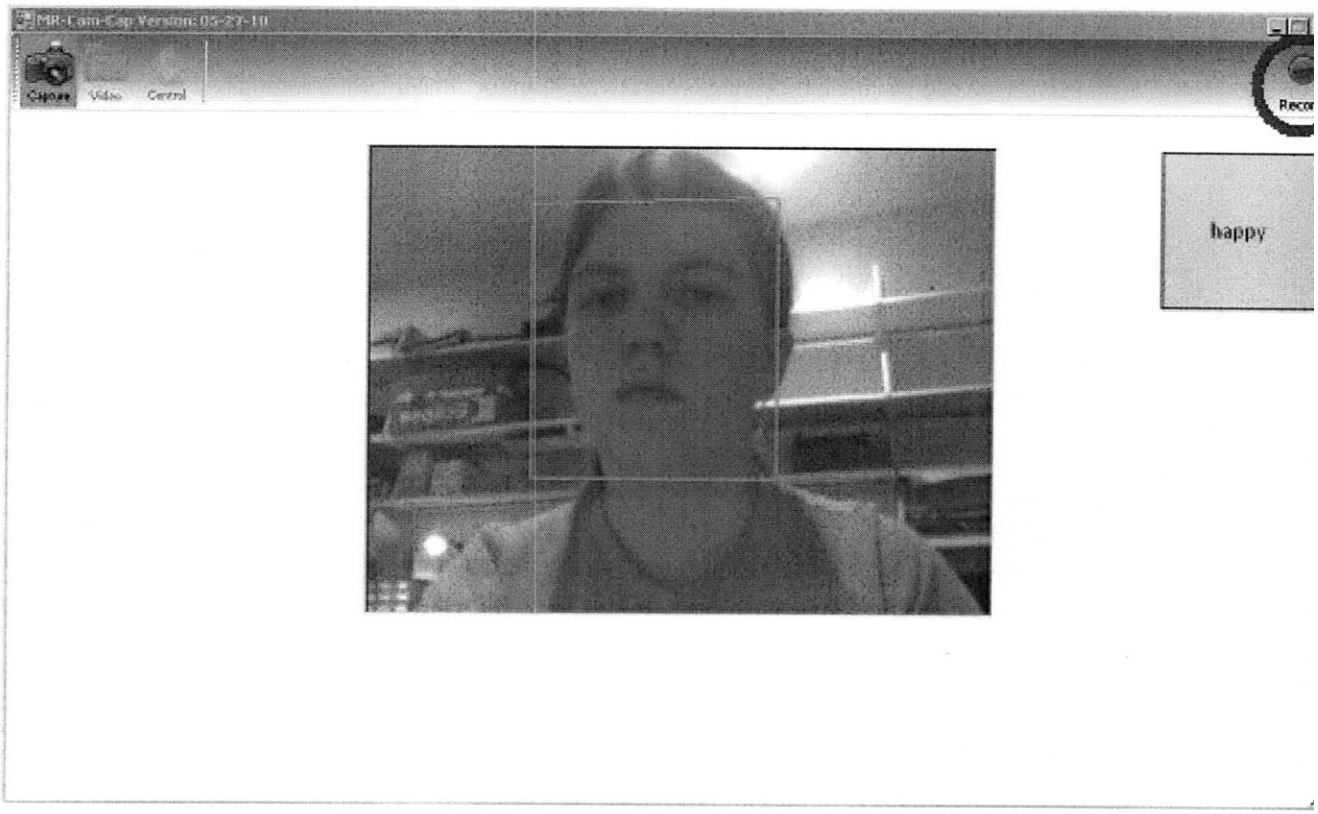
- When the program first starts, press Capture button to begin the capture process.



- Then press "No" on the pop-up that asks whether the camera is pointed toward the holder of the Samsung.



- If needed, review the scheduled emotion(s) with the student, discussing what the emotion looks like and when it occurs.
- Push the record button and play the "emotion hunt game" for 15-20 minutes. Find people who are showing the emotion(s) that you are working on that week. Make sure to hold the camera steady while recording. Use hand-over-hand guidance, if necessary. Assist the student to click on the proper emotion label. To optimize recordings, make sure subjects are as clear as possible: use good light, and avoid hats, glasses, and glare.



- When a face is selected in the video, the video display will go from black and white to color and the face will have a box around it. When the student is tagging, try to have the video be in color and a face be clearly visible.

- Stop filming by clicking "stop recording".
- Note on the log sheet what time the recording session ends.

- How to take good clips
  - push the buttons in the following order ONLY: Capture, Record, Stop recording
- How to ruin clips (avoid these!):
  - Pressing "record" before the video screen is fully loaded.
  - Pressing "record" before "capture" and not exiting out to try again.
  - Exiting while a clip is being recorded.
- Samsung Suggestions for Use
  - Stay next to the student
  - Provide prompts and guidance in use of the Samsung
  - Remind student to slow down and not keep clicking the screen, let the software respond to clicks. Cue the student to not rush through the program (i.e., say "slow down")
  - Prompt student to look at the faces, eyes, etc. for information about the emotion
  - Discuss each emotion, relate to real life scenarios, ask questions about emotional experiences, make connections about what learned with software and real life
  - When there is more than one emotion scheduled to review, talk about how the basic emotions are different from each other in how the eyes and mouth look, the head moves, and the situations and contexts in which it occurs
  - Use VidL to review videos for 5-10 minutes, and then play the emotion hunt game (recording videos) for 15-20 minutes each session.

9.4.4 Recording Session Daily Log

iSET Recording Session Daily Log Sheet

Please fill out a log sheet for each facial recognition session.
Emotion(s): _____
Week: _____, Session: _____
Date: _____
Participant Name: _____
Staff Name: _____
Session Start Time: _____, End Time: _____
Please check each activity that is completed and note start/end times:
___ *Example Videos:* Start Time: _____, End Time: _____
___ *Emotion Hunt Game:* Start Time: _____, End Time: _____
Please write any notes about the session below. For example, did anything unusual happen during or before the session? Did the participant say anything interesting about the emotions being worked on? Any other comments? This may be a nice place to take notes about emotion topics you discussed with your student that you can build on in later sessions.

9.4.5 Recording Session Fidelity Check

Observe the student for the duration of the intervention session. Answer the questions below. Gather software summary information at end.

Student ID: _____ Teacher ID: _____ Observer ID: _____ Date: _____ Week: _____ Session: _____
1. Did the adult sit/stand/walk next to the student during the session?
2. Did the adult provide prompts to slow down, listen carefully, watch and to help navigate through the program? Describe:
3. Did the adult and the student review example videos in VidL before recording on the Samsung?
4. Did the student use the emotion hunt game on the Samsung?

5. Did the adult and student discuss emotions and how they look and sound? Describe:

6. Did the adult and student discuss real life application of the emotions or process previous events and emotions involved? Describe:

7. Did the student label example videos for a minimum of 5 minutes?

8. Did the student record with the Samsung for a minimum of 15 minutes?

Other Comments:

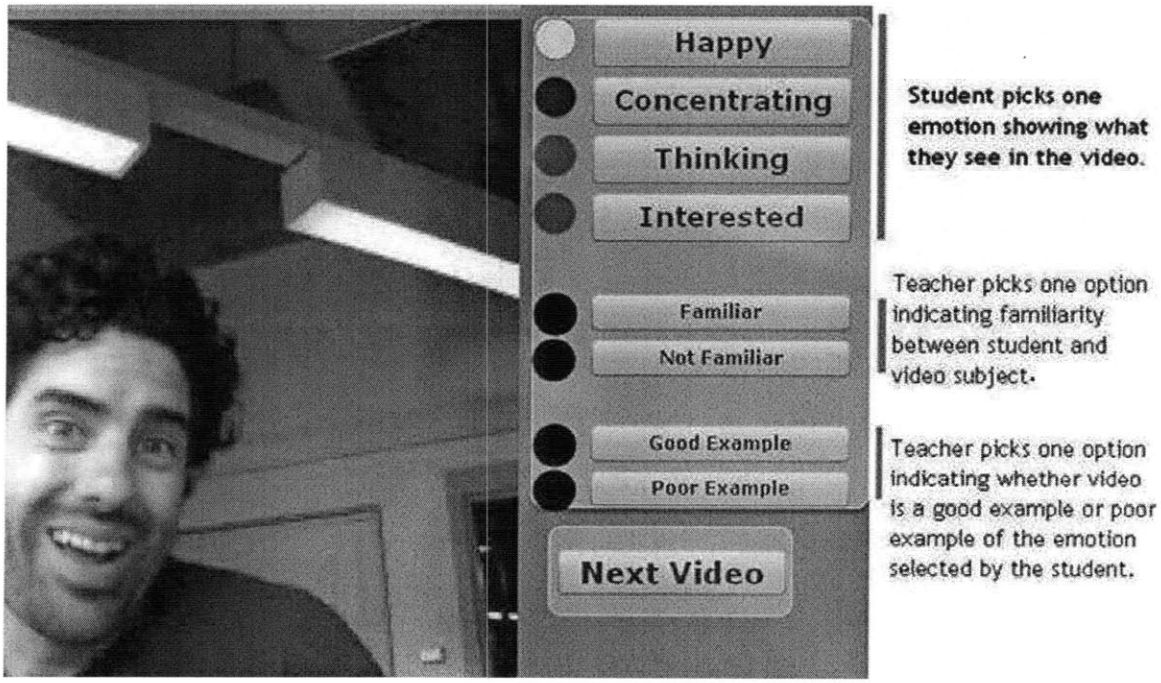
9.4.6 Viewing Session Instructions (example for Week 01, Happy)

iSET Intervention – Viewing Session Instructions – Weeks 01-08

Viewing Session Instructions (1-2 times weekly, 2nd and 4th sessions: do Examples for 5-10 minutes, Preapproved for 5-10 minutes, and Student-Recorded for 10-15 minutes)

- Please defer or close any pop-ups that appear on the Samsungs or laptops regarding VirusScan, automatic updates, etc, by clicking "Defer" or the X in the top-right corner of the window.
- Sign out the laptop and the Samsung at the location where they are kept. In the binder where you sign them out, make sure to mention time of sign-out and the staff who is signing out the equipment.
- Plug the laptop into the power cord and the ethernet cable (blue cable). Fill out log sheet in the binder regarding the times at which the session starts.
- Have the emotion descriptions and color associations available for use by student and teacher. On the velcroed list, remove any emotions that are not being worked on that week. The teacher should always have access to the full list of emotions.
- Review the scheduled emotion(s) with the student, discussing what the emotion looks like and when it occurs.
- During use of VidL, all mouse clicking should be performed by teachers. Students will verbally or gesturally indicate which emotion the teacher should click. (On student-recorded video, teachers will choose good example vs. poor example and familiar versus not familiar - i.e., a total stranger.)
- On the laptop, open Firefox by double clicking the icon on the desktop.
- During weeks 01-08 of the intervention, go to: http://tinyurl.com/VidL-iSET
- This is also bookmarked in Firefox. Put Firefox into full-screen mode by selecting the "view" menu and then selecting "full-screen". If the menu bar disappears, drag the cursor to the top of the screen
- Viewing Example Videos:
  - Note on the log sheet what time viewing examples begins and ends.
  - Login as:
    - User code: asdXX
    - Video type: weekYY-examples ... where XX is the student's user code (04-24) and YY is the current week of the intervention.
  - Click "next video". Label the video by clicking the emotion button, then click "next video" again. Label example videos for 5-10 minutes.
- Viewing Pre-Approved Videos:
  - Note on the log sheet what time viewing pre-approved videos begins and ends.
  - Refresh the VidL website on the menu bar, near the web address.
  - Log back in as:
    - *User code:* asdXX
    - *Video type:* weekYY-preapproved

- ... where XX is the student's user code (04-24) and YY is the current week of the intervention (01-12).
- Click "next video". Label the video by clicking the emotion button, then click "next video" again.
- Label preapproved videos for 5-10 minutes.

- Viewing Student-Recorded Videos:
  - Note on the log sheet what time viewing pre-approved videos begins and ends.
  - Refresh the VidL website on the menu bar, near the web address.
  - Log back in as:
    - *User code:* asdXX
    - *Video type:* weekYY-student-recorded
  - ... where XX is the student's user code (04-24) and YY is the current week of the intervention.
  - Click "next video". Label the video by clicking the emotion button, then click "next video" again.
  - Label student-recorded videos for 10-15 minutes, or until there are no more videos to label.
  - For each student-recorded video watched, the teacher should indicate level of familiarity of the subject ("familiar" or "not familiar" – use "not familiar" if the video subject is a total stranger to the student and "familiar" otherwise) and whether the video is a good example of the emotion shown ("good example" or "poor example" – choose "good example" if the video is a strong enough example of the emotion to be shown again, and "poor example" otherwise - if the video is blurry, unclear, or not representative of the emotion in question.)
  - 

  - Criteria for labeling a video "Poor Example":
    - 1. When more than one person in the video for more than half of the video clip.
    - 2. When the face-front view is not present for more than half of the video clip.
    - 3. When the emotion must be inferred from the activity, NOT from the face (e.g. twirling around).
    - 4. When the majority of the video is too blurry to see anything.
  - Suggestions for use during VidL labeling:
    - Sit next to the student
    - Provide prompts and guidance in use of the software

* Remind student to slow down and not keep clicking the screen; let the software respond to clicks. Cue the student to not rush through the program (i.e., "slow down")
* Prompt student to look at the faces, eyes, etc. for information about the emotion
* Direct student to use each part of the software (preapproved videos, students, videos, etc.) and follow the schedule
* Discuss each emotion, relate to real life scenarios, ask questions about emotional experiences, make connections about what learned with software and real life
* When there is more than one emotion scheduled to review, talk about how the basic emotions are different from each other in how the eyes and mouth look, the head moves, and the situations and contexts in which it occurs If you finish early, you can watch more "preapproved" videos or "example" videos, or you can further discuss the emotion(s) you're working on that week.

9.4.7 Viewing Session Daily Log

iSET Viewing Session Daily Log Sheet

Please fill out a log sheet for each facial recognition session.

Emotion(s): _____

Week: _____, Session: _____

Date: _____

Participant Name: _____

Staff Name: _____

Session Start Time: _____, End Time: _____

Please check each activity that is completed and note start/end times:

___ *Example Videos:* Start Time: _____, End Time: _____

___ *Pre-Approved Videos:* Start Time: _____, End Time: _____

___ *Student-Recorded Videos:* Start Time: _____, End Time: _____

Please write any notes about the session below. For example, did anything unusual happen during or before the session? Did the participant say anything interesting about the emotions being worked on? Any other comments? This may be a nice place to take notes about emotion topics you discussed with your student that you can build on in later sessions.

9.4.8 Viewing Session Fidelity Check

Observe the student for the duration of the intervention session. Answer the questions below. Gather software summary information at end.

Student ID: _____ Teacher ID: _____ Observer ID: _____ Date: _____ Week: _____ Session: _____

1. Did the adult sit next to the student during the session?
2. Did the adult provide prompts to slow down, listen carefully, & watch, and did the adult help navigate through the web interface? Describe:
3. Did the student label example videos?
4. Did the student label preapproved video?
5. Did the student label their own recorded video?
6. Did the adult and student discuss emotions and how they look and sound? Describe:
7. Did the adult and student discuss real life application of the emotions or process previous events and emotions involved? Describe:
8. Did the student label example videos for at least 5 minutes?
9. Did the student label preapproved videos for at least 5 minutes?
10. Did the student label their own video for at least 10 minutes? Other Comments:

9.4.9 Mind Reading DVD Session Instructions (example for Week 01, Happy)

(Images omitted for brevity.)

iSET Intervention - Mind Reading Session Instructions – Week 01 (Happy)

Mind Reading DVD Daily Schedule (3-4 times weekly: use for at least 20 minutes per session, looking at Rewards for no more than 5 minutes)

General suggestions for use of Mind Reading:

- Sit next to the participant
- Review the emotion(s) you are working on this week
- Provide prompts and guidance in the use of the software
- Remind participant to slow down and not keep clicking the mouse, let the software respond to mouse clicks – no rushing through the program
- Prompt the participant to look at the faces, eyes, etc. for information about the emotion
- Direct the participant to use each part of the software (emotions library, learning center, games, etc.) and follow the schedule
- Discuss each emotion, relate to real life scenarios, ask questions about emotional experiences, make connections about what learned with software and real life
- Talk about how the basic emotions are different from each other – in how the eyes look, mouth, movement, situations, contexts, etc.
- Try to use the program 3 to 4 times per week for at least 20 minutes per session
- Make sure the student only uses the Rewards sections for about 5 minutes per session
- Please defer or close any pop-ups that appear regarding VirusScan, automatic updates, etc.
- Mute sound on the laptop.

- Open the Mind Reading DVD application on the computer desktop.
- Click the screen to skip the opening animations. Choose the "Learning Center".
- Log in by selecting the name of the participant.
- For "Your Level", select 1.
- For "Your Helper", select "None".
- Click the "Start" button.
- Go into the "Rewards" section.
- The client can view currently earned rewards by clicking on them.
- Have clients choose rewards to be earned during the session: click "Collect New Reward" to choose a reward type. (Remember that total time spent in the Rewards sections should be no more than 5 minutes per session).
- On the "Reward Selection" page, click the type of reward that the client chooses.
- Click "Collect" to select that type of reward.
- Click "Back" to go back to the Learning Center.
- Select "Emotion Groups."
- Go to the second page in order to find "Happy."
- Select "Happy Group".

- Click on the face to play the video and read the text.
- Click the forward arrow.
- Select "Happy Situation".
- Watch the video which plays automatically, then click the forward arrow. (To rewatch the video, click the image again.)
- Select "Happy Group".
- Click on a face.
- Watch the video which plays automatically. Note: you can rewatch a video by clicking on the face again. Discuss the videos with the client. When you have watched and discussed the video with the client, click back arrow to return to the previous screen and pick another face. Repeat these steps until you have watched and discussed all videos.
- Click on the stories icon in the bottom left, then click on each number sequentially (1-6) and read each story that appears at the bottom of the screen aloud. Discuss each short story.
- When you're done reading the stories and looking at the videos, return to the main Learning Center area by clicking back three times.
- Click on "Lessons".
- Select "Build Your Own".
- Click the arrow next to "Afraid Group" to show the emotion group list. On the emotion group list, select Happy.
- Select "Happy" from the emotion list and click the Add button. This emotion should only be listed one time. Remove any other emotions other than the one you are focusing on. Click this emotion listed on the right side, then click the "remove" button.
- Click the "Start" button.
- Watch the video which plays automatically and read the text aloud. To watch the video again, click on the face. Discuss the text, then click the forward arrow.
- Select each of the faces and watch the video and read the accompanying story aloud.
- Then choose the back arrow to return to this screen and select the next face.
- When you have finished looking at all the faces (they will all be grayed out), click the forward arrow.
- Skip the following screen by clicking the forward arrow. Do not answer any questions!
- Watch the videos by clicking on the faces. The student should watch both videos. Signal the choice by clicking on the numbers for the client. (Click on the faces to replay the videos.)
- After the correct answer is chosen, click the forward arrow.
- When the lesson is finished, return to the Learning Center by clicking Back twice.
- Go into the Rewards section and explore the client's rewards with the clients. (Remember that total time spent in the Rewards sections should be no more than 5 minutes of a 20-25 minute Mind Reading session.) Unmute the computer and look at the rewards that were earned during the session. If you finish early, you can repeat the "Lessons" section to win more reinforcers or you can further discuss the emotion(s) you're working on that week.

9.4.10 Mind Reading DVD Session Daily Log

Mind Reading Session Daily Log Sheet

Please fill out a log sheet for each facial recognition session.

Emotion(s): _____

Week: _____, Session: _____

Date: _____

Participant Name: _____

Staff Name: _____

Session Start Time: _____, End Time: _____

Please check each activity that is completed and note start/end times:

___ *Learning Center Emotion Groups:* Start Time: _____, End Time: _____

___ *Learning Center Lessons:* Start Time: _____, End Time: _____

___ *Rewards Zone:* Start Time: _____, End Time: _____

Please write any notes about the session below. For example, did anything unusual happen during or before the session? Did the participant say anything interesting about the emotions being worked on? Any other comments? This may be a nice place to take notes about emotion topics you discussed with your student that you can build on in later sessions.

9.4.11 Mind Reading DVD Session Fidelity Check

Observe the student for the duration of the intervention session. Answer the questions below. Gather software summary information at end.

Student ID: _____ Teacher ID: _____ Observer ID: _____ Date: _____ Week: _____ Session: _____

1. Did the adult sit next to the student during the session?
2. Did the adult provide prompts to slow down, listen carefully, watch and to help navigate through the program? Describe:
3. Did the adult direct the student to use the different parts of the software? Describe:
4. Did the student use the Learning Center Emotion Groups?
5. Did the student use the Learning Center Lessons?
6. Did the student spend less than 33% of the total time in Rewards?
7. Did the adult and student discuss emotions and how they look and sound? Describe:
8. Did the adult and student discuss real life application of the emotions or process previous events and emotions involved? Describe:
9. Did the student use the Emotion Groups for a minimum of 7 minutes?
10. Did the student use the Lessons for a minimum of 7 minutes?
11. Did the student use the software for a minimum of 20 minutes? Other Comments:

9.4.12 Percentage of Videos Labeled by Teachers as "Good Examples"

Week #	Videos Labeled as "Good Examples"	Total Number of Videos	Percentage Labeled as "Good Examples"
1	316	1408	22.44%
2	144	850	16.94%
3	276	1041	26.51%
4	132	555	23.78%
5	174	684	25.44%
6	144	522	27.59%
7	57	198	28.79%
8	18	60	30.00%
9	33	105	31.43%

9.4.13 Increased Competence in Tracking

Week #	Percentage Clips Tracked
1	76.9%
2	57.3%
3	70.2%
4	74.4%
5	72.9%
6	73.9%
7	86.7%
8	93.3%

References

[1] Ashwin, E., Ashwin, C., Rhydderch, D., Howells, J., and Baron- Cohen, S. (2008). Eagle-Eyed Visual Acuity: An experimental investigation of enhanced perception in autism. Biol. Psychiatry 65, 17–21.

[2] Mottron, L., Dawson, M., Soulieres, I., Hubert, B. & Burack, J. Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. J. Autism Dev. Disord. 36, 27–43 (2006).

[3] Baron-Cohen, S. (2006). "The hyper-systemizing, assortative mating theory of autism." Progress in Neuro-Psychopharmacology & Biological Psychiatry 30: 865–872.

[4] Baron-Cohen S., Ashwin E., Ashwin C., Tavassoli T., Chakrabarti B. Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. Phil. Trans. R. Soc. B. 2009;364:1377–1383. doi:10.1098/rstb.2008.0337

[5] American Psychiatric Association. (2000). Pervasive developmental disorders. Diagnostic and statistical manual of mental disorders, DSM-IV-TR, (pp. 69-70). Washington, D.C.

[6] Baron-Cohen, S. (1995). Mindblindness, MIT Press Cambridge, Mass.

[7] Frith, U. (2003). Autism: explaining the enigma. (2nd ed.), . Oxford, Blackwell Publishing.

[8] Hobson, R. P. (1993). Autism and the development of mind. Hove, HK, Lawrence Erlbaum.

[9] Joseph, R. and H. Tager-Flusberg (1997). "An investigation of attention and affect in children with autism and Down syndrome." Journal of Autism and Developmental Disorders 27(4): 385-396.

[10] Klin, A., W. Jones, et al. (2002). "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism." Archives of General Psychiatry 59: 809-816.

[11] Pelphrey, K. A., J. P. Morris, et al. (2005). "Neural basis of eye gaze processing deficits in autism " Brain 128(5): 1038-1048.

[12] Volkmar, F. R. and L. Mayes (1991). "Gaze Behaviour in Autism." Development and Psychopathology 2: 61-69.

[13] Hill, E. L., S. Berthoz, et al. (2004). "Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives." Journal of Autism and Developmental Disorders 34: 229-235.

[14] Klin, A. and F. R. Volkmar (1995). "Asperger's Syndrome: Guidelines for Assesment and Diagnosis " Learning Disabilities Association of America.

[15] Wellman, H. M. (1992). The child's theory of mind, MIT Press Cambridge, Mass.

[16] Baron-Cohen S, Golan O, Wheelwright S, Hill J (2004) Mind-reading: the interactive guide to emotions. Jessica Kingsley, London (http://www.jkp.com/mindreading)

[17] Fletcher, R., Dobson, K., Goodwin, M.S., Eydgahi, H., Wilder-Smith, O., Fernholz, D., Kuboyama, Y., Hedman, E., Poh, M.Z., Picard, R.W., "iCalm: Wearable Sensor and Network Architecture for Wirelessly Communicating and Logging Autonomic Activity," IEEE Transactions on Information Technology in Biomedicine, vol.14, no.2, pp.215-223, March 2010.

[18] R.R. Morris, "Managing Sound Sensitivity in Autism Spectrum Disorder: New Technologies for Customized Intervention," Master's Thesis, Massachusetts Institute of Technology, 2009

[19] Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., Wheelwright, S. The Systemising Quotient: An investigation of adults with Asperger Syndrome or High Functioning Autism and normal sex differences. Philosophical Transactions of the Royal Society 2003, Series B, 358, 361–374.

[20] Golan, O., Baron-Cohen, S. Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. Development and Psychopathology (2006), 18:2:591-617 Cambridge University Press

[21] LaCava, P. G., Golan, O., Baron-Cohen, S., & Myles, B. S. (2007). Using assistive technology to teach emotion recognition to students with Asperger Syndrome:A pilot study. Remedial and Special Education, 28, 174-181.

[22] Teeters, A., El Kaliouby, R. and Picard, R.W. "Self-Cam: Feedback From What Would be your Social Partner," ACM SIGGRAPH, 2006 Research Posters, p. 138, July 30-Aug. 3, 2006, Boston, MA.

[23] Teeters, A. "Use of a Wearable Camera System in Conversation: Toward a Companion Tool for Social-Emotional Learning in Autism," MIT MS Thesis, September 2007.

[24] M. Eckhardt, M. Goodwin, and R.W. Picard, "Eyes Up: Interactive Tangible-Digital Puzzles for Learning about Eyes", Extended Abstract of IMFAR 2010, Philadelphia, PA, USA, May 20-22, 2010.

[25] Johnson, R. Boardmaker: The P.C.S. libraries on disk. Solana Beach, CA, Mayer Johnson, 1994.

[26] Wolf, J. M., Tanaka, J. W., Klaiman, C., Cockburn, J., Herlihy, L., Brown, C., South, M., McPartland, J., Kaiser, M. D., Phillips, R. and Schultz, R. T. (2008), Specific impairment of face-processing abilities in children with autism spectrum disorder using the Let's Face It! skills battery. Autism Research, 1: 329–340. doi: 10.1002/aur.56

[27] Gray, C. A. (1998). Social stories and comic strip conversations with students with Asperger syndrome and high-functioning autism. In E. Schopler & G. B. Mesibov (Eds.), Asperger syndrome or high-functioning autism? Current issues in autism (pp. 167–198). New York, NY: Plenum Press.

[28] Whalen, C., Moss, D., Ilan, A., Vaupel, M., Fielding, P., Macdonald, K., Cernich, S., Symon, J. Efficacy of TeachTown: Basics computer-assisted intervention for the Intensive Comprehensive Autism Program in Los Angeles Unified School District Autism May 2010 14: 179-197, first published on May 18, 2010 doi:10.1177/1362361310363282

[29] Golan, O., Baron-Cohen, S., Ashwin, E., et al. (2009). Enhancing emotion recognition in children with autism spectrum conditions: an intervention using animated vehicles with real emotional faces. J. Autism Dev. Dis.

[30] Asperger, H. (1991). "Autistic psychopathy" in childhood. In Frith, U. (Ed.), Autism and Asperger syndrome (pp. 37-92). Cambridge, U.K.: Cambridge University Press.

[31] el Kaliouby, R.: Mind-reading Machines: the automated inference of complex mental states from video. PhD thesis, University of Cambridge (2005)

[32] Baron-Cohen, S., S. Wheelwright, et al. (2002). The exact mind: empathising and systemising in autism spectrum conditions, Blackwell.

[33] Jolliffe, T. and S. Baron-Cohen (1997). "Are people with autism and Asperger syndrome faster than normal on the Embedded Figures Test." Journal of Child Psychology and Psychiatry 38(5): 527–534.

[34] Shah, A. and U. Frith (1983). "An islet of ability in autistic children: a research note." J Child Psychol Psychiatry 24(4): 613-20.

[35] Attwood, T. (1998). Asperger's Syndrome: A Guide for Parents and Professionals, Philadelphia: Jessica Kingsley Publishers.

[36] Adolphs, R. (2001). "The neurobiology of social cognition." Current Opinion in Neurobiology 11: 231-239.

[37] Frith, U., Ed. (1991). Autism and Apserger Syndrome, Cambridge University Press.

[38] Blackburn, J., K. Gottschewski, et al. (2000). A discussion about Theory of Mind: From an Autistic Perspective. Proceedings of Autism Europe's 6th International Congress, Glasgow.

[39] Grossman, J. B., A. Klin, et al. (2000). "Verbal bias in recognition of facial emotions in children with Asperger syndrome." Journal of Child Psychology and Psychiatry and Allied Disciplines 41: 369-379.

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

[41] Kanner, L. (1943). Autistic disturbances of affective contact. Nervous Child, 2, 217-250.

[42] M. Eckhardt, and R.W. Picard, "A More Effective Way to Label Affective Expressions," International Conference on Affective Computing and Intelligent Interaction (Demo Session), 2009.

[43] Madsen, M., el Kaliouby, R., Goodwin, M., and Picard, R.W., "Technology for Just-In-Time In-Situ Learning of Facial Affect for Persons Diagnosed with an Autism Spectrum Disorder," Proceedings of the 10th ACM Conference on Computers and Accessibility (ASSETS), October 13-15, 2008, Halifax, Canada.

[44] M. Madsen, R. el Kaliouby, M. Eckhardt, M. Hoque, M. Goodwin, and R.W. Picard, "Lessons from Participatory Design with Adolescents on the Autism Spectrum", Proc. Computer Human Interaction, CHI 2009.

[45] Kaufman, A. S., & Kaufman, N. L. (1990). Kaufman Brief Intelligence Test. Circle Pines, MN: American Guidance Service.

[46] Bagby RM, Parker JDA, Taylor GJ. The twenty-item Toronto Alexithymia Scale-I Item selection and cross validation of the factor structure. J Psychosom Res 1994; 38: 23-32.

[47] Bagby RM, Taylor GJ, Parker JDA. The twenty-item Toronto Alexithymia Scale: II. Convergent, discriminant, and concurrent validity. J Psychosom Res 1994; 38:33-40.

[48] Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). Autism Diagnostic Observation Schedule-WPS Edition. Los Angeles, CA: Western Psychological Services.

[49] Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The 'Reading the Mind in the Eyes' Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. J Child Psychol Psychiatry 2001; 42: 241–51.

[50] Constantino JN. The Social Responsiveness Scale. Los Angeles, Calif: Western Psychological Services; 2002.

[51] Benton, A. (1980) The neuropsychology of facial recognition. Am. Psychol. 35, 176–186.

[52] Ref: Madsen, M., Mahmoud, A., Kashef, Y. "iSET: enabling in situ & post hoc video labeling." Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS), Pittsburgh, Pennsylvania, USA, October 26-28, 2009.

[53] Dawson et al., 2007 M. Dawson, I. Soulieres, M.A. Gernsbacher and L. Mottron, The level and nature of autistic intelligence, Psychological Science 18 (2007), pp. 657–662.

[54] Chakrabarti, B., Bullmore, E., Baron-Cohen, s. 'Empathizing with basic emotions: Common and discrete neural substrates'. Social Neuroscience, 1747-0927, Volume 1, Issue 3, 2006, Pages 364 – 384

[55] Ekman P, FriesenW, Hager J. Facial action coding system. Salt Lake City, Utah: Research Nexus; 2002.

[56] "FaceSay: Social Skills Games That Work!" <http://www.facesay.com/>