Attention Modification to Attenuate Facial Emotion Recognition Deficits in Children with ASD

Andrea Trubanova Wieckowski

Dissertation submitted to the faculty of Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctorate of Philosophy

in

Psychology

Susan W. White, Ph.D., Committee Chair

Thomas H. Ollendick, Ph.D., Committee Member

Robin K. Panneton, Ph.D., Committee Member

John A. Richey, Ph.D., Committee Member

October 29, 2018

Blacksburg, VA

Keywords: Autism Spectrum Disorder; Facial Emotion Recognition; Eye-Tracking

Attention Modification to Attenuate Facial Emotion Recognition Deficits in Children with ASD

Andrea Trubanova Wieckowski

ABSTRACT

Prior studies have identified diminished attending to faces, and in particular the eye region, in individuals with Autism Spectrum Disorder (ASD), which may contribute to the impairments they experience with emotion recognition and expression. The current study evaluated the acceptability, feasibility, and preliminary effectiveness of an attention modification intervention designed to attenuate deficits in facial emotion recognition and expression in children with ASD. During the 10-session experimental treatment, children watched videos of people expressing different emotions with the facial features highlighted to guide children's attention. Eight children with ASD completed the treatment, of nine who began. On average, the children and their parents rated the treatment to be acceptable and helpful. Although treatment efficacy, in terms of improved facial emotion recognition (FER), was not apparent on task-based measures, children and their parents reported slight improvements and most parents indicated decreased socioemotional problems following treatment. Results of this preliminary trial suggest that further clinical research on visual attention retraining for ASD, within an experimental therapeutic program, may be promising.

GENERAL AUDIENCE ABSTRACT

Previous studies have shown that individuals with Autism Spectrum Disorder (ASD) show lower looking at faces, especially the eyes, which may lead to the difficulties they show with ability to recognize other's emotions and express their own emotions. This study looked at a new treatment designed to decrease the difficulties in emotion recognition and expression in children with ASD. The study looked at whether the treatment is possible, acceptable to children and their parents, and successful in decreasing the difficulty with emotion recognition. During the 10-session treatment, children watched videos of people making different expressions. The faces of the actors in the videos were highlighted to show the children the important area to look at. Eight children with ASD completed the treatment, of nine who started the treatment. On average, the children and their parents said that the treatment is acceptable and helpful. While the treatment was not successful in improving ability to recognize emotions on other's faces on several tasks, children and their parents reported slight improvements. In addition, most parents reported less problems with social skills and emotion recognition and expression after the treatment. These results suggest that more clinical research may be needed to evaluate usefulness of such attention retraining for children with ASD.

Acknowledgements

I am incredibly grateful to many supportive individuals during my doctoral studies. First, I would like to thank my doctoral advisor and committee chair, Dr. Susan White, who has provided invaluable mentorship, support, and guidance over the past five years. I would also like to thank my dissertation committee, Drs. Thomas Ollendick, Robin Panneton, and John Richey for their support and guidance during my studies. Your support and encouragement have been truly invaluable.

I would also like to thank my family and friends for their support during graduate school and this dissertation project. In particular, my parents, cohort mates, and my husband have provided incredible support and understanding throughout my graduate schooling. They have helped me every step of the way by providing unwavering encouragement.

My dissertation project would not have been possible without funding from Organization for Autism Research Graduate Research Grant and Routh Research and Dissertation Grant from Society of Clinical Child and Adolescent Psychology, APA Division 53. I would also like to thank members of the Psychosocial Interventions Lab as well as wonderful undergraduate research assistants, Megan Pickford and Rachel Harris, for their assistance with recruitment, data collection, and data entry.

Table of Contents

List of Tables

List of Figures

Attention Modification to Attenuate Facial Emotion Recognition Deficits in Children with ASD

Introduction

Recent studies have provided evidence showing similar social orienting between young infants with and without autism (e.g., Jones & Klin, 2013), but yet a difference in engagement in social interactions, atypical fixation patterns to the face, and poor recognition of facial information (e.g., facial expressions) later on in life. The developmental cascade of worsening social behavior, including deficits in facial emotion recognition seen in individuals with ASD, start with early deficits in social attention.

Recognition of nonverbally expressed emotion is fundamental for successful social communication and interaction (Ekman, 1992; Nuske, Vivanti, & Dissanayake, 2013). Emotion recognition emerges early in life, with newborns being able to discriminate happy, sad, and surprised facial expressions (Field, Woodson, Greenberg, & Cohen, 1982). In addition, by the preschool years, children can accurately verbally label most basic emotions (Widen & Russell, 2003). While emotion recognition develops early in childhood, young children with ASD experience more difficulty with recognition of certain expressions compared to their typically developing peers (Rump, Giovannelli, Minshew, & Strauss, 2009). Scambler et al. (2007) found that children with ASD between 17 and 34 months already show less response to facial expression than their typically developing peers. Also, by 10 years of age, children with ASD are worse than typically developing peers at labeling basic expressions (Lindner & Rosén, 2006; Tantam, Monaghan, Nicholson, & Stirling, 1989). Although most studies show that adolescents and adults with ASD are not impaired in recognizing basic, prototypical emotions (Capps, Yirmiya, & Sigman, 1992; Grossman, Klin, Carter, & Volkmar, 2000), they show impairment when stimuli are more subtle or complex, and presented briefly (Humphreys, Minshew, Leonard, & Behrmann, 2007), such as the natural emotions that are encountered in everyday social interactions. It is therefore important to utilize naturalistic social scenes, when evaluating emotion recognition difficulties in this population.

Deficits in both facial emotion recognition (FER) and facial emotion expression (FEE) have been established and, while the evidence is not conclusive, several studies show that difficulty in either recognition or expression greatly affects the quality of social interactions. These difficulties affect the ability of individuals to understand emotions of others as well as their ability to express emotions to others, both of which are essential for successful social interaction, including nonverbal communication (Nuske, Vivanti, & Dissanayake, 2013). Compared to FER, there is less research on FEE in ASD. However, children and adolescents with ASD differ in their quality and quantity of emotion expressions. Individuals with ASD display fewer nonverbal expressions of affect (Yirmiya, Kasari, Sigman, & Mundy, 1989). In addition, their facial expressions are often described as flat, peculiar, and at times exaggerated (Langdell, 1981; Yirmiya et al., 1989). This atypical emotion expression makes social communication challenging, even when their verbal abilities are intact. The differences in non-verbal expressivity have important impacts on children's social interactions. For example, Stagg, Slavny, Hand, Cardoso, and Smith (2013) found not only that adults rated children with ASD as being less expressive than typically developing children, but that children provided lower friendship ratings for these children with ASD compared to typically developing peers. In addition, typically developing observers have provided less favorable first impression rating for children and adults with ASD engaging in social behavior compared to controls (Sasson et al., 2017). These impressions were associated with reduced intention to engage in social interaction with the individuals.

In a meta-analysis on the relationship between emotion recognition abilities and everyday social functioning in ASD, Trevisan and Birmingham (2016) found moderate but significant association between ability to recognize facial expressions and important social behavior, including adaptive functioning, in several studies. The reported associations between deficits in emotion recognition and social competence in children with ASD suggests the importance of addressing emotion recognition deficits in this population. Given the advancements in technology and prior research suggesting that technology-based intervention might be more reinforcing to individuals with ASD (e.g., Goldsmith & LeBlanc, 2004), several technological interventions, including computer-based programs, have been developed to address FER deficits in ASD. Studies have shown that such interventions are promising, even though further research is needed in evaluation of their effectiveness (e.g., Kauo & Egel, 2016; Lee, Lam, Tsang, Yuen, & Ng, 2018).

While many studies have explored impairments in FER in individuals with ASD, more recently, research has focused on exploring the mechanism behind these impairments, utilizing eye-tracking and electroencephalography (EEG) in order to gain insights into the attentional and neurological correlates. Collectively, results from these studies are inconsistent but suggest impairment in some aspects of attention. In a review of eye tracking studies in ASD, Guillon and colleagues (2014) found that while the majority of the extant research indicates decreased visual attention to social stimuli in individuals with ASD, the degree of attention varies across contexts. For example, including several people interacting increases the likelihood of decreased attention to faces. Similarly, a recent systematic review by Black et al. (2017) found evidence from both eye tracking and EEG studies that suggests atypical attentional and cognitive processes in individuals with ASD. A number of studies reported atypical gaze to facial features during FER tasks while many EEG studies reported atypical modulation of the N170 component, a negative ERP which occurs between 130 and 200 milliseconds and is enhanced in response to faces (Eimer, Gosling, Nicholas, & Kiss, 2011), in individuals with ASD. Yet, several studies identified in the review failed to find any significant group (ASD versus typically developing) differences in attention to the facial features, especially for the children. Several studies, however, show that the eye gaze patterns of people with ASD are characterized by atypical fixation patterns, specifically more attending to inanimate objects and non-social features of the environment than to socially relevant stimuli (e.g., faces and, in particular, the eye region) (Chevallier et al., 2015; Hanley et al., 2015; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Pelphrey et al., 2002; Rice, Moriuchi, Jones, & Klin, 2012; Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009). Notably, the mixed findings reported (e.g., Guillon et al., 2014), might be due to variability in stimuli type, and therefore variability in ecological validity. Hanley and colleagues (2013) for example found that while individuals with ASD did not show diminished attention when facial stimuli were viewed in isolation, they showed diminished attention to the eyes in comparison to typical peers when the faces were viewed as part of social scenes. When exploring eye gaze as a potential mechanism behind impairment in facial emotion recognition, it is important to utilize ecologically valid stimuli that, to the extent possible, simulate situations that children encounter in their daily lives.

Greater attending to non-social environmental stimuli and therefore lower fixation to the eyes is associated with lower social competence (e.g., Jones, Carr, & Klin, 2008), and avoidance and reduced orientation have been shown to predict emotional recognition performance (Kliemann, Dziobek, Hatri, Steimke, & Heekeren, 2010). These results together suggest that individuals with ASD may have difficulties understanding others' emotions, if they do not attend to the important facial features. Two hypotheses have been presented to explain the reduced attention to other's facial features, specifically eyes, in individuals with ASD. One hypothesis is that individuals with ASD purposefully look away because they find the eyes to be aversive (i.e., gaze aversion; Hutt & Ounsted, 1966; Kliemann et al., 2010), while the other hypothesis is gaze indifference, or looking less due to viewing the stimuli as unengaging or uninformative (Davies, Dapretto, Sigman, Sepeta, & Bookheimer, 2011; Senju & Johnson, 2009). The mechanism behind the diminished attention to faces is important, as it should inform intervention. Moriuchi, Klin, and Jones (2016) explored the mechanism behind the diminished attention to eyes in 2-year-old children with autism by directly cuing children to look at the eyes. They found that children with ASD did not look away faster than typically developing children, therefore finding support that diminished eye-looking was consistent with the gaze indifference hypothesis, not finding any support for the gaze aversion hypothesis. However, the evidence for and against each of the two hypotheses is mixed. Whether the established lack of attention is due to gaze indifference or gaze aversion, interventions focusing on directing attention to the facial features may be successful in improving emotion recognition abilities in individuals with ASD, as long as it is successful in increasing attention to the facial region.

Forms of attention modification have been utilized with populations outside of ASD, mainly with individuals with social anxiety disorder. Attention Bias Modification Treatment (ABMT) has been shown to be successful in addressing threat bias seen in adults with social anxiety (e.g., Hakamata et al., 2010; Linetzky, Pergamin-Hight, Pine, & Bar-Haim, 2015; Schmidt, Richey, Buckner, & Timpano, 2009), by systematically redirecting attention away from the threat stimuli. Results from the social anxiety literature suggest that gaze patterns can be altered (e.g., Linetzky et al., 2015). However, alteration of eye gaze patterns has not been directly explored in children with ASD who show deficits in facial emotion recognition. Therefore, the goal of the study was to utilize attention modification to increase focus on social features (i.e., face area), and eye gaze analyses to target the potential mechanism behind the FER impairment.

We sought to develop and evaluate an attention modification program to address FER deficits in children with ASD. The primary therapeutic aim of this attention modification intervention was to alter the gaze patterns. The aspect of attention targeted, however, differs from that modified in ABMT for anxiety. Posner and Peterson (1990; 2012) divide the attention system into 3 interacting networks: 1) orienting toward stimuli, 2) alerting (readiness to receive or respond to information), and 3) executive attention (resolution of conflict). In the current program, we are not targeting bias in attention (e.g., ABMT), as the mechanism is not bias in disengagement from threat. In a review of attentional networks in ASD, Keehn et al. (2013) found that individuals with ASD are characterized by global attention deficits in all three networks - alerting, orienting, and executive attention. Consistently, multiple studies have reported greater visual latency to faces in individuals with ASD (e.g., Freeth, Foulsham, & Chapman, 2011; Riby & Hancock, 2009; Wilson, Brock, & Palermo, 2010), indicating difficulty with orienting toward the social stimuli (i.e., faces). Therefore, this study aims to increase attention in terms of orienting toward the facial region in order to explore its effect on FER ability.

In the current intervention protocol, we utilized ecologically valid stimuli (i.e., video clips of social interactions) to modify attention, in order to more closely simulate situations that children encounter in their daily lives. These videos, however, produce complexity regarding multi-modal presentation that may affect FER given the difficulty in disentangling the contribution of different modalities. Research into emotion recognition in ASD has generally focused on the visual modality only, and the few studies that investigated recognition of emotion using other modalities found discrepant results. While some studies indicate impairment across modalities for adults with ASD (e.g., Phillip et al., 2010), others found impairment in some (i.e., visual or tone of voice) but not other modalities (i.e., verbal content or combined modalities) for children with ASD (Lindner & Rosén, 2006). The use of more complex stimuli create complexity in terms of mix of compensatory strategies utilized during FER for children with ASD (e.g., Grossman et al., 2000). Therefore, while non-facial cues such as posture and tone of voice provide much of the information regarding the expressed emotion, the focus of this program is on outcome of facial emotion cues (utilizing both static and dynamic stimuli), in order to explore one aspect of social communication which makes social interactions difficult, in a controlled way.

We assert that greater attention to others' faces should translate to improved FER, which may ultimately translate to reduction of core symptoms of ASD (i.e., greater social competence) in children with ASD. The proposed approach therefore could be broadly effective, regardless of whether the root of the impairment is poor motivation (i.e., gaze indifference) or social aversion/arousal.

Present Study

The purpose of this study was to develop and assess the feasibility and acceptability of an attention modification intervention to attenuate deficits in facial emotion recognition and expression in individuals with ASD and to assess change in gaze to socially relevant cues following the intervention. In terms of feasibility, we aimed to enroll at least 8 participants who meet inclusion criteria within an 8-week recruitment period, demonstrating viability of the intervention and recruitment approach (which are important considerations for further, large-scale evaluation of the intervention). In addition, we anticipated low attrition (i.e., no more than 1 person dropping out) throughout the treatment and endpoint assessment session. In terms of acceptability, we hypothesized that at least 80% of the child participants and their parents would report 'high' or 'very high' acceptability ratings for the intervention. To assess the feasibility of the intervention to attenuate deficits in FER, we hypothesized that there would be a significant improvement in participants' facial emotion recognition ability from pre to post intervention. Lastly, we hypothesized that there would be a significant change in total fixation duration to facial features of the stimuli from pre to post intervention.

Method

Participants

Participants included children (total $n = 8$) between the ages of 9 and 12 years, inclusive. At the start of middle childhood (approximately age 9), facial features become more stable as there is more extreme change in facial dimensions during early childhood than middle childhood (Bishara, Peterson, & Bishara, 1984). In order to be eligible, all participants had to evidence difficulties with emotion recognition (based on parent report) and have at least low average intellectual abilities (IQ score greater than 75). In addition, participants were required to have received a clinical diagnosis of ASD, which was confirmed by the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012). No participants presented with severe co-occurring psychopathology, such as psychotic symptoms or severe aggression. In addition, children who were receiving psychosocial therapy aimed at addressing facial emotion recognition at the time of study were excluded. Demographic information for the participants is presented in Table 1. Participants were recruited through flyers in the local and surrounding community (e.g., stores, restaurants, libraries), existing research registry databases, local ASD support groups, and from prior studies, if they have agreed to be contacted about future research.

Measures

Affect Recognition subtest of the Developmental Neuropsychological Assessment (NEPSY-II: Korkman, Kirk, & Kemp, 2007). The Affect Recognition (AR) subtest of the NEPSY-II assesses a child's ability to recognize affect from colored photographs of children's faces. AR has been used in prior studies to measure deficits in FER in children with ASD (e.g., Williams, Gray, & Tonge, 2012). It provides age-based standard scores, with lower scores indicating poorer FER ability. AR subtest was administered to participants during the pre-treatment and post-treatment sessions.

Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012). The ADOS-2 is a semi-structured assessment of ASD characteristics and is considered the gold standard for assessment of ASD. Module 3, designed for verbally fluent children, was administered by a lead investigator who is research reliable on the administration and scoring of the measure. ADOS-2 was administered during the intake session only to assess for eligibility.

Child Behavior Checklist for Ages 6-18 (CBCL/6-18; Achenbach & Rescorla, 2001). CBCL/6-18 is a parent-report measure of child's behavior, comprised of 20 ratings of competence and 120 problem items. CBCL/6-18 yields scores on internalizing, externalizing, and total problems, in addition to scores on DSM-IV related scales. The CBCL/6-18 was completed by the parents of the participants during the intake and endpoint sessions.

Demographic Information Questionnaire. Parents of participants answered questions about the adolescents' gender, age, race, grade, and ability to express a range of facial expressions. In addition, they answered questions regarding both parent's (when available) educational information and income, and questions regarding their child's diagnoses, treatments, and medications. They were also asked to share if the child, the child's parents, child's grandparents, or child's siblings have any current medical or psychological diagnoses by endorsing a checklist of various disorders, one of which was ASD.

Emotion Regulation and Social Skills Questionnaire, Parent Version (ERSSQ; Beaumont & Sofronoff, 2008). ERSSQ is a parent measure designed to evaluate emotion regulation and social skills in children and adolescents with ASD. It consists of 27 items evaluated in terms of frequency of skill or behavior of the child. Parents completed the ERSSQ during the weekly treatment sessions, in order to evaluate potential changes in emotion regulation and social skills throughout the treatment.

Social Responsiveness Scale, Second Edition, Parent Version (SRS-2; Constantino & Gruber, 2012). The SRS-2 is comprised of 65 items measuring parent report of ASD-related social impairments. SRS-2 assesses social awareness, social cognition, social communication, social motivation, and restricted interests and repetitive behavior. SRS-2 was completed by the parents during the intake and endpoint sessions. While SRS-2 asks the parents to focus on the behavior during the past 6 months, at the endpoint session, parents were asked to rate their child's behavior since the start of the intervention.

Toronto Alexithymia Scale for Children (TAS-C, Rieffe, Oosterveld, & Meerum Terwogt, 2006). Adapted from the 20 item Toronto Alexithymia Scale developed for adults (TAS-20; Bagby, Parker, & Taylor, 1994), the TAS-C measures difficulties with understanding, processing, or describing emotions in children. The TAS-C produces three separate factor scores: Difficulty Describing Feelings subscale, Difficulty Identifying Feeling subscale, and Externally-Oriented Thinking subscale, which can be combined into a total score. The TAS-C was completed by the children during the pre-treatment and endpoint sessions.

Treatment Satisfaction Form: Both child and parent completed a treatment satisfaction form, which asked the child and their parent their thoughts regarding the treatment protocol. Parent and child rated the statements regarding whether they found the intervention to be helpful, acceptable in terms of time and commitment, whether they would recommend the program to a friend, and whether overall, they liked the program on a 5-point scale, ranging from "strongly disagree" to "strongly agree." In addition, parent and child were asked to rate whether the child's ability to recognize emotions has "greatly decreased", "slightly decreased", "stayed the same", "slightly increased" or "greatly increased." In addition, the child was asked to write about what they liked the about the program, what they disliked about the program, and what they would recommend to improve the program. Parents, on the other hand, were asked what changes they have noticed regarding their child's emotion recognition skills and what they would recommend to improve the program.

Video Emotion Recognition Task: In order to evaluate change in emotion recognition abilities, participants completed a video emotion recognition task, in which they viewed a video of an adult expressing one of six basic expressions (happiness, sadness, fear, anger, surprise, and disgust). After the presentation of each video, participants were asked to tell the examiner which emotion they saw portrayed in the video. Each expression was presented 6 times, for a total of 36 videos. Each video lasted 2.73 seconds (Wieckowski and White, 2017 provide further information regarding the task). The Video Emotion Recognition Task was administered to the participants during the pre-treatment and post-treatment sessions.

Video Emotion Expression Task: In order to evaluate change in emotion expression abilities, participants completed the expression task, which consisted of spontaneous and scripted conditions. In the first, spontaneous task, participants were asked to view a video of an adult expressing one of six basic emotions (happiness, sadness, fear, anger, surprise, and disgust). They were asked to respond to the video, using a facial expression. Each expression was presented 6 times for a total of 36 videos, each lasting 2.73 seconds. In the second, scripted condition, participants were asked to make an expression of a verbally presented emotion, without seeing a video. The examiner asked the participant to express each of the six basic expressions one time. The order of the emotions was randomized across participants (see Wieckowski and White, 2017 for more detail regarding the task). Video Emotion Expression Tasks were administered during the pre-treatment and post-treatment sessions.

Wechsler Abbreviated Scale of Intelligence, 2nd edition (WASI-2; Wechsler, 2011).

The WASI-2, a measure of cognitive functioning, was administered to all participants. Two subtests of the WASI-2 were administered (Vocabulary and Matrix Reasoning) to get an abbreviated full scale IQ-2. 4-subtest and 2-subtest WASI-2 scores correlate at .93. The WASI-2 was administered during the intake session only to determine eligibility.

Youth Top Problems (YTP). The YTP measure was administered to the parents, in addition to ERSSQ, during the intake session and during weekly treatment sessions, in order to evaluate changes in identified social difficulties during treatment. During the intake session, parents were asked to identify three behaviors related to nonverbal socioemotional interaction (e.g., facial emotion recognition) that are causing the child most difficulty. During the treatment sessions, the parents were asked to provide rating from 0 to 10, to indicate 'how big of a problem' the identified difficulties were for the child since the prior visit (i.e., within the last 72 hours).

Anxiety Measures

The goal of the experimental treatment was to train attention to others' faces in order to improve FER. It is possible that the heightening attention to previously under-noticed social stimuli could inadvertently increase social anxiety via greater awareness of others and their self-referential evaluations. As social anxiety can amplify the social problems that are characteristic of ASD (White, Schry, & Kreiser, 2014), it is important to assure that the intervention does not heighten social anxiety, which could become impairing. Therefore, assessment of change in social anxiety symptoms, and in particular monitoring for worsening anxiety, is necessary in any evaluation of treatment that focuses on socio-emotional functioning, including FER in the current study.

Brief Fear of Negative Evaluation Scale (BFNE; Leary, 1983). BFNE is a brief version of the full FNE scale (Watson & Friend, 1969), a self-report meaure that assesses amount of worry or fear about negative evaluation from others. Participants indicate how characteristic each of the 12 statements is for them on a 5-point scale (from "not at all" to "extremely"). Only the eight straightforward wording items were summed to create total BFNE score as these items have been found to yield the best diagnostic sensitivity and reliability (Carleton, Collimore, McCabe, & Antony, 2011). The BFNE was administered at the pre-treatment and post-treatment sessions in order to explore the potential of heightened anxiety following intervention.

Screen for Child Anxiety Related Disorders, Child and Parent Version (SCARED; Birmaher et al., 1997). The SCARED is a 41 item self-report and parent-report assessment measure used to assess various aspects of anxiety in youth, and includes scores for Panic Disorder, Generalized Anxiety, Separation Anxiety, Social Anxiety, and School Avoidance, in addition to a Total Score. The measure has shown acceptable test-retest reliability, internal consistency, and discriminant validity (Birmaher et al., 1997). The self-report and parent-report SCARED was administered at pre-treatment and post-treatment sessions along with BFNE to explore the potential of heightened anxiety following the intervention.

Apparatus and Stimuli

Eye gaze. Eye-tracking was completed using a Tobii T60 XL eye tracker. Participants sat approximately 60 cm from the eye tracker and they were instructed to look at the video stimuli on the screen. Prior to displaying each stimulus, a centered "X" (1.5 cm x 1.5 cm wide) appeared on screen in order to centralize the participants' attention. The stimuli were only presented after the participant successfully attended to the centered "X". The eye-tracking system was calibrated to each participant's eyes prior to data collection using a five-point calibration procedure (i.e., a red circle moving to five predefined locations across the screen, including the four corners and the center of the screen). The study investigator visually inspected the tracking to all five points and corrected any missing or excessive error tracking before advancing the participant to the task. Eye-tracking data were collected during the treatment session for the EU-Emotion Stimulus Set and the Intervention Task (described below) in addition to the Video Emotion Recognition Task and Video Emotion Expression Task during pre-treatment and post-treatment sessions. Participant's eye gaze patterns and fixations were collected through the Tobii studio and analyzed using a Matlab code (MatlabR2014b, Mathworks Inc., MA).

Emotion expression coding. The expressions of emotions from the Video Emotion Expression Task were recorded through a webcam camera located directly on top of the computer screen. Two undergraduate research assistants independently coded all videos from the Video Emotion Expression Task for both the scripted and the spontaneous conditions. The coders were masked as to whether the video was from the pre-treatment or post-treatment condition. They assigned three codes after viewing the full video, including 1) a rating for which emotion was portrayed (happy, sad, fear, disgust, surprise, or anger), 2) a clarity rating from 0 to 2 with 0 = indiscernible emotion, 1 = could not decide between two emotions, 2 = complete confidence in emotion, and 3) atypicality rating with 0 = emotion was recognizable and appeared typical in nature (i.e., no distorted or unusual features), 1 = expression was recognizable but a poor example because of some distortion or unusual feature, and 2 = expression was obviously odd, stereotyped, or mechanical in nature. After independently assigning all ratings, when discrepancy occurred, the two coders met to discuss the videos and came up with consensus ratings for all three criteria. For the scripted task, the two raters had 80.21% agreement for the emotion, and 75.00% agreement for the clarity rating. For the spontaneous task, the raters had 68.52% agreement for emotion and 61.11% agreement for the clarity rating.

EU-Emotion Stimulus Set (O'Reilly et al., 2016). The EU-Emotion Stimulus Set contains video clips of children and adults expressing different emotions and mental states. The set contains stimuli of facial expressions, body gesture scenes, and contextual social scenes. Only facial expressions were utilized for the current study, and only videos with at least 60% accuracy rating based on O'Reilly and colleagues (2016) were included. Twelve videos, two of each of the six basic emotions (i.e., fear, anger, disgust, happiness, sadness, and neutral) were shown to the participant during each session: one video of high intensity, and one of low intensity. The EU-Emotion Stimulus set was shown to the participants during the baseline periods as well during each treatment session. The order of the stimuli was semi-random, such that each session had one high intensity and one low intensity video per emotion. However, the videos were randomized across and within the sessions. The order for the EU videos was the same however for each participant for a specific session.

Stimuli for Intervention Task. Three undergraduate research assistants identified publicly available video clips in which at least one actor expressed one of the following emotions: happiness, sadness, anger, disgust, fear, and neutral expression. The clips were taken from television shows and movies not currently in mainstream child entertainment, in order to decrease chances of familiarity with the clip. Surprise, as an expression, was not shown, due to the emotion recognition task (Affect Recognition) not assessing for surprise expressions. A total of 290 clips were identified. More video clips were identified for disgust (n = 59), fear (n = 49), and neutral expression (n = 50) compared to happy (n = 43), sad (n = 45), and anger (n = 45) expressions given more difficulty in identification of these videos. Prior to use in this study, the 290 identified videos were validated to ensure they adequately captured the target emotion. The videos were randomly split into four surveys (to decrease completion time), which were administered via Qualtrics, to undergraduate students who received course credit for participation (n = 121). Participants were presented with the video and asked two questions: 1) "What emotion did the previous video represent?" with options of anger, happy, sad, neutral, fear, disgust and 2) "Rate how well the video represents the emotion you selected above" with options of very poor, poor, moderate, well, very well. Only stimuli with 60% or better agreement on the emotion shown were included as the treatment videos, except for three disgust videos. For disgust, because of lower overall mean agreement (M = 72.87%), a slightly lower threshold of 55% was used.

The final stimuli set for the intervention task consisted of 170 clips (17 clips per session) of social interactions where at least one character was expressing one of the six emotions. Seventeen clips were presented per session and each clip was only presented once during the ten-week treatment. In each treatment session, three clips were presented for each emotion (except for disgust, which had 2 clips per emotion, per session) and the order of the 17 presented stimuli remained constant within session across participants, so that all participants viewed the same order of the videos for each session. The videos were exactly five seconds in length followed by a black screen with options for the emotions written out in order to allow participants enough time to tell the examiner which emotion was displayed in the video. The order was semi-random in terms of each session having high/medium/low rated videos (based on agreement from a stimuli validation study). However, the videos were randomized within a session and across sessions. Each intervention session took approximately 15 minutes.

Procedure

Individuals interested in participating in the study were directed to contact the study investigator through phone or email. Before scheduling the participant, the study coordinator conducted a brief phone screener to assess for diagnostic status, age criteria, and other presenting concerns. If eligible, participants and their parents completed an in-person intake appointment. After providing assent (child) and parent permission (parent) for participation, all participants completed a series of measures. Following the intake appointment, participants were randomized to complete a baseline assessment consisting of either 4 time points, 5 time points, or 6 time points, before commencing intervention. We did not check to ensure stability within the baseline, as any manipulation of the baseline period would not allow for randomization of subjects to baseline periods, as is suggested for clinical and technology-based interventions (e.g., Kazdin, 1978; Dallery, Cassidy, & Raiff, 2013). The baseline periods occurred twice a week, to be consistent with session frequency during the intervention phase, therefore resulting in 2 week, 2.5 week, and 3 week baselines. To have consistent data points, we assessed FER using the same paradigm during each treatment session, after the administration of treatment. FER during baseline and treatment periods were assessed using the EU-Emotion Stimulus Set specifically designed for children with ASD (O'Reilly et al., 2015), described above. Methodologically, this allowed us to establish stability in the construct of interest prior to manipulation. After the baseline period, participants began the intervention, described below.

All participants attended 10 bi-weekly individual sessions, each one lasting approximately 20 minutes (including welcoming, treatment session, parent measures). This timing has been shown to be tolerable to participants for the ABMT protocols (e.g., Ollendick et al., 2018). In addition to the intake session, data were collected at the start of treatment (i.e., pre-treatment assessment) and at the end of treatment (i.e., post-treatment assessment). The schedule of measures is provided in Table 2. Participants were compensated $20 for completion of the intake, pre-treatment and post-treatment assessment sessions, in addition to $20 for completion of all treatment sessions, for a maximum total of $80 for completion of all assessments and treatments.

During the 10 treatment sessions, participants were seated approximately 60cm in front of the monitor, and were shown the Intervention Task Stimuli described above of an actor or actors displaying the emotion for 5 seconds. The faces of the actors in the video were surrounded by a square box (white dotted line) in order to draw attention to the socially relevant information (i.e., the social stimulus that conveys the target emotion). This procedure for highlighting the area of interest is consistent with prior research highlighting areas of interest (e.g., Richler, Floyd, & Gauthier, 2014) as well technological approaches being developed (e.g., Kinect system highlighting the face in a square) to “read” the facial expression. As the child proceeded through the 10-session program, the number of videos with a square box decreased so that for each subsequent session, the percentage of videos with the cue was less than in the prior session. Therefore, while in the first session all videos (100%) had the facial area highlighted, the last session included only two videos (12%) with the visual cue. This procedure follows the recommendation for basic intervention techniques for individuals with development delays, to eliminate need for continued prompting and therefore reduce prompt dependency (Lovaas, 2003). As participants watched each video, they were instructed to attend to the video and were informed that they were going to be asked questions regarding the video after each presentation. At the end of the video, the participants were asked which emotion was displayed in the clip, with options presented in print on the screen. The participants verbally labelled the emotion. While viewing the stimuli, participants’ eye gaze was recorded using a Tobii T60 XL eye-tracker, which tracked their eye-gaze in real time.

Data Analytic Plan

Eye-tracking Data

For the eye tracking data collected during the Video Emotion Recognition and Expression tasks and during the intervention sessions, Tobii Studio was used to analyze the data. The Areas of Interest (AOIs) included the face region, predefined using the oval-shaped AOI tool, and background region, predefined using a rectangle-shaped AOI tool, available in the Tobii T60 (Studio Professional) platform. In addition, the box highlighting the facial region was predefined, to allow for analyses regarding whether participants looked within the box, as instructed. Total duration of fixations made to these regions were calculated. For the eye-tracking data, any trial showing a major loss of tracking (i.e., less than 50% of the viewing time per stimulus) was excluded from data analyses. Eye gaze was analyzed within session and within person, in addition to across sessions in order to explore whether there is a significant change in eye gaze to face area for those individuals who improve on their facial emotion recognition. Duration of fixations made to these regions was calculated using an in-house Matlab code.

Statistical Hypothesis-Testing

Statistical analyses included a Reliable Change Index (RCI; Jacobson & Truax, 1991), which was calculated in order to determine the amount of change in facial emotion recognition (i.e., AR task from NEPSY-II) and other clinically relevant outcome measures (i.e., SRS-2, CBCL, TAS-C, ERSSQ) occurring from the treatment. RCI assesses for the amount of change and whether the change reflects more than just the fluctuation due to an imprecise measurement. The RCI accounts for imprecise measurement by dividing the difference of pre-treatment and post-treatment scores by the standard difference, to account for test-retest reliability of the measure. The following equation was used:

$$\text{RCI} = x_2 - x_1 / S_{diff} \text{ where } S_{diff} = \sqrt{2 \text{ x } [SD\sqrt{1 - rel_{test-retest}}]^2}$$

When test-retest reliability coefficient was not available, internal consistency coefficient (alpha) was utilized. While in general alpha coefficients tend to be greater than test-retest coefficients, they have both been utilized in prior studies (Iverson, 2011). The recommended cutoff of the RCI is 1.96, with scores higher than 1.96 indicating statistically significant change. The scores of the outcome measures were averaged for each time point separately, and then these scores were used to calculate the RCI for each distinct outcome measure. The test–retest reliabilities and standard deviations utilized were obtained from measure manuals and prior research, matching the sample in age, gender, and diagnoses, whenever possible.

A Wilcoxon Signed Rank Test, a non-parametric statistical hypothesis test, was used to compare initial and endpoint scores for all participants. The Wilcoxon Signed Rank Test is an alternative to a paired t-test, which assumes normal distribution. Given the small sample size, normal distribution cannot be assumed. The Wilcoxon Signed Rank Test provides a Z score with a significance test (p-value) to determine whether the change in scores across all participants is significant or not. Wilcoxon Signed Rank Test was used to compare scores for all outcome data, including those for which RCI analyses were not possible, due to no available standardized scores or test-retest reliabilities (i.e., for Accuracy on Intervention Video Task, Video Emotion Recognition Task, and Video Emotion Expression Task). Data were analyzed with IBM SPSS Statistics (SPSS 24.0). An effect size was calculated using Excel for all Wilcoxon Signed Rank Test analyses with the following formula: $r = Z/\text{sqrt}(N)$ where N is the total number of samples (Rosenthal, 1994). Given the small sample size and preliminary nature of the study, in addition to significance testing, we also focus on effect size interpretation. Per Rosenthal's guidelines (1994), a small effect size [r] is 0.1, medium effect size is 0.3, and large effect size is 0.5.

In addition, Simulation Modeling Analysis (SMA; Borckardt et al., 2008) was used to analyze single-case data for the EU-Emotion Stimulus Set Task for which baseline phase and intervention phase data were collected. SMA (www.clinicalresearcher.org) was designed for analyzing time-series data and it allows for the analysis of change in symptom level as well as the slope of symptom change from pre-treatment baseline to endpoint time point. SMA utilizes bootstrapping techniques to minimize autocorrelation effects, when measuring change across time points. More information regarding the SMA analyses is provided in the results section. Borckardt et al. (2008) recommend at least 5 data points per phase to conduct SMA. However, Borckardt tested the procedure using shorter baselines and found similar results, but noted that Type I error rates may exceed .05 and power may drop below .80 (personal communication noted in Pugliese, 2012). Therefore, since baseline lengths in this study ranged from three to five points, results utilizing SMA analyses should be interpreted with caution.

For the multiple time point data collected during the intervention phase, for which baseline data was not collected (i.e., ERSSQ, Intervention Task Accuracy), linear mixed effect model was fit. R package (Package, http://www.R-project.org/) was used to fit a linear mixed model which can accommodate both the variation among measurements within individuals and the individual-to-individual variation. Let Y_{ij} be the overall score on the measure, *j* be the time point (treatment 1 through 10), and *i* be the participant (ID 1 through 8). We can express our linear mixed effect model,

$$Y_{ij} = \beta_{0i} + \beta_{1i}\text{Time}_j + \varepsilon_{ij}$$

where $(\beta_{0i}, \beta_{1i})^T \sim N(0, \Sigma)$ and ε is designating random error with mean zero and some constant variance. The population mean and variance of random coefficient (β) were estimated. In addition, the test statistics were computed along with p values to detect any differences in equation coefficients using the t-test. The model was used to compare treatment effects across the intervention time points.

In addition, for the YTP, as each parent reported different problems and therefore the scores are not equivalent across participants, a linear model was fit to understand the relationship between time point and problem rating for each of the three problems that parents provided for each participant individually. Fit to the following linear model was tested: $Y_{ij} = \beta_{0i} + \beta_{1i}\text{Time}_j + \varepsilon_{ij}$, where Y_{ij} is the problem rating provided by the parent for Timepoint *j* for Participant *i.* An estimate and variance were calculated for each participant, for each problem rating, in addition to the t-test statistic and p value to detect change in rating over time.

Results

Feasibility – Enrollment (Hypothesis 1)

In terms of viability of the intervention and recruitment approach, our hypotheses were generally supported by the results. We aimed to enroll at least 8 participants who meet inclusion criteria within an 8 week recruitment period and anticipated low attrition (i.e., ≤ 1 person dropping out) for the study. Our results show that the study is viable to an extent, with successful enrollment of 8 participants who completed the study, with only one participant dropping after just one treatment session due to difficulty commuting a long distance to the center (~1 hour each way) for the treatment twice a week. In addition, one of the eight participants who completed the intervention sessions (ID 8) completed the study in a non-protocol timeframe with 3 weeks of treatment instead of 5, due to the child travelling out of the country for an extended period of time. The sessions were still kept on non-consecutive days, but this participant averaged three sessions, instead of two sessions, per week. Analyses were completed with and without this participant to see whether the abbreviated intervention timeframe effects the results. The other seven participants completed the program as designed, with two sessions a week, for five weeks, with no consecutive session days. No baseline or treatment sessions were missed for any of the participants. While we were successful at recruiting the projected 8 participants, the recruitment period was much longer than anticipated, with the recruitment spread over 10 months (June 2017 through April 2018). Table 1 displays descriptive statistics for the 8 participants that completed the intervention and Table 3 displays their pre-treatment and post-treatment scores on all measures.

Acceptability and Satisfaction (Hypothesis 2)

In terms of the intervention being acceptable to the participants, the results support the proposed hypotheses. On the Treatment Satisfaction form that the children and parents completed at the endpoint assessment, for the statement "This program is acceptable (in terms of time and work commitment)", one child gave the rating of "Neutral", while the other seven children rated it as "Agree" ($M = 3.88$, $SD = 0.35$). For parent rating of acceptability of the program, five parents provided a rating of "Agree" while the other three parents gave the rating of "Strongly Agree" ($M = 4.38$, $SD = 0.52$). Therefore, the results provide support that on average, the intervention is acceptable to children with ASD and their parents.

The Treatment Satisfaction form asked several other questions to assess satisfaction with the intervention program. For "Overall, this program seems helpful", five children provided the rating of "Neutral", while the other three children gave the rating of "Agree" ($M = 3.38$, $SD = 0.52$). For the statement "Overall, this program seems helpful in increasing my child's emotion recognition," two parents provided a rating of "Neutral", five parents gave the rating of "Agree", and one parent rated the statement as "Strongly Agree" ($M = 3.88$, $SD = 0.64$). For "I would recommend this program to a friend," three children rated the statement as "Neutral", four children rated it as "Agree," and one child rated it as "Strongly Agree" ($M = 3.75$, $SD = 0.71$). On the parent survey, one parent rated the statement as "Neutral", six parents rated it as "Agree," and one parent rated it as "Strongly Agree" ($M = 4.0$, $SD = 0.53$). For "Overall, I like the program," four children provided rating of "Neutral," three children gave the rating of "Agree," and one child gave the rating of "Strongly Agree" ($M = 3.63$, $SD = 0.74$). On the parent survey, seven parents gave the rating of "Agree," and one parent gave the rating of "Strongly Agree" ($M = 4.13$, $SD = 0.35$).

Facial Emotion Recognition Change (Hypothesis 3)

For the main analyses, our hypothesis of a significant improvement in participants' FER ability from pre to post intervention was supported for some, but not all measures. FER was assessed through several tasks and measures administered at the beginning and end of the intervention (i.e., AR subtest of NEPSY-II and Video Emotion Recognition Task), during the intervention (EU-Emotion Stimulus Set Task, Intervention Task), as well as child and parent report of change in FER at the end of intervention.

Affect Recognition subtest of NEPSY-II. A Wilcoxon signed-rank test indicated that scores on the AR task at the post-intervention assessment were not statistically higher than those in the pre-intervention assessment for the whole sample ($Z = -.17$, $p = .86$, $r = .04$). This result held when the non-protocol participant (ID 8) was excluded ($Z = -.65$, $p = .52$, $r = .17$). However, RCI analyses showed that one participant (ID 6) showed reliable increase in his score on the AR task (RCI = 2.57). None of the other seven participants showed a reliable change (Table 4).

Video Emotion Recognition Task. Table 3 shows accuracy scores on the Video Emotion Recognition Task (Pre-Tx: $M = 68.83$, $SD = 12.19$; Post-Tx: $M = 72.57$, $SD = 17.22$). A Wilcoxon signed-rank test indicated no significant difference between pre-treatment and post-treatment scores on the Video Emotion Recognition task, either for the whole sample ($Z = -.70$, $p = .48$, $r = .18$) or when the non-protocol participant (ID 8) was excluded ($Z = -.51$, $p = .61$, $r = .14$). In addition, no significant difference was observed for any of the individual emotions (all $p > .06$). No individual subject analyses were completed as no standardized norms exist for this measure.

EU-Emotion Stimulus Set Task. The EU task was the only task that was administered to the participants prior to the intervention during 2-week, 2.5-week, or 3-week baselines in addition to the intervention time points in order to evaluate change in FER across phases. Table 5 presents mean level symptom changes and mean level slope changes for SMA for the EU task. In addition, Figure 1 presents visual depictions of time series data for each subject. Overall, only two (25%; IDs 1 and 8) significant mean level differences were noted between the baseline and treatment phases using the SMA approach. Similarly, for only two participants (25%: IDs 4 and 8) significant changes were noted for slope analyses. SMA compares the data from the participants against 5 different slope vectors: 1) increasing baseline, decreasing treatment; 2) flat baseline, increasing treatment; 3) increasing baseline, flat treatment; 4) increasing from baseline through treatment; and 5) increasing during baseline with return to pre-treatment level at the start of treatment and increasing throughout treatment. Mathematically, these vectors can also be represented as: 1) 1 2 3 4 5 5 4 3 2 1 0 -1 -2 -3 -4; 2) 0 0 0 0 0 1 2 3 4 5 6 7 8 9 10; 3) 1 2 3 4 5 5 5 5 5 5 5 5 5 5 5; 4) 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15; 5) 1 2 3 4 5 1 2 3 4 5 6 7 8 9 10. A significant correlation indicates that the slope of the data matches the slope of the vector. For both participants who showed significant changes in slope (IDs 4 and 8), multiple slopes were significant ($p < .05$). While vector 2 is significant for both of these participants as would be predicted, the vector 2 slope is not best-fitting, as defined by greatest correlation. For participant 4, the best-fitting is slope vector 4 with increasing slope from baseline through treatment, while for participant 8, the best fitting slope vector is 3 with increasing baseline but flat treatment. Given the instability in baseline data (i.e., FER during baseline), it is difficult to draw conclusions regarding the effect of treatment relative to baseline.

Intervention Video Task. For the Intervention Task Accuracy, a Wilcoxon signed-rank test indicated that accuracy scores for the intervention videos at the last treatment session (treatment session 10) were statistically higher than those in the first treatment session (treatment session 1) for the whole sample ($Z = -2.23$, $p = .03$ with large effect $r = .56$) as well as when the non-protocol subject (ID 8) was excluded from analyses ($Z = -2.04$, $p = .04$, $r = .55$). Looking at analyses across all ten sessions, a linear mixed effect model showed a significant linear relationship between the accuracy on the task and time points across participants. Both intercept and slope fixed effects were significant ($\beta_0 = 10.87$, $SE = 0.81$, $t = 13.39$, $p < .001$; $\beta_1 = 0.23$, $SE =$

0.07, $t = 3.43$, $p < .01$). These results suggest that accuracy on the Intervention Video Task and time point have a significant positive linear relationship with slope across participants. This result stayed significant after exclusion of the non-protocol participant (ID 8). In order to evaluate the relationship between the accuracy on the Intervention Task and time point for each participant separately, a linear model was fit. As shown in Table 6, none of the individual participant relationships were significant, even though combining across all participants, the model was significant, accounting for the variability across subjects. The individual variability likely resulted in larger error term resulting in individual patterns to not be significant. However, combining across all participants, the model accounted for variability and showed significant pattern. The estimated linear functions for each participant are displayed in Figure 2.

Child and Parent Report of Change in FER. On the Treatment Satisfaction Form, both participants and their parents were asked to rate their change in FER. For the statement: "At this point, my ability to recognize emotions has:" two children said "Stayed the Same", five children said "Slightly Increased", while one child said "Greatly Increased" ($M = 3.88$, $SD = 0.64$). On the parent survey, when asked "At this point, my child's ability to recognize emotions has:" three parents provided rating of "Stayed the Same", while the other five parents completed the statement as "Slightly Increased" ($M = 3.63$, $SD = 0.52$).

For the question of "What changes (if any) have you noticed regarding your child's emotion recognition skills?" two parents did not provide an answer, one parent noted they have not noticed any differences, while the other five saw changes from "more facial expression done now", "seems better at noticing non-verbal cues", "is better able to recognize his own emotions and those of others", "begun to ask family members about their feelings more and react to other's reactions more appropriately", and "seems to be more thoughtful – seems to have a better recognition that (child) has separate, distinct emotions."

Facial Emotion Expression Change

Video Emotion Expression Task – Scripted. Table 3 displays the accuracy of facial expression prior to and following the treatment (Pre-Tx: $M = 91.67$, $SD = 15.43$; Post-Tx: $M = 95.83$, $SD = 11.78$). The change in score was not statistically significant as a Wilcoxon signed-rank test indicated no significant difference between pre-treatment and post-treatment accuracy scores ($Z = -1.00$, $p = .32$, $r = -.25$). While not statistically significant, the observed change in scores on clarity and atypicality scores were of medium effect: clarity ($Z = -1.90$, $p = .06$, $r = -.47$), atypicality ($Z = -1.60$, $p = .11$, $r = -.40$). The pattern of results remained after removal of the non-protocol subject (ID 8) for accuracy ($Z = -1.00$, $p = .32$, $r = -.27$), clarity ($Z = -1.63$, $p = .10$, $r = -.43$), and atypicality ($Z = -1.60$, $p = .11$, $r = -.43$). No individual subject analyses were completed as no standardized norms exist for this measure.

Video Emotion Expression Task – Spontaneous. Table 3 displays the accuracy of facial expression during the spontaneous task (Pre-Tx: $M = 71.88$, $SD = 8.84$; Post-Tx: $M = 75.35$, $SD = 11.06$). The change in score was not statistically significant as a Wilcoxon signed-rank test indicated no significant difference between pre-treatment and post-treatment accuracy scores ($Z = -.84$, $p = .40$, $r = -.21$). While not statistically significant, the change in scores on clarity and atypicality was of medium effect size: clarity ($Z = -1.61$, $p = .11$, $r = -.40$), atypicality scores ($Z = -1.48$, $p = .14$, $r = -.37$). This result remained after removal of the non-protocol subject (ID 8) for accuracy ($Z = -.68$, $p = .50$, $r = -.18$), clarity ($Z = -1.52$, $p = .13$, $r = -.41$), and atypicality ($Z = -1.48$, $p = .14$, $r = .39$). No individual subject analyses were completed as no standardized norms exist for this measure.

Other Clinically Relevant Outcome Measures

SRS-2. A Wilcoxon signed-rank test indicated that the total scores on the SRS-2 at the post-intervention assessment were statistically lower than those in the pre-intervention assessment for the whole sample (*Z* = -2.37, *p* = .02, *r* = .59) and also for when the non-protocol participant (ID 8) was excluded (*Z* = -2.21, *p* = .03, *r* = .59). However, RCI analyses show that only two participants (IDs 5 and 7) showed reliable decrease (corresponding to lesser social impairment) in their score (RCI = -1.97, -3.03, respectively). None of the other participants showed a reliable change (Table 4).

TAS-C. A Wilcoxon signed-rank test indicated that the total scores as well as the subscale scores on the TAS-C were not statistically different at the post-intervention assessment compared to the pre-intervention assessment for the whole sample (Total: *Z* = -.42, *p* = .67, *r* = .11; Difficulty Identifying Feelings: *Z* = -1.09, *p* = .28, *r* = -.27; Difficulty Describing Feelings: *Z* = -.68, *p* = .50, *r* = -.17; Externally Oriented Thinking: *Z* = -.68, *p* = .50, *r* = -.17). This pattern of results is unchanged with removal of the non-protocol participant (Total: *Z* = -.27, *p* = .79, *r* = .07; Difficulty Identifying Feelings: *Z* = -.74, *p* = .46, *r* = -.20; Difficulty Describing Feelings: *Z* = 0.00, *p* = 1.00; *r* = 0; Externally Oriented Thinking: *Z* = -1.59, *p* = .11, *r* = .43). However, RCI analyses show that two of the participants (IDs 2 and 8) evidenced a reliable decrease (corresponding to lesser alexithymia difficulty) in total TAS-C score from pre to post treatment (RCI = -3.15, -2.63, respectively). None of the other participants showed a reliable change (Table 4).

CBCL. A Wilcoxon signed-rank test indicated that the total scores on the CBCL at the post-intervention assessment were statistically lower than those in the pre-intervention assessment for the whole sample (*Z* = -2.24, *p* = .03, *r* = -.56) and also for when the non-protocol participant (ID 8) was excluded (*Z* = -2.03, *p* = .04, *r* = -.54). However, this change was not observed for the Social Problems subscale on the CBCL (*Z* = -.42, *p* = .67, *r* = -.11). Based on the RCI analyses, three participants (IDs 2, 6, and 7) showed a reliable decrease in total problems (RCI = -2.45, -2.15, -3.07 respectively). None of the other participants showed a reliable change for the Total Score, and none of the eight participants showed a reliable change for the Social Problems subscale (Table 4).

ERSSQ. For the ERSSQ, a Wilcoxon signed-rank test indicated that accuracy scores for the intervention videos at the last treatment session (treatment session 10) were not statistically higher than those in the first treatment session (treatment session 1) for the whole sample (Z = -1.86, p = .06, r = -.46) as well as when the non-protocol subject (ID 8) was excluded from analyses (Z = -1.57, p = .12, r = -.42). The effect sizes, however, suggest moderate effect of the intervention on the ERSSQ scores. In addition, RCI analyses revealed that 3 of the participants (IDs 5, 7, and 8) showed significantly improved accuracy from the first session to the last treatment session (RCI = 2.93, 7.55, 7.77, respectively; Table 4). In addition, looking at analyses across all ten sessions, linear mixed effect model showed a significant linear relationship between the scores on the ERSSQ and time point. Both intercept and slope fixed effects were significant (β_0 = 52.46, SE = 3.77, t = 13.90, p <.001; β_1 = 1.18, SE = 0.42, t = 2.78, p <.01). These results suggest that score on the ERSSQ and time point have a significant positive linear relationship across participants. This result stayed significant after exclusion of the non-protocol participant (ID 8). In order to evaluate the relationship between ERSSQ and time for each participant separately, linear model was fit. As shown in Table 7, the relationship between ERSSQ and time was significant for four participants (IDs 2, 5, 7 and 8). The estimated linear functions for each participant are displayed in Figure 3.

YTP. Each parent was able to identify three main difficulties related to social-emotional skills. The identified problems fell into four categories: 1) Recognition of emotions and social cues (e.g., Recognition of peer reaction, difficulty recognizing nonverbal cues, not recognizing emotions), 2) Emotion expression (e.g., flat affect, overexpression of emotion), 3) Emotion regulation (e.g., managing frustration, overreaction to other's words and emotions), and 4) Social skills (e.g., difficulty with back and forth conversation, unable to talk to strangers). Linear model results for each problem identified, for each parent, are displayed in Table 8. As can be seen from the table, all but two of the participants (IDs 1 and 5) show a significant relationship between problem rating and time point for at least one problem behavior identified by the parent, with three participants (IDs 3, 4, 8) showing significant relationship for all three identified problems. These results suggest that problems with social-emotional behavior improved as treatment proceeded.

Anxiety Change. In order to explore whether increasing focus on the social aspects of interactions heighten anxiety, parents of the participants filled out the SCARED-P, while the participants completed SCARED-C and BFNE. As can be seen from Table 3, although scores on these anxiety measures showed decline, on average, from pre-treatment to post-treatment, a Wilcoxon signed-rank test indicated that scores on the BFNE and SCARED-C/P post-intervention assessment were not statistically different from those in the pre-intervention assessment (BFNE: $Z = -1.46$, $p = .14$, $r = -.39$; SCARED-P Total: $Z = -1.89$, $p = .06$, $r = -.51$; SCARED-C Total: $Z = 0.00$, $p = 1.00$, $r = 0$). In addition, RCI analyses showed that while none of the participants showed reliable increase on these anxiety measures, several participants showed a reliable decrease in anxiety. On the BFNE measure, two participants (IDs 2 and 7) both showed a reliable decrease (RCI = -2.48, -2.13, respectively). On the SCARED-C, one participant (ID 2) showed a reliable decrease in the total score (RCI = -6.42). On the SCARED-P, three participants (ID 1, 7, and 8) showed a reliable decrease in the total score (RCI = -3.29, -3.47, -2.08, respectively). Standard deviation and internal consistency (alpha) values for the RCI for BFNE were taken from Capriola, Maddox, and White (2017) study and for SCARED from Stern, Gadgil, Blakeley-Smith, Reaven, and Hepburn (2014), since both these studies had samples of youth with ASD.

Change in Amount of Fixation Duration to Facial Features (Hypothesis 4)

In order to assess change in gaze to socially-relevant cues following the intervention, we looked at eye gaze change for several tasks for which eye tracking was collected (i.e., Video Emotion Recognition Task, Video Emotion Expression Task, EU-Emotion Stimulus Set Task). We hypothesized that there would be significant change in the amount of fixation duration to facial area from pre to post intervention.

All participants were successfully calibrated for the eye-tracking task, such that the calibration procedure showed detection of gaze within all five of the predefined areas with no missing data and no excessive scatter. However, not all of the participants showed acceptable on-task percentage scores. On-task percentage score above 50% is a common benchmark for inclusion of participants within ASD eye-tracking studies (e.g., Fischer et al., 2014; Swanson, Serlin, & Siller, 2013). Therefore, on-task percentage scores above 25% for the entire task and above 50% per stimulus were used.

Eye tracking was also utilized during the Intervention Video Task, to ensure that the participants were looking at the box, when a box was shown, as instructed. Across the six participants for whom tracking was available, on average, while looking at the screen when the box was present, participants looked within the box 84.93% of the time, suggesting that participants followed the instruction. This looking was fairly consistent across participants (ID1: M = 84.00, SD = 18.46; ID2: M = 75.18, SD = 35.10; ID3: M = 89.42, SD = 11.03; ID4: M = 88.41, SD = 13.09; ID6: M = 89.02, SD = 12.60; ID7: M = 83.58, SD = 20.80). No videos with a box were tracked for two participants (IDs 5 and 8).

Video Emotion Recognition Task Eye Tracking. For the Video Emotion Recognition Task, 3 out of 8 participants (IDs 2, 5, and 8) did not show on-task percentage scores above 50% for at least one stimulus during the task. In addition, one participant (ID 1) showed overall tracking of < 25% for the overall pre-intervention session. Therefore, only the four remaining participants (IDs 3, 4, 6, and 7) were analyzed. Table 9 displays the average fixation duration to the facial area during pre-treatment and post-treatment sessions for these participants. Wilcoxon signed-rank test indicated that the fixation duration to the face area at post-intervention assessment was not statistically higher than during the pre-intervention assessment (Z = -.73, p = .47, r = -.26).

Video Emotion Expression Task Eye Tracking. For the Video Emotion Expression Task spontaneous condition, 1 participant (ID 5) did not show on-task percentage above 50% for at least one stimulus and an additional 2 participants (IDs 2 and 8) showed overall tracking to fall below 25% for either the pre-treatment or post-treatment session. Therefore, only the five remaining participants (IDs 1, 3, 4, 6, and 7) were analyzed. Table 9 displays the average fixation duration to the face area during pre-treatment and post-treatment sessions for these participants. Wilcoxon signed-rank test indicated that the fixation duration to the face area at post-intervention assessment was not statistically higher than during the pre-intervention assessment (Z = -1.75, p = .08), despite a large effect size (r = -.55).

EU-Emotion Stimulus Set Task Eye Tracking. During the EU task, eye tracking was collected for all participants and every participant showed at least a 50% on-task percentage for at least one of the treatment sessions. Figure 4 shows the eye gaze duration to the face region for every participant. For the pre to post assessment analyses, data from the first session and from the last treatment session are used. Only one participant (ID 2) was excluded as this participant's last treatment session eye tracking fell below threshold. Table 9 displays the average fixation duration to the face region for the seven included participants. Wilcoxon signed-rank test indicated that fixation durations to the face area for the last treatment session (treatment session 10) were not statistically higher than those for the first treatment session (treatment session 1) for the whole sample (Z = -1.69, p = .09). However, the medium to large effect size (r = -.45) indicates that a change in eye gaze to face occurred but we may be underpowered to detect statistically significant change.

Similarly, looking at analyses across all sessions, the linear mixed effect model also did not show a significant linear relationship for the fixation duration to the face region (β_1 = -43.75, SE = 40.65, t = -1.08, p =.29). In order to evaluate the relationship between fixation duration to the face region and time for each participant separately, linear model was fit. The relationship between fixation duration to the face region and time was not significant for any of the participants (all $t < 1.83$, all $p > .13$). The estimated linear functions for each participant are displayed in Figure 4.

Discussion

This study sought to examine the feasibility and preliminary impact of a novel attention retraining program to attenuate facial emotion recognition deficits in children with ASD. Results suggest that the new attention modification program is feasible and acceptable to children as well as their parents. In terms of feasibility, 87.5% (n = 7) of children were able to complete the program in the specified time frame, with an additional child completing the program within a shorter time frame. Only one family withdrew, due to factors unrelated to the intervention (e.g., travel distance). In addition, the recruitment timeframe was much longer than expected, mainly due to difficulties with travel to and from sessions. As such, results suggest that the program may be most feasible to implement when travel barriers can be overcome (e.g., in the child's home or school), due to the twice weekly session commitment. Viability of sample ascertainment is a very important consideration for further, large-scale evaluation of the intervention (Leon, Davis, & Kraemer, 2011). For the participants who completed the study however, all but one child rated the program to be acceptable in terms of time and commitment and all parents noted that the program is acceptable. Similarly, for all other items on the treatment satisfaction forms, both parents and children provided either neutral or positive ratings regarding liking the program, the program being helpful, and recommending the program to a friend. Overall, based on parent and child reports, the developed attention retraining program is seen as acceptable to the families.

Within an experimental therapeutic approach, it is important to show that it is possible to change the most proximal mechanism, which theoretically leads to secondary changes. If the mechanism cannot be changed, the intervention will not be successful in leading to meaningful clinical improvement. In term of the eye gaze manipulation, when instructed to look at the box highlighting the facial features of the intervention videos, the participants generally did so (85% of the time), indicating that the manipulation was successful in getting the children's attention. Some variability is expected due to the complexity of the videos presented and therefore, the 85% looking time is acceptable. Assessment of the manipulation is an important consideration in evaluation of the program to assure that the manipulation is successful in producing the intended behavior, which for this study is looking at the area highlighting the face.

In addition, it is important to consider potential adverse effects following an intervention. Clinically, there is evidence that social anxiety may increase during social interventions, potentially due to the increased awareness of others' evaluations and perceptions (White, 2011). Therefore, we assessed for social anxiety and found that, based on self- and parent-report, social anxiety actually decreased or remained constant during intervention. Significant increases in social anxiety were not found for any children for any of the measures, indicating that the current intervention did not have untoward effect of increased social anxiety. The decrease in social anxiety for a few of the children is interesting and could be related to a decrease in their general impairment which may in turn decrease their anxiety regarding social interactions. This potential relationship and effect of the intervention on social anxiety was not directly tested and is only speculative.

In terms of the impact of this attention retraining intervention on FER, results are mixed. While we hypothesized that there would be a significant improvement in participants' FER ability from pre to post intervention, the results are not consistent across the different measures we used to assess FER ability. On the only standardized measure of FER utilized in this study (AR subtest of NEPSY-II), only one participant showed a reliable increase in FER score. Importantly however, this participant was the only child in the study to score below average on this measure prior to the intervention. All other participants scored in the average or above average range prior to the intervention, and therefore ceiling effects may have impacted ability to see movement on this measure.

For the other assessed FER tasks, results suggest variability based on the individual and the measure. On the group level, performance on the Video Emotion Recognition task following the intervention was not significant. On the other hand, while participants significantly increased their performance on the Intervention Video Task from the first intervention session to the last intervention session, and the overall average trajectory over the ten treatment sessions was significant, the individual participant trajectories showed that no participants showed a consistent increase in accuracy over time. This could be due to variable difficulty of emotion identification throughout the sessions or due to actual fluctuation of the scores (i.e., instability of the measure). Similarly, for the EU-Emotion Stimulus Set Task, only two participants showed increasing scores; however, as the scores increased prior to intervention (i.e., during the baseline period), it is difficult to draw conclusions regarding the effect of treatment relative to baseline. While the results suggest general improvement on the group level for some of the FER measures, the subject-level data suggest the intervention was not successful at increasing FER on these measures.

Although the majority of the participants did not show impairment in FER on the AR subtest, all parents of the participants indicated impairment in FER at the start of treatment. This discrepancy in parent report and child's performance on this standardized measure of FER suggests that the AR task may not be sensitive to detect FER impairments seen in the natural, daily setting observed by the parents. While parent report is not always accurate, our prior studies show that parents of children with ASD are variable in their reports of emotional difficulties. Not all parents see FER or FEE problems, even in confirmed ASD samples (e.g., White et al., 2018), suggesting usefulness of parent report in assessing for the deficits. Based on parent's report of their child's ability to recognize emotions following the intervention, five out of the eight parents noted slight increase in their child's FER, indicating potential positive impact of the intervention on FER from the parent's perspective.

Altogether, the results from this study highlight the variability in FER across measurement modality. These results suggest that FER as a construct may vary and present differently both across and within individuals, suggesting that more discrete behaviors related to FER (e.g., response time) are needed for single-subject research. Results from this study as well as variable findings in prior studies suggest that current FER measures may lack ecological validity, as all parents in the study reported their children showing FER impairment. However, it is also plausible that parents ascribe FER impairment when there is some other process, such as social apathy, that may account for the lack of appropriate emotion identification in the context of intact FER ability.

As FER is related to FEE, we evaluated the potential impact of the intervention on facial emotion expression abilities. For both scripted and spontaneous conditions, while the overall average score increased slightly, the increase was not statistically significant. This suggests that the children's ability to express emotions in terms of accuracy, clarity, and lack of atypicality was not impacted by the intervention.

Secondary to its impact on FER abilities, we aimed to explore preliminary impact of the attention retraining intervention on more distal and clinically relevant outcomes, including social competence. Interestingly, all eight subjects showed a decrease in the SRS-2 total score, resulting in a large effect size (r = -.59), indicating decreased social impairment; however, the decrease was reliable for only two of the participants. Similarly, a decrease in problem behaviors was seen on the CBCL for three of the participants, even though no participants showed a decrease on the social subscale, indicating more general decrease in problem behaviors outside of social skills. In addition, a decrease was observed for child report of alexithymia for two participants.

The most consistent outcome observed was for the two measures that parents completed at each treatment session over the five weeks. For the ERSSQ measure, which detects emotion recognition and social skills, four parents rated their children as increasing in these skills over the 10 sessions. In addition, on the YTP, which assesses parent identified behavioral change, six out of the eight parents rated at least two of the identified problems to have decreased over the course of the study. As with the FER measures, this result suggests the importance of incorporating parent ratings of problems they identify is most relevant to their children in addition to the normed measures, as the most consistent change occurred for the questions that were tailored to each child (i.e., identified by the parent). While most of these outcome measures showed a significant improvement for at least one child, it is important to point out that improvements were found for different children. For example, while participant 5 did not show improvement on the YTP, the child showed improvement based on the SRS-2 and ERSSQ.

Overall, seven out of the eight children showed improvement on at least one outcome measure, with participant 1 being the only participant to show no improvement on any clinically relevant outcome measures tested; however, the child's parent reported a slight increase in FER ability on the treatment satisfaction form. On the other hand, participant 7 showed improvement on SRS-2, ERSSQ, CBCL, and YTP, but no change in the FER or FEE measures or parent report of FER change. Therefore, while in general, children with ASD show improvements following the intervention in at least one assessed area, the areas are not consistent for participants and cannot be attributed to changes in FER. It is possible that the changes occur due to potential increased awareness of the FER difficulties given enrollment in the intervention. Further exploration with larger samples and a control condition is needed to more fully understand the factors leading to such changes.

Lastly, we assessed change in gaze to socially relevant cues, based on our hypothesis that the attention retraining would increase gaze toward faces (target engagement). Results suggest that change in gaze was not significant from pre to post intervention, or during the course of the treatment sessions (i.e., during EU-Emotion Stimulus Set Task). While the paradigm manipulation, which highlighted the facial area in the videos, was successful in terms of children looking within the area when the box was presented, the manipulation did not successfully alter the gaze during other viewing tasks. We failed to demonstrate target engagement (i.e., increased gaze to faces). This may be due to a host of factors including incomplete data or the stimuli used for the study. Perhaps, there is other mechanism behind change in FER that was reported by the parents, as the increase in FER reported by the parents for most of the participants cannot be attributed to changes in viewing patterns (i.e., greater looking at the faces). However, data loss severely limited ability to detect change in an already small sample and the medium to large effect size indicates that the change may be meaningful, but undetectable given the study's limited power.

Limitations

These findings should be considered in light of the limitations of this study. A primary limitation is the inclusion criterion of diminished FER based on parent report only. We found average to above average FER abilities for all but one participant on the standardized measure of FER (i.e., AR task on the NEPSY-II). Even though all parents of the participants reported FER difficulties, these difficulties were not observed on the standardized measure, limiting the ability to observe any significant change. This however is not a design failure per se, as there are no established measures that are sensitive to FER change in this population (Wieckowski et al., In preparation). In addition, a limitation for the current analyses was instability of the baseline for the EU-Emotion Stimulus Set Task for some of the participants, indicating variability and fluctuation of the measure, limiting ability to detect change due to the intervention. An important limitation is the loss of eye tracking data for many participants due to technical problems with the eye tracker and the stimuli used for the study. The stimuli used for the Intervention Task included dark backgrounds which resulted in lower gaze tracking compared to the tracking during the EU-Emotion Stimulus Set, which has stimuli presented on a white background. The contrast between the darker stimuli presented on lighter background resulted in greater overall tracking for all participants. The lack of this contrast for the Intervention Task videos however resulted in analyses regarding change in gaze to not be possible for half of the sample, limiting the ability to detect any significant changes in gaze following the intervention. Another limitation was the focus on accuracy coding of expression by undergraduate research assistants. The agreement between coders was low, given innate subjectivity in coding of emotions by people. Therefore, all codes were discussed and consensus codes were used for analyses. Lastly, in terms of participant characteristics, all but one participant was Caucasian and all participants had average to high average IQ, limiting generalizability. In addition, half of the participants had co-occurring psychopathologies, including anxiety, ADHD, and learning disorder, that may have affected the findings.

Future Directions

While the findings of this pilot study indicate feasibility and acceptability of a novel attention retraining intervention to improve FER in children with ASD, the impact on change in both FER (target engagement) and clinical outcomes is less clear. In addition to addressing the limitations highlighted above, given the variability in FER deficiency among children with ASD, it is important to ensure stability of FER impairment which may require adaptations or even creation of new measures to assess FER impairment in individuals with ASD that are able to detect actual change in the process following an intervention. Given the difficulty with sample ascertainment mostly due to travel concerns, the program could be adapted to be completed remotely in the future. In addition, prior to assessment of feasibility of attention retraining in individuals with ASD, it may be beneficial to assess the degree to which FER is modifiable in a tightly controlled experimental design to assure that the construct is able to be altered. Lastly, in future studies, it will be important to look at whether the proposed manipulation of highlighting the facial area increases processing of facial emotions or more general perception of social scenes (i.e., more attention to details in general) in order to better understand the mechanism behind the behavioral difficulties.

Conclusion

This study was the first attempt to develop and implement a brief attention retraining program to attenuate facial emotion recognition deficits in children with ASD. The results of this pilot study indicate the program is feasible to implement as planned and acceptable to the children and their parents. We endeavored to examine preliminary impact of the intervention on facial emotion recognition and social skills, with results suggesting that the program is successful at increasing FER based on parent and child report and decreasing parent identified socioemotional problem behaviors. However, change in FER as measured via behavioral tasks is not apparent, except when FER was considerably below average prior to the intervention. In addition, assessment of change in gaze revealed no significant change to socially-relevant cues following the intervention, which is inconsistent with the hypothesized mechanism behind the intervention in increasing gaze to socially relevant stimuli. Further exploration into FER as a construct and the best way to assess FER in an ecologically valid way is necessary due to the observed variability. Larger scale randomized controlled studies with youth with confirmed FER deficiency and use of treatment-naïve ratings of change in FER and socio-emotional functioning are needed to fully explore the impact of the attention modification program.

References

Achenbach, T. M., & Rescorla, L. A. (2001). Manual for the ASEBA school-age forms and profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The Twenty-item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, 38, 23-32.

Beaumont, R., & Sofronoff, K. (2008). A multi-component social skills intervention for children with Asperger syndrome: The junior detective training program. Journal of Child Psychology and Psychiatry, 49(7), 743–753. doi:10.1111/j.1469-7610.2008.01920

Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997). The screen for child anxiety related emotional disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(4), 545-553.

Bishara, S. E., Peterson, L. C., & Bishara, E. C. (1984). Changes in facial dimensions and relationships between the ages of 5 and 25 years. American journal of orthodontics, 85(3), 238-252.

Black, M. H., Chen, N. T., Iyer, K. K., Lipp, O. V., Bölte, S., Falkmer, M., ... & Girdler, S. (2017). Mechanisms of facial emotion recognition in autism spectrum disorders: insights from eye tracking and electroencephalography. *Neuroscience & Biobehavioral Reviews*, *80*, 488-515.

Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. American Psychologist, 63, 77-95.

Capps, L., Yirmiya, N., & Sigman, M. (1992). Understanding of simple and complex emotions in non-retarded children with autism. *Journal of Child Psychology and Psychiatry*, *33*, 1169–1182.

Capriola, N. N., Maddox, B. B., & White, S. W. (2017). No offense intended: Fear of negative evaluation in adolescents and adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *47*(12), 3803-3813.

Carleton, R. N., Collimore, K. C., McCabe, R. E., & Antony, M. M. (2011). Addressing revisions to the Brief Fear of Negative Evaluation scale: Measuring fear of negative evaluation across anxiety and mood disorders. *Journal of Anxiety Disorders*, *25*(6), 822-828.

Chevallier, C., Parish-Morris, J., McVey, A., Rump, K. M., Sasson, N. J., Herrington, J. D., & Schultz, R. T. (2015). Measuring social attention and motivation in autism spectrum disorder using eye-tracking: stimulus type matters. *Autism Research, 8*(5), 620-628.

Constantino, J. N., & Gruber, C. (2012). *The Social Responsiveness Scale – Second edition (SRS-2).* Torrance, CA: Western Psychological Services.

Dallery, J., Cassidy, R. N., & Raiff, B. R. (2013). Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of Medical Internet Research*, *15*(2).

Davies, M. S., Dapretto, M., Sigman, M., Sepeta, L., & Bookheimer, S. Y. (2011). Neural bases of gaze and emotion processing in children with autism spectrum disorders. *Brain and Behavior*, *1*(1), 1-11.

Eimer, M., Gosling, A., Nicholas, S., & Kiss, M. (2011). The N170 component and its links to configural face processing: a rapid neural adaptation study. *Brain Research. 1376*, 76–87.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169-200.

Field, T.M., Woodson, R.W., Greenberg, R., & Cohen, C. (1982). Discrimination and imitation of facial expressions by neonates. *Science, 218,* 179–181.

Fletcher-Watson, S., Leekam, S.R., Benson, V., Frank, M.C., & Findlay, J.M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. Neuropsychologia, 47, 248-257.

Freeth, M., Foulsham, T., & Chapman, P. (2011). The influence of visual saliency on fixation patterns in individuals with Autism Spectrum Disorders. *Neuropsychologia*, *49*(1), 156-160.

Goldsmith, T. R., & LeBlanc, L. A. (2004). Use of technology in interventions for children with autism. *Journal of Early and Intensive Behavior Intervention*, *1*(2), 166-177.

Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. R. (2000). Verbal bias in recognition of facial emotions in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *41*, 369–379.

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. Neuroscience and Biobehavioral Reviews, 42, 279-297.

Hakamata, Y., Lissek, S., Bar-Haim, Y., Britton, J. C., Fox, N. A., Leibenluft, E., ... & Pine, D. S. (2010). Attention bias modification treatment: a meta-analysis toward the establishment of novel treatment for anxiety. *Biological Psychiatry*, *68*(11), 982-990.

Hanley, M., McPhillips, M., Mulhern, G., & Riby, D. M. (2013). Spontaneous attention to faces in Asperger syndrome using ecologically valid static stimuli. *Autism*, *17*(6), 754-761.

Hanley, M., Riby, D. M., Carty, C., McAteer, A. M., Kennedy, A., McPhillips, M. (2015). The use of eye-tracking to explore social difficulties in cognitively able students with autism spectrum disorder: A pilot investigation. *Autism*, 1-6. doi:10.1177/1362361315580767

Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007). A fine-grained analysis of facial expression processing in high-functioning adults with autism. *Neuropsychologia*, *45*, 685–695.

Hutt, C., & Ounsted, C. (1966). The biological significance of gaze aversion with particular reference to the syndrome of infantile autism. *Behavioral Science*, *11*(5), 346-356.

Iverson, G. L. (2011). Reliable change index. In *Encyclopedia of clinical neuropsychology* (pp. 2150-2153). Springer, New York, NY.

Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching audlts predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry, 65*(8), 946-954.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature, 504*, 427-431.

Kauo, J. L., & Egel., A. L. (2016). The effectiveness of interventions in teaching emotion recognition to children with autism spectrum disorders. *Review Journal of Autism and Developmental Disorders, 3,* 254:265.

Kazdin, A. E. (1978). Methodological and interpretive problems of single-case experimental designs. *Journal of Consulting and Clinical Psychology*, *46*(4), 629.

Keehn, B., Müller, R. A., & Townsend, J. (2013). Atypical attentional networks and the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *37*(2), 164-183.

Kliemann, D., Dziobek, I., Hatri, A., Steimke, R., & Heekeren, H. R. (2010). Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. *The Journal of Neuroscience, 30*(37), 12281-12287.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (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. doi:10.1001/archpsyc.59.9.809

Korkman, M., Kirk, U., & Kemp, S. (2007). NEPSY-II: A developmental neuropsychological assessment. San Antonio, TX: The Psychological Corporation.

Langdell, T. (1981). Face Perception: An approach to the study of autism. *Doctoral Thesis*. University of London.

Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*(3), 371-375.

Lee, C. S., Lam, S. H., Tsang, S. T., Yuen, C. M., & Ng, C. K. (2018). The effectiveness of technology-based intervention in improving emotion recognition through facial expression in people with autism spectrum disorder: a systematic review. *Review Journal of Autism and Developmental Disorders*, *5*(2), 91-104.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*(5), 626-629.

Lindner, J. L., & Rosén, L. A. (2006). Decoding of emotion through facial expression, prosody, and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *36*, 769–777.

Linetzky, M., Pergamin-Hight, L., Pine, D. S., & Bar-Haim, Y. (2015). Quantitative evaluation of the clinical efficacy of attention bias modification treatment for anxiety disorders. *Depression and Anxiety*, *32*(6), 383-391.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism diagnostic observation schedule (ADOS-2)* (2nd ed.). Torrence, CA: Western Psychological Services.

Moriuchi, J. M., Klin, A., & Jones, W. (2016). Mechanisms of diminished attention to eyes in autism. *American Journal of Psychiatry*, *174*(1), 26-35.

Nuske, H.J, Vivanti, G., Dissanayake, C. (2013). Are emotion impairments unique to, universal, or specific in autism spectrum disorder? A comprehensive review. *Cognition & Emotion, 27(6).*

Ollendick, T. H., White, S. W., Richey, J., Kim-Spoon, J., Ryan, S. M., Wieckowski, A. T., ... & Smith, M. (2018). Attention bias modification treatment for adolescents with social anxiety disorder. *Behavior Therapy*.

O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., ... & Lundqvist, D. (2016). The EU-emotion stimulus set: a validation study. *Behavior Research Methods*, *48*(2), 567-576.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 32*, 249-261. doi:10.1023/A:1016374617369

Philip, R. C. M., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., ... & Hall, J. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological medicine*, *40*(11), 1919-1929.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25-42.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73-89.

Pugliese, C.E. (2012). A pilot study examining the feasibility and preliminary efficacy of problem solving therapy in college studies with autism spectrum disorders (Doctoral dissertation). Retrieved from https://vtechworks.lib.vt.edu/

Richler, J. J., Floyd, R. J., & Gauthier, I. (2014). The Vanderbilt Holistic Face Processing Test: A short and reliable measure of holistic face processing. *Journal of Vision*, *14*(11), 10-10.

Riby, D. M., & Hancock, P. J. (2009). Do faces capture the attention of individuals with Williams syndrome or autism? Evidence from tracking eye movements. *Journal of Autism and Developmental Disorders*, *39*(3), 421-431.

Rice, K., Moriuchi, J. M., Jones, W., & Klin, A. (2012). Parsing heterogeneity in autism spectrum disorders: Visual scanning of dynamic social scenes in school-aged children. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(3), 238-248.

Rieffe, C., Oosterveld, P., & Meerum Terwogt, M. (2006). An alexithymia questionnaire for children: Factorial and concurrent validation results. *Personality and Individual Differences, 40*, 123-133.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (231-244). New York, NY: Sage.

Rump, K. M., Giovannelli, J.L., Minshew, N.J., & Strauss, M.S. (2009). The development of emotion recognition in individuals with autism. *Child Development*, 80(5), 1434-1447.

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., & Grossman, R. B. (2017). Neurotypical peers are less willing to interact with those with autism based on thin slice judgments. *Scientific reports*, *7*, 40700.

Scambler, D. J., Hepburn, S., Rutherford, M. D., Wehner, E. A., & Rogers, S. J. (2007). Emotional responsivity in children with autism, children with other developmental disabilities, and children with typical development. *Journal of Autism and Developmental Disorders*, *37*(3), 553-563.

Schmidt, N. B., Richey, J. A., Buckner, J. D., & Timpano, K. R. (2009). Attention training for generalized social anxiety disorder. *Journal of Abnormal Psychology*, *118*(1), 5.

Senju, A., & Johnson, M. H. (2009). Atypical eye contact in autism: models, mechanisms and development. *Neuroscience & Biobehavioral Reviews*, *33*(8), 1204-1214.

Stagg, S. D., Slavny, R., Hand, C., Cardoso, A., & Smith, P. (2014). Does facial expressivity count? How typically developing children respond initially to children with autism. *Autism*, *18*(6), 704-711.

Stern, J. A., Gadgil, M. S., Blakeley-Smith, A., Reaven, J. A., & Hepburn, S. L. (2014). Psychometric properties of the SCARED in youth with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *8*(9), 1225-1234.

Tantam, D., Monaghan, L., Nicholson, H., & Stirling, J. (1989). Autistic children's ability to interpret faces: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *30,* 623–630.

Trevisan, D. A., & Birmingham, E. (2016). Are emotion recognition abilities related to everyday social functioning in ASD? A meta-analysis. *Research in Autism Spectrum Disorders, 32,* 24-42.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33*(4), 448.

Wechsler D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II)*. San Antonio, TX: NCS Pearson.

White, S. W. (2011). *Social skills training for children with Asperger syndrome and high-functioning autism*. Guilford Press.

White, S. W., Abbott, L., Wieckowski, A. T., Capriola-Hall, N. N., Aly, S., & Youssef, A. (2018). Feasibility of automated training for facial emotion expression and recognition in autism. *Behavior Therapy.*

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology, 39(*1), 114-128.

Wieckowski, A. T., Flynn, L. T., Richey, J. A., Gracanin, D., & White, S. W. (In Preparation). Measuring change in facial emotion recognition in individuals with Autism Spectrum Disorder: A systematic review.

Wieckowski, A. T., & White, S. W. (2017). Eye-gaze analysis of facial emotion recognition and expression in adolescents with ASD. *Journal of Clinical Child & Adolescent Psychology*, *46*(1), 110-124.

Williams, B. T., Gray, K. M., & Tonge, B. J. (2012). Teaching emotion recognition skills to young children with autism: a randomised controlled trial of an emotion training programme. *Journal of Child Psychology and Psychiatry*, *53*(12), 1268-1276.

Wilson, C. E., Brock, J., & Palermo, R. (2010). Attention to social stimuli and facial identity recognition skills in autism spectrum disorder. *Journal of Intellectual Disability Research*, *54*(12), 1104-1115.

White, S. W., Schry, A. R., & Kreiser, N. L. (2014). Social worries and difficulties: autism and/or social anxiety disorder?. In *Handbook of autism and anxiety* (pp. 121-136). Springer, Cham.

Yirmiya, N., Kasari, C., Sigman, M., & Mundy, P. (1989). Facial Expressions of Affect in Autistic, Mentally Retarded and Normal children. *Journal of Child Psychology and Psychiatry.* 30(5), 725-735.

Table 1

Demographic Characteristics (n = 8)

	Mean (*SD*)
Age (in years)	10.93 (1.29)
FSIQ-2	110.89 (13.60)
SRS-2	78.13 (6.75)
ADOS Total Score	9.11 (2.20)

	N (%)
Gender	
Male	5 (62.5)
Female	3 (37.5)
Race	
Caucasian	8 (100)
Hispanic/Latino	1 (12.5)
Diagnoses[1]	
ASD	8 (100)
Anxiety	1 (12.5)
OCD	0
ADHD	3 (37.5)
Depression	0
ID/LD	1 (12.5)
Receiving Services	5 (62.5)
Medications	
None	6 (75.0)
ADHD	1 (12.5)
Asthma/Allergy	1 (12.5)
FER/FEE Difficulties	
Difficulty Recognizing Emotions	8 (100)
Limited Range of Facial Expressions	8 (100)
Expressions not appropriate to situation	4 (50.0)

[1]Diagnoses were determined based on parent report on the demographics questionnaire.

Table 2

Assessment Schedule

Measure	Intake	Pre	During	Post
Demographic Form (parent)	X			
ADOS-2 (child)	X			
WASI-2 (child)	X			
AR (NEPSY-II) (child)		X		X
FER/FEE (child)		X		X
SRS-2 (parent)	X			X
TAS-C (child)		X		X
BFNE (child)		X		X
SCARED (child and parent)		X		X
CBCL/6-18 (parent)	X			X
ERSSQ (parent)			X	
YTP (parent)			X	
EU-Emotion Stimulus Set (child)			X	
Intervention Videos (child)			X	
Treatment Satisfaction (parent and child)				X

Table 3

Pre Treatment and Post Treatment Measure Scores (n = 8)

	Pre-Tx Mean (*SD*)	Post-Tx Mean (*SD*)	Sig. *Z* (*p*)
SRS-2			
Total T-Score	78.13 (6.75)	72.25 (9.54)	-2.37 (.02)*
NEPSY-II Affect Recognition			
Total Scaled Score	10.00 (3.12)	10.63 (1.51)	-0.17 (.86)
Video Emotion Recognition Task Accuracy			
Total (%)	68.83 (12.19)	72.57 (17.22)	-0.70 (.48)
Video Emotion Expression Task Scripted			
Total Correct (%)	91.67 (15.43)	95.83 (11.78)	-1.00 (.32)
Clarity (%)	73.96 (19.13)	84.38 (13.68)	-1.90 (.06)
Atypicality (%)	9.38 (14.39)	13.54 (13.32)	-1.60 (.11)
Video Emotion Expression Task Spontaneous			
Total Correct (%)	71.88 (8.84)	75.35 (11.06)	-0.84 (.40)
Clarity (%)	71.53 (8.66)	76.39 (10.07)	-1.61 (.11)
Atypicality (%)	3.13 (4.05)	0.52 (1.03)	-1.48 (.14)
TAS-C			
Total Score	14.50 (5.66)	13.00 (6.95)	-0.42 (.67)
Difficulty Identifying Feelings	4.13 (3.18)	3.00 (3.38)	-1.09 (.28)
Difficulty Describing Feelings	3.50 (2.39)	2.63 (2.83)	-0.68 (.50)
Externally Oriented Thinking	6.88 (2.70)	7.38 (2.00)	-0.68 (.50)
BFNE	16.13 (7.45)	11.63 (4.21)	-1.46 (.14)
SCARED-Child			
Total Score	17.13 (14.91)	12.50 (8.43)	0.00 (1.00)
SCARED-Parent			
Total Score	26.00 (18.21)	18.38 (14.51)	-1.89 (.06)
CBCL (T-Scores)			
Social Problems	62.88 (9.23)	62.13 (10.05)	-0.42 (.67)
Total Problems	66.88 (5.19)	62.38 (5.88)	-2.24 (.03)*

Note. Significance is based on Wilcoxon Signed Rank Test analyses across group only. * $p < .05$

Table 4

Reliable Change Indices for Outcome Variables From Baseline to Endpoint

ID		AR	SRS-2	TAS-C	ERSSQ	CBCL-T	CBCL-Social
1	BL	11	112	19	44	69	57
	EP	10	108	26	47	65	59
	RCI	-.37	-.30	1.84	.63	-1.09	.45
2	BL	12	114	23	64	69	59
	EP	13	107	11	73	60	52
	RCI	.45	-.53	-3.15*	1.89	-2.45*	-1.58
3	BL	9	108	11	49	65	65
	EP	9	102	13	51	61	70
	RCI	0	-.45	.53	0.42	-1.24	1.15
4	BL	10	74	10	52	64	69
	EP	12	73	13	48	66	69
	RCI	.90	-.08	.79	-.83	.61	0
5	BL	13	99	7	69	59	67
	EP	12	73	7	83	58	60
	RCI	-.45	-1.97*	0	2.93*	-.31	-1.61
6	BL	3	99	20	56	67	53
	EP	10	77	20	56	60	56
	RCI	2.57*	-1.67	0	0	-2.15*	.69
7	BL	10	92	11	42	65	53
	EP	9	52	9	78	55	51
	RCI	-.37	-3.03*	-.53	7.55*	-3.07*	-.46
8	BL	12	108	15	53	77	80
	EP	10	97	5	71	74	80
	RCI	-.90	-.83	-2.63*	3.77*	-.92	0

Note. BL = baseline, EP = endpoint, RCI = Reliable Change Index

Norms for RCI for AR obtained from Korkman et al., 2007, SRS-2 (Norms obtained from Constantino & Gruber, 2012), TAS-C (Norms obtained White et al. (unpublished)); alpha was used instead of test-retest), ERSSQ (Norms from Beaumont & Sofronoff, 2008; alpha was used instead of test-retest), CBCL-T = total subscale score for Child Behavior Checklist, CBCL-Social = Social subscale score for Child Behavior Checklist (Norms from Achenbach & Rescorla, 2011), * Reliable improvement (RCI ≤ or ≥ 1.96)

Table 5

SMA Analyses for EU-Emotion Stimulus Set Task

ID	Level Change	Slope Change				
		Vector1	Vector 2	Vector 3	Vector4	Vector 5
1 [4]	.68 (.02)*	.14 (.71)	.30 (.42)	.55 (.08)	.35 (.33)	.10 (.79)
2 [5]	-.14 (.60)	-.02 (.94)	-.08 (.76)	-.21 (.43)	-.13 (.63)	-.08 (.78)
3 [4]	.41 (.20)	-.17 (.62)	.29 (.40)	.45 (.15)	.34 (.32)	.24 (.48)
4 [6]	.55 (.06)	-.27 (.41)	.61 (.04)*	.70 (.01)*	.70 (.01)*	.61(.03)*
5 [6]	-.30 (.38)	.20 (.57)	-.13 (.72)	.04 (.91)	-.05 (.90)	.24 (.49)
6 [4]	.12 (.55)	-.20 (.32)	.26 (.19)	.25 (.21)	.29 (.15)	.33 (.11)
7 [5]	-.21 (.53)	-.26 (.43)	.09 (.78)	-.25 (.44)	-.00 (1.00)	.18 (.58)
8 [5]	.79 (.002)*	-.30 (.35)	.59 (.04)*	.75 (.003)*	.67 (.02)*	.38 (.22)

Note. Change in level and slopes with p-values designated in parentheses. The numbers in brackets by the ID indicate the number of baseline points. * = $p < .05$

Table 6

Relationship Between Accuracy on the Intervention Task and Time for Each Participant

ID	Estimate	SE	t-value	p-value
1	0.19	0.26	0.76	.47
2	.022	0.15	1.44	.19
3	0.14	0.10	1.40	.20
4	0.18	0.11	1.71	.13
5	0.09	0.13	.68	.52
6	0.29	0.28	1.03	.33
7	.43	0.27	1.61	.14
8	.27	0.15	1.73	.12

Table 7

Relationship Between ERSSQ Score and Time for Each Participant

ID	Estimate	SE	t-value	p-value

1	0.41	0.21	1.89	.10
2	.83	.27	3.10	.01*
3	-0.10	0.41	-.24	.82
4	-0.16	0.30	-.55	.60
5	1.36	0.28	4.84	<.01*
6	1.32	0.67	1.97	.08
7	3.01	0.72	4.19	<.01*
8	2.74	0.64	4.29	<.01*

Note. * indicates significant relationship

Table 8

Relationship Between Problem Behaviors Identified on the YTP and Time for Each Participant

ID	Problem	Estimate	SE	t-value	p-value
1	1. EE	0.00	0.00	1.73	.12

	2. Rec	-0.02	0.08	-0.24	.81
	3. SS	0.01	0.03	0.30	.77
2	1. ER	0.15	0.13	1.18	.27
	2. EE	-0.10	0.05	-1.98	.08
	3. Rec	-0.17	0.05	-3.30	.01*
3	1. Rec	-0.31	0.05	-6.53	<.001*
	2. Rec	-0.25	0.05	-5.11	<.001*
	3. EE	-0.18	0.05	-4.05	<.01*
4	1. ER	-0.16	0.04	-4.63	<.01*
	2. Rec	-0.15	0.06	-2.46	.04*
	3. Rec	-0.29	0.08	-3.49	<.01*
5	1. Rec	0.02	0.04	0.41	.69
	2. ER	-0.04	0.05	-0.71	.49
	3. Rec	0.03	0.05	0.55	.60
6	1. Rec	-0.18	0.06	-3.16	.01*
	2. EE	-0.08	0.09	-0.91	.39
	3. SS	-0.23	.08	-2.80	.02*
7	1. Rec	-0.33	0.04	-7.56	<.001*
	2. EE	-0.37	0.07	-5.40	<.001*
	3. Rec	-0.31	0.07	-4.15	<.01*
8	1. SS	-0.22	0.08	-2.60	.03*
	2. ER	-0.41	0.06	-6.33	<.001*
	3. EE	-0.11	0.08	-1.36	.21

Note. EE = emotion expression; Rec = recognition of emotions and social cues; ER = emotion regulation; SS = social skills. * designates significant relationship

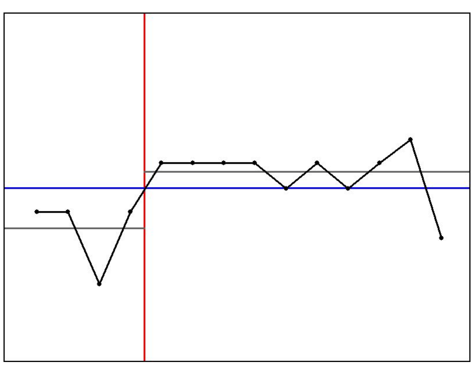
Table 9

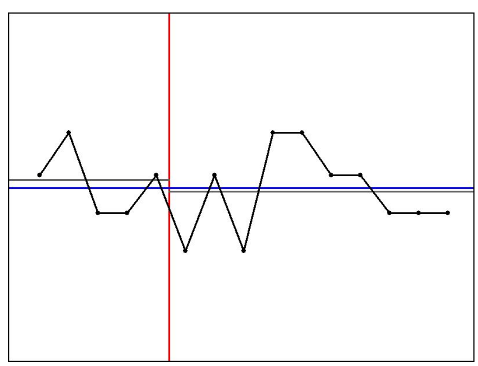
Fixation Duration (ms) to Face Region for Eye-Tracking Tasks

	Pre-Tx Mean (*SD*)	Post-Tx Mean (*SD*)
Video Emotion Recognition Task		
ID 3	1738.58	1548.09

ID 4	2623.21	1869.81
ID 6	1989.29	2106.59
ID 7	1771.47	1838.25
Overall Average	2030.64 (410.41)	1840.69 (228.89)
Video Emotion Expression Task		
ID 1	361.18	1389.17
ID 3	1366.01	1616.99
ID 4	2805.32	2598.56
ID 6	259.311	1921.40
ID 7	2127.51	2454.30
Overall Average	1383.87 (1105.03)	1996.08 (522.15)

	Tx 1 Mean (*SD*)	Tx 10 Mean (*SD*)
EU-Emotion Stimulus Set Task		
ID 1	3786.868	3345.113
ID 3	6459.625	5076.773
ID 4	6591.596	4844.908
ID 5	4380.506	5580.283
ID 6	4754.284	1708.675
ID 7	5444.839	5288.936
ID 8	4208.417	4059.145
Overall Average	5089.45 (1106.65)	4271.98 (1366.06)

1. 

2. 

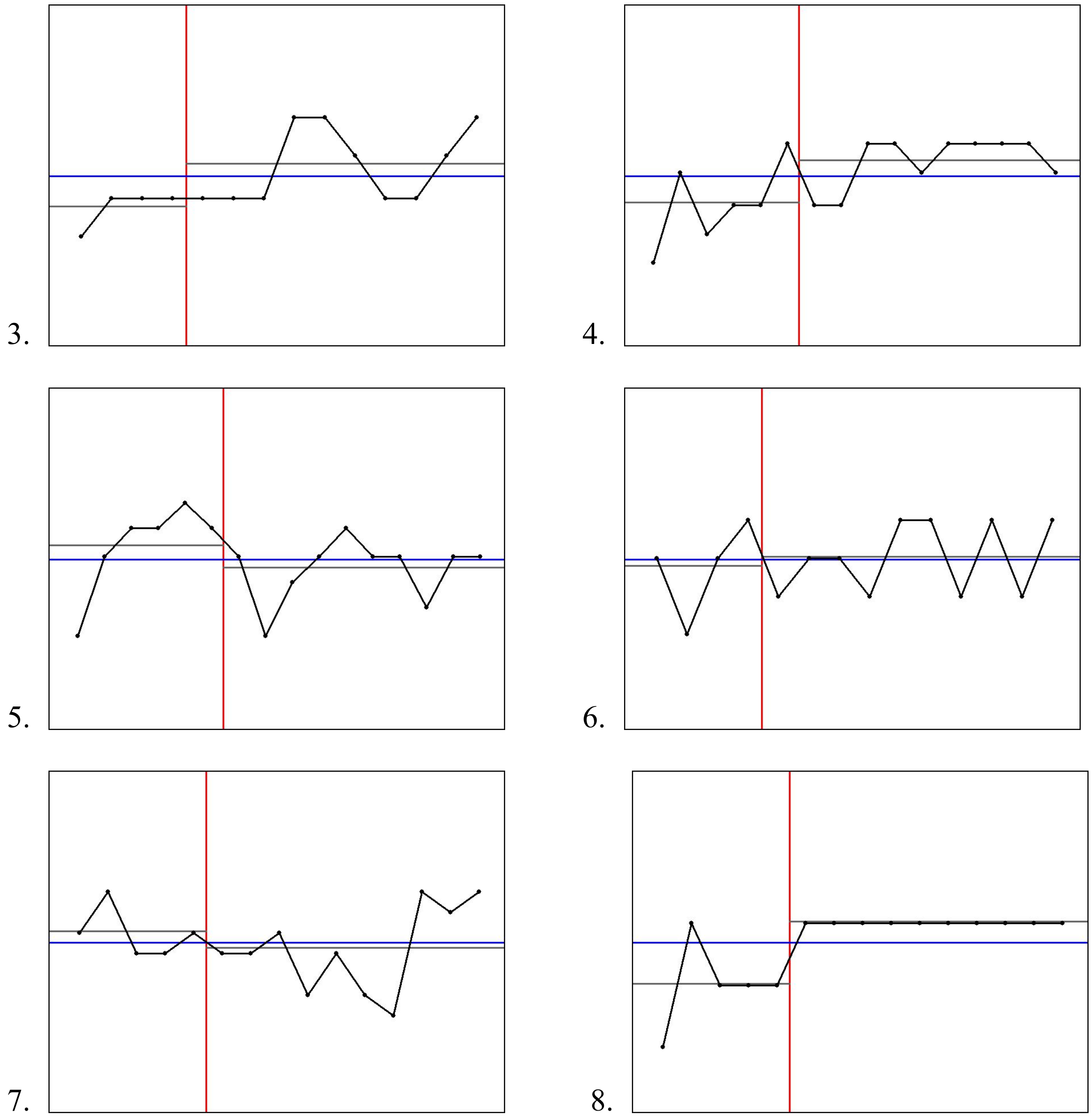


Figure 1. Baseline and treatment data series for IDs 1-8 for EU Task. Red line separates the baseline from treatment phase. Gray lines show average for each phase.

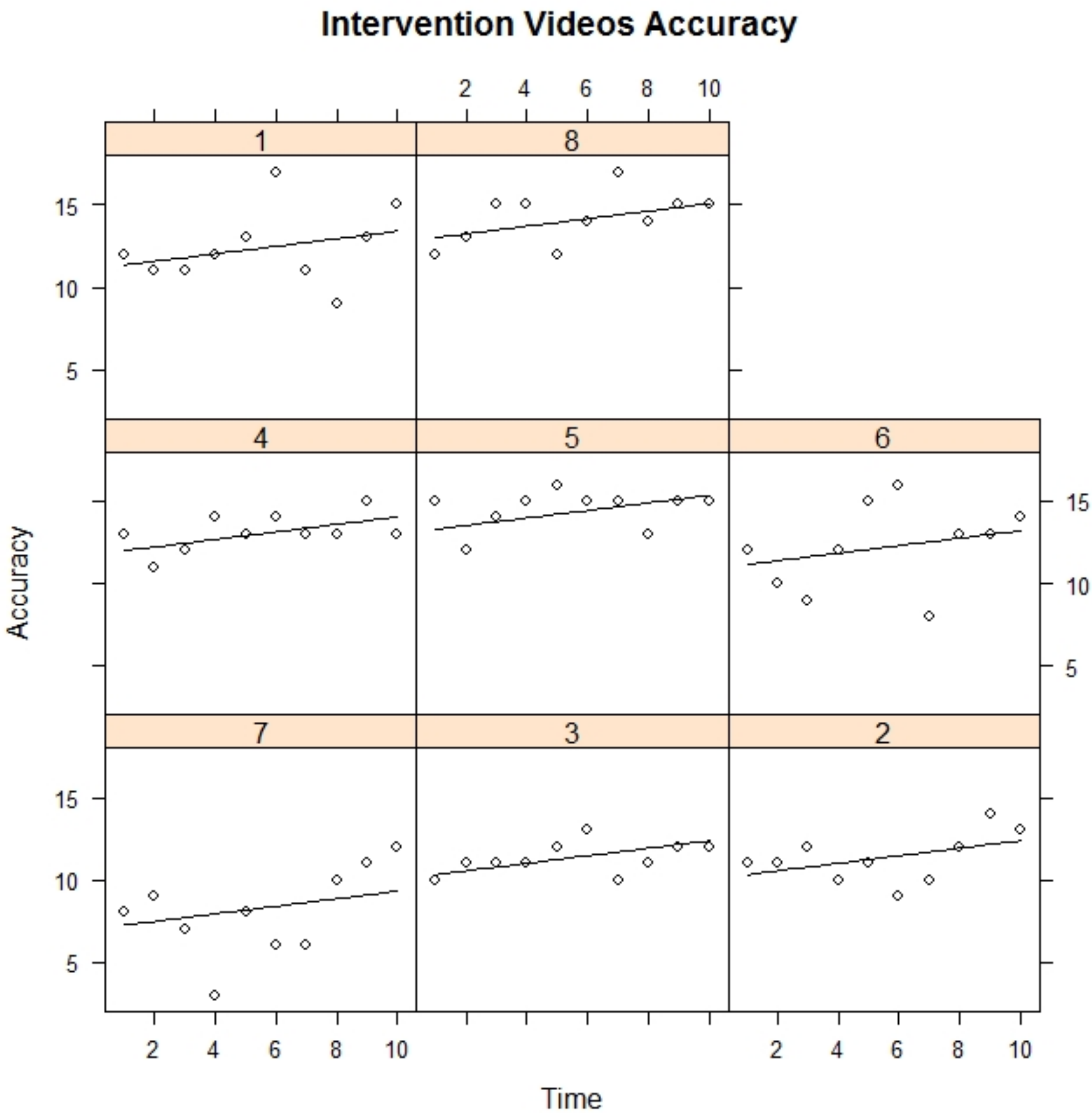


Figure 2. Estimated linear function for accuracy on the Intervention Videos Task for each subject.

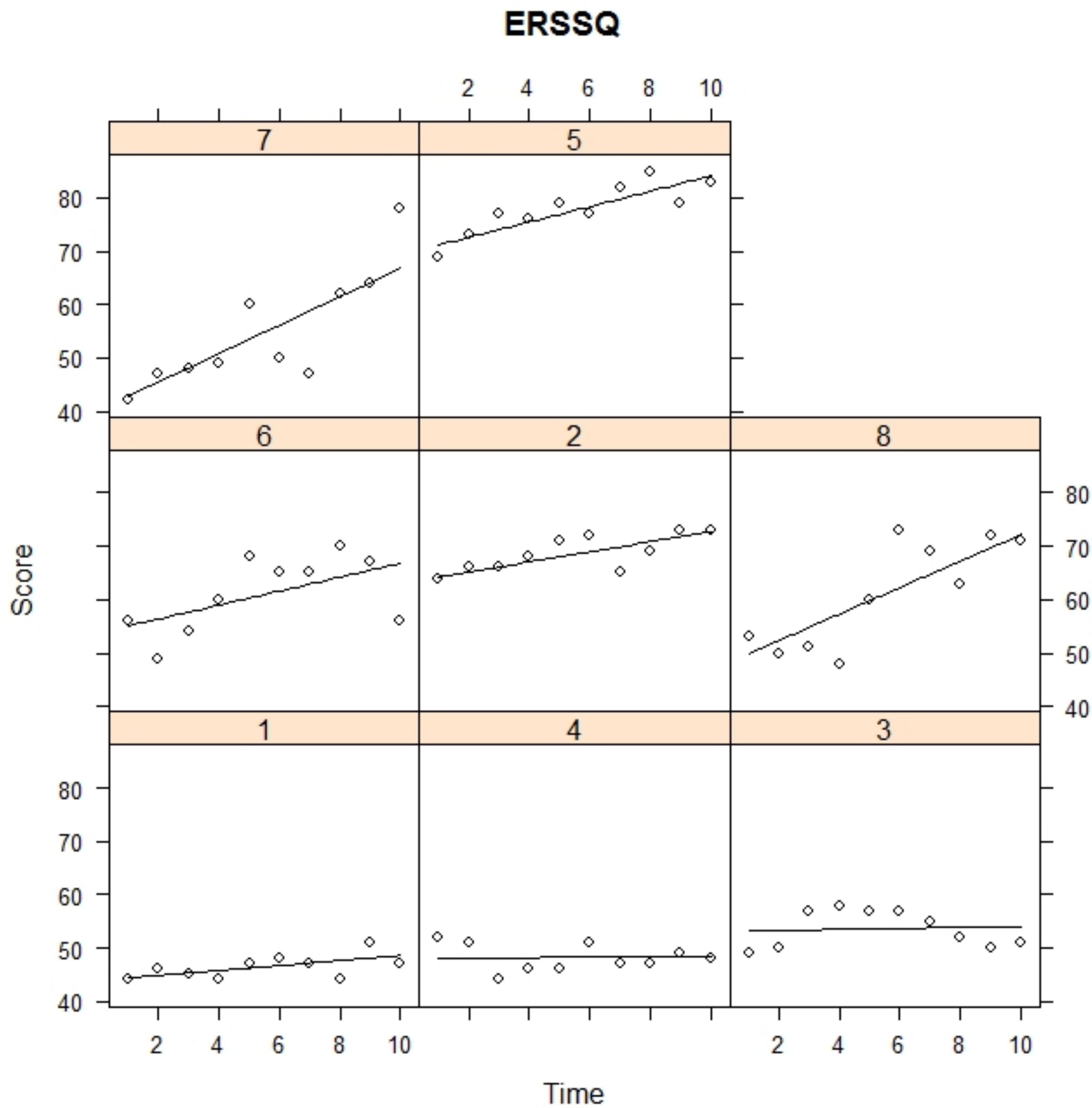


Figure 3. Estimated linear function for ERSSQ score for each subject.

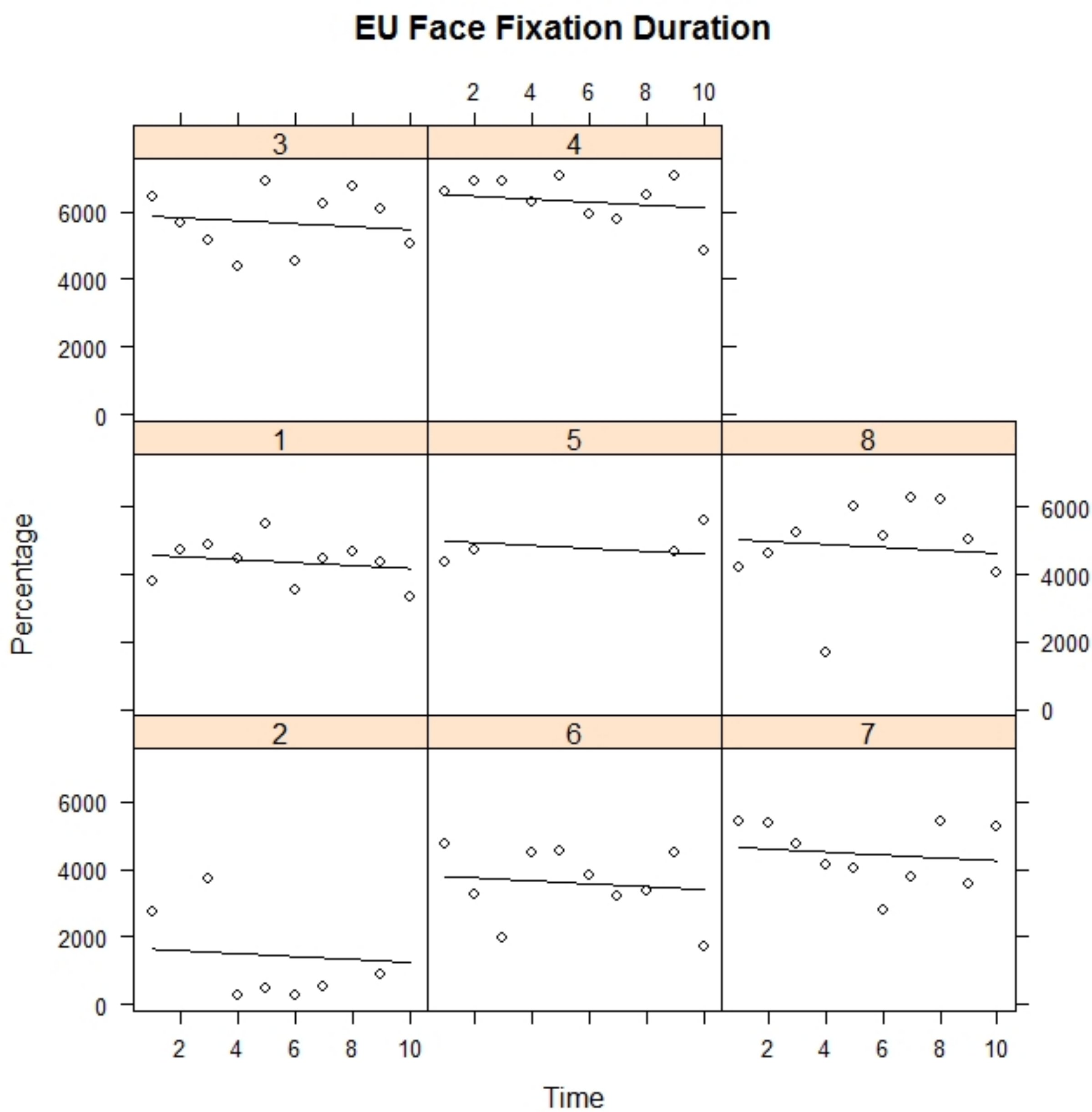


Figure 4. Estimated linear function for fixation duration to face region while viewing the EU-Emotion Stimulus Set for each subject.

Appendix A

Treatment Satisfaction Form Written Responses

ID	**Child: What do you like about the program?**	**Child: What do you dislike about this program?**	**Child: To improve the program, I would recommend:**	**Parent: To improve the program, I would recommend:**
1	How you get the money	I dislike the tests	More kids coming	Have a card with stickers to show child what they have accomplished.
2	I like helping the research	The repeated routine	To have different activities	N/A
3	The pay	Focusing the technology	Using a camera instead of the technology	Maybe points to or focus to re-iterate at home if there is something specific that the child is working on
4	N/A	N/A	N/A	Include ways for parents to be involved at home
5	It helps other people	Nothing	N/A	N/A
6	The magazines, the legos, the dot.	Nothing	More fun videos, breaks, magazines, prizes, and toys	It may benefit to have exercises and/or strategies to implement at home. Also, feedback as to the child's strengths/ weaknesses as discovered throughout the study would be appreciated.
7	Videos, dots, emotions, make different faces	Too long - took a while	Go quicker	N/A
8	It was fun	It took a long time	Nothing	More sessions - 3x week - for repetition in learning. Take-home materials? A description of how the treatment might work or be effective, so that it can be reinforced at home?

Appendix B

Demographics Form

Personal History

1. Your Child's Name: ____

2. Child's Age: __________________

3. Child's Date of Birth: ____/____/____

4. Address:

____
Street & Number   City   State   Zip

5. Phone number (for parent): ____

6. Email address (for parent) ____

7. Mother's name: ____

8. Father's name: ____

Demographics Form

9. Today's Date: ____/____/____

10. Child's Gender: ___ M ___ F

11. Your relationship to the child: ____

12. Child's race and/or ethnicity? (**Select all that apply**)

____ White/Caucasian, non-Hispanic, non-Arab
____ Black/African, African-American, non-Hispanic or African, non-Hispanic
____ Hispanic/Latino(a)
____ American Indian/Alaskan Native
____ Arab/Middle Eastern or Arab American
____ Asian/Asian-American
____ Pacific Islander
____ Other (Please specify ______________________)
____ I do not wish to provide this information

Biological mother: ___ Stepmother:___ Adoptive Mother:___ Foster Mother:___ Other:____

13. Mother's Education:
___ Completed 7th grade or less
___ Completed some high school
___ Graduated from high school
___ Graduated from trade school, business school or specialized training program
___ Completed an Associate degree
___ Graduated from college
___ Completed graduate school

Biological father: ____ Stepfather:___ Adoptive Father:__ Foster Father:___ Other:_______

14. Father's Education:
___ Completed 7th grade or less
___ Completed some high school
___ Graduated from high school
___ Graduated from trade school, business school or specialized training program
___ Completed an Associate degree
___ Graduated from college
___ Completed graduate school

15. How many siblings does your child have? _____________________________________
16. Income: What is your estimated gross annual family income?

___ < $20,000
___ $21,000 - $40,000
___ $41,000 - $60,000
___ $61,000 - $80,000

___ > $81,000

17. If currently in school, what grade is your child you in? __________________ (grade level)

18. Please check any of the following diagnoses that your child has been given:

_____ Anxiety Disorder (for example, Social Anxiety Disorder or Generalized Anxiety Disorder)
_____ Obsessive-Compulsive Disorder
_____ Attention-Deficit/Hyperactivity Disorder (ADHD)
_____ Autism Spectrum Disorder
_____ Depression
_____ Bipolar Disorder
_____ Personality Disorder (any type)
_____ Intellectual Disability/Mental Retardation
_____ Learning Disorder
_____ None of the above

19. Has your child ever been treated or received services for social skills or emotional problems?
Yes ____ No ____

Please check which of the following your child participated in, if any.

	Check if child participated and specify which services the child received
Behavioral Therapy (e.g., ABA, in-home, PECS,…)	
Pharmacological/Medication	
Augmentative (equine therapy, special diets, vitamins,…)	
Therapy (e.g., for anxiety, CBT,…)	
Social Interventions (e.g., group, lunch bunch,…)	

20. Does your child have any problems with his/her vision? Yes___ No___

If yes, please specify____

21. Does your child currently wear (**Select all that apply**)
____ Glasses (Type: ______________________________________)
____ Contacts

22. Has your child ever had any eye surgery? Yes___ No___

If yes, please specify ____

23. Does your child show a normal range of facial expressions? (for example, does s/he frown or pout or look embarrassed as well as laugh or cry? Yes___ No___

If No, please describe ____

Can s/he look guilty…or surprised…..or amused? Yes___ No___
Can you tell by her/his face when s/he is afraid or disgusted? Yes___ No___
Does s/he have the same range of facial expressions as other children? Yes___ No___

24. Does your child's facial expression usually seem appropriate to the particular situation as far as you can tell? Yes___ No___

If No, please describe ____

25. Medications: If your child is currently taking any medications, please specify the name of the medication and the following: dosage, time of day the medicine is taken, and reason for medication.

None ____________

ADHD medications ____

Anti-Depressant medications ____

Anti-Anxiety medications____

Anti-Seizure medications ____

Allergy/Asthma medications ____

Other medications ____

Other medications ____

26. Please put an X for any of the following diagnoses that **child's** father, mother, grandparents, or siblings have struggled with:

	Father	Mother	Grand-mother (maternal)	Grand-father (maternal)	Grand-mother (paternal)	Grand-father (paternal)	Sibling(s)
Anxiety Disorder							

Obsessive-Compulsive Disorder (OCD)							
Attention-Deficit/Hyperactivity Disorder (ADHD)							
Autism Spectrum Disorder (ASD)							
Depression (include post-partum)							
Bipolar Disorder							
Personality Disorder (any type)							
Intellectual Disability/ Mental Retardation							
Learning Disorder							
Other (please specify)							

Appendix C

Treatment Satisfaction Survey (Child)

1. **Overall, this program seems helpful.**

Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree

2. **This program is acceptable (in terms of time and amount of measures).**

Strongly | Disagree | Neutral | Agree | Strongly

Disagree Agree

3. At this point, my ability to recognize emotions has:

Greatly Decrease	Slightly Decreased	Stayed the Same	Slightly Increased	Greatly Increased

4. I would recommend this program to a friend.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I like the program.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

6. What do you like about the program?

____

____

____

7. **What do you dislike about the program?**

____

____

____

8. **To improve the program, I would recommend:**

____

____

____

Appendix D

Treatment Satisfaction Survey (Parent)

1. Overall, this program seems helpful in increasing my child's emotion recognition.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. This program is acceptable (in terms of time and work commitment).

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. **At this point, my child's ability to recognize emotions has:**

Greatly Decreased	Slightly Decreased	Stayed the Same	Slightly Increased	Greatly Increased

4. **I would recommend this program to a friend whose child has difficulty with emotion recognition.**

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. **Overall, I like the program.**

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

6. **What changes (if any) have you noticed regarding your child's emotion recognition skills?**

____
____
____
____

7. **To improve the program, I would recommend:**

____
____
____
____