RESEARCH ARTICLE

Comparing Automatic Eye Tracking and Manual Gaze Coding Methods in Young Children with Autism Spectrum Disorder

Courtney E. Venker, Ron Pomper, Tristan Mahr, Jan Edwards, Jenny Saffran, and Susan Ellis Weismer

Eye-gaze methods offer numerous advantages for studying cognitive processes in children with autism spectrum disorder (ASD), but data loss may threaten the validity and generalizability of results. Some eye-gaze systems may be more vulnerable to data loss than others, but to our knowledge, this issue has not been empirically investigated. In the current study, we asked whether automatic eye-tracking and manual gaze coding produce different rates of data loss or different results in a group of 51 toddlers with ASD. Data from both systems were gathered (from the same children) simultaneously, during the same experimental sessions. As predicted, manual gaze coding produced significantly less data loss than automatic eye tracking, as indicated by the number of usable trials and the proportion of looks to the images per trial. In addition, automatic eye-tracking and manual gaze coding produced different patterns of results, suggesting that the eye-gaze system used to address a particular research question could alter a study's findings and the scientific conclusions that follow. It is our hope that the information from this and future methodological studies will help researchers to select the eye-gaze measurement system that best fits their research questions and target population, as well as help consumers of autism research to interpret the findings from studies that utilize eye-gaze methods with children with ASD. ***Autism Res*** *2020, 13: 271–283.* © 2019 International Society for Autism Research, Wiley Periodicals, Inc.

Lay Summary: The current study found that automatic eye-tracking and manual gaze coding produced different rates of data loss and different overall patterns of results in young children with ASD. These findings show that the choice of eye-gaze system may impact the findings of a study—important information for both researchers and consumers of autism research.

Keywords: eye tracking; methodology; data quality; children; language processing; autism

Introduction

Eye-gaze methods—including automatic eye-tracking and manual coding of eye gaze—have been widely used to investigate real-time cognitive, linguistic, and attentional processes in infants and young children [Aslin, 2007, 2012; Fernald, Zangl, Portillo, & Marchman, 2008; Oakes, 2012]. Eye-gaze methodology has also become increasingly popular in studies of children with autism spectrum disorder (ASD) in recent years [Chita-Tegmark, Arunachalam, Nelson, & Tager-Flusberg, 2015; Falck-Ytter, Bölte, & Gredebäck, 2013; Kaldy, Kraper, Carter, & Blaser, 2011; Potrzeba, Fein, & Naigles, 2015; Swensen, Kelley, Fein, & Naigles, 2007]. Eye-gaze techniques offer advantages in autism research because they provide a window into complex cognitive processes simply by measuring participants' gaze to visual stimuli on a screen. Furthermore, they have limited behavioral response demands, do not require social interaction, and are appropriate for participants with a wide range of ages, cognitive skills, and language abilities.

Given the growing popularity of eye-gaze methods in autism research, it is important to consider methodological issues that may impact the data from which we draw our inferences [Nyström, Andersson, Holmqvist, & van de Weijer, 2013; Oakes, 2012; Venker & Kover, 2015; Wass, Forssman, & Leppänen, 2014; Wass, Smith, & Johnson, 2013]. For example, eye-gaze methods are vulnerable to data loss—periods of time in which participants' gaze is not (or appears not to be) directed to the stimuli of interest. Data loss is problematic because it can threaten the validity of dependent variables, diminish statistical power, limit the generalizability of findings, and produce inaccurate results [Wass et al., 2014]. In addition, limiting data loss will increase the likelihood of attaining rigorous and reproducible results, as emphasized by the National Institutes of Health [Collins & Tabak, 2014]. Though a certain amount of data loss is

From the Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin (C.E.V.); Waisman Center and Department of Psychology, University of Wisconsin-Madison, Madison, Wisconsin (R.P., J.S.); Waisman Center and Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, Wisconsin (T.M., J.E., S.E.W.)

Courtney E. Venker is now at the Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan. Jan Edwards is now at the Department of Hearing and Speech Sciences, Language Science Center, University of Maryland-College Park, Maryland.

Received April 4, 2019; accepted for publication September 22, 2019

Address for correspondence and reprints: Courtney E. Venker, Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan. E-mail: cvenker@msu.edu

Published online 17 October 2019 in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.2225

© 2019 International Society for Autism Research, Wiley Periodicals, Inc.

unavoidable, some eye-gaze systems may be more vulnerable to data loss than others and different systems may even produce different patterns of results. In the current study, we investigated these issues by directly comparing data gathered from young children with ASD using two different eye-gaze systems—automatic eye tracking and manual gaze coding.

Studies of children with ASD typically measure gaze location using one of two systems: automatic eyetracking or manual coding of eye gaze from video. Both eye-gaze systems determine where children are looking, but they do so in different ways. Automatic eye trackers determine gaze location using a light source—usually near-infrared lights—to create corneal reflections that are recorded by cameras within the eye tracker (Tobii Technology, Stockholm, Sweden). Gaze location is based on three pieces of information: corneal reflection, pupil position, and location of the participant's head relative to the screen [Wass et al., 2014]. Calibration is required to maximize the accuracy of gaze location measurements [Nyström et al., 2013], and processing algorithms are applied to the raw gaze coordinates to map gaze locations to areas of interest (AOIs) on the screen. Eye-tracking methods based on corneal reflection have been used to measure gaze location for decades (see Karatekin, 2007, for a historical review). Tobii Technology AB (2012) describes a test specification for validating the spatial measurements of eye-tracking devices. These tests quantify the accuracy and precision of gaze measurements by having viewers (and artificial eyes) fixate on known screen locations under various viewing conditions. Timing measurements can also be validated by comparing an eye tracker's output to a video recording [e.g., Morgante, Zolfaghari, & Johnson, 2012].

Although some eye trackers involve head-mounted equipment, we focus here on remote eye trackers because they do not require physical contact with the equipment and therefore are more appropriate for young children with ASD [Falck-Ytter et al., 2013; Venker & Kover, 2015]. Remote eye trackers are robust to a certain degree of head movement, but they require some information about the location of the child's head in 3D space. Thus, children's heads must remain within a 3D "track box" in order to determine gaze location. For example, the eye tracker used in the current study—the Tobii X2-60—allows for head movements of 50 cm (width) × 36 cm (height), with the participant positioned between 45 and 90 cm from the eye tracker (Tobii Technology, Stockholm, Sweden).

Manual gaze-coding systems determine gaze location quite differently from automatic eye tracking. In manual gaze coding, human coders view a video of the child's face that was recorded during the experiment [Fernald et al., 2008; Naigles & Tovar, 2012]. Coders determine gaze location for each time frame, based on the visual angle of children's eyes and the known location of AOIs on the screen. Coders must complete a comprehensive training process prior to coding independently [Fernald et al., 2008; Naigles & Tovar, 2012; Venker & Kover, 2015]. As with other types of behavioral coding, coders also need to participate in periodic lab-wide reliability checks to prevent drift from the original training procedures over time [Yoder, Lloyd, & Symons, 2018]. It is customary for studies using manual gaze coding to report inter-coder agreement for a subset of videos that were coded independently by two different coders [Fernald et al., 2008; Naigles & Tovar, 2012].

Because of its automated approach, eye tracking offers several advantages over manual gaze coding. It is objective, efficient, and has relatively high temporal and spatial resolution [Dalrymple, Manner, Harmelink, Teska, & Elison, 2018; Hessels, Andersson, Hooge, Nyström, & Kemner, 2015]. Manual gaze coding, on the other hand, is subjective, requires extensive reliability training, is labor-intensive (e.g., about 1 hr for a 5-min video), and has more limited spatial and temporal resolution [Aslin & McMurray, 2004; Wass et al., 2013]. As a result of its increased temporal and spatial precision, automatic eye tracking is capable of capturing certain dependent variables that manual gaze coding cannot (e.g., pupil size or discrete fixations within an AOI), opening up exciting new areas of inquiry [Blaser, Eglington, Carter, & Kaldy, 2014; Oakes, 2012; Ozkan, 2018]. Thus, in some studies, automatic eye tracking may be required to capture the dependent variables to answer a particular research question. In other studies, however, either automatic eye-tracking or manual gaze coding would be capable of capturing the dependent variables of interest.

One experimental design that can be used with either automatic eye-tracking or manual gaze coding is a "2-large-AOI" design, in which visual stimuli (e.g., objects, faces) are presented simultaneously on the left and right sides of the screen [Fernald et al., 2008; Tek, Jaffery, Fein, & Naigles, 2008; Unruh et al., 2016]. Because of its flexibility, the 2-large-AOI design has been used to study constructs as diverse as memory [Oakes, Kovack-lesh, & Horst, 2010], spoken language comprehension [Brock, Norbury, Einav, & Nation, 2008; Goodwin, Fein, & Naigles, 2012], visual preferences [Pierce, Conant, Hazin, Stoner, & Desmond, 2011; Pierce et al., 2016], and social orienting [Unruh et al., 2016]. In this type of study, gaze location during each moment in time is typically categorized as directed to the left AOI, the right AOI, or neither (e.g., between images, away from the screen). From this information, researchers can derive numerous dependent variables, including relative looks to each AOI, time to shift between AOIs, and the length and location of the longest look. Though the 2-large-AOI design differentiates two relatively broad AOIs, as opposed to discrete fixations within a given AOI, it is possible for either system to produce inaccurate results. In manual gaze coding, for example, a human coder could judge a child to be looking at an image AOI, when in fact

the child is looking slightly outside the boundaries of the AOI (e.g., off-screen or at a non-AOI part of the screen). The same type of error could occur in automatic eye tracking when internal processing algorithms estimate gaze location inaccurately [Dalrymple et al., 2018; Niehorster, Cornelissen, Holmqvist, Hooge, & Hessels, 2018]. Because both systems are capable of reporting gaze location inaccurately, we do not consider inaccuracy to be a disadvantage unique to either system. We return to this issue in the Discussion.

Despite the clear benefits of automatic eye tracking, manual gaze coding may offer at least one substantive methodological advantage over automatic eye tracking: lower rates of data loss. Because manual gaze coding involves judging gaze location from video of children's faces, it is relatively flexible: coders can determine gaze location as long as children's eyes are clearly visible on the video. Automatic eye tracking, on the other hand, requires multiple pieces of information to determine gaze location. If any piece of information is missing, the eye tracker will be unable to report gaze location, resulting in data loss—even if the child's gaze was directed toward one of the AOIs. The eye tracker may also need time to "recover" before it regains the track after the eyes have moved off-screen and then back on-screen again [Oakes, 2010]. Thus, automatic eye tracking may be especially affected by behaviors such as fidgeting, excessive head movement, watery eyes, and changes in the position of the child's head [Hessels, Cornelissen, Kemner, & Hooge, 2015; Niehorster et al., 2018; Wass et al., 2014]. Considering the impact of such behaviors on data quality is especially important in studies of children with ASD, where behaviors such as squinting, body rocking, head tilting, and peering out of the corner of the eye are likely to occur.

The goal of the current study was to determine whether automatic eye-tracking and manual gaze coding systems produced different rates of data loss or different overall results. Young children with ASD participated in a screen-based semantic processing task that contained two conditions: Target Present and Target Absent. Target Present trials presented two images (e.g., *hat*, *bowl*) and named one of them (e.g., *Look at the hat!*). Target Absent trials presented two images (e.g., *hat*, *bowl*) and named an item that was semantically related to one of the objects (e.g., *Look at the pants!*). During a given experimental session, children's eye movements were simultaneously recorded both by an eye tracker and by a video camera for later offline coding. Prior to conducting the analyses, we processed and cleaned the data from each system, following standard procedures. In this way, we only examined trials that would typically be included in published analyses, thereby maximizing the relevance of the results.

Our first research question was: Do automatic eye-tracking and manual gaze coding produce different rates of data loss, as indicated by the number of trials contributed per child or by the amount of looking time to the images per trial? Based on the inherent vulnerability of eye tracking to data loss, we predicted that automatic eye tracking would produce significantly more data loss than manual gaze coding across both metrics. Our second research question was: Do automatic eye-tracking and manual gaze coding produce different patterns of results? To address this question, we conducted a growth curve analysis modeling looks to the target images over time and tested the impact of the eye-gaze system (eye tracking vs. manual gaze coding) on children's performance. Though the lack of empirical data in this area prevented us from making specific predictions, we were particularly interested in whether the analyses revealed any significant interactions between eye-gaze system and condition, as such a finding would indicate a difference, by system, in the relationship between the two conditions. We also conducted post hoc analyses of the data from each system separately, to determine how the overall findings may have differed if we had only gathered data from a single system.

Method

Participants

Participants were part of a broader research project investigating early lexical processing. The project was approved by the university institutional review board, and parents provided written informed consent for their child's participation. Children completed a 2-day evaluation that included a battery of developmental assessments and parent questionnaires, as well as several experimental eye-gaze tasks. The current study focused on one eye-gaze task that was administered on both days. This task is described in more detail below; results from the other tasks will be reported elsewhere.

Research visits were conducted by an interdisciplinary team that included a licensed psychologist and speech-language pathologist with expertise in autism diagnostics. Each child received a DSM-V diagnosis of ASD [American Psychiatric Association, 2013], based on results of the *Autism Diagnostic Interview-Revised* [Rutter, LeCouteur, & Lord, 2003], the *Autism Diagnostic Observation Schedule, Second Edition* [ADOS-2; Lord et al., 2012], and clinical expertise. Based on their age and language level, children received the Toddler Module (no words/younger n = 11; some words/older n = 8), Module 1 (no words n = 15; some words n = 14), or Module 2 (younger than 5 n = 3). Two subscales of the *Mullen Scales of Early Learning* [Mullen, 1995] were administered: Visual Reception and Fine Motor. Based on previous work [Bishop, Guthrie, Coffing, & Lord, 2011], age equivalents from these two subscales were averaged, divided by chronological age, and multiplied by 100 to derive a nonverbal Ratio IQ score for each child

Table 1. Participant Characteristics

	Mean (*SD*) Range
Age in months	30.80 (3.35) 24–36
Nonverbal ratio IQ (MSEL)	65.74 (15.99) 31–102
ASD symptom severity (ADOS-2)	8.10 (1.66) 4–10
Auditory comprehension (PLS-5)	59.25 (12.12) 50–98
Expressive communication (PLS-5)	73.14 (10.45) 50–100
Total language (PLS-5)	64.02 (10.40) 50–95

Note. MSEL = Mullen Scales of Early Learning; ASD = Autism Spectrum Disorder; ADOS-2 = Autism Diagnostic Observation Schedule, Second Edition; PLS-5 = Preschool Language Scales, Fifth Edition. Standard scores were used for the ADOS-2 and PLS-5.

(mean of 100, *SD* of 15). The *Preschool Language Scale, 5th Edition* [PLS-5; Zimmerman, Steiner, & Pond, 2011] was administered to assess receptive language (Auditory Comprehension scale) and expressive language (Expressive Communication scale). The PLS-5 yields standard scores for both receptive and expressive language (mean of 100, *SD* of 15).

Participants were 51 children with ASD who contributed usable data (see "Eye-Gaze Data Processing") from both the automatic eye-tracking and manual gaze coding systems (see Table 1 for participant characteristics). Forty children were male and 11 were female. Forty-seven children were reported by their parent or caregiver to be White, and four children were reported to be more than one race. Five children were reported to be Hispanic or Latino and 46 children were reported not to be Hispanic or Latino. ADOS-2 comparison scores [Gotham, Pickles, & Lord, 2009] provided a measure of autism severity. The mean score was 8, indicating that on average children demonstrated a high level of autism-related symptoms. Forty-eight of the 51 participants displayed clinical language delays based on PLS-5 total scores at least −1.25 *SD* below the mean. Nonverbal ratio IQ scores were below 70 for 61% of the sample (31/50; an IQ score could not be computed for one child).

Semantic Processing Task

Children completed two blocks of a looking-while-listening task designed to evaluate semantic representations of early acquired words. Children sat on a parent or caregiver's lap in front of a 55-in. television screen (see Fig. 1). Parents were instructed not to talk to their child or direct their attention. Parents wore opaque sunglasses to prevent them from viewing the screen and inadvertently influencing their child's performance. The audio was presented from a central speaker located below the screen. During an experimental session, children's eye movements were simultaneously recorded both by a video camera (for later offline coding) and by an automatic eye tracker. The video camera was mounted below the screen and recorded video of the children's faces at a rate of 30 frames per second during the experiment for later manual gaze coding. The eye tracker, a Tobii X2-60 (Tobii Technology, Stockholm, Sweden), was placed on the end of a 75 cm extendable arm below the screen and recorded gaze location automatically at a rate of 60 Hz. Participants were seated so that their eyes were approximately 60 cm from the eye tracker (the standard distance recommended for optimal tracking). Positioning the eye tracker in this way—between the participant and the screen—allowed us to capture looks to the entire 55-in. screen while remaining within the 36° of visual angle (from center) recommended for optimal tracking. Specifically, the visual angle was 24° from the center to the lower corners of the screen and 33.6° from the center to the upper corners of the screen.

Figure 1. Visual depiction of the experimental setup. Children sat on their parent's lap in the chair while viewing the task. The video camera was placed directly below the screen. The automatic eye tracker was placed on the end of the extendable arm to ensure appropriate placement.

The experimental task was developed and administered using E-Prime 2.0 (version 2.0.10.356) and the data were analyzed in RStudio (vers. 1.1.456; R vers. 3.5.1; R Core Team, 2019). Prior to the task, children completed a 5-point Tobii infant calibration, which presented a shaking image of a chick with a trilling sound. If calibration was poor (i.e., the green lines were not contained in the circles for at least 4 of the 5 points), the experimenter reran the calibration. If the child failed calibration after

multiple attempts, the task was run without the eye tracker for later manual coding. (In the current study, six children were unable to complete calibration. See Eye-Gaze Data Processing for more information.)

Target Present trials presented two images (e.g., *hat*, *bowl*) and named one of them (e.g., *Look at the hat!*). Target Absent trials presented two images (e.g., *hat*, *bowl*) with an auditory prompt naming an item that was semantically related to one of the objects (e.g., *Look at the pants!*). Trials lasted approximately 6.5 sec. Analyses were conducted on the window of time from 300 to 2,000 msec after the target noun onset. Filler trials of additional nouns were included to increase variability and maximize children's attention to the task, but these trials were not analyzed. Children received two blocks of the experiment with different trial orders. Each block included eight Target Present and eight Target Absent trials, for a maximum of 16 trials per condition.

Eye-Gaze Data Processing

Eye movements were coded from video by trained research assistants (using iCoder, Version 2.03) at a rate of 30 frames per second. The coders were unable to hear the audio, which prevented any bias toward coding looks to the named image. Each frame was assigned a code of "target" or "distractor" or a code of "shifting" or "away" for frames in which gaze was between images or off the screen [Fernald et al., 2008]. Independent coding by two trained coders was completed for 20% of the full sample; frame agreement was 99% and shift agreement was 96%.

To allow a direct comparison of the automatic eye-tracking data set and manual-coded data set, we equated the sampling rates of each system to 30 Hz. This required downsampling the automatic eye-tracking data, which was originally collected at a sampling rate of 60 Hz. To mirror the procedures used in manual coding, segments of missing data due to blinks were interpolated for periods of time up to 233 msec assuming the AOI (left or right) was the same at the beginning and end of the period of nonimage time segments.

For the purposes of the current study, data loss included instances when a look to an image AOI was not recorded due to limitations of the eye-gaze system ("technical" data loss) as well as instances when a look to an AOI was not recorded because children's gaze was actually directed outside the AOIs ("true" data loss). We use the term data loss to refer to both types of occurrences because they produce the same outcome: periods of time in which gaze location is not recorded as directed to an AOI, and thus contribute no data to the analyses. Because both eye-gaze systems treat looks away from the AOIs—true data loss—similarly, any differences between automatic eye-tracking and manual gaze coding are most likely due to technical data loss.

The full sample had initially included 70 children with ASD. Because this study compared eye-tracking and manual gaze coding, children were excluded from the analyses if they failed to contribute usable data from both systems. We defined "usable data" for a given system as four or more trials per condition with at least 50% looking time to the images during the analysis window (300–2,000 msec after noun onset).[1] Eight children were excluded because they did not contribute gaze data from both sources on at least 1 day, leaving 62 participants. (Six of these eight children were excluded because they were unable to complete calibration, a topic we return to in the Discussion.) Next, we removed all trials in which children looked away from the images more than 50% of the time during the analysis window. Children were excluded if they did not have at least four trials remaining in both conditions for each Source. Six children had too few trials in the manual data set. Eleven children had too few trials in the eye-tracking data set. (Note that the 11 children who had too few trials in the eye-tracking data set included the six children who had too few trials in the manual-coded data set, plus five additional children.) To ensure a level playing field, it was critical that the analyses include only the children who had contributed data from both sources. Thus, the 11 children who had too few trials in either Source were removed, leaving 51 participants who contributed data to the primary analyses.

Analysis Plan

Our first goal was to determine whether automatic eye-tracking and manual gaze coding produced different rates of data loss. To address this question, we constructed two linear mixed-effects models using the lme4 package [vers. 1.1-17; Bates, Machler, Bolker, & Walker, 2015]. The dependent variable in the first model was the number of trials per child. The dependent variable in the second model was the proportion of frames on which children were fixating the target or distractor object out of the total number of frames during the analysis window (300–2,000 msec after target word onset). Both models included Source as a fixed effect (contrast coded as –0.5 for manual gaze coding vs. 0.5 for automatic eye tracking). Random effects included a by-subject intercept and slope for Source.

To determine whether automatic eye-tracking and manual gaze coding produced different results overall, we used mixed-effects growth curve analysis to quantify changes in the time course of children's fixations to the target object during the critical window [Mirman, 2014]. The dependent variable was the proportion of frames on which children fixated the target object out of the frames they fixated the

[1]We selected the 300–2,000 msec analysis window because it is similar to time windows used in previous work and because it contained the average rise and plateau of looks to target across conditions.

target or distractor object for each time frame during the critical window 300–2,000 msec after the onset of the target noun. To accommodate the binary nature of the data (i.e., fixations to the target or distractor) and instances in which a child always fixated the target or the distractor object, we transformed this proportion to empirical logodds. Our fixed effects included Condition (contrast coded as −0.5 for Target Present vs. 0.5 for Target Absent), Source (contrast coded as −0.5 for manual gaze coding vs. 0.5 for automatic eye tracking), four orthogonal time terms (intercept, linear, quadratic, and cubic), and all the 2-way and 3-way interactions. Models were fit using Maximum Likelihood estimation. As recommended by Barr, Levy, Scheepers, and Tily [2013], random effects were permitted across participants for all factors and all interactions (i.e., a full random effects structure). The significance of *t*-scores was evaluated assuming a normal distribution (i.e., *t*-values > ±1.96 were considered significant). This assumption is appropriate given the large number of participants and data points collected.

Results

Our first research question asked whether automatic eye-tracking and manual gaze coding produced different rates of data loss, as indicated by the number of usable trials per child and the proportion of looking time to the images per trial. Given our exclusionary criteria, this analysis involves a level playing field. First, participants who were missing too much data from one system were also excluded from the other system (e.g., the five participants who had excessive missing data only with automatic eye tracking). That is, we compared the number of useable trials only for those participants with enough data to be included in the final sample. Second, we excluded trials with too much missing data (i.e., without fixations to either object for more than 50% of the frames). That is, we compared the proportion of looking times to images only on those trials in which children were attentive and tracked. We did not include Condition or its interaction with Source in these models. Although children's accuracy may differ in each Condition, the amount of useable data should not. Moreover, including Condition and the interactions would have overfit the models (i.e., using three effects to fit two data points per participant). As illustrated by the means and *SD*s reported below, the amount of data loss per Conditions was highly similar for a given Source, confirming our assumption.

We first examined the mean number of trials contributed per child for each source (see Fig. 2). In the manual-coded data set, participants contributed 12.41 trials in the Target Absent condition (*SD* = 3.48, range = 4–16) and 12.53 trials in the Target Present condition (*SD* = 3.58, range = 5–16). In the automatic eye tracking data set, participants contributed 10.49 trials in the Target Absent condition (*SD* = 3.78, range = 4–16) and 10.75 trials in the Target Present condition (*SD* = 3.80, range = 4–16). As predicted, children contributed significantly more trials in the manual-coded data set than in the automatic eye tracking data set, $t(50) = -4.85$, $P < 0.001$.

We next examined the proportion of looking time to the images during the analysis window (300–2,000 msec after target onset, see Fig. 3). In the manual-coded data set, children looked at the images 90.05% of the time in the Target Absent condition (*SD* = 4.70, range = 76.28–98.56) and 91.22% of the time in the Target Present condition (*SD* = 5.05, range = 74.62–97.70). In the automatic eye tracking data set, children looked at the images 87.88% of the time in the Target Absent condition (*SD* = 6.15, range = 71.64–99.52) and 88.73% of the time in the Target Present condition (*SD* = 5.57, range = 71.70–97.53). Consistent with our predictions, the proportion of looking time to the images was significantly higher in the manual-coded data set than in the automatic eye tracking data set, $t(50) = -3.88$, $P < 0.001$.

Our second research question asked whether automatic eye-tracking and manual gaze coding produced different patterns of results. Visual examination of the data revealed overall similarities in mean looks to the target image across the two systems (see Fig. 4). Children's looks to the target increased over time, and the mean curves were higher (in an absolute sense) for the Target Present condition than for the Target Absent condition. Next, we statistically tested the effect of Source (automatic eye tracking vs. manual gaze coding) in a growth curve analysis modeling looks to the target image over time. The full model results are presented in Table 2. We will first discuss the model results across Sources and will then discuss the Source by Condition interactions.

Collapsing across Source and Condition, there was a significant effect of all time terms (*P*'s < 0.017). This indicates that children's fixations to the target object were significantly greater than chance (intercept), increased from the beginning to the end of the window (linear time), reached a peak asymptote and then declined (quadratic), and were delayed in increasing from baseline (cubic). As expected, there was a significant effect of Condition on the intercept ($P = 0.029$) and quadratic time ($P = 0.017$), indicating that children looked at the target image more and had a steeper peak asymptote in accuracy in the Target Present condition than in the Target Absent condition, regardless of Source. Condition did not have a significant effect on linear time ($P = 0.300$) or cubic time ($P = 0.545$). There was also a significant effect of Source on the linear time term: across Conditions, the average slope of the increase in looks to Target over time was significantly smaller for the manual-coded data than the eye-tracking data ($P = 0.027$).[2] Source did not have a

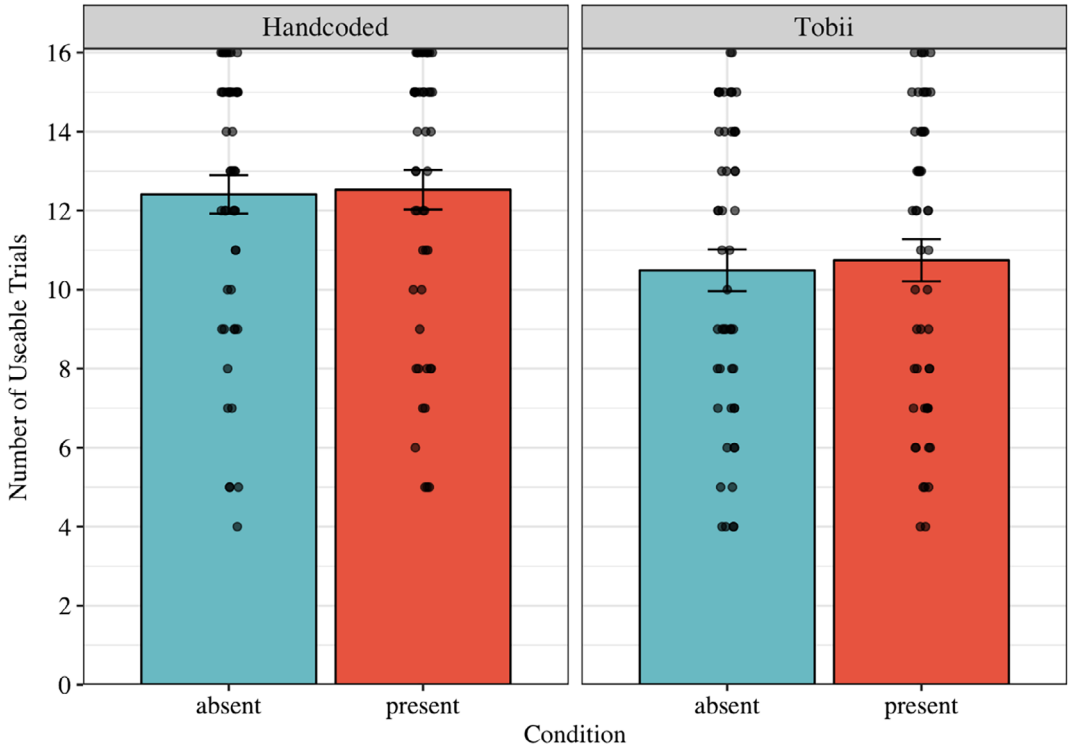

Figure 2. Mean number of usable trials per child, separated by source (Hand coded = manual gaze coding; Tobii = automatic eye tracking) and condition (Target Absent vs. Target Present). The dots represent means for individual children. The bars represent ±1 *SE* above and below the mean.

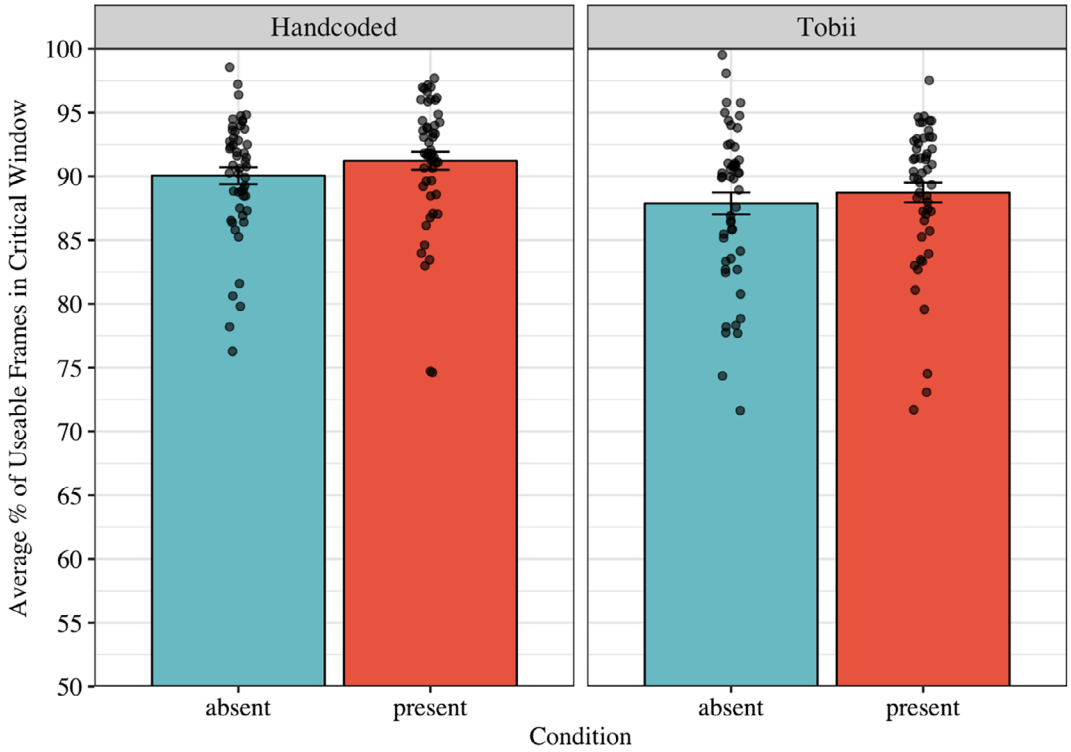

Figure 3. Mean proportion of time points during the analysis window (300–2,000 msec after noun onset) in which children were looking at the images, separated by source (Hand coded = manual gaze coding; Tobii = automatic eye tracking) and condition (Target Absent vs. Target Present). Dots represent the mean for each child. The bars represent ±1 *SE* above and below the mean.

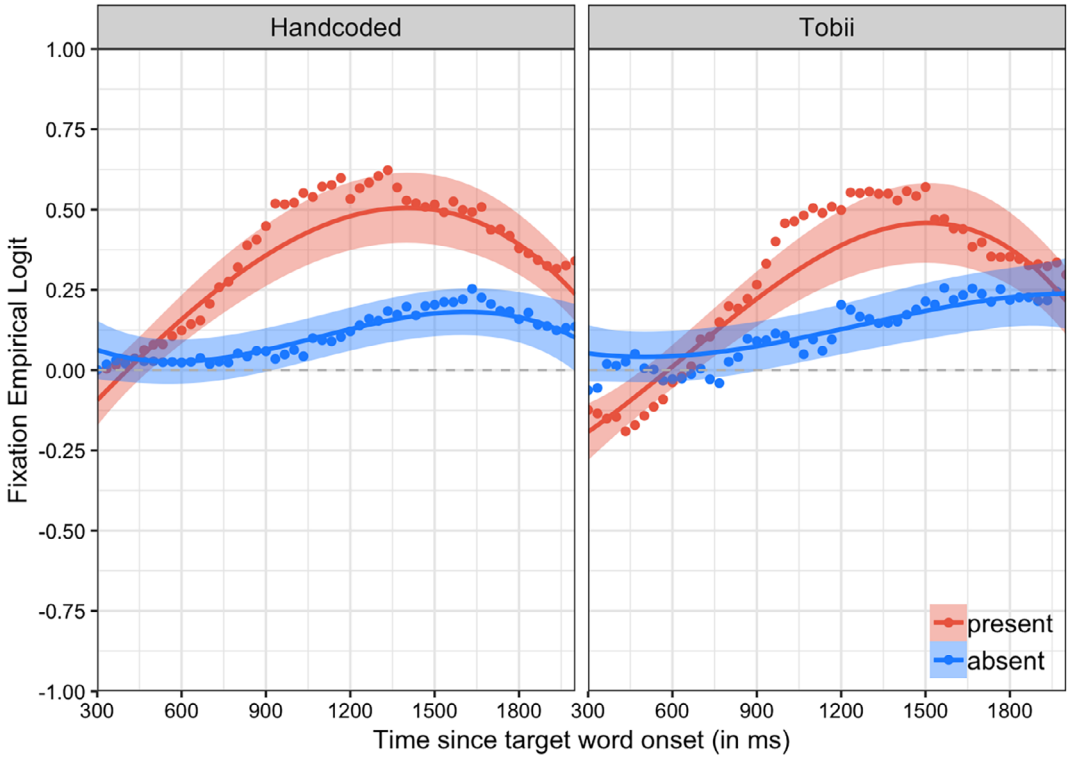

Figure 4. Probability of looking to the target during the analysis window (300–2,000 msec after noun onset), separated by source (Hand coded = manual gaze coding; Tobii = automatic eye tracking) and condition (Target Absent vs. Target Present). Dots represent group means for raw data. Solid lines represent growth curve estimates of looking probability. Shaded bands reflect ±1 *SE* around the mean.

significant effect on intercept ($P = 0.149$), quadratic time ($P = 0.065$), or cubic time ($P = 0.648$).

We were particularly interested in the presence of any significant Source by Condition interactions, which would indicate that the difference between the Target Absent and Target Present conditions was larger for one Source than for the other. There were no significant effects of the Condition by Source interaction on linear, quadratic, or cubic time (all Ps > 0.099). However, there was a significant effect of the Condition by Source interaction on the intercept ($P = 0.015$), indicating that the size of the Condition effect (i.e., the difference in overall accuracy on Target Absent vs. Target Present trials) was significantly different between the two Sources. In other words, although children looked significantly less at the target image in Target Absent than Target Present trials overall, the decrease in accuracy from Target Present to Target Absent trials was significantly larger for manual gaze coding than for automatic eye tracking. This pattern is evident in Figure 4, where the gap between the red curve (Target Present) and the blue curve (Target Absent) is larger for manual gaze coding than for automatic eye tracking.

[2]This difference may have been driven by the fact that children's accuracy started lower (below chance), thereby allowing more room for growth of the linear time term.

To complement the findings from the previous analyses, we conducted separate post hoc analyses for automatic eye-tracking and manual gaze coding to determine

Table 2. Model Results for Both Systems

	Estimate	*SE*	*t* value	*P* value
(Intercept)	0.210	0.052	4.052	<0.001*
ot1	0.894	0.239	3.744	<0.001*
ot2	−0.535	0.180	−2.972	0.003*
ot3	−0.187	0.079	−2.386	0.017*
Condition	−0.234	0.107	−2.190	0.029*
Source	−0.037	0.026	−1.443	0.149
ot1:Condition	−0.471	0.454	−1.037	0.300
ot2:Condition	0.850	0.355	2.393	0.017*
ot3:Condition	0.129	0.213	0.605	0.545
ot1:Source	0.268	0.121	2.214	0.027*
ot2:Source	0.150	0.081	1.846	0.065
ot3:Source	−0.042	0.092	−0.457	0.648
Condition:Source	0.100	0.041	2.423	0.015*
ot1:Condition:Source	−0.207	0.299	−0.694	0.488
ot2:Condition:Source	0.037	0.195	0.188	0.851
ot3:Condition:Source	0.249	0.151	1.653	0.099

Note. The independent variable was Time and the dependent variable was the log odds of looking to the target image. ot1 = linear time. ot2 = quadratic time. ot3 = cubic time. Condition (Target Present vs. Target Absent) and Source (manual gaze coding vs. automatic eye tracking) were contrast coded using −0.5 and 0.5. Thus, the overall Condition and Source results reflect average findings across both Conditions and/or Sources.

*Significance at $P < 0.05$.

Table 3. Model Results for Manual Gaze Coding

	Estimate	*SE*	*t* value	*P* value
(Intercept)	0.228	0.045	5.098	<0.001*
ot1	0.652	0.211	3.089	0.002*
ot2	−0.485	0.164	−2.952	0.003*
ot3	−0.189	0.089	−2.130	0.033*
Condition	−0.241	0.961	−2.511	0.012*
ot1:Condition	−0.428	0.424	−1.010	0.313
ot2:Condition	0.780	0.311	2.505	0.012*
ot3:Condition	−0.041	0.177	−0.231	0.818

Note. The independent variable was Time and the dependent variable was the log odds of looking to the target image. ot1 = linear time. ot2 = quadratic time. ot3 = cubic time. Condition (Target Present vs. Target Absent) was contrast coded using −0.5 and 0.5.

*Significance at $P < 0.05$.

what the results would have been if we had collected data from only a single system. The data processing and modeling approaches were identical to those used in the previous analyses, with two exceptions. First, children were not excluded if they failed to contribute adequate data for both systems. Instead, children were retained in the analyses for a given system if they contributed adequate data from that system alone, which resulted in a different sample size for each system. Second, Source was not entered into the models because each analysis included data from only one system.

The manual gaze coding data set included 65 of the original 70 children. Full model results for manual gaze coding are presented in Table 3. As in the previous analyses, there was a significant effect of Condition on the intercept ($P = 0.012$) and quadratic time ($P = 0.012$), indicating that children looked at the target image more and had a steeper peak asymptote in accuracy in the Target Present condition than in the Target Absent condition. The effect of Condition on linear time ($P = 0.313$) and cubic time ($P = 0.818$) remained nonsignificant. The automatic eye tracking data set included 53 of the original 70 children. Full model results for automatic eye tracking are presented in Table 4. Consistent with the previous analyses, there was a significant effect of Condition on quadratic time ($P = 0.013$) and no significant effect of Condition on linear time ($P = 0.190$) or cubic time ($P = 0.482$). In contrast to the previous analyses, however, there was not a significant effect of Condition on the intercept ($P = 0.164$), indicating no significant difference in the amount of time children spent looking at the target image in the Target Present and the Target Absent conditions. In sum, the overall findings for linear, quadratic, and cubic time for both systems were similar to the results in the previous analysis. However, Condition effects differed between the two systems; in this analysis, the effect of Condition was only significant for the manual gaze coding data set.

Table 4. Model Results for Automatic Eye Tracking

	Estimate	*SE*	*t* value	*P* value
(Intercept)	0.184	0.054	3.427	<0.001*
ot1	0.952	0.264	3.608	<0.001*
ot2	−0.485	0.178	−2.730	0.006*
ot3	−0.171	0.922	−1.858	0.063
Condition	−0.153	0.110	−1.391	0.164
ot1:Condition	−0.669	0.510	−1.311	0.190
ot2:Condition	0.938	0.378	2.479	0.013*
ot3:Condition	0.171	0.244	0.703	0.482

Note. The independent variable was Time and the dependent variable was the log odds of looking to the target image. ot1 = linear time. ot2 = quadratic time. ot3 = cubic time. Condition (Target Present vs. Target Absent) was contrast coded using −0.5 and 0.5.

*Significance at $P < 0.05$.

Discussion

To our knowledge, the current study is the first to directly compare data from automatic eye-tracking and manual gaze coding methods gathered simultaneously from the same children, during the same experimental sessions. As predicted, manual gaze coding produced significantly less data loss in young children with ASD than automatic eye tracking, as indicated by two different metrics: the number of usable trials and the proportion of looks to the images per trial. Anecdotal observations have suggested that manual gaze coding may be less vulnerable to data loss than automatic eye tracking, and the current empirical evidence supports these observations. This finding is important because limiting data loss increases the likelihood that the data on which we base our interpretations are valid and reliable. Maximizing validity and reliability is particularly important in studies of individual differences in children with ASD, which require accurate measurements at the level of individual participants. Thus, although eye tracking offers several clear advantages over manual gaze coding (e.g., automaticity, objectivity), manual gaze coding offers at least one advantage: lower rates of data loss in young children with ASD [Venker & Kover, 2015].

In addition to data loss, we directly compared the overall results from automatic eye-tracking and manual gaze coding by entering the data from both systems into a single model. There were numerous similarities in findings across the two systems, suggesting that automatic eye-tracking and manual gaze coding largely captured similar information. Regardless of system, children looked significantly more at the target image in the Target Present condition than in the Target Absent condition—an unsurprising finding, since the named object was visible only in the Target Present condition. Despite these similarities, results revealed one notable discrepancy between the two systems: the difference in overall accuracy between the two conditions was significantly larger for manual gaze coding than for automatic eye tracking. Though we had expected the two systems to differ in terms of data loss, we did not expect to find a change in

the pattern of results. This finding did not appear to be attributable to differences in statistical power related to the numbers of participants, as both data sets contained only the 51 participants who had contributed data from both systems. It is possible that having more trials per child and more data per trial in the manual gaze coded data decreased within-child variability and provided a more robust representation of children's performance.

Given the discrepancy between the patterns of results emerging from the two systems, we next asked: What would the results have been, and how might the conclusions have differed, if we had only gathered data from a single system? After all, most research labs use either one system or the other—not both. Post hoc analyses (on separately cleaned data sets for each system) revealed that the results of the manual gaze coding analysis mirrored those in the previous analysis. Specifically, children looked significantly more at the target image in the Target Present condition than in the Target Absent condition ($P = 0.012$). In contrast, the results of the automatic eye-tracking analysis indicated no significant difference ($P = 0.164$) in the amount of time children spent looking at the target image across the two conditions. Either of these findings—a significant difference between conditions, or a nonsignificant difference—may have important potential theoretical and clinical implications. Because the two systems yielded different conclusions, however, the implications of one set of results have the potential to be strikingly different from the implications of the other set of results. Thus, these findings suggest that the eye-gaze system used to address a particular scientific question could alter a study's results and the scientific conclusions that follow. In addition, these results provide additional context for the previous finding that manual gaze coding yielded a larger effect size between conditions than automatic eye tracking—namely, that the manual gaze coding data may have been driving the results in the primary analysis.

The post hoc analyses also revealed meaningful information about differences in data loss at the level of individual children. Following separate data cleaning for each system, the manual gaze coding data set included 65 (of the original 70) children, and the automatic eye tracking data set included 53 children. Thus, in addition to more trial loss and less looking time overall, more children were excluded (in an absolute sense) from the automatic eye-tracking data set than the manual gaze coding data set. Excluding participants is undesirable because it reduces statistical power and limits the generalizability of findings. Issues of generalizability are even more concerning when participants are excluded systematically, on the basis of child characteristics. In the current study, for example, the 17 children who were excluded from the automatic eye-tracking data set (but included in the manual gaze coding data set) had significantly higher autism severity ($M = 9.18$, $SD = 1.07$, range = 6–10) than the 53 children who were retained in the automatic eye-tracking data set ($M = 8.17$, $SD = 1.67$, range = 4–10; $P = 0.006$).[3] Our study is not the first to find a link between autism severity and child-level exclusion in an eye-gaze study. For example, Shic, Bradshaw, Klin, Scassellati, and Chawarska [2011] also found that toddlers with ASD who were excluded due to poor attention had significantly more severe autism symptoms than toddlers who were retained in the analyses. Given that children with high autism severity may have difficulties with language processing [Bavin et al., 2014; Goodwin et al., 2012; Potrzeba et al., 2015], it is critically important to consider how participant exclusion impacts study findings.

Though the current data cannot unambiguously answer this question, it is useful to consider why the manual gaze coding and automatic eye-tracking systems yielded different results, both in the primary analysis of children with data from both systems and the post hoc analyses of each system alone. Because the separate post hoc analyses contained different numbers of children (65 children for manual coding and 53 for eye-tracking), they may have been impacted by differences in statistical power. Also, recall that the children excluded from the separate automatic eye tracking data set had higher autism severity than those who were retained. Thus, differences in child characteristics may also have played a role in the post hoc analyses, since the manual gaze coding model represented a broader range of severity than the automatic eye-tracking model [Bavin et al., 2014; Shic et al., 2011]. However, the primary analyses could not have been affected by differences in participant exclusion since they contained only the 51 children who contributed data from both systems.

One potential explanation for the discrepancy in both the primary and post hoc analyses is a difference in accuracy—in other words, whether a child was truly fixating a given image at a given moment in time. Although the current findings do not speak directly to accuracy, either system could have produced inaccurate results. Manual gaze coding has been described as being more vulnerable to inaccuracy than automatic eye tracking because it is based on human judgment [e.g., Wass et al., 2013]. Indeed, human coders can certainly make incorrect decisions about gaze location. However, the fact that automatic eye tracking is based on light reflections and automated algorithms instead of human judgment does not mean it is always accurate. A growing number of studies have begun to identify concerns regarding the accuracy of automatic eye trackers, especially in populations that may demonstrate considerable head and body movement [Dalrymple et al., 2018; Hessels, Andersson, et al., 2015;

[3]The groups did not significantly differ in age, nonverbal IQ, or receptive or expressive language skills (all $Ps > 0.383$).

Hessels, Cornelissen, et al., 2015; Niehorster et al., 2018; Schlegelmilch & Wertz, 2019].

Dalrymple et al. [2018] examined the accuracy of an automated Tobii eye tracker and found that data from toddlers with typical development had poorer accuracy and precision than data from school-aged children and adults. In fact, the mean accuracy for toddlers fell outside the accuracy range described in the eye-tracker manual. The accuracy of remote eye-tracking systems appears to be particularly compromised when participants adopt nonoptimal poses, such as tilting their heads or rotating their heads to the left or right side [Hessels, Andersson, et al., 2015; Niehorster et al., 2018]. This is concerning, given that individuals with ASD often examine visual stimuli while adopting nonstandard head orientations, such as turning their heads and peering out of the corners of their eyes. High-quality calibration increases accuracy, but it can be difficult to achieve [Aslin & McMurray, 2004; Nyström et al., 2013; Schlegelmilch & Wertz, 2019; Tenenbaum, Amso, Abar, & Sheinkopf, 2014]. Some children may be unable to complete calibration [Dalrymple et al., 2018], and time spent on calibration (and re-calibration) decreases the likelihood that children will remain engaged in the remainder of the task [Aslin & McMurray, 2004]. It can also be difficult to tell whether poor calibration occurs because of the measurement error of the system, or because a child did not actually fixate the intended target.

The current study had several limitations. Our findings were based on one eye-tracking system and one manual gaze coding system, and other systems may produce different results [Hessels, Andersson, et al., 2015; Niehorster et al., 2018]. We focused on one set of data cleaning criteria, which in our experience are representative of those commonly used in published research. However, changes in trial-level and child-level cleaning criteria could have different effects. In addition, it is important to note that the current findings are most relevant to studies in which both manual gaze coding and automatic eye tracking are a potentially viable option—likely those using a 2-large-AOI design. The question of disproportionate data loss across children with ASD and children with typical development is an additional question that warrants future investigation.

Conclusions

As recently as 15 years ago, the use of automatic eye tracking in infants and young children was rare [Aslin, 2007; Aslin & McMurray, 2004]. Since that time, however, automatic eye tracking has become increasingly common in research labs studying young children, including those with neurodevelopmental disorders. Despite the clear methodological advantages of automatic eye tracking, manual gaze coding may limit rates of data loss in young children with ASD. Furthermore, the choice of eye-gaze system has the potential to impact statistical results and subsequent scientific conclusions. Given these findings, our research teams have continued to use manual gaze coding for studies in which the design and dependent variables allow for either type of system. It is our hope that the findings from the current study will allow autism researchers to make more informed decisions when selecting an eye-gaze system, whether either system would be appropriate. The information from this and future methodological studies will help researchers to select the eye-gaze measurement system that best fits their research questions and target population, as well as help consumers of autism research to interpret the findings from studies that utilize eye-gaze methods with children with ASD. In addition, these findings highlight the importance of continuing to develop more robust eye-gaze methods to maximize scientific progress in autism research.

Acknowledgments

We thank the families and children for giving their time to participate in this research. We thank Liz Premo for her critical role in data collection, Rob Olson for his technical assistance and expertise, and Jessica Umhoefer, Heidi Sindberg, and Corey Ray-Subramanian for their clinical expertise. We also thank the members of the Little Listeners Project team for their input and assistance. A portion of this work was presented in a talk at the American Speech-Language-Hearing Association conference in November, 2016. This work was supported by NIH R01 DC012513 (Ellis Weismer, Edwards, Saffran, PIs) and a core grant to the Waisman Center (U54 HD090256).

References

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Washington, DC: American Psychiatric Association.

Aslin, R. N. (2007). What's in a look? Developmental Science, 10, 48–53. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted

Aslin, R. N. (2012). Infant eyes: A window on cognitive development. Infancy, 17, 126–140. https://doi.org/10.1111/j.1532-7078.2011.00097.x.Infant

Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. Infancy, 6(2), 155–163. https://doi.org/10.1207/s15327078in0602_1

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Keep it maximal. Journal of Memory and Language, 68(3), 1–43. https://doi.org/10.1016/j.jml.2012.11.001.Random

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67, 1–48.

Bavin, E. L., Kidd, E., Prendergast, L., Baker, E., Dissanayake, C., & Prior, M. (2014). Severity of autism is related to children's language processing. Autism Research, 7, 687–694. https://doi.org/10.1002/aur.1410

Bishop, S. L., Guthrie, W., Coffing, M., & Lord, C. (2011). Convergent validity of the Mullen Scales of early learning and the differential ability scales in children with autism spectrum disorders. American Journal on Intellectual and Developmental Disabilities, 116, 331–343. https://doi.org/10.1352/1944-7558-116.5.331

Blaser, E., Eglington, L., Carter, A. S., & Kaldy, Z. (2014). Pupillometry reveals a mechanism for the Autism Spectrum Disorder (ASD) advantage in visual tasks. Scientific Reports, 4, 4301. https://doi.org/10.1038/srep04301

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. Cognition, 108, 896–904. https://doi.org/10.1016/j.cognition.2008.06.007

Chita-Tegmark, M., Arunachalam, S., Nelson, C. A., & Tager-Flusberg, H. (2015). Eye-tracking measurements of language processing: Developmental differences in children at high-risk for ASD. Journal of Autism and Developmental Disorders, 45, 3327–3338. https://doi.org/10.1007/s10803-015-2495-5

Collins, F. S., & Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. Nature, 505, 612–613.

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018). An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood. Frontiers in Psychology, 9, 1–12. https://doi.org/10.3389/fpsyg.2018.00803

Falck-Ytter, T., Bölte, S., & Gredebäck, G. (2013). Eye tracking in early autism research. Journal of Neurodevelopmental Disorders, 5, 1–11. https://doi.org/10.1186/1866-1955-5-28

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. Fernandez, & H. Clahsen (Eds.), Developmental psycholinguistics: on-line methods in children's language processing (pp. 97–135). Amsterdam: John Benjamins.

Goodwin, A., Fein, D., & Naigles, L. R. (2012). Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. Autism Research, 5(2), 109–123. https://doi.org/10.1002/aur.1220

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. Journal of Autism and Developmental Disorders, 39, 693–705. https://doi.org/10.1007/s10803-008-0674-3.Standardizing

Hessels, R. S., Andersson, R., Hooge, I. T. C., Nyström, M., & Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. Infancy, 20, 601–633. https://doi.org/10.1111/infa.12093

Hessels, R. S., Cornelissen, T. H. W., Kemner, C., & Hooge, I. T. C. (2015). Qualitative tests of remote eyetracker recovery and performance during head rotation. Behavior Research Methods, 47, 848–859.

Kaldy, Z., Kraper, C., Carter, A. S., & Blaser, E. (2011). Toddlers with Autism Spectrum Disorder are more successful at visual search than typically developing toddlers. Developmental Science, 14, 980–988. https://doi.org/10.1111/j.1467-7687.2011.01053.x

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. Developmental Review, 27(3), 283–348. https://doi.org/10.1016/j.dr.2007.06.006

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule, Second Edition (ADOS-2) Manual (Part 1): Modules 1–4. Torrence, CA: Western Psychological Services.

Mirman, D. (2014). Growth curve analysis and visualization using R. Boca Raton, FL: CRC Press.

Morgante, J. D., Zolfaghari, R., & Johnson, S. P. (2012). A critical test of temporal and spatial accuracy of the tobii T60XL eye tracker. Infancy, 17(1), 9–32. https://doi.org/10.1111/j.1532-7078.2011.00089.x

Mullen, E. M. (1995). Mullen scales of early learning. Minneapolis, MN: AGS Edition.

Naigles, L. R., & Tovar, A. T. (2012). Portable intermodal preferential looking (IPL): Investigating language comprehension in typically developing toddlers and young children with autism. Journal of Visualized Experiments, 70, e4331. https://doi.org/10.3791/4331

Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. Behavior Research Methods, 50, 213–227. https://doi.org/10.3758/s13428-017-0863-0

Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. Behavior Research Methods, 45, 272–288. https://doi.org/10.3758/s13428-012-0247-4

Oakes, L. M. (2010). Infancy guidelines for publishing eye-tracking data. Infancy, 15, 1–5. https://doi.org/10.1111/j.1532-7078.2010.00030.x

Oakes, L. M. (2012). Advances in eye tracking in infancy research. Infancy, 17, 1–8. https://doi.org/10.1111/j.1532-7078.2011.00101.x

Oakes, L. M., Kovack-lesh, K. A., & Horst, J. S. (2010). Two are better than one: Comparison influences infants' visual recognition memory. Journal of Experimental Child Psychology, 104(1), 124–131. https://doi.org/10.1016/j.jecp.2008.09.001.Two

Ozkan, A. (2018). Using eye-tracking methods in infant memory research. The Journal of Neurobehavioral Sciences, 5, 62–66.

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. Archives of General Psychiatry, 68(1), 101–109. https://doi.org/10.1001/archgenpsychiatry.2010.113

Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C. C., & Malige, A. (2016). Eye-tracking reveals abnormal visual preference for geometric images as an early biomarker of an ASD subtype associated with increased symptom severity. Biological Psychiatry. 79, 657–666. https://doi.org/10.1016/j.biopsych.2015.03.032

Potrzeba, E. R., Fein, D., & Naigles, L. (2015). Investigating the shape bias in typically developing children and children with autism spectrum disorders. Frontiers in Psychology, 6, 1–12. https://doi.org/10.3389/fpsyg.2015.00446

Rutter, M., LeCouteur, A., & Lord, C. (2003). Autism diagnostic interview-revised. Los Angeles: Western Psychological Service.

Schlegelmilch, K., & Wertz, A. E. (2019). The effects of calibration target, screen location, and movement type on infant

eye-tracking data quality. Infancy, 24(4), 636–662. https://doi.org/10.1111/infa.12294

Shic, F., Bradshaw, J., Klin, A., Scassellati, B., & Chawarska, K. (2011). Limited activity monitoring in toddlers with autism spectrum disorder. Brain Research, 1380, 246–254. https://doi.org/10.1016/J.BRAINRES.2010.11.074

Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). Processes of language acquisition in children with autism: evidence from preferential looking. Child Development, 78, 542–557. https://doi.org/10.1111/j.1467-8624.2007.01022.x

Tek, S., Jaffery, G., Fein, D., & Naigles, L. R. (2008). Do children with autism spectrum disorders show a shape bias in word learning? Autism Research, 1, 208–222. https://doi.org/10.1002/aur.38.Do

Tenenbaum, E. J., Amso, D., Abar, B., & Sheinkopf, S. J. (2014). Attention and word learning in autistic, language delayed and typically developing children. Frontiers in Psychology, 5, 490. https://doi.org/10.3389/fpsyg.2014.00490

Unruh, K. E., Sasson, N. J., Shafer, R. L., Whitten, A., Miller, S. J., Turner-Brown, L., & Bodfish, J. W. (2016). Social orienting and attention is influenced by the presence of competing nonsocial information in adolescents with autism. Frontiers in Neuroscience, 10, 1–12. https://doi.org/10.3389/fnins.2016.00586

Venker, C. E., & Kover, S. T. (2015). An open conversation on using eye-gaze methods in studies of neurodevelopmental disorders. Journal of Speech, Language, and Hearing Research, 58, 1719–1732. https://doi.org/doi:10.1044/2015_JSLHR-L-14-0304

Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. Infancy, 19, 427–460. https://doi.org/10.1111/infa.12055

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. Behavior Research Methods, 45, 229–250. https://doi.org/10.3758/s13428-012-0245-6

Yoder, P. J., Lloyd, B. P., & Symons, F. J. (2018). Observational measurement of behavior (2nd ed.). Baltimore, MD: Paul H. Brookes.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). Preschool language scales (5th ed.). San Antonio, TX: The Psychological Corporation.