AUTISM SYMPTOM PRESENTATION IN TODDLERS DURING BRIEF OBSERVATIONS: WHAT CAN WE SEE IN TEN MINUTES?

by

Terisa Poulsen Gabrielsen

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Educational Psychology

The University of Utah

August 2012

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Terisa Poulsen Gabrielsen**

has been approved by the following supervisory committee members:

Judith S. Miller , Chair **4/27/2012**
Date Approved

William R. Jenson , Co-Chair **4/27/2012**
Date Approved

Elaine Clark , Member **4/27/2012**
Date Approved

Paul S. Carbone , Member **4/27/2012**
Date Approved

Michael K. Gardner , Member **4/27/2012**
Date Approved

and by **Elaine Clark** , Chair of
the Department of **Educational Psychology**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Presentation of symptoms of Autism Spectrum Disorders (ASD) can vary within individuals and over the lifespan. This may make it difficult for a clinician with limited autism expertise to identify subtle signs of ASD at various ages. Furthermore, some individuals with ASD can exhibit appropriate social behaviors some of the time. There has been recent emphasis on ASD screening in primary care settings. However, the structure of a typical primary care appointment may not elicit an accurate sample of social behavior; this affects diagnostic impressions and subsequent referrals, which in turn can delay treatment during critical developmental periods. Typical primary care appointments are 10-20 minutes long, which may not be sufficient time to form an accurate clinical impression of need for referral. This study aimed to identify differences in symptom presentation during brief clinical observations among children with and without ASD. Participants included 3 groups of children between the ages of 15-33 months: (1) children with early signs of ASD, (2) children with suspected language delay, and (3) typically developing children. A clinical evaluation for early signs of ASD (including the Autism Diagnostic Observation Schedule [ADOS]) was videotaped. Social behaviors displayed by the children were measured across the first 10 minutes of the evaluation, and again 30 minutes into the evaluation for 10 minutes. Frequencies of atypical and typical behaviors and changes in behavior over time were compared among the 3 groups. Findings indicated low rates of atypical behavior relative to typical behavior in all three groups. Although the ASD group showed statistically higher rates of atypical behavior, and lower cognitive development and adaptive behaviors on standardized tests, expert clinical impressions based on 10-minute observations were incorrect in the ASD group 39% of the time. Brief observations likely do not provide enough of a behavioral sample to make a correct referral. Standardized screening identified more children correctly as needing referral for ASD evaluation than clinical impression alone.

To my husband, who followed my dream.

To the children he gave me and the grandchildren they have given us.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Atypical
AAP	American Academy of Pediatrics
ADOS	Autism Diagnostic Observation Schedule
ANOVA	Analysis of Variance
ASD	Autism Spectrum Disorder
CCD	Council on Children with Disabilities
CSHCN	Children with Special Health Care Needs
DVD	Digital Video Disc
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision
EI	Early Intervention
ERP	Event Related Potential
ESAT	Early Screening of Autistic Traits
HOME	Health Outcomes Medical Excellence
HSD	Honestly Significant Difference
IRB	Institutional Review Board
ITC	Communication and Symbolic Behavior Scales Developmental Profile Infant Toddler Checklist
JM	Judith Miller, PhD

M	Mean
M-CHAT	Modified Checklist for Autism in Toddlers
MRI	Magnetic Resonance Imaging
PDD-NOS	Pervasive Developmental Delay-Not Otherwise Specified
PIR	Partial Interval Recording
PPV	Positive Predictive Value
SD	Standard Deviation
SES	Socioeconomic Status
SORF	Systematic Observation of Red Flags
SPSS	Statistical Package for Social Sciences
SSD	Suspected Speech Delay
T	Typical
TDC	Typically Developing Control
TG	Terisa Gabrielsen
U	Unclear
X	Nonoccurrence

ACKNOWLEDGMENTS

There are many highly dedicated and talented people who have made this work possible. First is my mentor, Judith Miller, PhD, who has been ever generous in time and talent and to her family for the time she has spent away working with me. Megan Farley, PhD, and Leslie Speer, PhD devoted their incredible expertise and exponentially more time than they had agreed to for this project. Thank you also to their families for the time they spent coding. The invaluable research team of Michele Villalobos, Joe Viskochil Brandon Segura, and Natalie Wahmhoff tirelessly conducted evaluations of toddlers for years to get this data. Paul Carbone, MD has been a champion of not only this project, but autism services in Utah and the entire country. I have been extremely fortunate to have his guidance and support. Courtney Baker, PhD, from the Children's Hospital of Philadelphia was a critical statistical consultant in the final preparation phases of this manuscript. Finally, the highest respect and thanks to the physicians, staff, and patients' families from Granger Medical Pediatrics and the surrounding area.

INTRODUCTION

Autism Spectrum Disorders (ASDs) are a growing concern as prevalence rates are reported to be increasing (Centers for Disease Control and Prevention [CDC], 2007; 2009; 2012). Causes and cures have yet to be identified, so early intervention is critical to improve the outcome of a child with ASD (Dawson et al., 2004; Nadel & Poss, 2007; Rogers, 1996). Early identification is the key to getting children enrolled in early intervention programs during prime periods of language, cognitive, and social development. This is not easily done, as early signs of ASDs can be subtle and difficult to identify at early ages. A formal screening opportunity is one method of identifying children with ASDs that can prompt a referral for evaluation and intervention earlier than waiting for parents or clinicians to become concerned on their own (Miller et al., 2011; Yirmiya & Charman, 2010). Unfortunately, adoption of formal universal screening is difficult, despite professional recommendations and access to free or low cost screening instruments. Health care providers report reliance on their own clinical judgment to determine if a child is exhibiting early signs of ASD or not (American Academy of Pediatrics [AAP], 2003). This study takes the first step of trying to understand this clinical judgment by examining what kind of behaviors can reasonably be observed during a brief interaction such as a pediatric visit.

Autism Spectrum Disorders

An autism spectrum disorder (ASD) is a lifelong condition of functional impairment affecting reciprocal social interaction, communication and language. ASDs are also characterized by restricted or repetitive interests, behavior, or activities. The level of affectedness or impairment varies broadly, as do the manifestations of symptoms.

Within the spectrum of autism disorders, Autistic Disorder is diagnosed when there are multiple symptoms across all of the diagnostic domains: qualitative impairment in reciprocal social interaction, qualitative impairment in communication, and the presence of restricted interests and repetitive behaviors. To meet diagnostic criteria, the symptoms must begin to appear before the age of 3 years (APA, 2000), making it one of the earliest emerging neurodevelopmental conditions (Yirmiya & Charman, 2010). Asperger Disorder and Pervasive Developmental Delay-Not Otherwise Specified (PDD-NOS) have similar symptoms, but to a lesser degree or with fewer symptoms in total. Age of onset is not specified. Furthermore, to meet the diagnostic criteria for Asperger Disorder, no impairment or delay in language, cognitive, or adaptive skill is permitted. Together, these three disorders are commonly referred to as autism spectrum disorders (ASDs; Johnson, Myers, & Council on Children with Disabilities [CCD], 2007.)

Autism spectrum disorders have become a topic of growing concern in health care, education, psychology, and public policy, as the numbers of children diagnosed with autism is rapidly rising and these systems are tasked with responding to increasing needs. With this growth, more research is being funded and conducted and our knowledge about autism spectrum disorders is growing rapidly. There is still much that is not known, however. Notably, there is not yet a practical or reliable physiological method of identifying ASDs before behavioral symptoms begin to appear. Parents, health care providers, and educators are becoming more aware of the early behavioral signs of autism, but symptoms can be subtle in the early stages and early symptomatic presentations can be confusing. It is nevertheless very important that ASDs be identified as early as possible in order for intervention to begin during critical developmental periods of language, cognitive and social development.

One focus of research worldwide has been on standardized behavioral screening of toddlers to identify children with early signs of ASD. Although screening studies show promise, there are many obstacles to overcome before screening is universally adopted. Clinical judgment is still the predominant method of identification among primary care providers, but there are also challenges in making accurate clinical judgments. Many factors exist that can affect a clinician's perception of a child's symptomatic presentation.

Parents, health care providers, early childhood educators, and policy makers need research data about current practices, specifically the use of clinical judgment to identify early signs of ASD for purposes of early intervention referrals. As part of evidence based practice for all concerned systems, research regarding the symptomatic presentation of toddlers during brief evaluations may yield important information that will guide decision making about early identification practices and policies.

ASDs Are an Increasing Public Health and Education Concern

Growing Prevalence Rates of ASDs

Less than one generation ago, autism spectrum disorders were thought to be rare, with epidemiological studies reporting a median rate of 5 per 10,000 (1 in 2,000), and a range of 2-20 per 10,000 for Autistic Disorder (APA, 2000). However, a recently published study with community samples from 14 U.S. states indicated that 11.3 per 1,000 (approximately 1 in 88) children at age 8 were identified from health and/or educational records as having symptoms that meet criteria for an ASD (CDC, 2012). This compares to rates of 9 per 1,000 (approximately 1 in 110: CDC, 2009) and 6.7 per 1,000 (approximately 1 in 150: CDC, 2007) that were previously reported by the Autism and Developmental Disabilities Monitoring (ADDM) Network. Each of the studies was conducted using the same methodology, which allows meaningful comparisons of rates across years. The 13 communities that participated in both studies reported an increase in prevalence of 78%

between the 2002 and 2008 cohorts (CDC, 2012). The authors of the CDC study caution that the ADDM Network sites are not representative of the United States as a whole, so the rates themselves cannot be generalized, but the increases in rates reported by these studies confirms that the prevalence of ASDs within the surveillance populations is increasing. The new rate of 1 in 88 is consistent with a rate calculated from parent reports in a 2007 survey, reporting prevalence of 11 in 1,000 (approximately 1 in 91; Kogan et al., 2009).

Reports from the ADDM network use a methodology that does not rely on official diagnoses of ASDs. Instead, educational and/or health records of individuals are examined for descriptions of ASD behaviors, and case status can be assigned even if the diagnosis or educational classification was not given. This method of investigation controls for the effects of selection bias for or against an ASD diagnosis by studying all eligible children in the geographic region of interest, rather than just those with a clinical diagnosis or just those who are willing to participate in an evaluation. The Utah prevalence rate that was reported in the ADDM study of the 2008 cohort (2,123 children in Utah) was 1 in 47. The Utah sample population was noted to be smaller in 2008 than in 2002, which may have an impact in the representative nature of the 2008 rate (CDC, 2012). The rate from the 2002 cohort was 1 in 133 (CDC, 2007).

A separate Utah study compared the same 2 cohort years (2002 and 2008) using the different approach of determining administrative prevalence (case ascertainment by reported health diagnoses and/or educational classifications only). According to this methodology, the prevalence rate for autism spectrum disorders in Utah doubled between 2002 and 2008, with the rate among 8-year olds in 2008 calculated to be 1 in 77. This estimate was based on cases found in a sample population of 33, 757 children in Davis, Salt Lake, and Utah counties (Pinborough-Zimmerman et al., 2011). These and other states' prevalence rates make ASD now among the most common forms of severe developmental disability (Zwaigenbaum et al., 2009).

Early Intervention: Systemic Response to Increasing Needs

Early intensive intervention results in better outcomes for children with ASDs (Dawson et al., 2004; Nadel & Poss, 2007; Rogers, 1996, Rogers & Vismara, 2008). One method of quantifying the benefits is in terms of cost savings attributable to early intervention. Lifetime incremental societal cost for a person with an ASD in the United States is estimated to be $3.2 million (Ganz, 2007). A study in the United Kingdom estimates the annual costs of autism to that nation to be £28 billion sterling (approximately $44.24 billion U.S., based on exchange rate of 1 British Pound Sterling to 1.58 U.S. dollars), with individual annual costs for children estimated to be £2.7 billion ($4.26 billion U.S.). They further calculated lifetime costs per person to be £1.23 million ($1.94 million U.S.) for a person with ASD and intellectual disability and £ 0.80 million ($1.26 million U.S.) for a person with ASD who does not have an intellectual disability (Knapp, Romeo, & Beecham, 2009). With early identification and early intervention, improved functionality and outcome can result in an estimated lifetime savings of $1.8 million. Net savings for the educational system alone are estimated to be $364,896 to $335,618 per child (figures adjusted for inflation: Jacobson, Mulick, & Green, 1998).

United States public policymakers have responded to the increase in ASDs with two legislative mandates. One is the Combating Autism Act, first enacted in 2008 and reauthorized in 2011 as the Combating Autism Reauthorization Act (CARA: Public Law 122-32, 112th Congress, 2011). This legislation provides funding for research, screening, and intervention (Interagency Autism Coordinating Committee, 2010).

The other legislation is The Individuals with Disabilities Education Improvement Act of 2004 (IDEIA: U.S. Dept. of Education, 2004). IDEIA is a broad education legislation requiring that individuals with disabilities must have access to a free and appropriate public education, which has been specifically outlined to include both early intervention (age 0-3 years) and education (age 3-21 years). Part C of the Individuals with Disabilities Education Act (IDEA), as amended by IDEIA, provides federal funds to states to make available Early Intervention services for infants and toddlers (from birth to age 3) with disabilities and their families. Part B of the Act covers services for children ages 3 to 21 years old.

The structure and funding of Early Intervention services varies from state to state, but its purpose is always to provide therapeutic services to clients from birth to 3 years old who have developmental delays or disabilities, including ASDs. The aim of early intervention is to help children to grow and develop (National Dissemination Center for Children with Disabilities [NICHCY], 2011; U.S. Department of Education, 2004). While Early Intervention services exist in all states, the quality of services in terms of time spent with the child, therapies received, and access to services can vary by income, language status, and location, especially in rural areas (Peacock & Lin, 2012).

Formal diagnosis is not needed to qualify for Early Intervention services in the United States, as these programs qualify children for services based on documented delays in specific areas of development. If early symptoms are not noticed or identified by either parents or professionals, however, intervention (public or private) may not begin until symptoms have become more pronounced and the child's developmental trajectory is significantly different from other children the same age.

Early Intervention Program Components

Effective components of an early intervention program include, but are not limited to the following: (1) entry into intervention as soon as an ASD is suspected rather than waiting for formal diagnosis; (2) intensity of 25 hours per week; (3) inclusion of a family component; (4) interactions with typical peers; and (5) generalization of skills learned (Johnson et al., 2007). Rogers and Vismara (2008) indicated additional critical components are (6) building functional communication skills and (7) addressing challenging behaviors with positive behavioral supports. Although the intensity of 25 hours per week is ideal, early intervention has also been shown to be effective even if the amount of time spent in direct intervention is as little as one hour per week (Vismara, Columbi & Rogers, 2009).

There is no single model for Early Intervention programs, and even within a single state, several agencies are contracted for Early Intervention services, which introduces an element of variability in the characteristics of programs. A review of published research on early intervention programs found 59% to include a family component, 37% conducted in natural settings, and that the all reviewed programs were reported to use systematic interventions. The number of service hours varied from 1 to 40, and did not include hours of intervention provided by parents (Schertz, Baker, Hurwitz, & Brenner, 2011).

Schwartz and Sandall (2010) discuss the realities of Early Intervention programs compared to the recommended components, citing Hebbler et al. (2007) in stating that the average number of service hours provided by Early Intervention programs to children with ASD is 90 minutes weekly. Disparities in service levels exist between geographical areas with higher levels of education and better access to regional autism centers compared to areas without these advantages. Increasing the infrastructure of the Part C system is recommended as the remedy for these disparities, rather than decreases in service delivery.

Evidence for Effectiveness

For specific programs, a systematic review of early intensive interventions by the Vanderbilt Evidence-Based Practice Center concluded that Lovaas types of intervention and the Early Start Denver Model (Rogers & Vismara, 2008) show evidence of some improvements in cognitive performance, language skills, and adaptive behavior skills in some young children with ASD. Some of the studies are rated as good (Early Start Denver Model, Randomized Control Trial), some rated as fair (non-Randomized Control Trial Early Start Denver Model and 7 of the Lovaas studies), and some rated as poor (14 Lovaas studies; Warren et al., 2011). Rogers and Vismara (2008) also reviewed early intensive interventions and found Lovaas' techniques in general to be well-established interventions.

Variability in programs does not necessarily mean treatment effectiveness is also variable. A comprehensive behavioral treatment for young children has been determined to be a treatment with an Established level of evidence by the National Standards Project in the age groups of 0-2 and 3-5 and older children. Skills increased with this type of comprehensive intervention include communication skills, higher cognitive functions, interpersonal skills, motor skills, personal responsibility skills, and play skills (Horner, Carr, Strain, Todd, & Reed, 2002; National Autism Center, 2009; Schreibman, 2000). Using evidence based interventions, one new direction in research is to move beyond a comprehensive program to a highly individualized program that can incorporate different treatment approaches to address the variety of treatment needs in children with ASDs. Research to establish specific methodologies for matching interventions to individual needs is in the early stages, however (Stahmer, Shreibman, & Cunningham, 2010).

The Role of Child Find in Early Intervention

IDEIA legislation in 2004 revised IDEA to emphasize Child Find (a federal mandate to states requiring search activities to identify, locate, and evaluate children with disabilities) for underserved populations of infants and toddlers (U.S. Dept. of Education, 2004). The rationale behind Child Find is that the state has an active role and responsibility to search for children who may need intervention before they enter the educational system, as not all parents may be aware of developmental problems or available services. This mandate applies to ASDs as well as other types of developmental delay or disability. The law does not state specific methodologies for identifying children with disabilities, however.

Each state administers its own Child Find programs. As a result, systems vary by state. For example, in Utah, Baby Watch is the State Health Department organization that oversees Child Find activities in the 0 to 3-

year-old age range. During the federal fiscal year of 2009-2010, Utah Baby Watch Early Intervention Programs (BWEIP) reported serving 1.96% of Utah children aged 0 – 2 years. While this level of service exceeded the target rate of 1.92%, the national baseline percentage rate for states with similarly narrow eligibility criteria was 2.67% (Utah BWEIP, 2010). This figure compares with results from the national Early Childhood Longitudinal Study–Birth Cohort stating that at 24 months of age, nearly 14% of children have developmental delays that are likely to make them eligible for Early Intervention services, indicating there are many children who are not identified for intervention by age 2. (Radecki, Sand-Loud, O'Connor, Sharp, & Olson, 2011).

In the Federal Fiscal Year 2009-2010, health care providers were the largest referral source for children in Early Intervention, making 50% of referrals. Thus health care providers are a natural target for Child Find efforts to increase the identification of the number of children eligible to be served in Early Intervention (Utah BWEIP, 2010). Utah BWEIP employs outreach strategies such as distribution of bookmarks in Spanish; building relationships with agencies such as Foster Care, local hospitals, the family and pediatric practice community; and homeless outreach to increase referrals (Utah BWEIP, 2008). Pediatric health care providers are the only professionals who have consistent and comprehensive access to children in the critical toddler age range, also indicating the need for targeting pediatric health care providers in efforts to identify children with ASDs as early as possible (Robins & Dumont-Mathieu, 2006).

Early identification of children with ASD can convey many benefits to the child and his or her family. These benefits include timely information about educational, social, communication, and behavioral interventions. Benefits also include early identification of co-morbid medical, developmental, and psychiatric conditions. Understanding a child's profile of strengths and difficulties, coupled with appropriate intervention, is likely to reduce the risk of developing secondary behavioral difficulties (Dover & LeCouteur, 2007). Early identification is considered essential to ensure that children receive optimal intervention services (CDC, 2007; Johnson et al, 2007).

Early Identification of ASD Is Complicated, Which Delays Intervention

There are many aspects of early signs of ASD that can make early identification of symptoms difficult to identify. Most ASDs have an idiopathic etiology, making it difficult to determine if a child is at risk. Some physiological signs are beginning to show promise as possible early markers of ASDs, but none have been established as reliable indicators of ASD risk or early predictors. Even if symptoms are rising to a level of concern, parents may not act on concerns for several months and health care providers may focus on other conditions as explanations for symptomatic presentation and delay treatment. Early signs can be subtle, and there may be a regression of skills that complicates early symptomatic presentation. There are several influences that can impact clinical judgment of the need for referral to early intervention services, including unintentional bias. All of these factors can delay identification of early signs of ASDs and therefore delay entry into early intervention.

Etiology

ASDs are generally understood to be biologically based neurodevelopmental disorders that are highly heritable. Both genetic susceptibility and protective genetic factors may be involved (Limperopoulos, 2009). Approximately 6-10% of individuals with ASD have genetic or and metabolic conditions that are known to be associated with ASDs (Cohen et al., 2005; Johnson et al., 2007). These conditions include Fragile X Syndrome, Rett Disorder [Syndrome], Tuberous Sclerosis Complex, Phenylketoniuria, Prader-Willi or Angelman Syndrome, Trisomy 21, Neurofibromatosis, Adenylosuccinate Lyase Deficiency, Smith-Magenis Syndrome, San Filippo Syndrome, and Smith-Lemli-Opitz Syndrome. Because these are frequently also associated with severe developmental delays, they are often identified early in life. At least 17 genes have been identified as candidates for genetic causes of syndromic ASDs. However, the majority of individuals with ASDs have an idiopathic or nonsyndromic type, and will not have an identifiable “secondary” cause. (Caglayan, 2010). Some consensus views of ASD etiology are that it is the result of three vectors of influence: (1) an underlying biologic vulnerability in individuals who experience (2) varying degrees of exogenous stressors (3) during a critical period of brain development (Limperopoulos, 2009).

This complicated picture of multifactorial pathogenesis is problematic both for identification and for discovery of etiology of ASDs. There is no universal age at which the behaviors of idiopathic ASDs manifest, nor a specific developmental presentation or course associated with idiopathic ASD. Rather, idiopathic ASD is associated with a wide variety of behavioral presentations which, at some unspecified point in development, reaches a level of concern to parents, physicians, or teachers. Research is ongoing to better characterize genetic and environmental influences in the etiology of ASDs (Interagency Autism Coordinating Committee, 2010). As children with an ASD grow, their developmental trajectories diverge further and further away from typical development. Earlier identification of ASDs may aid research into the etiology of ASD by facilitating research that can occur closer and closer to the actual onset of the disorder (Johnson et al., 2007).

Early Symptomatic Presentation of ASDs

The Earliest Signs of ASD

Research into the very earliest signs of autism is relatively new, but the growing literature base is beginning to shed light on how ASDs unfold early in development. Physiological changes are beginning to be identified as early as 6 months, before behavioral symptoms become apparent (Wolff et al, 2012). However, type and severity of symptoms at onset do not necessarily predict functional outcome (Shumway et al., 2011), and symptomatic presentation can vary widely in children between 6 months and 2 years old (Yirmiya & Ozonoff, 2007). Children in this age range are not yet expected to have developed peer relationships or conversation skills, so signs of problems with social development as indicated by diagnostic criteria in the Diagnostic and Statistical Manual, Fourth Edition, Text Revision (DSM-IV-TR; APA, 2000) may not always be evident. While the core symptom of ASD is a deficit in social relating, recent research suggests that autism disrupts multiple aspects of early development (Rogers, 2009). Even in children without clear difficulties in the first year of life, by age 18 to 24 months, symptoms of ASD are generally apparent (Zwaigenbaum et al., 2009).

Early Physiological Signs

Among studies to find physiological early signs is a multisite study of infant siblings of children with autism (Wolff et al., 2012). Autism is highly heritable, so younger siblings of children already identified with ASDs are at greater risk. The prevalence rate for autism in siblings of a child with an ASD is estimated to be 4- 10% (Szatmari, Jones, Zwaigenbaum, & MacLean, 1998; Sumi, Taniai, Miyachi, & Tanemura, 2006). Infant siblings of children already identified with ASD, then, are important research targets. The chance of observing an ASD unfold is much higher due to the higher genetic risk. The opportunity to witness ASD unfold as close to birth as possible means research into the actual onset and developmental course (rather than retrospective accounts of developmental differences) can progress.

One such study is the Infant Brain Imaging Study (IBIS) network, which uses Diffusion Tensor Imaging (a type of Magnetic Resonance Imaging, or MRI) technology to observe changes in babies' brains as they develop. Infants enter the study for a series of evaluations that involve developmental testing, biometric measurements, parent interviews about adaptive skill development, language developmental testing, and evaluation for autism symptoms. These visits begin at age 6 months and are repeated at 12, 24, and 36 months. Intermediate visits at 18 and 30 months can also be made.

To date, the network of sites (Center for Autism Research at the Children's Hospital of Philadelphia, The University of North Carolina, University of Utah, Washington University in St. Louis, University of Washington, McGill University, and the University of Alberta) have released data on studies of 92 infants in the study. Differences in brain development (white matter fiber tract organization) have been found in the 28 infants (30%) who were identified as having an ASD by the 24-month evaluation (Wolff et al., 2012). While these findings suggest that early signs of ASD may be discernible by technology such as MRI, there are limitations. The findings in these siblings may not be representative of the ASD population as a whole, and results from MRI screening are a long way from informing screening and identification practices.

Early Behavioral Signs Noticed by Parents

Parents often report noticing developmental concerns between 14 and 19 months of age. First symptoms noticed are delays in speech and language developments, followed by abnormal social responsivity, medical problems, and nonspecific difficulties related to sleeping, eating, and attention. At this very young age, parents typically do not report concerns regarding unusual interest and stereotyped behaviors. This may be due to later onset of these symptoms or milder manifestations at earlier ages (Chawarska, et al., 2007).

Other studies indicate an age of recognition by parents in the age range of 15 to 18 months, but also indicate that parents may delay discussing their concerns with their child's physician for several months (Johnson et al., 2007). Even if concerns are noted by parents at 17 to 19 months, parents may not seek professional advice until the child is 21 to 25 months old (Chawarska et al., 2007; Gupta et al., 2007; Johnson et al., 2007). A common finding in each of these studies, however, is a delay of months to years between first recognition of symptoms and formal identification of ASDs.

Retrospective Video Studies of Early Signs

The increasing use of home video to record a child's first years have provided researchers with methods of investigating early behavioral signs of

ASDs without reliance on parent report which may be biased or incomplete. A review of these studies indicates that differences in behavior can be seen as early as 4 – 6 months. Some differences observed include motor development, joint attention, orientation to name, communication, language development (babbling), eye contact and gaze. These differences were conceptualized as dyadic abnormalities as well as reduced amounts of time paid to social stimuli (Yirmiya & Charman, 2010). Although these differences are evident through detailed coding by trained professionals and use of statistical comparison methods, they may not be sufficient to cause early concern in parents as they are subtle and consist primarily of the absence of normal social developmental signs, not the presence of atypical or concerning behavior.

Regression

Another variation in symptom presentation that may complicate early identification is the occurrence of regression. As many as 27% of children with ASDs are considered by their parents to be developing normally until the age of 15 to 24 months, at which point some children who had been communicating with words and gestures and/or had been making eye contact and responding to praise cease to display these skills (Johnson et al., 2007; Lingam et al., 2003, as cited in Ozonoff et al., 2010). One retrospective video study found that children with a regression showed higher abilities in babbling compared to nonregressive cases of ASD before the regression

(Yirimiya & Charman, 2010). Research suggests that regression may occur in most if not all children with ASD, but at different times and rates (Ozonoff, 2009).

Research into the early physiological signs of ASDs suggests that it might be possible to detect early signs of regression, but no reliable methods have yet been established. Nordahl et al. (2011) found increased head circumference to be associated with regression in males with a divergent pattern beginning at 4 to 5 months of age. The cumulative effect of this accelerated growth was 5% difference from typical development and non-regressive ASDs at 18 months (the age where regressions begin to be noticed), and 6% difference at age 3. These differences occur gradually and are so slight as to not cause much alarm at a regular well-baby exam or even be detected upon casual examination by parents or health care providers.

Parents may not be accurate observers of regression. In examining this question of parents' ability to express concern about early onset of symptoms, Ozonoff et al. (2011) found that low or decreased levels of social communicative behaviors, and social-communicative behaviors that developed, but then stopped developing (reached a plateau) could be observed in home videos, but these patterns of symptom onset had poor correspondence with parent reports of symptom development. In a prospective study, Ozonoff et al., (2010) found that social-communicative behavior declines were significantly evident by 12 months, but that 83% of parents failed to report regression symptoms at that time.

Thus, while some forms of regression appear to be significant enough to be noticed by parents, decelerations or developmental plateaus may not be very noticeable, or perhaps not noticeable at the time they are first becoming apparent. Furthermore, subtle signs of ASD may exist and go undetected before a regression occurs. This makes it difficult to rely on parents or provider observations of regression, as the timing and severity of a regression is widely variable. It also makes it difficult to know the earliest age at which ASD features can be reliably observed in community practice (outside of focused research studies). If a child is screened or evaluated for early signs of autism too early, a regressive presentation may be missed—if evaluation is postponed, it may be too late and the opportunity for earlier identification and intervention is missed.

How Early Are ASDs Diagnosed?

Given the importance of early identification and intervention, the age of diagnosis of ASDs can be examined as an indicator of how early ASDs are currently being identified. There is a lower limit to the age at which a stable diagnosis can be made based on a behavioral presentation. Research has established multiple disparities in health care in the United States based on race, ethnicity, language, and SES (Smedley, Stith, Nelson, National

Institute of Medicine, 2002). Unfortunately, such disparities are also found in how early an ASD is diagnosed.

Lower Limit of Diagnostic Stability

Diagnoses of ASDs made as early as age 2 have been shown to be feasible (Baird, Charman, Baron-Cohen, Cox, Swettenham, Wheelwright, & Drew, 2001; Lord et al., 2006; Robins, Fein, Barton & Green, 2001; Stone, Coonrod, & Ousley, 2000). Diagnoses of children at younger ages have not yet been found to be a stable predictor of maintaining an ASD diagnosis as an older child. This may be because early diagnosis of ASD is challenging in children younger than 2 years old due to the subtlety and variety of symptoms (Zwaigenbaum et al., 2009). Attempts to find symptomatic predictors of ASD in infants as young as 6 months have been made, but differences in symptoms (e.g., reduced eye gaze, lack of response to name) have not yet proven to be reliably predictive of development of an ASD (Nadig et al., 2007; Young, Merin, Rogers, & Ozonoff, 2009). While diagnostic difficulties exist within the first 2 years of a child's life, these years are also the years in which intervention is extremely important. Even if a definitive diagnosis is not obtained during these early years, the most important action that can be taken is that parents be referred to appropriate intervention services and parent support resources if there is a concern (Zwaigenbaum et al., 2009).

Differences in Average Age of Diagnosis

Average age of diagnosis varies in different populations according to several variables. In Hong Kong, for example, the average age of diagnosis is 3 years (Wong et al., 2004). The average age of ASD diagnosis in the United States, however, is approximately age 4 for the general population and later for disadvantaged populations (Mandell, Ittenbach, Levy & Pinto-Martin, 2007; Mandell et al., 2009; Rice, 2009; Zwiegenbaum et al., 2009). Examination of records of 2,568 children in one of the earlier ADDM prevalence studies (13 U.S. states participating, children who were 8 years old in 2002) found the average age of diagnosis in this cohort to be 5.7 years (Shattuck et al., 2009). In the most recent ADDM prevalence study, median age of earliest diagnosis ranged from 48 months for Autistic Disorder to 75 months for Asperger Disorder. PDD-NOS or ASD subtypes were reported with a median age of 53 months for earliest diagnosis (CDC, 2012). Average age of diagnosis for children enrolled in Medicaid was found to be 5 years, 4.9 months (Mandell et al., 2010). Some systemic and/or cultural factors such as race and ethnicity and SES have been found to be associated with delayed or earlier ages of ASD diagnosis.

Racial Disparities in Age of Diagnosis

Delayed Diagnosis

Mandell et al. (2007) reported African American children were 2.6 times less likely than non-Hispanic Caucasian children to receive an ASD diagnosis on their first specialty care visit. Similar disparities among African American, Hispanic, and children of other ethnicities and races compared to white children were found in a larger study (Mandell et al., 2009). Research is now being conducted to see if differences in presentation of symptoms may actually differ between White and African American children as one explanation for these disparities. One study shows significant differences in inflexible adherence to nonfunctional routines/rituals, persistent preoccupation with parts of objects, abnormal motor development, and odd responses to sensory stimuli, with these symptoms higher in White children than African American children. No differences in any of the recorded core social symptoms of ASD were found between the two groups, however. The study was conducted by reviewing records, so the documentation of symptoms by various clinicians may yet have been influenced by bias (Sell, Giarelli, Blum, Hanlon, & Levy, 2012).

Earlier Diagnosis

Racial differences can also correlate with earlier diagnoses. Asian children enrolled in Medicaid 2002--2004 were diagnosed at earlier ages than the other cultural and racial groups enrolled in Medicaid (56.7 months for autistic disorder, 61.3 months for other ASDs). Latino children in this large sample (28,722 total) were also found to be diagnosed earlier than other groups (58.8 months for autistic disorder and 62.7 months for other ASDs). Even though these ages are earlier than the average age for the sample as a whole (64.9 months), they still represent diagnoses that are given much later than is optimal (Mandell et al., 2010).

Disparities According to Lower SES

Mandell et al. (2007) reported families with lower income levels were 3.4 times more likely to receive a misdiagnosis before their child was eventually diagnosed with ASD at a later date. Social class bias exists across studies of ASD (i.e., higher rates reported in higher social classes) perhaps due to differences in case ascertainment and inequalities in access to health care (Bhasin & Schendel, 2007). Children enrolled in Medicaid were diagnosed at a late age (64.9 months on average), and this disparity was found to be unrelated to governmental policies (i.e., Medicaid enrollment) or systemic factors (i.e., access to health care). Rather, the disparity seemed to be associated with child-level factors. For instance, children who qualified for Medicaid through poverty were diagnosed with ASD, on average, earlier than children who were enrolled in Medicaid due to foster placement, disability, or other reasons. Medicaid enrollment itself was less of a determining factor than the reason a child was qualified for Medicaid. The study also found children in large urban or rural areas to be diagnosed at later ages than children in small urban or suburban areas. This suggests that even though health care options may be more available in a large urban area, an individual child within that urban area may still not have adequate access to health care due to other factors (Mandell et al., 2010).

What Can Be Done to Identify ASDs Earlier?

To improve the outcome of children with ASDs through earlier intervention, advocacy organizations have sponsored efforts to increase public and health care provider awareness of early signs. Efforts to promote early identification in a more systematic manner have focused on screening for ASDs as a way to identify ASDs earlier and to also gather epidemiological information about the prevalence of ASDs in various populations.

Increasing Awareness of Early Signs

Efforts to identify ASDs earlier have included public awareness campaigns such as "'Learn the Signs. Act Early." (CDC, 2010) and "Red Flags" (First Signs, 2012). Both include websites and advertising campaigns to educate parents on signs to watch for typical and atypical development. Another public service ad campaign was launched in 2005 by Autism Speaks, a large private, nonprofit autism advocacy organization in conjunction with the Ad Council, another nonprofit organization dedicated to improving the lives of the American people (Autism Speaks, Inc., 2012).

"Learn the Signs. Act Early."

The "Learn the Signs. Act Early." campaign posts developmental milestones for parents to watch for and a list of "Act Early Warning Signs," on the Internet. The milestones include a list of behaviors that, if not present, should be brought to the attention of a health care provider (see

Appendix A). This campaign targeted three key audiences: (1) parents of children younger than 4 years old; (2) health care professionals;, and (3) early childhood educators and childcare providers. Preparatory research for the campaign discovered only 37% of parents knew what behaviors were suggestive of ASDs and only 22% knew the best time to seek help was before age 2. Educators reported that they needed more information to talk with parents about developmental concerns (Daniel, Prue, Taylor, Thomas, & Scales, 2009).

At baseline (before the campaign was launched), health care providers revealed that 95% agreed autism interventions were most effective if initiated prior to school age, yet 30% recommended that concerned parents wait and see if their child's development progressed when symptoms were first brought to their attention. Only 27% of health care professionals believed that health care professionals in general were well informed about developmental disorders and 41% agreed that they (personally) had necessary resources to educate parents.

Two years following the launch of the campaign, 57% of parents knew the early warning signs, and 42% knew the best time to get help was before age 2. Health care professionals who were familiar with the campaign (43%) said they had more resources for parents of children to monitor their child's development. These health care professionals were also less likely to take a "wait-and-see" approach, yet 19% still said they would advise parents to wait (Daniel et al., 2009).

"First Signs" Red Flags

All First Signs' information and screening tools are available from its website (www.firstsigns.org), but formal partnerships between First Signs and individual state health departments are the main emphasis of the program. The First Signs program is active in only eight U.S. states: Alabama, Delaware, Minnesota, New Jersey, Ohio, Pennsylvania, Texas, and Wisconsin. In these states, health care providers are the targeted audience, but cooperation with Early Intervention agencies is also a critical component of First Signs programs. Budgets for targeted campaigns are tailored to each state with a cost range of $50,000 to $500,000. Campaigns include screening kits for health care providers, educational mailings, parent brochures, public service announcements, and training (First Signs, Inc., 2012).

The "Red Flags" portion of the campaign focuses on dissemination of information about atypical or "absent" social behavior that may signal the presence of an ASD, such as lack of appropriate eye gaze, lack of warm, joyful expressions, lack of sharing interest or enjoyment, and lack of response to name as impairments in social interaction (see Appendix B). Communication red flags include lack of showing gestures and a lack of coordination of nonverbal communication. Another communication red flag is unusual prosody, including little variation in pitch, abnormal volume, odd intonation, irregular rhythm, and unusual voice quality. Red flags for repetitive behaviors and restricted interest at a very young age include repetitive movements with objects and repetitive movements or posturing of body, arms, hands, or fingers (First Signs, 2012). These red flags are the central focus of posters, media campaigns, and training in the First Signs programs.

"Learn the Signs" Autism Speaks Ad Council Campaign

A third approach aimed at a mass market audience has been the "Learn the Signs" national media campaign, consisting of print, billboard, and television ads. These ads feature celebrities such as Tommy Hilfiger, Jamie McMurray, Ernie Els, and Toni Braxton. In each ad, the odds of the celebrity's various career milestones are given as perspective for the much higher chances of having a child with autism. Each of the celebrities has a family member with autism who appears in the ad as well, showing that ASDs can happen to anyone. These odds are the only information given in the ads. The message at the end of each ad is to go to the Autism Speaks website to learn the signs of autism and make "a lifetime of difference" through early identification (Autism Speaks, 2012). The Ad Council reports that parents who viewed the ads are 160% more likely to be aware of the high prevalence of autism, 55% more likely to think of ASD as it relates to their own children, and 93% more likely to be able to identify at least one of the early signs of autism (Ad Council, 2011).

Limitations of Public Awareness Campaigns

Awareness campaigns are certainly an important element in the overall efforts to identify early signs of autism. They are targeted at specific populations who are most likely to make the early identification (i.e., health care providers, parents, early educators and Early Intervention agencies) or targeted at mass audiences for maximum penetration of the message. It is not known, however, if earlier diagnoses are occurring as a result of these campaigns. Research data from "Learn the Signs, Act Early" is qualitative data that indicate a degree of change in perceptions and attitudes, but it does not provide any information about changes in actual practice or about effects of the campaign on age of identification and diagnosis. The First Signs' Red Flags program does not have as wide an appeal by its nature as a targeted program. It is not adopted in all states, is tied to a financial partnership with a commercial enterprise, and has not yet provided evidence of effectiveness in the literature. Autism Speaks' Ad Council campaign research data is interesting, but details of the research methodology and effect on actions related to early identification are not currently available. These campaigns may have raised awareness of autism early signs, but they do not specifically address the question of whether a particular child meets criteria for risk of an autism spectrum disorder.

Early Identification of ASD Through Screening

The American Academy of Pediatrics has recommended universal screening for ASDs as part of routine health care for toddlers between 18 and 24 months. There is some debate whether ASD screening tools are yet feasible or reliable enough to be adopted as a universal screening practice, however. Feasibility and reliability of systematic screening approaches to identifying ASDs have been the focus of research in many countries. Effectiveness data regarding screening instruments, screening processes, or the prevalence of ASDs in a particular population are the outcomes of these research efforts. In countries with nationalized health care systems, screening is often tested in large populations as part of free health care visits either to a clinic or as part of home visits.

In the absence of a nationalized health care system, studies in the United States include lower participant numbers and are subject to some selection bias in that groups with an interest in ASDs (families, designated treatment agencies, or health care providers) typically make the effort to participate in screening. These studies have focused on either general developmental screening or autism screening using a variety of screening instruments. Methodology and results vary across these studies, and research is ongoing regarding the effectiveness of individual screening tools, but screening in general is shown to have some effectiveness in identification of ASDs that had not previously been identified.

Studies of Screening Tools and Methods

Checklist for Autism in Toddlers (CHAT)

The first autism screening study of a large population took place in England in 1992-1993 using the Checklist for Autism in Toddlers (CHAT; Baron-Cohen et al., 2000). It intention was to test the effectiveness of a screening instrument to detect emerging signs of the disorder and identify cases before clinically significant symptoms had been recognized by parents or professionals (Yirmiya & Charman, 2010). In this study, 16,235 children were screened as part of free well-baby checkups at 18 months of age, representing 39.8% of the target population. Although the positive predictive value from the CHAT was high at 83% (indicating 83% of those identified by the CHAT as being at-risk were found to have autism), and specificity was also high (99.9% -100% of children who did not have autism did not screen positive), the sensitivity was low, with only 38% of those ultimately diagnosed with ASDs correctly identified by the CHAT. In 2005, the United Kingdom National Screening Committee did not recommend routine developmental screening following the study's publication, citing the lack of a reliable, sensitive, and specific instrument for screening as the reason (Allison et al., 2008).

Quantitative Checklist for Autism in Toddlers (Q-CHAT)

A modified version of the CHAT, the Q-CHAT (Quantitative Checklist for Autism in Toddlers) was subsequently used to screen 754 toddlers (representing 32% of questionnaires mailed out to parents) in the United Kingdom. Results of the screening were preliminary in terms of validation of the Q-CHAT, and prevalence rates were not reported (Allison et al., 2008). Further study of the Q-CHAT items has been conducted in an effort to identify the most predictive items for a shortened screener. With just 10 items, a cut-point of three failed questions was found to have sensitivity of .91 and specificity of .89, with positive predictive value of .58. The study did not include an unselected population, however, so the Q-CHAT is not yet ready for use as a general population screener (Allison, Auyeung, & Baron-Cohen, 2012).

CHAT-23

Another version of the CHAT, the CHAT-23, was tested in Hong Kong, where there is a health care system similar to that in the United Kingdom, including frequent well child checks provided at maternal child health centers. Screening in this study was also not conducted in a population-based sample. Researchers administered screens for 212 children (whose diagnostic status had already been determined) in a two-step screening process involving 23 questions from the CHAT as the first step. If the first step was failed, the second step was a 5-minute in-person evaluation by trained clinicians, using four of the observational items from the original CHAT (i.e., eye contact, gaze monitoring, pretend play, and protodeclarative pointing). Sensitivity and specificity were calculated using three different cutoff scores. If 6 of the 23 parent questions were failed, sensitivity was .84 with specificity of .85. If 2 of the 4 observational items were failed, specificity was .74 with sensitivity of .91. The authors also identified 7 critical questions among the 23 based on discriminant function analysis. If 2 of these 7 were failed, sensitivity was .93 with specificity of .77. The authors of the study concluded the parent questionnaire was the first choice for screening due to its simplicity and ease of administration. (Wong et al., 2004).

Early Screening of Autistic Traits (ESAT)

In the Netherlands, a two-stage screening process was employed in the national free well-baby clinics using a system called Early Screening of Autistic Traits (ESAT: van Daalen et al, 2009). Clinicians were trained to administer an in person screening following failure of screening through parental questions. This study had a control group, with one region of the country screening all toddlers in the free clinics and another region with no universal screening. Earlier identification of ASDs was found in the region with screening. The mean age of diagnosis, however, was 63.5 months, with only 30% of those identified under the age of 36 months. Screening was found to benefit children with lower cognitive abilities, so children with ASD and higher cognitive abilities may have been missed (Oosterling et al., 2010).

Social Attention and Communication Study

In Victoria, Australia, a government-funded program of childhood health care provides periodic well child visits to Maternal and Child Health nurses from birth to age 6 (at no cost to families). These nurses were specifically trained for ASD in person screening and conducted screening at well child visits. Any children found with atypical behavioral symptoms were referred to a diagnostic center for full evaluation by age 24 months. Of the 22,168 children screened, 89 met criteria for ASD after a full diagnostic evaluation. This identification rate of 1:249 compared to a U.S. prevalence rate of 1:110 suggests that not all children with ASDs were identified through this screening process (Barbaro, Ridgeway, & Dissanayake, 2011).

Possible Effects of Culture on Screening

Another study in South Korea screened 7 – 12 year olds in the general population and also screened some children from a disability registry. Screening of the general population was done through schools. By this method (sending information out to parents through the schools) the participation rate was 63% of parents contacted (*n* = 23,234). Typically, epidemiologic studies that are least subject to bias have a participation rate approaching 80%, so it is possible that some selection bias affected the final results. There were 1,214 children with positive screening results, with 869 parents (72%) agreeing to a full evaluation. Of those that agreed, however, only 33% completed the evaluation. Within this smaller sample, the rate of ASDs found was 1.89%, or 1:53. In this group of identified children, 67% with ASD had previously been undiagnosed and untreated.

Cultural differences in recognition of symptoms in the study were controlled for by translation and back-translation of screeners with cultural sensitivity considerations. Even though cultural considerations were made for screening and diagnosis, the authors of the study emphasize the cultural stigma that exists in South Korea for ASDs as a factor in the low participation rate and in the high number of children who were previously undiagnosed and untreated (Kim et al., 2011).

Each of these studies was conducted outside of the United States under conditions of government-supported health care systems or schools. Conclusions from these studies are useful in understanding how effective screening can be in identifying ASDs in very young children. Application of these findings to the U.S. health care system, specifically in primary care settings, was not addressed. Participation rates (given as percentages of target populations) in some of these studies were low and prevalence rates were not consistent across studies or with prevalence studies such as the ADDM network research results. Generalization of findings, then, to community pediatric practice is limited.

ASD Screening in Pediatric Health Care Settings

Because of the variable onset of ASD, the subtle nature of early signs, and importance of earlier intervention, there has been a push for toddlers to be formally screened for ASDs. Pediatric primary care providers are the only professionals who interact with preschoolers and their parents on a continual basis. As such, pediatric health care visits are the most logical screening point for ASDs.

Screening tools are now widely available and have sensitivity that is that is on par with other universal screening tools. There are criticisms of universal screening programs for ASD and obstacles to implementing screening in pediatric practice. As a result, health care providers still rely on clinical judgment of ASD symptoms even though screening is available and recommended. There is some concern that a pediatrician's office may not be the ideal setting for a clinical evaluation of ASD symptoms. Children with or without ASDs may behave in an atypical manner in the environment of a medical examination room and busy pediatric clinic, which is an argument for using parent-report screeners to inform referral decisions (Robins, & Dumont-Mathieu, 2006).

Parent Report ASD Screening Questionnaires

One autism screening tool was specifically designed to be administered in a pediatric health care setting. The Modified Checklist for Autism in Toddlers, or M-CHAT (Robins, Fein, Barton, & Green, 2001), is a shortened version of the CHAT with only 23 yes/no questions for a caregiver. Intended for the age range of 16-30 months, it has reported sensitivity rates of .85 and specificity of .93 for ASDs (Dumont-Mathieu & Fein, 2005; Robins et al., 2001).

The M-CHAT has been studied specifically within a pediatric primary care setting by Robins (2008). This study screened 4,797 children through pediatric primary care providers in Atlanta, Georgia. This was not a population-based study, so no prevalence rates were reported, but the positive predictive value of the M-CHAT was reported to be .57, and it was noted that only 4 of the 21 children identified through screening had previously been flagged for developmental concerns by their pediatrician. The prevalence rate within this sample, then, was 1:228, less than the ADDM prevalence rates of 1:110 (CDC, 2009) to 1:88 (CDC, 2012), suggesting that all children with ASD may not have been identified by the implementation of M-CHAT screening in this study or that some form of selection bias in the study may have resulted in a reduced number of children found to have early signs of ASD.

For children younger than 16 months, the Communication and Symbolic Behavior Scales Developmental Profile Infant Toddler Checklist (ITC: Wetherby & Prizant, 2002) has been studied as a screening tool for autism spectrum disorders. Although not originally designed to be an autism screener, the ITC has reported sensitivity to ASDs of .94 and specificity of

.88. The authors of the study did not determine a prevalence rate as their aim was primarily to test the properties of the ITC as a screening tool for ASDs (Wetherby et al., 2004).

American Academy of Pediatrics Recommends Repeated Screening

In recognition of the importance of early identification of early signs of ASD and the pivotal role of a pediatric health care provider in identifying early signs and making referrals, in 2007 the American Academy of Pediatrics (AAP) published guidelines for screening all children at ages 18 and 24 months using an autism-specific screening questionnaire. A previous technical report published in 2001 was updated and replaced by two clinical reports detailing best practice in evaluation, identification, and management of ASDs. The 2007 report on evaluation and identification includes an algorithm for identifying risk factors in a child of any age (Johnson et al., 2007).

The algorithm describes a practice of scoring 1 point for any of the following risk factors during a preventative care visit or an extra visit for autism-related concern: (1) sibling with ASD; (2) parental concern; (3) other caregiver concern; and/or (4) pediatrician concern. If the child's total of risk factors is 0, the recommendation is that the child be screened for ASDs at the 18- and 24-month preventative care visit or the next visit. If the total of risk factors is 1, the child should be screened for social and communicative skills if under 18 months old, or screened with an autism-specific screening tool if 18 months or older. (Examples of screeners given in the article were the M-CHAT and ITC.) If the results of this screening are negative, the recommendation is to provide parental education pamphlets published by the AAP and schedule an extra follow-up visit within one month (Gupta et al., 2007; Johnson et al., 2007).

If the screens are positive, or if the child's total of risk factors is 2 or more, parental education is provided and referrals for comprehensive ASD evaluation, early intervention/early childhood education services, and audiologic evaluation are made. A follow-up visit is also recommended.

Adoption of Screening for ASDs in Pediatric Practice

Despite the AAP recommendations for screening, not all providers endorse the need for screening (Al-Qabandi, Gorter, & Rosenbaum, 2011). Although there is a growing research literature base on ASD screening tools and methods (Johnson et al., 2007), the actual decision making process of primary health care providers with regard to referral for an autism evaluation is not yet known. It is thought to be based on knowledge, prior training, and experience with ASD, which implies that many individuals who warrant a referral may not receive one due to lack of knowledge or experience on the part of the health care provider (Allison, Auyeung, & Baron-Cohen, 2012).

The most recently published survey of pediatricians reported that only 8% of primary care pediatricians routinely screened for ASDs (Dosreis, Weiner, Johnson, & Newschaffer, 2006). Another study (unpublished) reported rates of using standardized screening measures for autism specifically to be as high as 42% of responding pediatricians, with at least 28% still relying solely on clinical observations for referral (Zeiger, 2009).

Developmental Screening in Pediatric Practice

More research has been done on general developmental screening (as opposed to ASD-specific screening) using standardized screening tools. A 2002 AAP survey of its members regarding the use of general developmental screening tools and referrals to Early Intervention found that only 23% of pediatricians consistently used effective standardized screening instruments to assess their patients for developmental problems (Radecki et al., 2011). Other surveys report that 70% of pediatricians identify potential developmental problems via clinical assessment without the use of a screening instrument or checklist (AAP, 2003).

These data have been updated with a new survey of Pediatric Fellows. This 2009 survey indicates that 47% of pediatricians report that they use at least one standardized developmental screening tool routinely in clinical practice (Radecki et al., 2011). While these new data reflect improvement, over half of pediatric care providers still do not use standardized developmental screening instruments and rely instead on clinical judgment alone. Clinical observations without structured screening instruments may not be accurate in the case of ASDs, however (Johnson et al., 2007; Robins & Dumont-Mathieu, 2006).

In North Carolina, Pinto-Martin et al. (2005) launched a large initiative to improve the rate of general developmental screening in pediatric practice. Within the framework of a quality improvement initiative, leadership from key agencies formed a state advisory group to implement the project. Researchers focused on aspects of office process in order to increase the screening rates. These efforts resulted in an average screening rate of 75%, which includes some pediatric practices with higher rates. The authors of the study reported the necessity of identifying a physician champion within each practice to promote the project. Also mentioned as critical to improvement of screening rates were mapping office workflow, conducting staff orientations, and identifying existing system supports for children who screen positive for ASDs or developmental delays.

Obstacles to Screening in Primary Care

Surveys of primary pediatric care providers also reveal that the biggest obstacle to adoption and use of formalized screening for developmental delays is lack of time (82% of respondents). Lack of staff to implement screening (48%), and inadequate reimbursement for screening (44%) were also identified as primary obstacles to screening (AAP, 2003).

Criticisms of Universal Screening for ASDs

Although progress has been made in research to establish the effectiveness of universal screening for ASDs, and the AAP recommendation (Johnson et al., 2007) still stands, criticisms of universal screening efforts are also found in the literature. Criticisms include the lack of a single effective treatment for ASDs, (contraindicating a universal screen), insufficiently high sensitivity and specificity of screeners, adverse psychological impact of false positive ASD screens, the incapacity of the U.S. healthcare system to implement universal screening, and the unknown compliance of parents with treatment recommendations as reasons to not recommend universal screening for ASDs (Al-Qabandi, Gorter, and Rosenbaum, 2011).

Criticisms also are found from researchers outside the U.S., where nationalized health care systems can facilitate universal screening. These criticisms include the cost of screening, delays in intervention after false-negative screens, and the potentially costly increase in demand for diagnostic and intervention services (Allison, Auyeung & Baron-Cohen, 2012). Yirmiya and Charman (2010) urge caution for recommendation of universal screening when previous studies have not included systematic follow up to determine if the screening was universally specific and sensitive. While these criticisms adequately enumerate the costs of screening, they do not balance or weigh these costs with the benefits of screening in terms of earlier identification and intervention.

To answer one of the criticisms of autism screening as not yet sensitive, specific, or predictive enough to warrant universal screening, an examination of other universal screening measures shows ASD screening to be comparable in sensitivity at least. Using data from the two screeners previously mentioned as representative of ASD screening tools, sensitivity of the M-CHAT is .85 and specificity is .93, with positive predictive value of .57. The ITC has sensitivity to ASDs of .94 and specificity of .88. By comparison, universal newborn hearing screening programs have sensitivity of .80-.98 and specificity of .90, with positive predictive values calculated between .067 and .43 (Keren, Helfand, Homer, McPhillips, & Lieu, 2002).

Examination of other existing universal screening programs shows sensitivity and specificity to be at levels similar to ASD screening. Lead screening sensitivity is .88, with specificity of .94 (Binns, LeBailly, Fingar, & Saunders, 1999). Sensitivity of cystic fibrosis screening is reported to be between .86 and .99 with positive predictive values between .082 and .098 (Comeau et al., 2004; Grosse et al., 2004). Congenital hypothyroidism is also universally screened in the United States with a sensitivity of .70 to >.90 and specificity of .99 (AAP, Rose, American Thyroid Assoc., Brown, Lawson Wilkins Pediatric Endocrine Society, 2006). These figures suggest that the sensitivity of some autism screeners is on par with sensitivity of other, already established, screening tools for other conditions. Specificity of the two ASD screeners in this comparison is slightly below that of the biological specimen screens, but higher than that of newborn hearing screening. It should be noted that direct comparisons of positive predictive value figures is problematic due to the inclusion of prevalence rates in the calculation. Since prevalence rates vary by disease and also by sample, a highly effective screening tool may have a low positive predictive value due to low population prevalence, and a high positive predictive value may be obtained from a high risk sample, but will not generalize to a population-based sample (Allison, Auyeung, & Baron-Cohen, 2012).

How Does Clinical Judgment Compare to Screening?

Al-Qabandi et al. (2011) called for research to show that screening was more effective than surveillance (ongoing and systematic collection of data relevant to the identification of a disorder over time by an integrated health system: Baird et al. 2001). Surveillance in this context can also be thought of as clinically monitoring for signs of disease (ASD). As previously stated, screening studies have found that screening can identify early symptoms of ASDs in children who had not yet been identified through clinical judgment, but the reasons why the children had not yet been identified through clinical judgment were not explored (Kim et al., 2011; Miller et al., 2011; Oosterling et al., 2010; Robins, 2008). A direct comparison of screening to clinical judgment for each child identified was also not included in any of the analyses. The reasons why clinical judgment may not be as effective as screening have been explored in general, but not specifically studied in direct comparison to screening thus far.

Potential Problems with Clinical Judgment Alone

Several possible avenues of exploration to answer the question of how clinical judgment compares to screening have been identified in the literature. Difficulties associated with differential diagnosis of co-morbid conditions, sources of unintentional bias, differences in perception, varying levels of experience and expertise, and the possibility of something inherent in the symptomatic presentation of ASDs in a brief evaluation are all possible reasons for the disparity between clinical judgment and screening effectiveness in identifying children who need referrals for ASD evaluations.

Co-morbid Conditions Complicate Differential Diagnosis

Some symptoms of autism can appear similar to some symptoms of severe and profound general developmental delay, language disorder, and hyperkinetic and attentional disorders. Differential diagnosis among these disorders is difficult in very young children, but it is also possible that an autism spectrum disorder is co-morbid with hyperkinetic and attentional disorders, obsessive compulsive disorder, and developmental disorders of motor function, as well as specific and general learning problems. Differential diagnosis of autism spectrum disorders in young children with severe and profound developmental disability and in children with superior intelligence can be particularly difficult (Baird et al., 2001). If co-morbid disorders are present, ASD symptoms may be difficult to separate from other symptoms and may be missed or dismissed as part of the other presentation.

Perceptions May Be Influenced by Other Impairments

Early identification of an ASD involves detection of differences in social interaction and communication that can sometimes be very subtle. Detection and correct interpretation of symptoms can be affected by a variety of factors, including being overshadowed by more common developmental problems that may also be present.

For example, impairments in hearing or vision, or severe developmental delays other than ASD may complicate the screening and evaluation process. These conditions are more readily apparent to health care providers and parents, and are serious concerns that may overshadow concerns about social and communication abilities. Standardized screening and evaluation instruments have shown limited specificity in the presence of these other impairments (Zwaigenbaum et al., 2009).

Another symptom that may overshadow the perception of social and overall communication difficulties is speech delay. The most common ASD symptom parents report as their first concern is speech delay (Chawarska et al., 2007). Given this, and the diagnostic criteria of speech delay in Autistic Disorder, it is also important to note that social skills deficits are far more specific to ASDs than speech delay. When a concern for speech delay is first discussed between a health care provider and parents, focus on the speech delay might be an obstacle to further exploration of social deficits, possibly due to assumptions of deficits being caused by the speech delay itself. Speech can develop at irregular rates, so a health care provider may subscribe to a wait-and-see approach for speech development without considering the possibility of ASD. If this occurs, it may delay a referral for ASD screening. Also, base rates for some of these other disorders are higher than ASD, so it is understandable that health care providers would take the approach of focusing on ruling out more common conditions first.

Bias in Clinical Judgment Versus Screening May Exist

Bias in evaluation of children from minority groups has also been shown to be a factor in clinical judgment of the need for referral for ASD. Physicians in one study evaluated written vignettes with details of majority and minority demographic status. The physicians recommended ASD evaluations for majority status children at twice the rate of minority children. When given standardized questions to guide the assessment rather than clinical judgment based on the vignettes alone, however, majority and minority children were referred at rates that were closer to equal (Begeer, El Bouk, Boussaid, Terwogt, & Koot, 2009).

An older study concluded that although no bias was found in ASD diagnosis according to race or ethnicity, there was a significant difference in ASD diagnoses between diagnostic vignettes including higher SES components (higher rate of ASD diagnoses) and those with lower SES components (lower rates of ASD diagnoses). In discussing their results, the authors theorized that clinicians may attribute behavioral symptoms to a variety of factors associated with low SES, but if a child is from an "advantaged" background, there were few explanations for his behavior other than ASD (Cuccaro et al., 1996).

Experience and Expertise May Be Lacking

Given the complexity of symptomatic presentation and heterogeneity of symptoms of ASD, extensive experience and expertise in ASD symptoms is required for effective early identification by clinical judgment alone. Health care providers may not see enough patients with ASD in their individual practices to gain sufficient experience in early ASD symptoms. Continued professional education has been suggested by Mandell et al. (2009) as one solution to this problem. A pilot training program to provide expertise to pediatricians resulted in some success in accurate diagnostic referrals based on a one-hour evaluation, but had the side effect of overdiagnosis under a forced choice paradigm (Warren, Stone & Humberd, 2009).

Brief Clinical Evaluations May Not Be Accurate

Anecdotally, parents report frustration with brief clinical evaluations where a diagnosis is given or ruled out after a few minutes. For example, parents may report that a physician determined their child could not have an

ASD because he “made eye contact” with the physician, or spoke to the examiner. Likewise, parents may report that when an important diagnosis like an ASD is made after a brief interaction, they do not feel like the full range of strengths and weaknesses of their child have been taken into consideration. However, it is not currently known whether diagnoses made or ruled out on the basis of brief observations are actually less accurate than results from a more comprehensive assessment. Because this and other aspects of differences between clinical judgment and screening discussed above are not well understood, this gap in the research literature became the focus of this dissertation.

Development of Research Questions

Is There a Novelty Effect in Brief Evaluations?

In a previous study screening toddlers for ASDs (Miller et al., 2011), the study team (Judith Miller [JM], Terisa Gabrielsen [TG], and Michele Villalobos) observed differences in the presentation patterns of symptoms in toddlers with and without ASD. For many toddlers ultimately judged to have significant early signs of ASD, diagnostic impression at the beginning of the evaluation was that the child was typically developing. This impression changed as the evaluation continued and the child’s behavioral presentation changed. One symptomatic presentation trajectory in children with early signs of ASD was observed as children engaged in apparent social behavior (eye contact, response to name, etc.) at the beginning of the evaluation, but then showed distinctive features of ASD (lack of social responding, lack of initiation) as more time elapsed. An opposite trajectory was often observed in children with suspected speech delay. Often these children refused to engage with the examiner in the first few minutes of the evaluation, but eventually warmed up and became increasingly social and engaged with the examiner and toys as the evaluation continued.

From these observations, the key variable seemed to be elapsed time. This suggested the possibility of a novelty effect in which a child with autism was interested in a new environment and people temporarily due to novelty, but then reverted to a more typical symptomatic presentation as he became more comfortable or familiar with the environment and people.

Research on Novel Stimuli and Autism

No research has yet been found to specifically address the question of a possible novelty effect related to a pediatric health care visit (or any health care visit). Some literature does exist on novelty and people with ASD, however. It has been established that individuals with ASD react differently to novel stimuli than individuals without ASD. Each of these studies suggests how a child will respond to a variety of novel stimuli. The majority of studies address the issue of attention and how attention to novel stimuli is accelerated or delayed compared to typically developing control (TDC) groups. One study looked at how behavior changes in the presence of a novel therapist.

Attention to Novel Stimuli

One study of the differences in how people with ASDs respond to novel stimuli suggests that there may be a delay in paying attention to a novel stimulus. This study looked at physiological differences that may be either causes or effects of having an ASD. Response to novelty was studied by means of recording event-related potentials (ERPs) occurring in the brain when novel stimuli were presented under conditions such as the oddball paradigm (asking the person to respond to a target symbol that is presented among distractor symbols, some of which are frequently presented, others which are seldom presented, or novel). In a study with older children and young adults, those with ASD showed a prolonged latency to novel distractors, especially in the right hemisphere. The authors concluded that people with ASD may over-process or spend a longer time than necessary on novel stimuli. These differences were considered by the authors to be reliable enough to suggest that ERP evaluations in a novelty task may be useful as an outcome measurement in biobehavioral treatment of ASDs (Sokhadze et al., 2009).

Other aspects of attention to novel stimuli have been studied. Baron-Cohen and colleagues (Gomot, Belmonte, Bullmore, Bernard, & Baron-Cohen, 2008) found a faster reaction time for novel auditory stimuli in children, but did not examine how long attention was fixed on the novel stimuli. A study with adults discovered that individuals with ASD have a better ability to detect novel stimuli (based on unique features) amongst similar stimuli by identifying the novel stimuli faster than other groups (Plaisted, O'Riordan, and Baron-Cohen, 1998). In contrast, Landry and Bryson (2004) found that young children with ASD are impaired in their ability to disengage attention and were slower to shift their attention away to a new stimulus. Others showed impaired attentional prioritization to novel stimuli in favor of recognition of features of familiar stimuli (Keehn & Joseph, 2008). These findings present several possibilities for attention to novelty that may influence what a child's behavior, particularly their gaze, may look like in a novel situation.

Social and Nonsocial Novel Stimuli

Yet other studies have examined types of stimuli, showing that children with ASD will orient more readily to a novel non-social stimulus such as the sound of a toy than to the social stimulus of a smile or calling of the child's name (Dawson, Meltzoff, Osteling, Rinaldi, & Brown, 1998; Dawson et al., 2004). If the two types of stimuli occur together (e.g., introduction of a novel object such as medical exam equipment occurring at the same time as social interaction), it is conceivable that a response to a nonsocial stimulus may be misinterpreted as a response to a social stimulus. This, along with the different reactions to novel stimuli in laboratory conditions, presents a complicated range of possible reactions to the novel situation of a medical appointment or evaluation for ASD. It is possible that attention to novel stimuli within the first few minutes of an evaluation may disguise the expected presentation of ASD symptoms (i.e., lack of social response, lack of eye contact, lack of joint attention) in a very young child.

Effects of Novel Therapists on Behavior

In contrast to the research above, some research suggests that children with autism should show increased levels of atypical behavior in the presence of novel individuals. A hallmark of effective behavioral intervention is to systematically introduce new therapists to children with ASD, as it is well known clinically that children with ASD do not automatically generalize skills learned with one therapist to other therapists. Runco, Charlop and Schreibman (1986) demonstrated this and found that if a novel therapist was introduced to a child with an ASD, the child's behavior and level or previous gain in skills was temporarily disrupted. The focus of this study was on conditions that elicited increases in self-stimulation behaviors, and the presence of an unfamiliar therapist was one of the conditions found to significantly increase the amount of self-stimulatory behaviors. In addition, because this study focused on the child's ability to demonstrate learned behavioral skills, rather than interaction in an unstructured interaction, it is not known whether this same increase in atypical behavior would be expected in an encounter such as a medical exam.

Lack of Existing Research on Brief Evaluations

To date, no study has examined the brief duration of a health care appointment as a variable in the accuracy of ASD diagnostic judgment and appropriate referral. Visits to a pediatric clinic may involve as few as 10 minutes with the pediatric care provider. A survey of parents found 33.6% spent less than 10 minutes with a clinician, 47.1% spent 11 – 20 minutes, and only 20.3% spent more than 20 minutes with their care provider. In this first category (visits shorter than 10 minutes) there were disproportionate representations of uninsured (43.7% vs. the 33.6% occurrence of 1- to10-minute visits in the sample). Across the four categories of visit duration, female clinicians were represented at rates of 22% in 1-10 minutes, 54% between 11 and 20 minutes, and 20% over 20 minutes. Male clinicians were more heavily represented in the shortest visit interval, with 38% in the 1-10 group, 43% between 11 and 20 minutes, and 18.7% over 20 minutes (Halfon, Stevens, Larson, & Olson, 2011). In this short time frame, the clinician, activities, and environment are novel for children who have not had previous significant health care difficulties. It is not yet known what types of behaviors might be displayed during this brief window, or whether those behaviors are representative of the child's repertoire. If the behavioral sample is not representative of a child's typical behavior, health care providers may be at a disadvantage in making accurate referral decisions.

If research can establish the extent to which symptoms of ASD present within a brief evaluation, health care providers will be better informed when making these decisions about referral for ASD evaluations. Dissemination of research describing these effects may lower the age of ASD diagnosis by encouraging health care providers to consider screening instruments or expression of parent concern as sufficient evidence for referral to supplement clinical judgment for decision making.

Background Summary

Recent increases in the prevalence rate of autism spectrum disorders is a cause of concern for health care, education, and public policy systems. In response to the increasing identification of children with ASD, increased emphasis has been placed on early intervention and early identification of signs of ASD.

Early Intervention programs are entitlements that exist throughout the United States. The specific components of these programs may vary, as may the time spent in direct intervention. Programs that are comprehensive and are systematically based on evidence based treatment techniques are effective in promoting development of skills. If children with disabilities such as ASDs are identified at an early age through Child Find activities, through systematic screening and/or surveillance, Early Intervention programs have the potential to greatly reduce the lifetime cost of ASDs in individuals. Costs of Early Intervention programs are far outweighed by savings to the education system in terms of reduced need for specialized services during school age years of a child's life.

Early identification of ASDs is critical for maximum effectiveness of treatment through early intervention. Parents may identify early signs of autism at a younger age than the average age of professional diagnosis, but these concerns are not always communicated or acted on in a timely manner. Earliest signs of ASDs are subtle and may develop gradually or skills may slowly disappear in a regression. Although ideally autism is diagnosed at age 2, and the average age of diagnosis in some populations is as young as 3 years old, the average age of diagnosis in general in the United States is reported to be as late as 4, 5, and almost 6 years old. Public awareness campaigns have been launched to hopefully lower the age of symptom recognition among parents, health care providers, early childhood educators and childcare providers, but it is not known if these campaigns have changed parent or professional behaviors related to early identification.

Identification of ASD in very young children is problematic in that several issues impact the accuracy of diagnostic judgments. Regression of skill levels, co-morbid impairments and delays, variability in development of very young children, diagnostic perceptions and bias, lack of experience and expertise with ASDs, and time constraints on health care visits can all have adverse effects on identification of children at risk for ASD by a pediatric health care provider if clinical judgment is the only data used to determine the need for ASD evaluation referral.

To identify children with ASDs earlier, screening with standardized screening tools for ASDs has been recommended by the AAP in addition to surveillance for all children at well child visits to pediatric health care providers. Uptake of screening for general developmental concerns is still less than half of pediatric health care providers, however. Uptake of screening for ASD is even less. Criticisms of universal screening programs include the impact of false positive and false negative results, concerns about sensitivity, specificity, and positive predictive value of screeners, and the prediction that the American health care and Early Intervention systems are not prepared for a possible deluge of children identified as being at risk for ASDs. These criticisms have not been substantiated by research and have not been weighed against the benefits of universal screening, however.

As pediatric health care providers are the professionals with the best access to the preschool population, they are the *de facto* gatekeepers of ASD referrals. It is important to examine diagnostic impression issues in the framework of pediatric health care visits. Based on observations of a community-based sample of toddlers, it is possible that a novelty effect or differences in behavior during the short duration of most pediatric health care visits could affect accuracy of clinical judgment. Under novel conditions, a brief evaluation may result in social behavior that appears to be typical in initially, which may further complicate diagnostic impressions and prevent early referrals for ASD evaluation and treatment. Although some research exists on differences in reaction to novelty for children with an ASD, no existing research has examined what differences in symptomatic presentation may exist under novel conditions and a brief evaluation. For the majority of pediatric health care providers who rely on clinical judgment alone as an informal screen for symptoms of ASD, this is an area of research with important implications that has not yet been explored.

Purpose of This Study

This study had the aim of investigating aspects of symptomatic presentation of toddlers with ASDs under conditions of a brief evaluation and novel setting and evaluators. This examination was meant to determine if any yet unidentified factors exist that may explain why clinical judgment in these brief circumstances is less effective than standardized screening in identifying early signs of ASD. Based on clinical observations, it is possible that children with ASDs exhibit more typical behavior at the beginning of a novel encounter than they might show at later time points. Description of symptomatic presentation of early ASDs as might be seen in a brief (10-minute) exam in a pediatric health care provider's clinic has not yet been found in the literature. This information has importance in the evidence base for best practices in pediatric medicine, but also has implications for the educational system and psychology professions as early identification of

ASDs and subsequent early intervention improves the outcomes and ability levels of children with ASDs preparatory to their entrance into the school system and their life course.

Research Questions

The following research questions were examined for this dissertation:

(1) Do children with early ASD, language delays, or typical development show different rates of typical and atypical behavior during the first 10 minutes of a novel situation compared to 30-40 minutes?

(2) If differences exist, are they across all types of behaviors or specific to a select category of behaviors?

(3) Is behavior in the earlier or later time interval more consistent with a child's subsequent diagnosis?

(4) Is the impression of need for further referral accurate after a 10-minute observation? 30-40 minutes into an observation?

METHODS

Participants

Toddlers Showing Significant Early Signs of ASD

Granger Medical Pediatrics

Toddlers showing early signs of ASD were recruited through a three level screening process conducted in a large, independent, pediatric practice in West Valley City, Utah (Granger Medical Pediatrics). This population is unique among screening studies in that the entire population in this group comes from a single pediatrics practice catchment area. All providers within the practice participated in the screening study, and an overall screening rate of 80% was achieved, so possible sources of bias such as provider enthusiasm for ASD screening or self-selection of participants by provider were minimized. All parents of children born in 2006 were offered participation in the study regardless of their previous knowledge of or interest in ASD.

Level 1 (universal) screening consisted of two parent-report screeners (The Modified Checklist for Autism in Toddlers, [M-CHAT: Robins, Fein, Barton, & Green, 2001] and Infant Toddler Checklist [ITC: Wetherby & Prizant, 2002]). Children whose screening results indicated concern were advanced to Level 2 (at risk) screening by telephone interview to review failed items with parents in more detail. If concern remained after Level 2 screening, participants were invited to the University of Utah for a 2-hour, in person screening for ASD, which was the Level 3 (indicated) screen. If results of the Level 3 screen were positive for ASD, families were given referrals for Early Intervention and evaluations and children were assigned to the ASD group.

Toddlers Showing Signs of Suspected Speech Delay

Granger Medical Pediatrics

Toddlers in the suspected speech delay group were recruited as part of the same screening process as toddlers with early signs of ASD from Granger Pediatrics. After Level 3 screening, parents were given feedback indicating the child was showing signs of possible speech delays, but not early signs of ASD. Referrals for Early Intervention and Speech and Language evaluations were given to families and children were assigned to the Suspected Speech Delay (SSD) group. Neither the ASD nor the SSD group participants were identified as such until the conclusion of the three-stage screening process. Thus, they are not part of a selected high risk population, but represent the range of symptom severity that may been seen in a typical pediatric practice within a single catchment area.

Toddlers with Typical Development Control Group

Granger Medical Pediatrics Catchment Area

As part of the screening conducted at Granger Medical, follow up phone calls were made to several parents who had not completed all of the screening questionnaires during their visit. These children had not specifically failed their Level 1 screening, but screening was incomplete. At the end of the follow up phone call, if Level 1 and/or Level 2 screens were both negative, some parents were invited to bring their children for a Level 3 screening as typically developing control subjects. Unfortunately, only two parents responded to this invitation. A third child from the original recruitment was evaluated due to positive screens at Levels 1 and 2, but was found to have no symptoms of ASD or speech delay at Level 3 evaluation. This child was assigned to the TDC group as well. At the commencement of the current project, the original participant pool members were older than the target age for screening, so additional recruitment of participants for the TDC group was necessary.

TDC recruitment was conducted by way of flyer advertising in a small geographical radius surrounding Granger Medical Pediatrics. Flyers were posted at day care centers, recreation centers, grocery stores, and libraries. Parents contacted research staff by phone and children were subsequently screened over the phone using the same procedures outlined for ASD and SSD participants (e.g., M-CHAT and ITC, Level 1 screens and Level 2 follow-up screens if necessary). Parents whose children screened negative were invited to bring their children in for a Level 3 evaluation to participate as controls in the study. Participants recruited as typically developing controls in this manner received a small cash compensation ($50) for completing the evaluation. Four children in the TDC recruited group had visited Granger Medical Clinic or were patients of Granger Medical Pediatrics. Thus, half of the TDC group (7 of 14) had at some point been involved with Granger Medical Clinic or Pediatrics. The remaining half of the group was recruited from the immediate vicinity of Granger Medical Pediatrics.

Two children recruited as TDCs were excluded after Level 3 evaluation due to concerns about language development. These children were referred to Early Intervention and the Utah State Health Department clinic for Children with Special Health Care Needs for further evaluation.

Inclusionary Criteria and Qualifying Measures

Initial Screening Measures

The Modified Checklist for Autism in Toddlers

The initial screening consisted of two parent questionnaires. The Modified Checklist for Autism in Toddlers (M-CHAT: Robins, Fein, Barton & Green, 2001) is an autism-specific questionnaire consisting of 23 questions that can be answered "Yes" or "No." Six of these questions are considered to be critical items. M-CHAT questionnaires that yielded an initial score of 3 or more total failed responses, or 2 or more failed critical responses were followed up with a phone interview to clarify responses. This follow up interview has been shown to reduce the number of false positive results on the M-CHAT (Robins et al., 2001). The M-CHAT is designed for children between the ages of 16 to 30 months, but was used on the entire screening sample (age range of 15 – 33 months). As previously stated, the M-CHAT has reported sensitivity rates of .85 and specificity of .93 for ASDs (Dumont-Mathieu & Fein, 2005; Robins et al., 2001) and positive predictive value of the M-CHAT reported to be .57 (Robins, 2008). The M-CHAT is copyrighted, but is offered for free use on the Internet by the authors. It is available from several autism early identification websites and by a direct link to the official M-CHAT site (http://www2.gsu.edu/~psydlr/DianaLRobins/Official_M-CHAT_Website.html). There is also an online version that includes the follow up interview and generates a printed report of the results for parents or professionals (www.m-chat.org). It is also available in 47 languages, including Spanish, so screening for this study (including the follow-up interview) was conducted in both English and Spanish. Spanish follow up interviews were conducted by a psychology graduate student who was fluent in Spanish and had lived in a Spanish speaking country for several years. Some follow up interviews were conducted by a native Spanish speaker who was also a licensed psychologist. Others were conducted by a life long bilingual Spanish/English speaker who was a post baccalaureate psychology technician working at the Utah Autism Research Project.

CSBS DP Infant Toddler Checklist

The second screening questionnaire used was the Communication and Symbolic Behavior Scales Developmental Profile Infant Toddler Checklist (ITC: Wetherby & Prizant, 2002). This questionnaire consists of 24 questions about the frequency of behaviors related to social communication, speech development, and symbolic communication skills. Parents choose from responses of "Not Yet," "Sometimes," or "Often" for most questions. Other questions have responses in a quantifiable range (i.e., number of words spoken, number of blocks stacked, etc.) The ITC is designed for children from 6 to 24 months of age. The ITC is not designed specifically for the identification of autism in toddlers, but it does identify deficits in domains that are considered key in ASD identification. The ITC authors reported sensitivity to ASDs as .94 and specificity as .88 (Wetherby et al., 2004).

The domains measured by the questionnaire (social communication, speech and symbolic communication) are the domains most likely to be identified as indicative of risk of autism spectrum disorder at the young age of the study population (Chawarska et al., 2007). The authors' manual for the ITC includes guidelines for determining risk of developmental delay which state that failure of any domain on the questionnaire other than speech indicates the need for ASD follow-up. (The authors' recommendation if speech domain is the only failed domain is to monitor and rescreen in 3 months.) Cutoff scores according to age (in one month increments from 6 to

24 months) are given by the authors based on scores that are at or below the 10th percentile of standardized norms (Wetherby & Prizant, 2002). The ITC is copyrighted, but the questionnaire and scoring information is offered free for use on the Internet from the authors and copyright holders (http://firstwords.fsu.edu/index.php/early-identification-of-communication-delays and www.brookespublishing.com).

Internal consistency reliability for the ITC is .95, test-retest reliability has an effect size of 0.01. Interrater reliability is in a range between .90 and .97. Mean scores in the standardization sample (*n* = 2188) show a pattern of score increases consistent with developmental progression. The effect size of differences between score for males and females is .15. The ITC also has strong predictive validity (correlations of .80 for 12-17 months of age and .88 for 18-24 months) with the in-depth caregiver questionnaire that is part of the Communication and Social Behavior Scales measurement system. This caregiver questionnaire was not used, but a follow up interview with parents based on the original questions was used in the study. Positive predictive value of 57.4% and negative predictive value of 91.6% are reported when the ITC is compared with 2-year-old follow-up with the Mullen Scales of Early Learning. The overreferral and underreferral rates in this comparison are 13.5% and 5.7%, respectively (Wetherby & Prizant, 2002). The ITC was offered in Spanish and English. Spanish follow up interviews were conducted by the Spanish speaking research staff described previously.

In-Person Screening Qualifying Measures

Autism Diagnostic Observation Schedule (ADOS)

The ADOS is intended to complement information obtained from developmental tests and a caregiver history in the diagnostic process. The ADOS has reported sensitivity and specificity of 94% for Module 1 (appropriate for children who are not yet using phrase speech) and sensitivity of 94% and specificity of 87% for Module 2 (appropriate for children with phrase speech; Gotham, Risi, Pickles, & Lord, 2007). The Autism Diagnostic Observation Schedule-Toddler Version (ADOS-T: Lord, Rutter, DiLavore, & Risi, in press) has been shown in preliminary studies to have a sensitivity and specificity of 95% in a population between ages 12 and 30 months who were nonverbal. In verbal children, the ADOS-T has yielded sensitivity of 93% and specificity of 95% (Luyster et al., 2009). Algorithm scores have acceptable internal consistency and excellent inter-rater and test–retest reliability. The ADOS scoring algorithm, using both the formal cutoff and the ranges of concern, has excellent diagnostic validity for ASD versus non-spectrum conditions (Luyster et al., 2009).

The ADOS-G Modules 1 and 2 were administered by researchers who reached research reliability in administration and scoring with Judith Miller, PhD, a licensed clinical psychologist who specializes in ASD diagnosis. Dr. Miller is, at this writing, a Senior Scientist and Training Director at the Center for Autism Research at the Children's Hospital of Philadelphia.

Permission to use the prepublication version of the ADOS-T for research was granted by the publisher. When a participant's dominant language was Spanish (according to parent report), administration of the ADOS modules by Ms. Villalobos was in Spanish with an additional interpreter in the room for the child and his parent. Dr. Miller supervised the ADOS administrations.

Additional Measures at In Person Screening

Vineland Adaptive Behavior Scales, Second Edition

Parents were interviewed during Level 2 screening by a graduate research assistant to complete the Vineland Adaptive Behavior Scales, Second Edition, Survey Interview Form (Vineland II: Sparrow, Cicchetti & Balla, 2005). The Vineland II is an adaptive measure used to ascertain the child's level of age-appropriate functioning. The Vineland II yields domain scores for Communication (receptive and expressive--written domain not applicable in this age range), Daily Living Skills (personal, domestic, community), Socialization (interpersonal relationships, play and leisure time, and coping skills), and Motor Skills (gross and fine). An optional Maladaptive Behavior Index was not used. The Vineland II also yields an overall adaptive behavior composite score.

Internal consistency reliabilities for the Vineland II at ages birth to 5 years show an average in the mid to low .80s. Test-retest reliability across domains ranges from .75 to .96, with Adaptive Behavior Composite reliabilities at .96 in the 0-2 year age range. Interrater reliability for ages birth through 6 years for the Adaptive Behavior Composite is .87, with domain reliabilities ranging from .48 to .75 (Sparrow et al., 2005).

Evidence of validity for the Vineland II is given in the form of a developmental progression of scores and mean scores that are very similar across education levels of the parent. Additional evidence is given for similar mean scores across ethnic groups. Factor loadings for the Vineland II individual domains correlate to the Adaptive Behavior Composite in a range from .90 to .95. Validity of the Vineland II in an autism population was done outside of the age range of the current study sample, but showed significant differences from the non-ASD sample (Sparrow et al., 2005). The Vineland II was administered in Spanish by graduate psychology students to parents whose dominant language was Spanish. In some of these administrations, the Spanish Version of the Vineland-II was used.

Mullen Scales of Early Learning (Mullen)

The Mullen Scales of Early Learning (Mullen, 1995) was administered to the children during Level 2 (in-person) screening to assist clinical judgment in the domains of expressive language, receptive language, gross motor development, fine motor development, and visual reception. The Mullen is designed for children ages 0 to 68 months of age. If the primary language of the child was Spanish, the Mullen was administered in Spanish either directly by one of the Spanish-speaking graduate students or with an interpreter for the child.

The Mullen results include report of an Early Learning Composite Standard Score as an estimate of overall cognitive development. This index score reflects subdomain scores for Visual Reception, Fine Motor, Receptive Language, and Expressive Language. The Gross Motor subdomain is reported separately and is not included in the Early Language Composite score.

Visual Reception measures a child's ability to recognize and match items and shapes (including some memory components) with increasing levels of detail. It also measures a child's visual problem solving abilities. Receptive Language includes a variety of tasks requiring recognition and response to verbal directions (i.e., point to body parts, identify colors, follow directions, demonstrate understanding of spatial orientation words, and general knowledge). Expressive Language tasks measure language use during the Mullen, including the ability to identify pictures, count, repeat prompts, answer questions, and use elements of grammar such as pronouns.

In the Mullen Gross Motor subtest, children are asked to walk up steps, kick a ball, run, walk on a line, and jump, among other gross motor tasks. Fine Motor tasks present a child with tasks of drawing, writing, cutting, and manipulating small objects with precision.

Internal consistency reliability for the Mullen has median values ranging from .75 to .83, with the internal reliability of the composite Mullen score having a median value of .91. Test-retest reliability ranges from .71 to

.96, and shows a high degree of stability over time in children 24 months old and younger. Inter-rater reliability ranges from .91 to .99. Evidence of validity includes a developmental progression of scores. Concurrent validity shows the Mullen to have higher correlations with other cognitive measures (.53-.59) than psychomotor measures (.21 to .52).

The Mullen Early Learning Composite score has a correlation of .70 with other mental development measures and .43 with psychomotor development measures, indicating its validity as a measure of global cognitive ability. The Mullen has correlations ranging from .65 to .82 with other measures of language ability in the relevant age range (Mullen, 1995).

Group Assignment

Group assignment was made according to Level 3 evaluation results. As stated in Zwaigenbaum et al. (2009), diagnosis of ASD in toddlers is based on clinical judgment, taking into account information from the child's developmental and medical history; observed and parent-reported ASD symptoms; and cognitive, language, and adaptive skill measures.

Clinical judgment regarding the presence of significant early signs of ASD was made by Judith Miller, PhD, a licensed psychologist specializing in ASD diagnosis. At the end of the 2-hour Level 3 evaluation, parents were given specific feedback on the child's development and any early signs of ASD or Suspected Speech Delay (SSD). Parents of Typically Developing Control participants were given feedback indicating no signs of ASD or SSD were observed. Because of the young age of the children, parents were not given a formal diagnosis after Level 3 evaluations, but were told whether their child was showing significant signs of ASD that warranted intervention and further evaluation.

Exclusionary Criteria

Early ASD and Suspected Speech Delay Groups

All children who showed early signs of ASD at the conclusion of Level 3 evaluations were eligible for inclusion in the study. Subsequent examination of the quality of video recording of Level 3 evaluations eliminated one child in the ASD group from inclusion due to an unacceptably high proportion of off-camera time, however. All children identified at Level 3 evaluations with SSD were eligible for inclusion in the study. Some SSD children were not included in the sample because they did not match participants in the other two groups in age or gender. Exclusionary criteria of hearing or vision loss was not required as none was reported for any of the participants in any of the groups.

Typically Developing Control Group

Participants in the TDC group were matched by age and gender to the other groups at the time of Level 3 screening, so children were excluded from participation if their age and gender was already sufficiently represented in the control group. Some families responded to the advertisement flyers and completed all three levels of screening to qualify as TDC, but were not included in the final sample. In order to preserve the unique character of the sample, children were excluded if they did not live or patronize businesses in the catchment area of Granger Medical Pediatrics.

Settings

Level 1 (universal) screening for participants in the early ASD and suspected speech delay groups took place at Granger Medical Pediatrics during February through September of 2008 as part of pediatric health care visits. Level 1 screening for three TDC participants also took place at Granger Medical Pediatrics. Level 1 screening for recruited TDC participants was by telephone at the University of Utah in the Spring of 2010. Level 2 screening was done exclusively by telephone by graduate students at the University. Level 3 (indicated) screening for all three groups took place at the University of Utah at the Utah Autism Research Project Spring, 2008 through Summer, 2010. Level 3 evaluations were conducted in a room furnished with table and chairs, a couch, and toy bin. A large one way mirror (with an observation room on the other side) constituted one wall of the room.

Procedures

Preliminary Procedures

Institutional Review Board Approval

The Institutional Review Board of the University of Utah approved the initial recruitment and screening phases of the study from 2007 through 11/10/2010. An amendment to the research protocol for recruitment of typically developing participants was approved 2/10/2010 (IRB_00015927—The EACH CHILD Study). Informed consent was obtained from all parents of participants for screening and evaluations.

Video Recording of Level 3 Evaluations

Video recordings of the early ASD and speech delay groups were obtained through the EACH CHILD Study (Miller et al., 2011) conducted with participants from Granger Medical Pediatrics. Video recordings of three TDC participants were also obtained from this study. Video recordings for the remaining TDC participants were obtained as they were recruited into the current study. Parents of all participants explicitly gave informed consent for use of the video recordings by The EACH CHILD Study. Parents were given the option of not giving consent for videotaping without any effect on participation. All parents gave consent for video to be used within the study.

Videos were prepared for coding by breaking up the target video segments (0 - 10 minutes and 30 - 40 minutes) into 10-second intervals interspersed with 4-second transition slides labeling the next interval with a time stamp. Separate DVDs for each 10-minute video were prepared. Videos were prepared using iMovie '09, Version 8.0.6 © 2007-2010, Apple, Inc. No identifying information was provided to the raters with the exception of age of the child. One identifying piece of information may have been presented to the raters in the form of the child's first name being called by the examiner on the video as part of the evaluation process. This identification was unavoidable due to the need to code the child's response to his name being called. Disclosure of first name, however, did not give raters any information about the child's diagnostic group.

Design

A simple experimental design between and within groups was used to structure the study. Repeated measures analysis was based on coding of behaviors from video recordings of toddlers in a novel, play-based situation (administration of an ADOS) that was consistent in format and setting. The purpose of the analysis was to determine if the frequencies of specific Typical behaviors (behavior that is consistent with typical development with the possible exception of speech delay) and/or Atypical behaviors (consistent with ASD) were observed at different rates as a function of elapsed time. Coding of behaviors according to coding criteria was done using 10-second partial interval recording across two observation sessions—a 10-minute interval at the beginning of the evaluation and again between the 30- and 40-minute marks.

Partial Interval Recording

Partial interval recording (PIR) was chosen as the coding method based on research showing it to be more sensitive to behavioral change (Harrop, Daniels & Foulkes, 1990; Harrop & Daniels, 1986), and recommendation of PIR over momentary time sampling when estimating frequencies (Meany-Daboul, Roscoe, Bourret, & Ahearn, 2007).

In partial interval recording, within each 10-second interval, the coder marked the presence of each type of behavior (Atypical behavior, Typical behavior, or Unclear behavior) during the interval. The absence of any codable behaviors was marked with an “X” or nonoccurrence code. Raters were not asked to make a forced choice of behavior within a single partial interval. It is possible that the rater marked both Atypical and Typical behaviors as present within the same partial interval, but in data collection the interval was scored as Atypical in that case to maximize sensitivity to Atypical behaviors.

Raters

Raters were licensed psychologists with extensive experience in ASD diagnosis who are reliable on ADOS administration and scoring. One rater, (Leslie Speer, PhD) is a licensed psychologist formerly with the University of

Utah Child, Adolescent and Adult Behavioral Health Clinic. Dr. Speer conducted evaluations for ASD in the Autism Diagnostic Clinic of the Child, Adolescent, and Adult Behavioral Health Clinic at the University of Utah under the supervision of Judith Miller, PhD. Dr. Speer also has clinical experience with The Children's Center (a mental health treatment facility for 2- to 6-year olds) and a local school district's special education preschool. She is currently conducting autism evaluations as a pediatric psychologist at the Center for Pediatric Behavioral Health Autism Center, part of Cleveland Clinic Children's Hospital.

The other rater, Megan Farley, PhD, conducts autism evaluations at the University of Utah Child, Adolescent and Adult Behavioral Health Clinic and the Neurobehavioral HOME Clinic and is an investigator in autism research at the Utah Autism Research Project. She is a licensed psychologist and Assistant Professor in the Department of Psychiatry at the University of Utah. Dr. Farley has been formally trained on the Bayley Scales of Infant Development ®, Third Edition (Bayley, 2005) and conducted infant and toddler developmental evaluations as part of another study at the University. Thus, both raters were familiar with ASDs in young children and typical and atypical development in toddlers and preschoolers.

Both raters were blind to the hypothesis of the study, blind to participant group membership, blind to what diagnostic groups were included, and to the proportion or probability of ASD within the sample.

They were also blind as to whether the video they were viewing was from Time 1 or Time 2. Raters were not aware of Time 1 and Time 2 intervals, but were told they may see the same child more than once. Neither coder had any previous contact with any of the study participants they viewed.

Dependent Variables

Atypical Behaviors

Information from the literature about behaviors generally considered to be early signs of autism was gathered primarily from the ADOS, the Systematic Observation of Red Flags (SORF: *Wetherby et al., 2004),* and the diagnostic criteria for Autistic Disorder, Asperger's Disorder, and Pervasive Developmental Delay, Not Otherwise Specified (PDD--NOS) as listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR: American Psychiatric Association, 2000). These three elements sufficiently represent the sum of the literature on very early signs of ASD (see Table 1).

The ADOS is considered to be the "gold standard" as an observational component of ASD diagnostic evaluations (Papinikolaou et al., 2009). The SORF has been shown to differentiate between ASD and other developmental delays in very young children (Wetherby et al., 2004). The DSM-IV-TR criteria are the basis of a clinical diagnosis of an ASD. Behaviors from these sources were collapsed into five broad categories: Responding, Response to Name, Sounds and Words, Play and Stereotypical Movement, and Initiating.

Table 1. Behaviors Indicative of ASD in Very Young Children

Behavior Categories for Coding	***SORF[1] Red Flags***	***ADOS Algorithm[2]***	***DSM-IV-TR[3]***
Social Initiating	Lack of appropriate gaze Lack of warm, joyful expressions with gaze Lack of coordination of gaze, facial expression, gesture, and sound Lack of pointing Lack of showing	Unusual Eye Contact Less Directed Facial Expressions, Lower Quality of social overtures Lower Amount of Social Overtures to parent/caregiver Less integration of gaze and other behaviors during social overture Less Requesting Less spontaneous initiation of joint attention Less showing, pointing Few Gestures	Marked impairment in eye to eye gaze, facial expression, body posture, and gestures to regulate social interaction Lack of showing, bringing, or pointing out objects of interest
Social Responding	Lack of sharing enjoyment or interest Lack of response to contextual cues	Less shared enjoyment in interaction Lower overall quality of rapport Less response to joint attention	Lack of social or emotional reciprocity
Unconventional Sounds and Words	Unusual prosody Lack of vocalizations with consonants	Odd Intonation of Vocalizations and Verbalizations Lower Frequency of vocalization directed at others Echolalia	Delay in development of or total lack of spoken language, not compensated for by using gestures Stereotyped and repetitive use of language, or idiosyncratic language
Response to Name	Lack of response to name when called	Less Response to name	

Table 1. Continued

Behavior Categories for Coding	*SORF[1] Red Flags*	*ADOS Algorithm[2]*	*DSM-IV-TR[3]*
Play with Objects and People, Stereotypy	Repetitive movements or posturing of body, arms, hands, or fingers Repetitive movements with objects Lack of playing with a variety of toys	Less imagination/creativity Less functional play with objects Unusual sensory interest in play material/person Hand and finger movements/posturing Unusually repetitive interests or stereotyped behavior Use of other's body to communicate	Encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal in either intensity or focus Persistent preoccupation with parts of objects Stereotyped and repetitive motor mannerisms Apparent inflexibility in adherence to specific nonfunctional routines or rituals Lack of varied, spontaneous make-believe, or social imitative play appropriate to developmental level

[1] *Systematic Observation of Red Flags, Wetherby et al, 2004*

[2] *Autism Diagnostic Observation Schedule, Modules 1, 2 and T,* Lord, Rutter, DiLavore, & Risi, 1999 and in press

[3] *Diagnostic Statistical Manual for Mental Disorders, Fourth Edition, Text Revision,* APA, 2000. Not included due to age range: Marked impairment in ability to initiate or sustain conversation with others, Failure to develop peer relationships appropriate to developmental level.

Broad categories were chosen for coding and reporting. This choice accomplished three goals: (1) it more closely resembled the level of distinctions likely to be made by general pediatricians; (2) it maximized the chances that behaviors would be coded, especially if the behavior observed was not anticipated and listed under a specific code; and (3) aggregated results would be more meaningful to interpret in broader terms rather than very small amounts of data across a large number of categories.

Typical Behavior

Within each of the coded behavior categories, behaviors that were not associated with the qualities seen in ASD were rated as Typical behavior. Whereas the SORF items are expressed primarily as "lack" of certain skills, the presence of those skills would be considered Typical behavior. For example, lack of pointing or showing when a child encounters something new and interesting would be coded as an Atypical behavior, while pointing and showing would be coded as a Typical behavior. Typical Play was coded as using or playing with toys according to their obvious function or an imaginary one. Atypical play would be repetitively banging, dropping, or carrying a toy with little variety in engagement or in choice of toys. Stereotypical movements were also coded in the category of Atypical Play. Some specific behaviors were identified in early previews of video recordings that were not easily codable according to Table 1. Thus, the following clarifications were provided to the raters.

For Responding:

- intentional gaze was a Typical Response,
- a blank stare was Atypical;
- walking away from a social bid was Atypical;
- fear of an activity was Atypical if it was an extreme reaction not matched to the situation.

The following additional definitions were developed for Sounds:

- Sound could be Atypical in volume and intonation;
- breathing and coughing were not codable sounds;
- the default classification of Sounds behavior is Typical unless clearly Atypical.

The following additional definitions were developed for Play:

- the third time an action is repeated, it became Atypical, even if the action in isolation is Typical (i.e., banging, throwing, wandering, holding a toy but not playing with it);
- eating appropriately was coded as Typical play; Typical play was dependent on the object (i.e., banging with a hammer was Typical, but banging with a bottle was

Atypical, shaking a jar was Typical, but shaking a pen was Atypical);

- mouthing a toy was Unclear play.

The following additional definitions were developed for Initiating:

- Intentionally rolling a car to the examiner was an Initiation;
- Initiations were actions that maintain an interaction, even if the interaction was originally initiated by the examiner;
- Initiating must be social—social gaze must clearly have the intent of looking at the person.

Unclear Behaviors

Because the evaluations of participants were play-based, and the children were free to move about the room most of the time, there were moments in the recordings where view was obscured, the child was temporarily off camera, or the child's face was not clearly visible. There were also some behaviors that could not be determined as clearly typical or clearly atypical. A code of "unclear" was used to accommodate these occurrences.

Nonoccurrences or No Opportunity

Within each behavioral category, an interval could be marked with an "X" code to indicate there was no occurrence of that behavior in that interval. For example, if the examiner was not socially engaged with the child (e.g., was talking with parents instead), the interval could be marked as having "No Opportunity" for Responding. If the child was not initiating, but there was nothing in the interval that may have been expected to have elicited an initiation (such as a desirable object), that interval was coded as a Nonoccurrence. In the Response to Name category, nonoccurrence codes were used for all intervals in which the Examiner did not call the child's name. In the Sounds and Words category, if the child made no sound during an interval, the nonoccurrence code was used.

Clinical Judgment of Need for Referral

After the final codes were recorded, raters were asked "Based on this observation alone, would you recommend this child for an ASD evaluation? Yes/No". No other instructions were given.

Independent Variable

Elapsed Time During Observation

The Time 1 video recording sample of behavior was the first 10 minutes of an evaluation, when the novelty effect, if present, was most likely to be in evidence. The Time 2 behavior sample was another 10-minute observation beginning after 30 minutes of video recording had elapsed. Measurement of the first 10-minute interval began as the child entered the room or the first image of the child in the room. The second time interval began 30 minutes after the child entered the room or was observed in the room. Very minor (1/10th of a second) clips were lost in transition to the 4-second interval transition between each coding interval. Over the course of a 10-minute video, this could add up to as much as 6 seconds cumulatively, but since the same technology was used to create all of the videos, this possible 6-second loss would have been consistent across all videos and should not have affected the overall amount of behavior coding. On several occasions, video recording was stopped so that parents could take the child out of the room for diaper changes. This elapsed time was not counted towards elapsed time on the video recording. Raters had the ability to rewind the video if necessary to watch intervals again, thus maximizing opportunities to observe any atypical behaviors that occurred in the video clips.

Establishing Reliability of Raters

Raters were trained to use the behavior coding system using videos of Level 3 autism evaluations administered to children from a pilot phase of the The EACH CHILD Study (IRB approval, informed parental consent and permission to use videos for training were obtained). These videos included participants from a different pediatric practice and an Early Intervention agency. Raters were required to establish 80% agreement or higher with each other according to exact agreement before coding videos of current study participants.

To ensure ongoing interrater reliability, 20% of videos were coded by both raters and coding sheets were compared by comparing intervals in each category. Videos were chosen for reliability at random within each diagnostic group, so although videos were chosen at random, the balance of ASD, SSD, and TDC children was equal between the two raters. Balance of Time 1 and Time 2 videos as also equalized in assigning videos for reliability. Raters were blind as to which of the videos were reliability videos.

Reliability was monitored throughout the study. Reliability videos were dispersed from beginning to end among the 51 videos each coder was assigned. If overall reliability by exact agreement between raters dropped substantially below 80%, retraining sessions were held until reliability on training videos was again at 80% or higher.

Assignment of Videos for Coding

Because raters were blind to the study hypothesis, bias regarding coding of behaviors at Time 1 versus Time 2 was minimized. Videos were initially assigned to raters randomly, but were then also counterbalanced so that each coder viewed the same proportion of Time 1 and Time 2 videos, as was done in assignment of reliability videos. Each coder also had the same proportion of videos from each of the diagnostic groups to view. Because videos were assigned randomly with the exceptions mentioned above, occasionally a coder would see both videos of a child, but videos were never identified as Time 1 or Time 2 to the raters. In most cases, a child was viewed by a coder only once. Estimates of symptomatic severity were not made or controlled for.

Coding of Video Samples

Raters were asked to view each 10-minute video clip. A coding tally sheet was provided with space to mark observed behavior as being either Atypical (A), Typical, (T), Unclear (U) or Nonoccuring (X). See Appendix C for coding sheet. Videos were structured with 10-second clips followed by 4-second breaks that indicate the time stamp for the next clip so that raters were always able to identify which clip they were viewing and could match it accurately to the same interval on the coding sheet (i.e., 0:10 for the 1st 10-second interval. 0:20 for the second, etc.). The video could be paused during the 4-second break if necessary. During reliability training, raters preferred to watch each video clip three times—once to code Responses and Response to Name, once to code Sounds and Play, and a final time to code Initiations. Raters reported that they maintained this pattern for coding (three passes) for study videos as well.

Data Collection and Analysis

Coding sheets were delivered, mailed or emailed to TG for entry into a spreadsheet for data collection using Microsoft ® Excel ® 2008 for Mac.

Codes were entered as “A,” “T,” “U,” or “X.” Data sheets were double entered and checked for accuracy by computerized comparison functions within the spreadsheet. Any errors found were verified against original coding sheets and corrected. Statistical analyses were completed using the Statistical Package for the Social Sciences (SPSS), 19th and 20th editions.

Preliminary Analyses

Demographic characteristics were analyzed for significant differences to determine if groups were well matched. Nominal variables such as race and ethnicity, sex, language and SES were compared using chi-square analysis. Ordinal variables such as age were examined using ANOVA. Results of independent measures (Mullen and Vineland) were analyzed using ANOVA and Tukey’s HSD to characterize group differences in cognitive abilities and adaptive behaviors.

Reliability between raters was calculated by exact agreement (number of intervals with agreement divided by the total of agreement and disagreement). Percentage of exact agreement was calculated in each behavior category as well as overall agreement. Positive agreement and negative agreement were also calculated. Kappa was used as a measure of reliability for comparison to exact agreement.

Choice of Analyses

Statistical tests of differences between groups in behavior coding was by repeated measures ANOVA for research questions involving Time 1 and Time 2. For questions without reference to Time 1 or Time 2, ANOVA was used to determine if group means were significantly different. Post-hoc analysis utilizing Tukey's HSD was used on all analyses where significant differences were found among the three groups. Analysis of correct and incorrect clinical judgments was by chi-square.

RESULTS

Preliminary Analyses

Participants

Participants included 42 children divided into the three groups by diagnostic category (ASD $n = 14$, SSD $n = 14$, TDC $n = 14$), and represented diversity of race, ethnicity, language, and socioeconomic status (SES). Five different racial groups and Hispanic ethnicity are represented in the sample. Demographic characteristics were not significantly different across groups. Groups did not significantly differ with regard to age, $F(2,39) = .45$, $p = .6$, sex, $\chi^2\ (2) = 2.10$, $p = .35$, or race or ethnicity, $\chi^2\ (10) = 8.00$, $p = .63$. Table 2 shows means and standard deviations for age and the composition of the groups in terms of sex, racial ethnicity, and SES. Type of insurance coverage was used as its own variable and as a possible proxy for SES. Three categories of insurance were used in the analysis: Private insurance, government subsidized insurance such as Medicaid or CHIP, and self-insurance. There were more children with private insurance in the ASD group than in the other two groups (12 in the ASD versus 9 each in the SSD

Table 2

Demographic Characteristics of Sample

Characteristic *n* (%)	ASD	SSD	TDC	Total
Gender				
Male	12 (86)	9 (64)	9 (64)	30 (71)
Female	2 (14)	5 (36)	5 (36)	12 (29)
Race/Ethnicity				
Caucasian, Non-Hispanic	8 (57)	8 (57)	8 (57)	24 (57)
African-American	1 (7)	--	--	1 (2)
Native American	--	--	1 (7)	1 (2)
Hispanic, no race specified	4 (29)	4 (29)	4 (29)	12 (29)
Pacific Islander	1 (7)	2 (14)	--	3 (7)
Asian	--	--	1 (7)	1 (2)
Spanish Speaking	3 (21)	--	3 (21)	6 (14)
Insurance				
Private Insurance	12 (86)	9 (64)	9 (64)	30 (71)
CHIP/Medicaid	2 (14)	5 (36)	2 (14)	9 (21)
Self-Insured	--	--	3 (21)	3 (7)
Mean age in months (*SD*)	22.7 (4.8)	23.0 (5.5)	24.5 (5.5)	23.4 (5.2)
Age range in months	16.3-33.0	16.0-33.2	15.4-32.5	15.4-33.2

ASD = Early Signs of Autism Group, SSD = Suspected Speech Delay Group, TDC = Typically Developing Control Group, CHIP = Children's Health Insurance Plan, *SD* = standard deviation and TDC groups). Chi-square analyses revealed a trend towards a significant difference between groups, $\chi^2(4) = 8.60$, $p = .07$.

All screening procedures, including evaluations, were available in both English and Spanish. Three participants in the ASD group and 3 in the TDC group received evaluations in Spanish. There were no children in the SSD group who required evaluations in Spanish. Differences on native language ability between groups, however, were not statistically significant, $\chi^2(2) = 3.5$, $p = .17$.

Cognitive and Adaptive Measures

Cognitive Development (Mullen)

Children in the TDC group obtained mean scores in the Average range (standard scores in the range from 85 to 115; *t*-scores in the range of 40 – 60) across all domains of the Mullen. As might be expected given the nature of developmental delays such as ASD and speech delay, children in the SSD and ASD groups obtained lower mean scores than children in the TDC group in all domains. See Table 3 for means and standard deviations of Mullen composite and subdomain scores. ANOVA revealed significant differences for the Mullen overall Early Learning Composite scores, $F(2,39) = 22.35$, $p<.001$. Significant differences also existed in the domains of Visual Reception, $F(2,39) = 8.700$, $p = .001$, Receptive Language, $F(2,39) = 27.11$, $p<.001$, and Expressive Language, $F(2,39) = 26.16$, $p<.001$. With regard to

Table 3

Cognitive Development

	Mean Scores (Standard Deviations)			Tukey's HSD Post-hoc analysis comparisons		
Mullen Scales of Early Learning	ASD	SSD	TDC	ASD and SSD	ASD and TDC	SSD and TDC
Early Learning Composite Standard Score	66.14 (12.61)	77.86 (16.94)	102.64 (14.38)	$p = .10$	$p < .001$	$p < .001$
Subdomain *t*-scores						
Visual Reception	30.00 (11.92)	39.21 (11.16)	48.71 (12.49)	$p = .11$	$p < .001$	$p = .10$
Receptive Language	26.07 (6.64)	39.43 (12.64)	53.93 (9.90)	$p = .01$	$p < .001$	$p < .001$
Expressive Language	28.64 (9.06)	32.79 (8.53)	53.07 (10.93)	$p = .49$	$p < .001$	$p < .001$
Gross Motor	41.69 (12.00)	51.21 (13.48)	58.57 (9.47)	$p = .04$	$p = .001$	$p = .32$
Fine Motor	35.64 (12.92)	43.71 (13.60)	45.50 (11.67)	$p = .23$	$p = .11$	$p = .93$

HSD = Honestly Significant Difference, ASD = Early Signs of Autism group, SSD = Suspected Speech Delay group, TDC = Typically Developing Control group motor development, Mullen Gross Motor scores were also different, *F*(2,39) = 8.01, *p* = .001, but Fine Motor scores did not meet the threshold for statistical significance, *F*(2,39) = 2.373, *p* = .11. Figures 1 and 2 represent the means and standard errors of Mullen scores by group.

Post-hoc analysis utilizing Tukey's HSD on the Mullen scores were conducted to determine differences between individual groups. On the Early Learning Composite scores, both ASD and SSD group means were significantly lower than the TDC group means, but not significantly different from each other. Significance data for paired comparisons of the Early Learning Composite means are also shown in Table 3.

Each domain of the Mullen was then analyzed using Tukey's HSD post hoc paired comparisons (see Table 3 for means and significance data). In the area of Receptive Language, the ASD group's average score was in the Very Low range, which was significantly lower than the SSD (Low Average range), which in turn was significantly lower than the TDC group (Average range). In the area of Expressive Language, the ASD and SSD groups both scored significantly lower than the TDC group, but not different from each other. In the area of Visual Reception, the ASD group scored significantly lower than the TDC group, with the SSD group scores falling between the means of ASD and TDC, and not significantly different from either group. Finally, in the area of Gross Motor development, the ASD group scores were significantly lower than both SSD and TDC mean scores. SSD and TDC

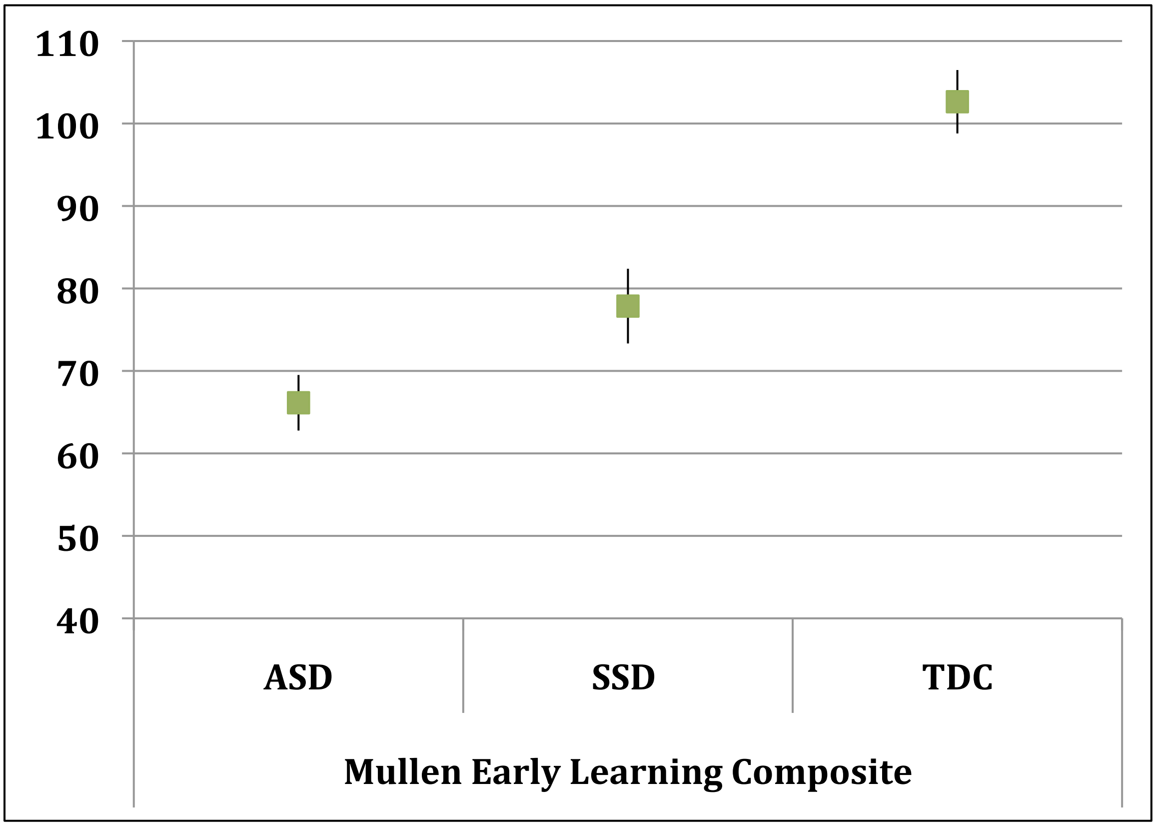

Figure 1. Mullen Early Learning Composite Mean Standard Scores with Standard Error.

ASD = Early signs of Autism Spectrum Disorders group, SSD = Suspected Speech Delay group, TDC = Typically Developing Control group.

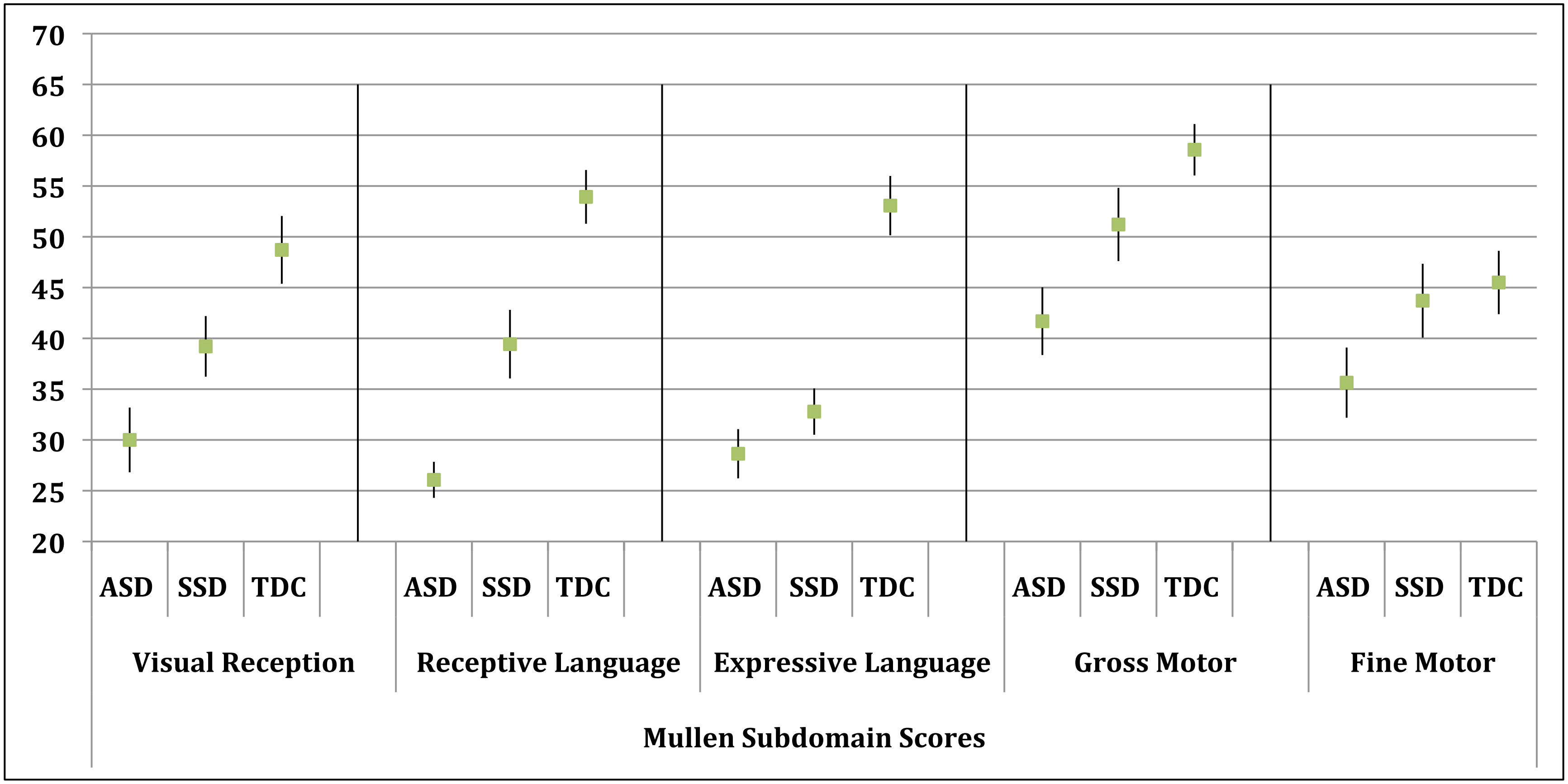

Figure 2. Mullen Subdomain Mean *t*-scores and Standard Error.

ASD = Early signs of Autism Spectrum Disorders group, SSD = Suspected Speech Delay group, TDC = Typically Developing Control group

group means for Gross Motor were not significantly different from each other, however. In all paired comparisons of Receptive Language mean scores, there were significant differences between groups. In the Expressive Language subdomain, however, both ASD and SSD groups were significantly different from the TDC means.

Adaptive Behavior (Vineland)

Differences in adaptive behavior skill levels were also anticipated due to the nature of communication and social deficits inherent in ASD. One of the children in the SSD group did not have a completed Vineland for analysis. Thus, the Vineland comparison group total was 41 instead of 42. ANOVA revealed significant differences between groups on the overall Adaptive Behavior Composite mean scores, $F(2,38) = 30.23$, $p<.001$. Significant differences were also found across Communication, $F(2,38) = 24.93$, $p<.001$, Daily Living, $F(2,38) = 11.60$, $p<.001$, Socialization, $F(2,38) = 18.964$, $p<.001$, and Motor domains on the Vineland, $F(2,38)$, $= 10.51$, $p<.001$. See Table 4 for means and standard deviations of Vineland composite and domain scores. As expected, children in the TDC group obtained mean scores in the Average range (standard scores in the range between 85 and 115) across all domains of the Vineland, and children in the SSD and ASD groups obtained lower mean scores on all domains.

Table 4

Adaptive Behavior

	Means, (Standard Deviations)			Tukey's HSD Post-hoc Comparisons		
Vineland Standard Scores	ASD	SSD	TDC	ASD and SSD	ASD and TDC	SSD and TDC
Adaptive Behavior Composite	84.14 (8.06)	90.69 (9.55)	110.14 (9.89)	$p = .17$	$p <.001$	$p <.001$
Communication	79.07 (13.36)	87.00 (11.34)	108.71 (9.87)	$p = .19$	$p <.001$	$p < .001$
Socialization	86.86 (8.62)	92.69 (10.10)	109.29 (11.08)	$p = .29$	$p <.001$	$p < .001$
Motor	88.43 (8.82)	92.62 (10.45)	103.14 (6.68)	$p = .44$	$p <.001$	$p = .009$
Daily Living	92.21 (10.70)	97.54 (10.74)	113.00 (13.75)	$p = .48$	$p <.001$	$p = .005$

HSD = Honestly Significant Difference, ASD = Early signs of Autism Spectrum Disorders group, SSD = Suspected Speech Delay group, TDC = Typically Developing Control group

Post-hoc analysis utilizing Tukey's HSD on the Vineland scores showed the ASD group mean scores to be similar to the SSD group on the Adaptive Behavior Composite and all domains. Both ASD and SSD group means were lower than and significantly different from the TDC group on the Adaptive Behavior Composite and all domains. Table 4 shows Tukey's HSD significance data. Means and standard errors of the Vineland scores are shown in Figure 3.

Behavior Coding

Interrater Reliability

Eighteen of the videos (20%) were coded by both raters. Within these 18 videos, 5,391 interval codes were directly compared (9 codes were blank). Percentage of exact agreement was calculated on each coding item (number of intervals with exact agreement divided by total intervals of agreement and disagreement) to be 82%. For individual behavior categories, percentage of agreement was highest for Response to Name (94%), Sounds (88%), and Play, (82%), followed by Responding (72%) and Initiating (72%). Exact agreement is the method that gives the clearest practical sense of consistency between raters (Repp, Deitz, Boles, Deitz, & Repp, 1976).

Although kappa is a commonly used calculation of reliability, it has limitations. In a critical review of statistical techniques for comparing raters, Ludbrook (2002) remarked that because categorical measurement inherently

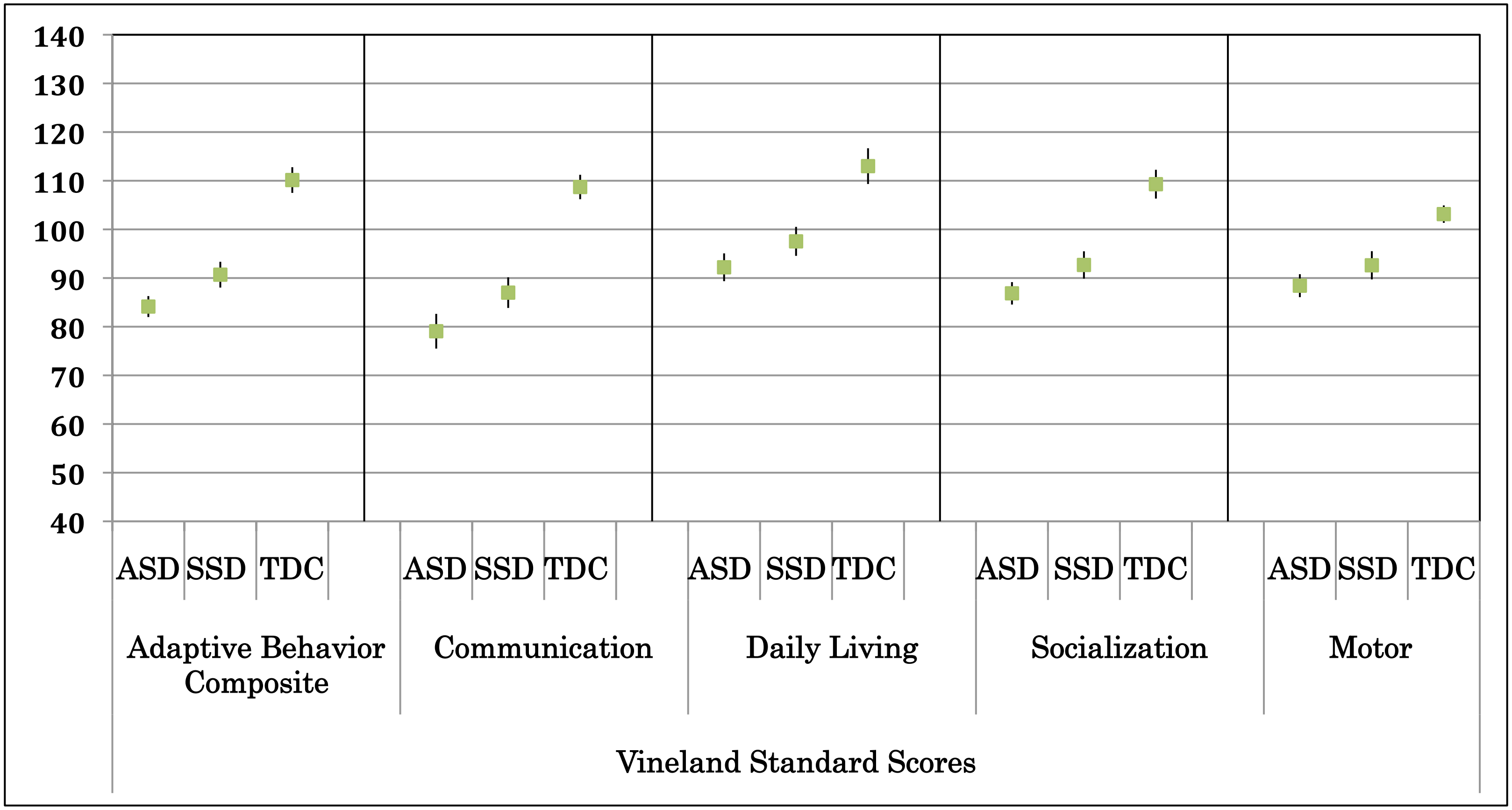

Figure 3. Vineland Mean Standard Scores and Standard Error.

ASD = Early signs of Autism Spectrum Disorders group, SSD = Suspected Speech Delay group, TDC = Typically Developing Control group involves subjective measurement or rating, bias is assumed, especially in a clinical setting. He concludes that the kappa technique is thus not an optimum method of determining reliability for clinical observation. Another limitation is that when the prevalence of a behavior or disease is very high or very low, a high observed proportion of agreement may exist, yet the value of kappa may indicate poor reliability (Cunningham, 2009; Viera & Garrett, 2005).

The behavior coding data in this study include a wide variety of base rates for behaviors, including one (Response to Name) that was very low (6%). The overall occurrence of Atypical behaviors (2%) and Unclear codes (4%) across behaviors were also very low. Because of these low rates, kappa may not be the best indicator of reliability (Cunningham, 2009, Ludbrook, 2002). Overall kappa for these data was .67 which is characterized as "good," "substantial," or "lenient" by various categorization scales (Landis & Koch, 1977; Altman, 1990; and Fleiss, 1981; all as cited in Byrt, 1996; Watkins & Pacheco, 2000), but again is likely not the best way to present reliability for this study.

One recommended way of presenting reliability data when the kappa may not be the best fit for the data is to also report the proportions of specific agreement, which include positive and negative agreement (Byrt, 1996). Because there were four possible codes in this case, there are two comparisons of behavior that are relevant. One comparison is to consider behavior that was either present (positive: A + T codes) or absent (negative: U + X codes). The other is to limit the comparison to present, coded behaviors (A or T codes) only.

Positive agreement is the number of agreements on present behavior times 2 (for each coder) divided by the total of all present behavior codes given (agreements times 2 plus disagreements). Negative agreement is the number of agreements on absent behavior times 2 divided by the total of all absent behavior codes given (agreements times 2 plus disagreements: Cicchetti & Feinstein, 1990, as cited in Uebersax, 2009). Positive agreement on the occurrence of present behavior (A+T) was .84 and negative agreement that behaviors were absent (X+U) was .87.

For the comparison of Atypical and Typical codes, agreement for Atypical codes was the number of agreements on Atypical behavior times 2 divided by the total of agreements times 2 plus disagreements on Atypical behavior. Likewise agreement for Typical codes was the number of agreements times 2 divided by the number of agreements times 2 plus disagreements for Typical codes. Agreement on the low-occurring Atypical behavior was .35, whereas agreement that Typical behavior was present was .97. Typical behavior codes constituted 42% of all codes whereas Atypical behavior codes constituted 2% of all codes.

Anomalous Data

Anomalies were identified in the reliability data. Three videos had scores that were clear anomalies: one Initiating agreement of 35%, one Responding agreement of 43%, and one Play agreement of 43%. These scores were from separate participant videos, and reliability in other behavior categories on these same videos was not anomalous. TG and JM reviewed each of these videos and discovered one systematic error in each case that accounted for the low percentage of agreement. In these cases, of the two codings performed for reliability on these videos, the coding with errors was excluded. With the exception of these three instances, random assignment was used to decide which of the two raters' results to include in analysis.

Excluded Data

Examination of Unclear codes from all videos revealed some additional anomalies. In two videos in the ASD group, the child was off camera for 33% or more of the time. These were clear anomalies due to a technical error in camera work. One was at Time 1 (first 10 minutes), the other was at Time 2 (30 minutes). Thus, these data were excluded pairwise from analysis in relevant behavior categories (e.g., a child being off camera did not interfere with accurate coding of Sounds, so they were not excluded from that analysis).

Research Questions

Question 1: Do Children with Early ASD, Language Delays, or Typical Development Show Different Rates of Typical and Atypical Behavior During the First 10 Minutes of a Novel Situation Compared to 30-40 Minutes?

The sum of Atypical behavior codes was calculated for each participant at Time 1 (0-10 minutes) and Time 2 (30-40 minutes). Sums were similarly calculated for Typical, Unclear, and Nonoccurrence behaviors, and used to determine whether diagnostic groups showed different rates of behaviors at Time 1 and Time 2. Table 5 shows the number of Atypical, Typical, Unclear and Nonoccurrence codes for the entire sample at Time 1 and Time 2. Figure 4 is a graphical representation of the occurrence rates of the four codes. Each behavior type was first examined for skewness before testing for differences between groups and then within groups.

Atypical Behaviors

The distribution of intervals coded as Atypical was significantly positively skewed at both time points (Time 1 skewness statistic = 2.128, se_{skew} = .365; Time 2 skewness statistic = 2.679, se_{skew} = .365), and was

Table 5

Total Codes

	All Codes, All Videos, All Groups	
	Time 1	Time 2
Total Atypical Behavior Codes	283	253
Total Typical Behavior Codes	5,053	5,422
Total Unclear Behavior Codes	632	412
Total Nonoccurrence Codes	6,632	6,513
Total Codes	12,600	12,600

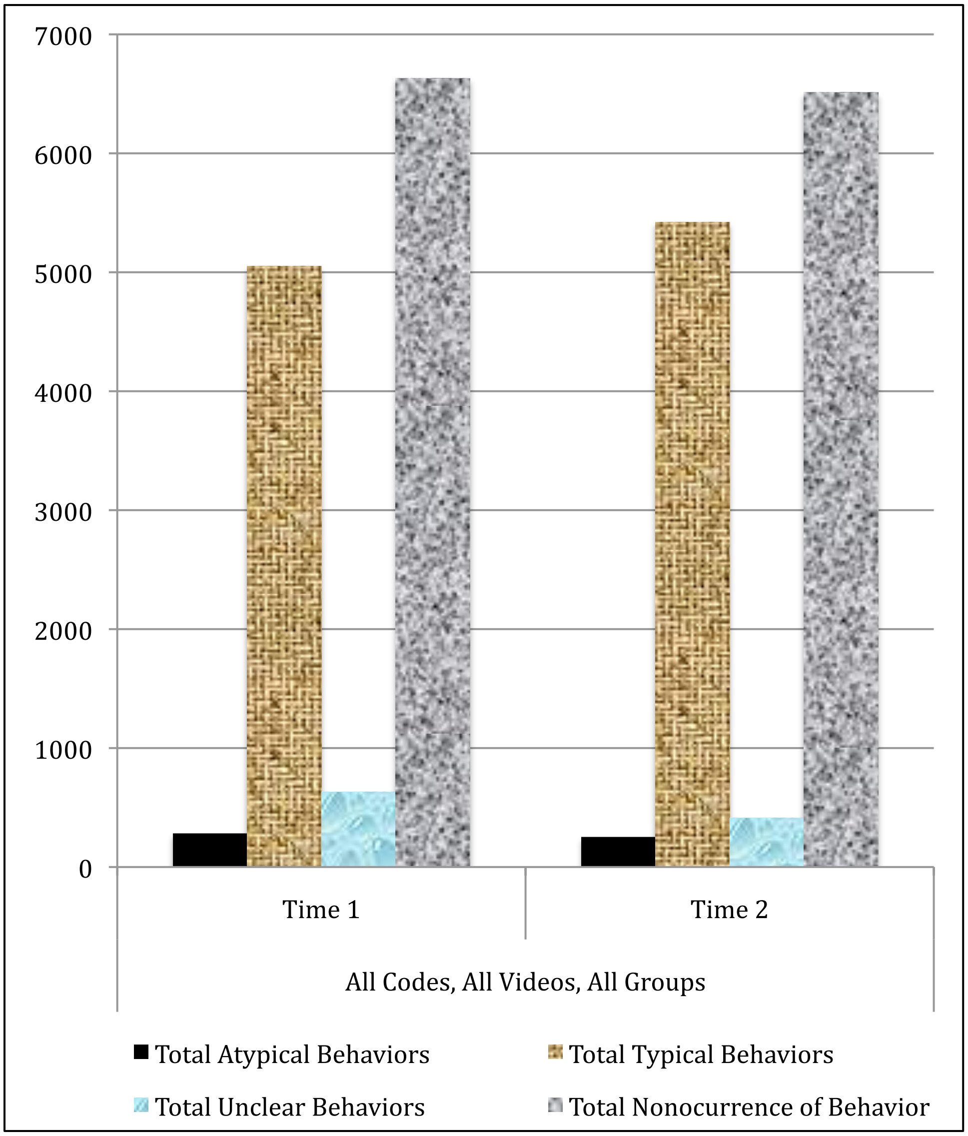

Figure 4. Occurrence of Observed Behaviors.

therefore corrected with a square root transformation. Thus, all analyses of Aytpical codes utilized the transformed version of this variable.

Repeated measures ANOVA revealed a main effect of group ($F(2,37) = 7.674$, $p = .01$). Post-hoc analysis utilizing Tukey's HSD showed the ASD group (transformed variable [square root] means and SD: Time 1 $M = 2.45$, $SD = 2.43$; Time 2 $M = 3.27$, $SD = 2.15$) exhibited significantly more Atypical behavior than the SSD group (Time 1 $M = .85$, $SD = 1.27$; Time 2 $M = .96$, $SD = 1.07$), $p = .01$, and the TDC group (Time 1 $M = 1.52$, $SD = 1.79$; Time 2 $M = .66$, $SD = 1.14$), $p = .01$. The SSD group and TDC groups were not significantly different from each other in Atypical behavior, ($p = .94$).

Within subject differences were then examined to determine if time had any effect on the number of Atypical behaviors observed between Time 1 and Time 2. No significant effect of time was found, ($F(2,37) =.01$, $p = .93$). Effect size of the differences over time independent of group was calculated by η^2_p (partial eta squared) to be .01, which is below practical significance (Barnette & McClean, 2006; Ferguson, 2009). Observed power according to this effect size was calculated to be .10, which means that this repeated measures ANOVA had only a 10% chance of detecting the interaction if the interaction exists in the population.

There was a trend toward an interaction of time within groups ($F(2,37) = 2$. Effect size of the differences over time within groups was calculated by

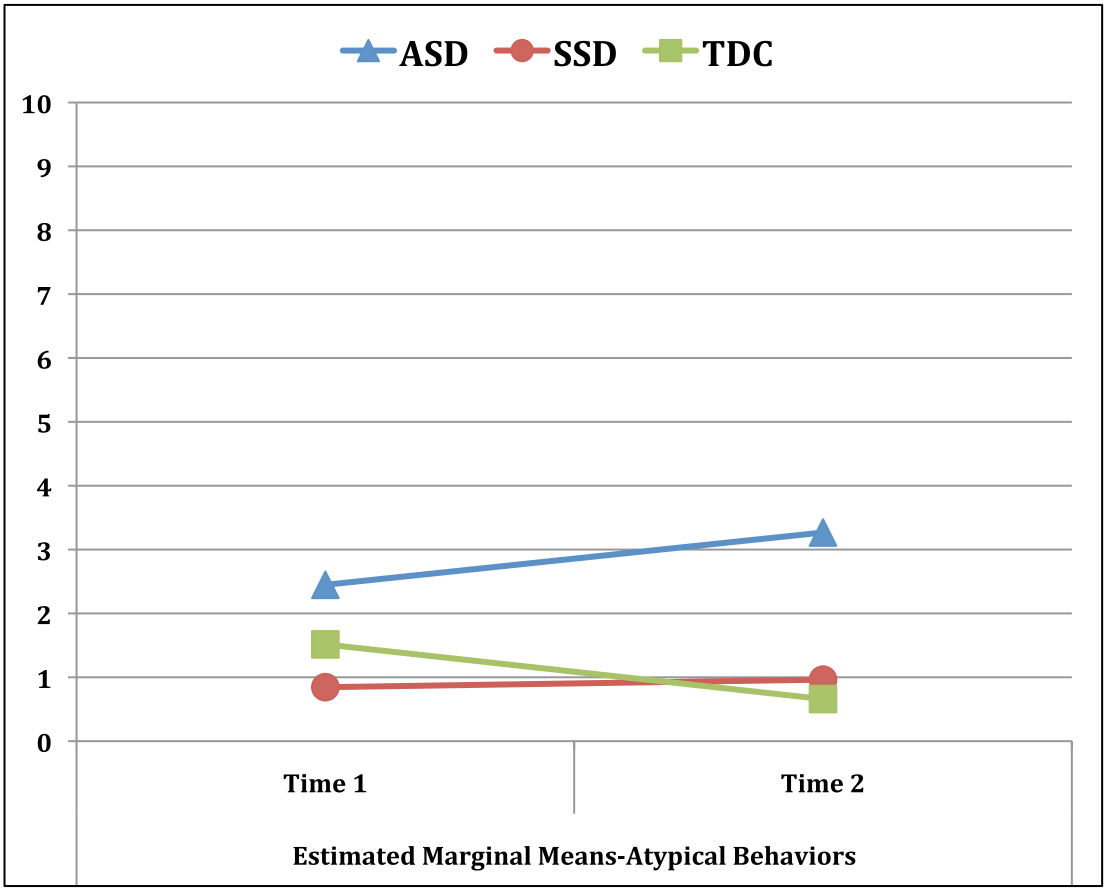

Figure 5. Estimated Marginal Means of Atypical Behavior Time 1 and Time 2

ASD = Early Signs of Autism, SSD = Suspected Speech Delay, TDC = Typically Developing Control

Note: Scale of behavior count is 0 – 300. Only the bottom section of scale (0 - 10) shown here for clarity.

η^2_p (partial eta squared) to be .118, which is considered a medium effect size by Cohen (.06 is medium, .14 is large), but may be more helpful to interpret as above minimum, but less than moderate (Barnette & McClean, 2006; Ferguson, 2009). Observed power according to this effect size was calculated to be .47, which means that this repeated measures ANOVA had only a 47% chance of detecting the interaction if the interaction exists in the population.

Typical Behaviors

The distribution of intervals coded as Typical was found to be acceptable in terms of skewness (skewness statistic = -.533, se_{skew} = .365 for Time 1; skewness statistic = -.348, se_{skew} = .365 for Time 2). The two videos from the ASD group with more than 33% off-camera codes were also excluded from this analysis.

Repeated measures ANOVA revealed significant differences in Typical behaviors between the three groups ($F(2,37) = 4.513$, $p = .018$). Post-hoc comparisons utilizing the Tukey HSD test indicated that the mean score for the ASD group (Time 1 M = 111.08, SD = 35.05; Time 2 M = 107.83, SD = 37.11) and TDC groups were significantly different, $p = .02$, but that the SSD group (Time 1 M = 120.14, SD = 24.74; Time 2 M = 139.21, SD = 20.14) was not significantly different from the ASD group, $p = .10$, or the TDC group (Time 1 M = 134.64, SD = 23.93; Time 2 M = 140.57, SD = 30.81), $p = .67$.

Contrary to the hypothesis that children with early signs of ASD would show significantly less Typical behavior at Time 2 compared to Time 1, no significant interaction of time was found independent of group $F(2,37) = 2.17$, $p = .15$. Effect size of the differences over time was calculated by η^2_p to be .06, which is above minimum, but the observed power was only .30.

No interaction of time by groups was found, $F(2,37) = 1.711$, $p = .20$, Figure 6 shows plots of change over time for each group. Effect size of the differences over time within groups was calculated by η^2_p to be .085 which is above minimum, but less than moderate (Barnette & McClean, 2006; Ferguson, 2009). Observed power according to this effect size was calculated to be .34.

Unclear and Nonoccurence Codes

The distribution of intervals coded as Unclear was significantly positively skewed at both time points (Time 1 skewness statistic = 4.41 $se_{skew} = .365$; Time 2 skewness statistic = 2.522, $se_{skew} = .365$), and was therefore corrected with a square root transformation. The two videos from the ASD group with more than 33% off-camera codes were also excluded from this analysis.

Repeated measures ANOVA revealed a main effect of group ($F(2,37) = 3.918$, $p = .029$) on Unclear behavior. Post-hoc analysis utilizing Tukey's HSD showed the ASD group had the highest rate of unclear behavior (Time 1 $M = 3.32$, $SD = 2.05$; Time 2 $M = 3.29$, $SD = 1.74$), and was significantly

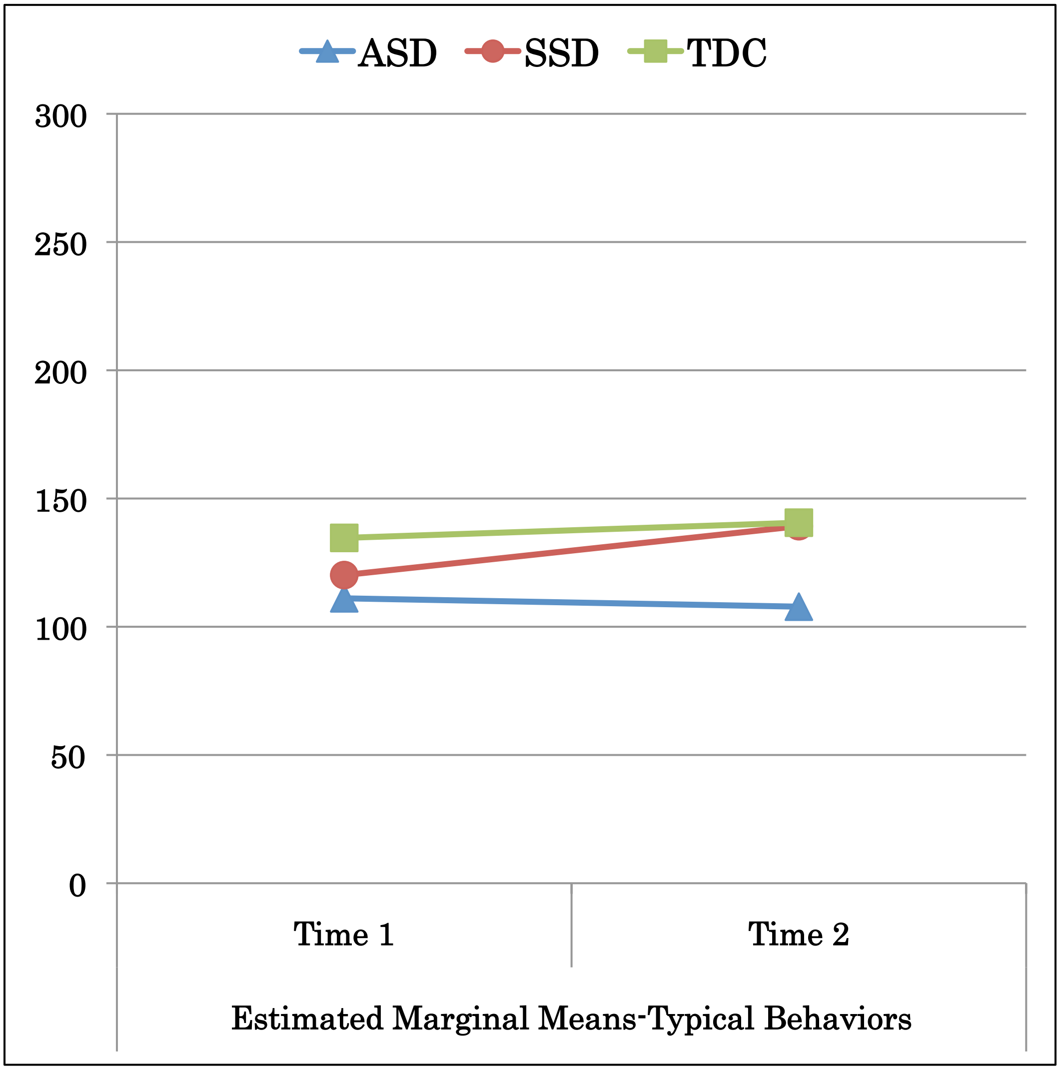

Figure 6. Estimated Marginal Means of Typical Behavior Time 1 and Time 2

ASD = Early Signs of Autism, SSD = Suspected Speech Delay, TDC = Typically Developing Control different from the TDC group (Time 1 *M* = 2.71, *SD* = 1.72; Time 2 *M* = 1.68, *SD* = 1.45), $p = .02$. The ASD group was not significantly different from the SSD group (Time 1 *M* = 3.18, *SD* = 1.39; Time 2 *M* = 2.31, *SD* = 1.36), $p = .35$. The SSD group was also not significantly different from the TDC group, *p* = ,33.

There was no main effect of time, $F(2,37) = 2.57$, $p = .12$. Effect size calculated by η^2_p was .06, which is above minimum. The observed power was .35.

The interaction of time by group was not found to be significant. ($F(2,37) = .56$, $p = .58$). Effect size calculated by η^2_p was .03, which is below minimum. The observed power was .14. Figure 7 shows plots of change over time by group. Effect size of the differences over time within groups was calculated by η^2_p to be .029, which is considered to be below the practically significant effect size of .04 (Ferguson, 2009). Observed power according to this effect size was calculated to be .14.

Nonoccurrence codes were found to be not significantly different between groups ($F(2,37) = .814$, $p = .415$). There was no main effect of time either, $F = .70$, $p = .41$. Effect size calculated by η^2_p was .02 which was below minimum, with observed power of .13. Interaction of time by group was also not significantly different ($F(2,37) = 1.901$, $p = .164$). Effect size

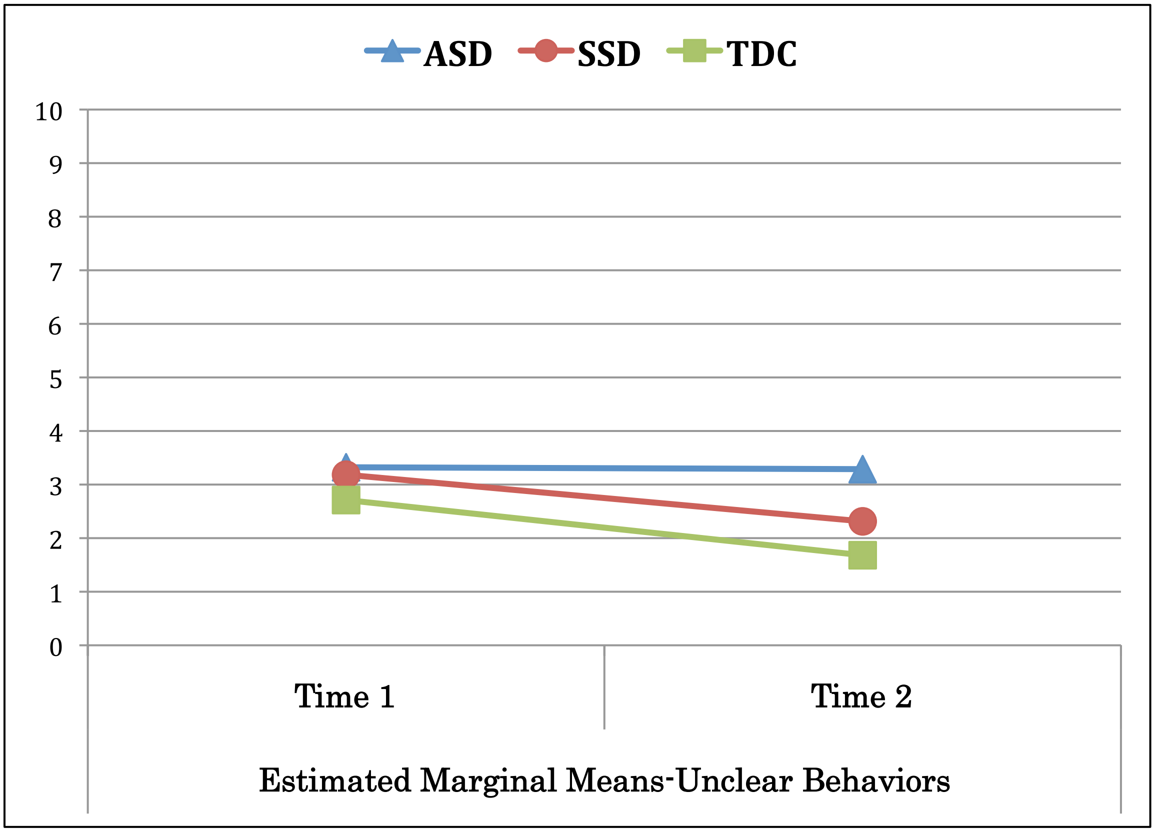

Figure 7. Estimated Marginal Means of Unclear Behavior Time 1 and Time 2

ASD = Early Signs of Autism, SSD = Suspected Speech Delay, TDC = Typically Developing Control

Note: Scale of behavior count is 0 – 300. Only the bottom section of scale (0 - 10) shown here for clarity.

calculated by η^2_p was .09 which was above minimum, and with observed power of .37. Figure 8 shows plots of estimated marginal means for nonoccurrence codes.

Combining Time 1 and Time 2 Codes for Remaining Analyses

Because no significant difference was found between Time 1 and Time 2, Time 1 and Time 2 data were combined for remaining analyses of behavior codes.

Question 2: If Differences Exist, Are They Across All Types of Behaviors or Specific to a Select Category of Behaviors?

Rates and Ratios of Behavior

Because Nonoccurrence and Unclear codes did not denote behaviors that could be qualitatively rated, no further analysis was conducted on those codes. The remaining codes for Atypical and Typical behavior were examined to determine if the proportions of Atypical Behavior and Typical Behavior codes were similar across all three groups when each behavior category was examined separately. As in the analysis of differences over time, two videos in the ASD group in which the child was off-camera for 33% or more of the time were excluded for analyses of Responding, Response to Name, Play, and Initiating. However, all 42 cases were examined for differences in behavior

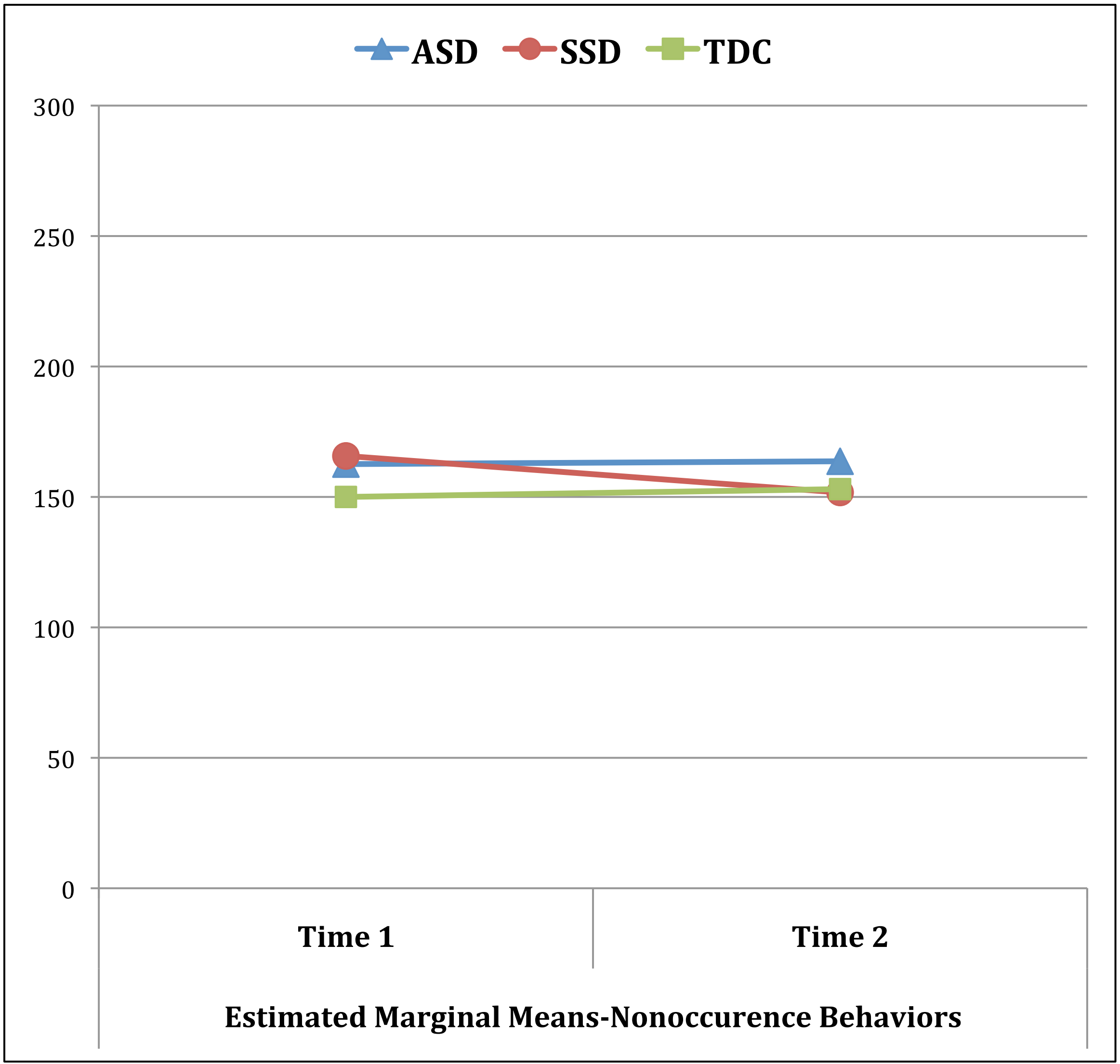

Figure 8. Estimated Marginal Means of Nonoccurrence Behaviors Time 1 and Time 2

ASD = Early Signs of Autism, SSD = Suspected Speech Delay, TDC = Typically Developing Control in the Sounds category. For each of the analyses, the number of occurrences for each child for Time 1 and Time 2 were added together, then divided by 120 (the total number of intervals across both Time 1 and Time 2 in each behavior category). The resulting percentage score was then used in the analysis of differences.

Rate of Occurrence--Atypical Behaviors

As with the distribution of intervals coded as Atypical, the distributions of the percentage of intervals with Atypical behavior was significantly positively skewed. The rate of total Atypical behaviors (skewness statistic = 1.97, se_{skew} = .365) and the rate of Atypical behaviors in Response to Name (skewness statistic = 1.814, se_{skew} = .365) were borderline for skew. Responding (skewness statistic = 3.53, se_{skew} = .365), Sounds (skewness statistic = 2.82, se_{skew} = .365), Play (skewness statistic = 2.70, se_{skew} = .365) and Initiating (skewness statistic = 3.26, se_{skew} = .365) were significantly skewed. Thus, all of the percentage scores were corrected with a square root transformation, and the transformed version of the variable was used in all analyses. The two videos from the ASD group with more than 33% off camera codes were excluded from this analysis in all categories except Sounds.

Total Rates of Atypical Behaviors

ANOVA revealed significant differences between groups in the rates of Total Atypical behaviors $F(2,37) = 7.938$, $p = .001$. Post-hoc analyses utilizing Tukey's HSD showed the ASD group ($M = .19$ $SD = .11$) to have significantly higher rates of Atypical behaviors than both the SSD ($M = .06$ $SD = .06$), $p = .001$, and TDC groups ($M = .07$ $SD = .08$), $p = .01$, in terms of Atypical behaviors observed. The SSD and TDC groups were not significantly different from each other, $p = .91$.

Rates of Atypical Responding

Significant differences also existed between groups in Responding, $F(2,37) = 5.14$, $p = .01$. Post-hoc analysis utilizing Tukey's HSD indicated the ASD ($M = .21$ $SD = .18$) group showed significantly lower rates of responding compared to the SSD ($M = .08$ $SD = .09$), $p = .04$, and TDC groups ($M = .06$ $SD = .09$), $p = .01$ in terms of the rate of Atypical responding behaviors observed. The SSD and TDC groups were not significantly different, however, $p = .91$.

Rates of Atypical Response to Name

Significant differences also existed in Atypical Response to Name behaviors observed, $F(2,37) = 7.76$, $p = .01$. Post-hoc analysis using Tukey's HSD showed the ASD group ($M = 1.12$ $SD = .82$), to have significantly lower rates of responding to name than both the SSD ($M = .19$ $SD = .35$), $p = .04$

and TDC groups (*M* = .49 *SD* = .60), *p* = .01. The SSD and TDC groups were not significantly different; however, *p* = .91.

An opportunity to respond to name was defined as any interval in which the examiner called the child's name to get his or her attention. These opportunities were observed in only 6% of 10 second intervals across all videos. Some of these occasions were formal presses for response, as in the ADOS. Other opportunities were informal bids for attention. The number of opportunities to respond to name was not controlled for, and was not equivalent between groups. The examiner called the name of a child in the ASD group an average of 5 times per 10-minute interval (*SD* = 3.6). Children in the ASD group responded to their names an average of 2.8 times of the 5 (*SD* = 2.4). In contrast, in the SSD group, the examiner called the child's name an average of 2.4 times (*SD* = 2) and the child responded in a typical manner an average of 2 times (*SD* = 1.7). In the TDC group, the examiner called the child's name 2.9 times on average (*SD* = 2), and the children responded in a typical manner on average 2.3 (*SD* = 1.8) times (see Appendices D and E for Atypical and Typical Responses to Name).

All children in the ASD group responded to their names at least once, even though they may have failed to respond on all other occasions. Also, 57% of children in the SSD group and 50% of children in the TDC group failed to respond at least once (Appendices D and E).

Rates of Atypical Sounds

In the analysis of rates of Atypical behavior in the category of Sounds, the two videos with bad camera work were not excluded as sounds were clear enough to rate (very few unclear codes were recorded) even when the child was off camera. The rate of Atypical Sounds was significantly different between groups, $F(2,39) = 5.33$, $p = .01$. Post-hoc analysis using Tukey's HSD showed the ASD group ($M = .18$ $SD = .15$), to have a significantly higher percentage of atypical sounds than the SSD group ($M = .03$ $SD = .08$), $p = .01$, but did not meet the threshold for significant difference from the TDC group ($M = .07$ $SD = .15$), $p = .06$, in terms of the rate of Atypical sounds observed. The SSD and TDC groups were not significantly different from each other; however, $p = .72$.

Rates of Atypical Play

Play behaviors were also significantly different $F(2,37) = 5.09$, $p = .011$. Post-hoc analysis using Tukey's HSD showed the ASD group ($M = .15$ $SD = .19$) to have a significantly higher percentage of atypical play behaviors and stereotypical behaviors than the SSD groups ($M = .01$ $SD = .03$), $p = .01$, but did not meet the threshold for significance in comparison to the TDC group ($M = .04$ $SD = .09$), $p = .06$. The SSD and TDC groups were not significantly different, $p = .76$.

Rates of Atypical Initiating

The rate of Atypical behavior in the Initiating category was not significantly different between groups, *F*(2,37) = .406, *p* = .67.

Rate of Occurrence – Typical Behaviors

Rates of Total Typical Behaviors

The rates of total Typical behaviors across categories were significantly different between groups *F*(2,37) = 4.513, *p* = .02. Post-hoc analysis utilizing Tukey's HSD indicated the ASD group (*M* = .36, *SD* = .10) showed a significantly lower percentage of typical behavior than the TDC group (*M* = .46, *SD* = .07), *p* = .02. The ASD group was not significantly different from the SSD group (*M* = .43 *SD* = .06), *p* = .10, however. The SSD and TDC groups were also not significantly different from each other, *p* = .67.

Rates of Typical Responding

Significant differences existed between groups in Responding, *F*(2,37) = 10.47, *p*<.001. Post-hoc analysis using Tukey's HSD showed the ASD group (*M* = .53, *SD* = .16) had a significantly lower percentage of typical responses than both the SSD (*M* = .73 *SD* = .11), *p* = .002 and TDC groups (*M* = .75, *SD* = .09), *p*<.001. The SSD and TDC groups were not significantly different; however, *p* = .86.

Rates of Typical Initiating

Significant differences also existed in Typical behavior of Initiating between groups, *F* (2,37) = 8.25, *p* = .001. Post-hoc analysis using Tukey's HSD showed the ASD group (*M* = .22, *SD* = .14) to have significantly lower percentages of typical initiating behaviors than both the SSD (*M* = .40, *SD* = .12), *p* = .01 and TDC groups (*M* = .41 *SD* = .12), *p* = .01. The SSD and TDC groups were not significantly different; however, *p* = .99.

Rates of Typical Play

Differences in Typical play behavior between groups were not significant, *F*(2,37) = 1.20, *p* = .31.

Rates of Typical Response to Name

Differences in Typical behavior in the categories of Response to Name *F*(2,37) = .908, *p* = .41, and Play *F*(2,37) = 1.198, *p* = .31, were not significant.

Rates of Typical Sounds

Analysis of Sounds was completed using all videos because technical difficulties with camera work did not interfere with rating of sounds. No significant differences were found between groups in the rate of Typical Sounds, *F*(2,39) = 1.01, *p* = .37.

Question 3: Is Behavior in the Earlier or Later Time Interval More Consistent With a Child's Subsequent Diagnosis?

The proposed analysis was to determine predominance of Atypical or Typical behavior at Time 1 and Time 2 for analysis. However, as discussed above, no significant differences were found over time. In addition, the overwhelmingly predominant behavior code was Typical (see Table 5). Thus, the relationship between Atypical and Typical behavior within each behavior category was explored to answer the modified question, "What is the ratio of Atypical to Typical behaviors across behavior categories?"

Typical behavior was more likely to be observed in all categories across all groups, but Atypical behavior was present in all categories across all groups as well. Table 6 displays ratios of Atypical behaviors to Typical behaviors by behavior categories and diagnostic group.

Another way to consider differences and relationships between Typical and Atypical behaviors observed is to look at the relative proportions of each as a percentage of the total of Atypical and Typical behaviors combined. To do this, the number of Atypical behaviors across groups (in each behavior category) was divided by the total of Atypical plus Typical behaviors in each behavior category. This procedure was repeated for Typical behaviors. The resulting percentages are mirrored proportions of 100%, shown in Tables 7 and 8. See Figure 9 for graphical representation of results.

Table 6

Ratios of Atypical to Typical Behavior

	Atypical: Typical		
Behavior	ASD	SSD	TDC
Responding	1:7.8	1:63	1:50.8
Response to Name	1:1.2	1:4	1:5.2
Sounds and Words	1:5.6	1:13	1:31
Play	1:12.6	1:88	1:659.2
Initiating	1:18.2	1:136	1:67.4

Table 7.

Atypical Behaviors

	ASD		SSD		TDC	
Behavior	Total Count	% Qualitatively Codable Behaviors	Total Count	% Qualitatively Codable Behaviors	Total Count	% Qualitatively Codable Behaviors
Responding	113	11.33	24	1.93	20	1.56
Response to Name	62	44.60	11	16.18	16	20.00
Sounds and Words	95	15.18	12	3.22	41	7.14
Play	90	7.35	2	0.15	15	1.13
Initiating	20	5.21	10	1.46	5	0.73
Total	380	11.3	59	1.6	98	2.5

Table 8.

Typical Behaviors

	ASD		SSD		TDC	
Behavior	Total Count	% Qualitatively Codable Behaviors	Total Count	% Qualitatively Codable Behaviors	Total Count	% Qualitatively Codable Behaviors
Responding	884	88.67	1,220	98.07	1,260	98.44
Response to Name	77	55.40	57	83.82	64	80.00
Sounds and Words	531	84.82	361	96.78	533	92.86
Play	1,135	92.65	1,319	99.85	1,316	98.87
Initiating	364	94.79	674	98.54	680	99.27
Total	2,991	88.72	3,631	98.40	3,853	97.54

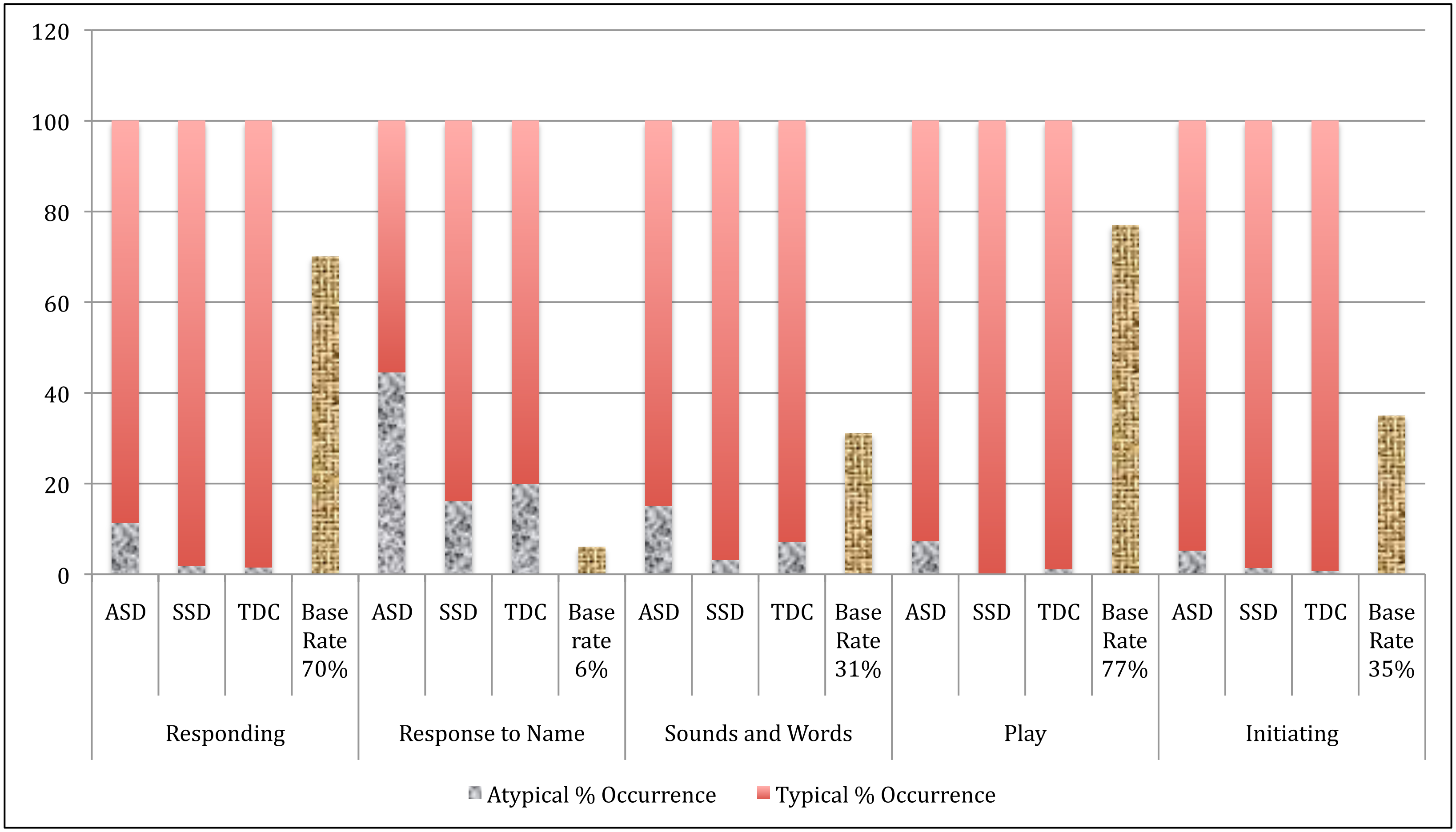

Figure 9. Relationship of Atypical to Typical Behaviors as Percentages of Atypical and Typical Total Behaviors and Base Rates of Codable Behavior Occurrence

When examining the frequency of either Atypical or Typical behavior occurring, it is critical to consider the base rate of Atypical and Typical behaviors. For example, Atypical Response to Name was observed 44.6% of the time a codable behavior was observed, but opportunities to respond (or not) occurred only 6 % of the time, so the actual occurrence rate is 44.6% of 6% of intervals, or 2.7% of total intervals observed. Base rate was calculated as the percentage of all intervals in which a behavior was observed as either Atypical or Typical, in other words, a clearly observable behavior that could be qualitatively coded. The number of Atypical and Typical (combined) behaviors was divided by the total number of opportunities (5,040 in each category) to calculate the base rate of the behavior. Base rates for each behavior are also shown in Figure 9.

Question 4: Is the Impression of Need for Further Referral Accurate After a 10-minute Observation or 30-40 Minutes into an Observation?

At the end of each coding session, raters answered the question, “Based on this observation alone, would you refer this child for an ASD evaluation?” The “Yes” or “No” answer was characterized as “Correct” or “Incorrect” according to diagnostic group (e.g., “Yes” was “Correct” for ASD, but “Incorrect” in SSD or TDC groups). Rater judgment was most inaccurate for the ASD group (61% correct). Of the 28 separate observations made of children in the ASD group, rater judgment was correct for 17 video behavioral samples, but incorrect for 11 of the samples. In the SSD group, 7 of the 28 were incorrectly referred for an ASD evaluation. In the TDC group, raters incorrectly recommended referral for 3 of the 28 for an ASD evaluation. See Figure 10 for correct and incorrect rater judgments by diagnostic group.

Differences Between Time 1 and Time 2

Comparison of Time 1 and Time 2 rater judgments (Correct or Incorrect) showed no statistically significant differences $F(1,82) = .562$, $p = .46$. In order to better understand the differences in rater judgment accuracy, differences between the raters, differences attributable to age of the child, and differences attributed to diagnostic group were examined in addition to the variable of time.

Differences Between Raters

From the reliability data, we know that raters recorded exactly the same codes 82% of the time on the videos examined for reliability, indicating substantial overall agreement between raters in terms of identifying behaviors. Although there were no statistically significant differences between correct and incorrect judgments between Time 1 and Time 2, it cannot be assumed that no significant differences existed between raters at Time 1 and Time 2 in terms of their rater judgments. Therefore, correct and

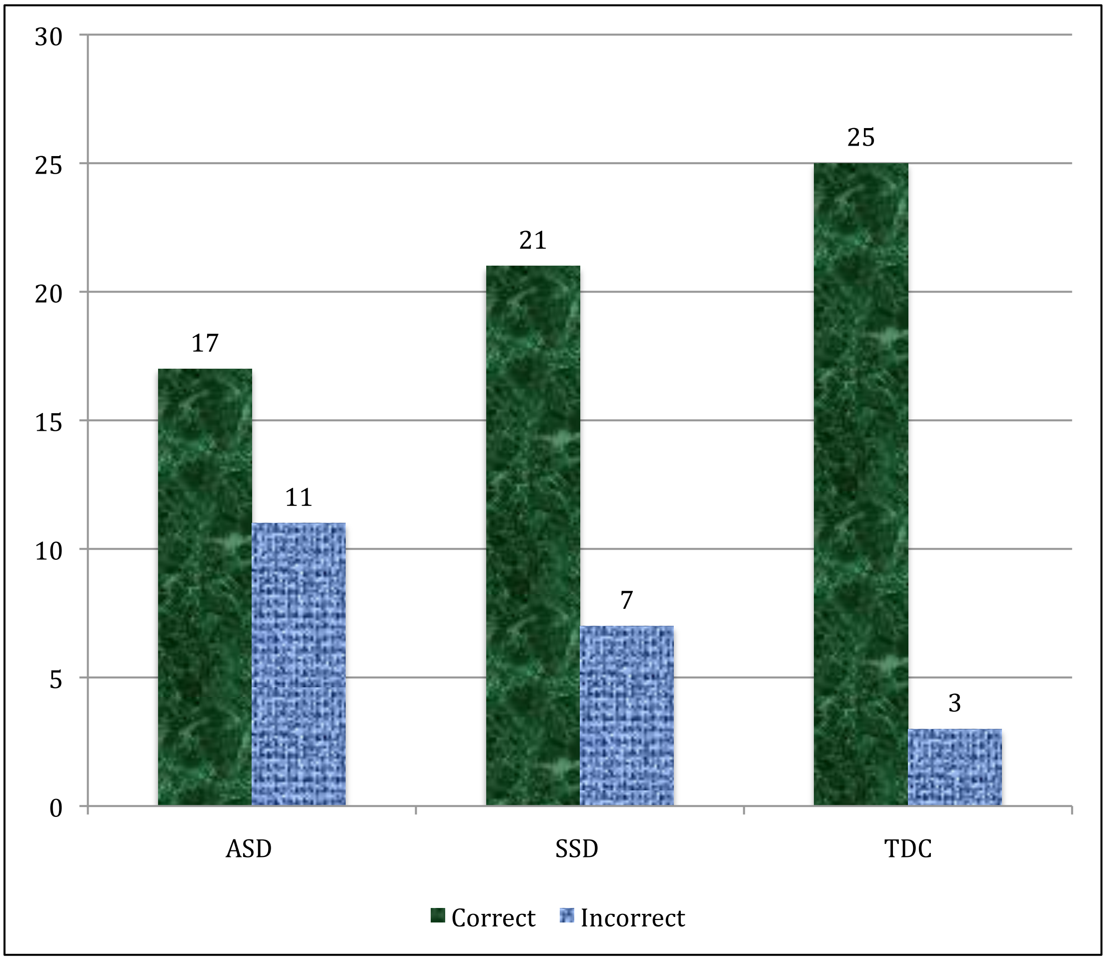

Figure 10. Correct and Incorrect Referrals by Diagnostic Group

ASD = Early Signs of Autism Spectrum Disorder, SSD = Suspected Speech Delay, TDC = Typically Developing Control incorrect judgments were analyzed separately for Time 1 and Time 2 to ascertain if any differences existed between raters by time.

Time 1

For all groups combined, there were no significant differences between raters, $\chi^2(1) = .154$, $p = .70$. Accuracy of coding was also examined between raters for each diagnostic group. No significant differences between raters were found in the ASD group, $\chi^2(1) = .219$ $p = .64$; the SSD group, $\chi^2(1) = .729$, $p = .39$; or the TDC group, $\chi^2(1) = 2.33$, $p = .13$.

Time 2

Analysis of accuracy of rater judgment at Time 2 produced the same results. For all groups combined, there were no significant differences between raters, $\chi^2(1) = .42$, $p = .52$. Consistent with Time 1, analysis of each group found no significant differences at Time 2 in the ASD group, $\chi^2(1) = .93$, $p = .33$; the SSD group, $\chi^2(1) = .14$, $p = .71$; or the TDC group, $\chi^2(1) = .43$, $p = .51$.

Individual Views

With no significant differences found between raters at Time 1 or Time 2, videos were collapsed into a single group for further analysis. Individual views of each video (e.g., Time 1 and Time 2 of the same child constitute 2 individual views) were examined for differences in accuracy of rater judgment. Because 20% of the videos were viewed by both raters for reliability, some children were viewed twice by a rater. To avoid variance attributable to individual children in repeated views, only one of these two views was included in the analysis. Random numbers were generated to determine which of the two views was included. The remaining videos (n = 67) were analyzed for accuracy of rater judgment between raters.

Overall, there were no significant differences in accuracy of rater judgment between the two raters, $\chi^2(1) = .25$, $p = .61$. Analysis by diagnostic group also yielded no significant differences between raters in the ASD group, $\chi^2(1) = .135$, $p = .71$; SSD group, $\chi^2(1) = .00$, $p = 1.0$; or TDC group, $\chi^2(1) = .1.69$, $p = .19$.

Differences Between Groups

Because no significant differences were found in rater judgment at either Time 1 or Time 2, and no differences between Time 1 and Time 2 were found between raters, both of these variables were collapsed for further analysis. With raters combined and Time 1 and Time 2 combined, a trend toward significance was found among the three diagnostic groups in terms of correct or incorrect rater judgments, $\chi^2(1) = .5.79$, $p = .06$.

The research question asked if observers were able to make an accurate judgment about a referral question for an autism evaluation, so further analysis was performed to determine if the two non-ASD groups (SSD and TDC) could be collapsed into a homogenous group that did not merit a referral for ASD evaluation. No significant differences in correct or incorrect rater judgments were found between the SSD and TDC groups, $\chi^2(1) = 2.21$, $p = .14$. These two groups were collapsed for further analysis with the combined group referred to as the non-ASD group.

A significant difference was found in accuracy of rater judgment between the ASD group and the non-ASD group, $\chi^2(1) = 4.08$, $p = .04$, with accuracy of rater judgments in the ASD group lower than accuracy of rater judgments of children without early signs of ASD.

Differences Attributable to Age of the Child

To determine if age could be a factor in predicting the accuracy of rater judgment, the mean ages of the videos coded as incorrect and correct were examined for significant differences. The age range of the participants was 15.4-33.2 months, with no significant differences between group mean age, $F(2,39) = .45$, $p = .6$. Time 1 and Time 2 groups were reestablished for this analysis under the hypothesis that age of the child may be related to a longer "warm up" period for younger children which may affect rater judgment if only one 10-minute view of the child was available. If this was the case, Time 2 views would show more correct judgment across age ranges.

Contrary to this hypothesis, comparison of means at Time 1 showed a significant difference between incorrect and correct rater judgments based on age, $F(1,40) = 7.218$, $p = .01$. The mean age of children's videos with incorrect rater judgments was 20.21 months (SD = 4.01months). Mean age of the children's videos with correct rater judgments was 24.66 months (SD = 5.14).

When the analysis was repeated for Time 2, no significant difference was found, *F*(1,40) = .019, *p* = .890. Mean age of the incorrect group was 23.17 months (*SD* = 5.38) and the correct group mean age was 23.45 (*SD* = 5.25).

DISCUSSION

Symptomatic Presentation

This study set out to examine typical and atypical behaviors that might be observed during a brief observation of toddlers with ASD, suspected speech delay, and toddlers who are typically developing. This was conceptualized as a way to gain insight into the experience of a front line clinical provider charged with making the first referral for a developmental evaluation. While children with early signs of ASD in the study showed significantly fewer typical behaviors than children with speech delays and children who were typically developing in some behavior categories (e.g., Responding and Initiating), it was surprising to find that the overwhelming majority of behaviors shown by all children, including the children with early signs of ASD, were grossly typical across all behavior categories. This is the first study to detail ratios and rates of typical and atypical behavior in brief observations, which may help us understand what factors affect decisions made to refer for further evaluation or not. The study adds to the research literature on early identification of ASD symptoms and evidence-based best practice in pediatric health care.

Screening Versus Clinical Judgment

There are currently no reliable or consistent tests for physiological markers to determine risk for development of ASD, although some early studies have shown promise (Nordahl et al., 2011; Sokhadze et al., 2009; Wolff et al., 2012). Given the absence of objective methods of identification, behavioral observation and screening are the best available tools we have to identify ASDs in the earliest stages (Johnson et al., 2007).

This study shows that the information available to a clinician within a 10-minute observation consists primarily of typical behaviors, with very few atypical behaviors, even in children with early signs of ASD. Considering that substantial differential diagnostic information is not available through behavior observation in a brief evaluation, differences in symptomatic presentation may be captured more accurately through the use of an autism screening tool as part of an autism screening process. This dilemma of accuracy of clinical judgment being sometimes at odds with statistical prediction has been present in the literature for several decades. Resolution of the dilemma as stated in the classical work of Paul Meehl, *Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of the Evidence* (1954, 1996, as cited in Grove, 2005) is to favor the method that is most likely to be accurate.

ASD screening measures have been rightly criticized for the risk of false positive or false negative results in population-wide screening (Allison et al., 2012; Al-Qabandi et al, 2011; Oosterling et al., 2010). Clinical judgment has also been criticized for low sensitivity after screening identified more children with ASD (Kim et al., 2011; Miller et al., 2011; Oosterling et al., 2010; Robins, 2008) Results from our study suggest that clinical observation of ASD symptoms during a 10-minute period would identify only 61% of the children known to need an ASD referral. However, this observation also resulted in false positives for children who needed only a speech evaluation (25%) and typically developing children (11%). Apprehension about false positives from universal screening programs has been one factor that may affect the slow uptake of universal screening. Our results suggest, however, that clinical judgment would be significantly less sensitive, and that it would not necessarily avoid the problem of false positives that may be upsetting to parents (Al-Qabandi, Gorter, & Rosenbaum, 2011).

This is also the first study to examine how the time constraints of community based pediatric practice could influence decision making about referral for ASD evaluation. Previous research has indicated that clinical judgment is not as effective as screening for identification of ASD risk (Robins, & Dumont-Mathieu, 2006), and factors such as bias have been explored as possible reasons why clinical judgment may be less effective (Begeer et al., 2009; Mandell et al., 2007, 2009). This is the first study to examine what a pediatric care provider or other evaluator may observe in a brief 10-minute evaluation in terms of typical and atypical behaviors.

Understanding the extent to which typical and atypical behaviors manifest within a brief observation will allow pediatric health care providers to make more informed decisions about the possible need for ASD evaluation referrals.

Underlying Differences in Cognitive and Adaptive Abilities Are Not Reliably Apparent in Brief Observations

Differences between the groups—ASD and TDC—existed in almost every comparison of scores on measures of cognitive ability on the Mullen, and adaptive skills as measured by the Vineland. On both measures, across all domains but Mullen Fine Motor, mean scores from the ASD group were significantly lower than those of the TDC group.

Mullen Receptive Language was the only domain in which significant differences were found across all groups, including SSD. On some measures, (e.g., Mullen Early Learning Composite, Mullen Visual Reception, Mullen Expressive Language, and all Vineland skill domains), ASD and SSD children achieved scores that were not significantly different from each other (both lower than TDC). In other domains, (e.g., Mullen Visual Receptions and Mullen Gross Motor), SSD and TDC groups were not significantly different from each other (both higher than ASD). In only one domain (Mullen Fine Motor) were there no significant differences between groups.

Despite these very strong differences in measured abilities, and ability levels that were characterized as Very Low and Moderately Low in the ASD group, these differences in ability did not seem to be apparent enough during the brief behavioral observation to influence rater judgment about whether to refer for an ASD evaluation. Underlying differences in abilities and skills, then, may not be reliably apparent in brief social observations.

Does a Novelty Effect Exist?

Behavioral data were obtained by having two raters view 10-minute video clips of children. The first of these clips was the first 10 minutes of an evaluation, the second was at 30 – 40 minutes to test for the presence of a novelty effect in which atypical behaviors became more evident over time. A trend towards significance was seen in the change in Atypical behaviors, but the extremely low rate of occurrence of Atypical behaviors (2.0% at Time 1 – 2.2% at Time 2) makes it possible that a novelty effect will not be found at these two time points even with a larger sample and more power. The data suggest a difference, but not enough Atypical behavior was observed to definitively answer the research question of a novelty effect. The finding of greater interest is the low occurrence rate and the overwhelming predominance of typical behaviors that were seen in the ASD group. Even if the hypothesized novelty effect had been found, it may or may not have been noticeable enough to the clinical examiner to overcome the amount of typical behaviors observed.

Low Occurrence of Atypical Behaviors Affects Detection of Differences

In each time sample, raters recorded Atypical and Typical behaviors observed within five broad behavioral categories. These five categories were defined by compiling behavioral signs of ASD from the ADOS, the DSM-IV-TR, and the Systematic Observation of Red Flags (SORF). The five categories encompassed general themes of early signs of ASD that might be both known to pediatric providers and observable to providers and parents. These categories included general social initiating and responding, responding to name, quality of sounds and words, and quality of play. While autism diagnostic instruments such as the ADOS make much finer behavioral distinctions, during brief observations, broad categories are likely to be what parents and general providers consider when looking for early signs of an ASD.

Although occurrence rates were very low, there were significantly more Atypical behaviors evident in the ASD group in all behavior categories except Initiating. Differences were statistically significant in Responding, Response to Name, Sounds, and Play, but accuracy of rater judgment about when to refer did not reliably reflect these differences. This may be a case of statistical significance that does not rise to the level of clinical significance. Our results suggest that with such a low occurrence of Atypical behaviors to begin with, it was not possible for clinicians to reliably detect the statistically higher rate of atypical behaviors displayed by the children with ASD.

Response to Name

Typically developing children learn to respond to their names by around age 6 months (CDC, 2011) and by 12 months, failure to respond to name is a strong, but not universal early indicator for developmental delays in general and for ASD (Nadig et al., 2007). In this sample, there was a significant difference in the rate of failure to respond to name between the ASD group and the SSD and TDC groups. The number of opportunities to observe this behavior complicates interpretation of these findings, however. While the ADOS has a specific press regarding response to name, behavior samples used in the study were taken from the ADOS based on elapsed time, not on matched presses for behaviors and responses. Thus, coding of Response to Name may have reflected some specific presses, but also naturally occurring attempts to get the child's attention by calling his name. All children in the study had at least one opportunity to respond to their names (see Appendices D and E). While children with ASD had a high rate of failure to respond to their name (45%), the ratio of atypical to typical responding was still 1.2:1, suggesting that children with ASD may respond to their names some of the time.

The very specific nature of response to name behavior makes it an appealing press to use in a brief observation, but these data suggest that the children may respond well enough (55% chance of response) that it might not be the most sensitive indicator of ASD symptoms. Consideration must also be given to the number of opportunities to respond to his or her name. In this study, the number of opportunities for responding to name was not controlled for. Children in the ASD group had, on average, more opportunities to respond than SSD and TDC children, which may have affected the higher rate of atypical responses (more opportunities to show an atypical response). It may also have reflected additional presses provided to a type of child who fails to respond the first time, or a child who is generally less in tune with the adult which results in more specific requests for attention.

These results are, however, consistent with a previous study of 12-month-old children in which response to name was able to distinguish between groups at high and low risk for ASD, yet responses after one or two presses had a sensitivity of only .50 for ASD (50% of those who ultimately were identified as having ASD passed the test and responded to their names at 12 months, 40% responded at 18 months, and 40% responded at 24 months). The authors of the study concluded that a failure to respond to name is indicative of developmental difficulties as early as 12 months and could be used as a screen in a pediatric setting. Based on the low specificity of response to name, however, they also concluded that typical, or intact response to name behavior should not preclude a more comprehensive evaluation in at risk populations or children who present with other clinical concerns (Nadig et al., 2007). Thus, a response to name press as observed in the literature and the current study may not be sensitive enough to accurately inform the decision to refer.

Sounds

Sounds as a behavior category resulted in nonsignificant differences in Typical and Atypical behaviors. Atypical sounds were defined as echolalia, odd prosody, pitch, or volume, or babbling without consonant sounds. Judging the quality of the sound produced during a brief observation might be difficult in some cases, as was reflected by one child in the TDC group who was coded as having a very high number of atypical sounds during the first 10 minutes. Both raters correctly judged the child as not needing a referral for an ASD evaluation, however.

Another difficulty with detecting differences in Sounds and Words within a 10-minute observation is that a child may not produce enough sounds to make a clear judgment. In this sample, sounds were observed (heard) in only 35% of coding intervals. Although there were no statistically significant differences among the three groups for typical sounds, there were more atypical sounds in the ASD group than the other two groups. Proportionally, when sounds were made, the ratio of typical to atypical sounds in all groups ranged from 5.6:1 in the ASD group to 13:1 in the SSD group and 31:1 in the TDC group.

Play

Play behaviors had the highest frequency of behaviors in our study (77% of 10-second intervals). Play with objects and stereotypical or unusual sensory seeking behaviors were both considered play behaviors; the former as "typical" and the latter as "atypical." Based on this gross categorization, children in the ASD group were much more likely to appear engaged in typical behavior than atypical behavior (ratio of 12.6:1). This may be surprising to some, given that stereotypical behaviors might be expected to be seen when a child with early signs of ASD experiences the stress of a novel environment. However, the emergence of repetitive behaviors in the early course of ASD is not well understood, and it may be that stereotyped behaviors are simply not as apparent in this young age range compared to preschool age children (Chawarska et al., 2007). Differences in play or stereotypical behavior were not found to be significant between groups in the sample. This is in contrast to a study of play behaviors in 18-month-old siblings of children with ASD based on a 4-minute observation with toys. This group of siblings at risk for ASD were found to have more atypical play behaviors than typically developing controls (Christensen et al., 2010). ASD sibling populations are at high risk for ASD symptoms, and include children who will go on to develop ASD, whereas the current study was based on a community sample of children who were not siblings of children with ASD.

Responding

Responding opportunities and behaviors (other than Response to Name) were observed at a relatively high rate in the sample (70% of 10-second intervals). This may be because eliciting a response from a child is, by its nature, a behavior that is easier than many of the other behaviors for an examiner to control in terms of opportunity. Use of responding as an informal screening behavior alone might result in a high rate of false negatives, however. Children in the ASD group showed significantly fewer typical responses than both the SSD and TDC groups, but still responded in a typical manner in a 7.8:1 ratio for each atypical response (e.g., not responding to a social bid). For coding, Responding was defined as a response to context cues, response to joint attention, expression of shared enjoyment, or accepting an object from the examiner.

Initiating

Initiating behaviors were observed to occur at a lower rate than Responding (35% of 10-second intervals). Typical initiating was defined as initiating or maintaining an interaction, eye gaze, sharing enjoyment or interest, joint attention, and pointing. Atypical initiating was defined as an unusual form of initiation (e.g., using the examiner's hand as a tool) or failing to maintain an interaction. For example, if the examiner started a back and forth interaction to which the child responded (e.g., by catching the ball rolled by the examiner), but did not initiate further interaction by reciprocating, (e.g., did not roll the ball back to the examiner), that was coded as a Typical Response, and Atypical initiating behavior. Other types of Initiating behaviors may have been more difficult to capture in coding as they represented the lack of a typical behavior. This may also be true of clinical judgment in other settings. By analyzing the rate of Typical and Atypical behaviors in all behavior categories, factors such as varying base rates and difficulties with "absent" behavior coding were equalized in the analysis of rates and ratios. The rates of Typical behavior were significantly different between the ASD group and both the SSD and TDC groups. In the ASD group, there were, on average, 18 or more typical initiations observed for every one time an initiation was clearly pressed for (such as maintenance of a back and forth interaction) and did not occur.

Typical Behaviors Differed in Social Interaction Behaviors Only

Typical behaviors were not significantly different across groups in the categories of Response to Name, Sounds, and Play. In the categories of social interaction (Responding and Initiating), Typical behaviors differed with lower rates of interaction in the ASD group. This is consistent with the diagnostic criteria of ASD as being primarily a disorder of social interaction (APA, 2000). It may be easier to identify atypical behaviors in the discrete category of Response to Name, in Sounds due to the obvious lack of talking, odd qualities of speech, or in Play due to nonfunctional play behaviors.

Symptomatic presentation in Responding and Initiating, however, as shown in these data, involves reduced, but not absent interactions, which may be more difficult to discern in a busy pediatric exam.

High Rates of Typical Total Behaviors May Overshadow the Significance of Different Atypical Behaviors

The behavior categories of Response to Name, Sounds, and Play were not found to be significantly different for Typical behaviors, nor was Initiating different in Atypical behaviors. When all behaviors were collapsed into a single total for Typical or Atypical behaviors, however, both Typical and Atypical totals were found to be significantly different between groups. The proportion of Atypical behaviors in the ASD group was significantly higher than it was in the TDC groups.

The overall rate of Typical behaviors in the ASD group was 88.7% (2,991 Typical codes/3371 total Atypical and Typical codes). This very high rate of Typical behavior may have overshadowed the rate of Atypical behaviors, which was 11.3% in the ASD group. That is, while this is statistically higher than the rate for the SSD (1.6%) and TDC (2.5%) groups, it might be difficult for a clinician to *perceive* this difference within a brief observation. Appendices F and G show rates and ratios of all categories of behavior in the three groups.

Accuracy of Clinical Judgment May Be Affected by Symptomatic Presentation

Findings of the study indicate that clinical judgments based on brief observations alone would miss approximately 39% of toddlers who need to be referred for ASD evaluations. Within the sample, raters were more accurate in identifying children who did not need a referral for an ASD evaluation. Rater judgment for referral was accurate in 75% of cases in the SSD group. Accuracy in the TDC group was even higher, at 89%. Given the very low rates of atypical behavior observed in all groups compared to the rates of typical behaviors across all groups, it might be hypothesized that this preponderance of typical behavior was a determining factor in rater judgment. However, this would not explain why some children in the SSD and TDC groups that were identified as needing an ASD referral. Correct and incorrect rater judgments were seen in cases where no Atypical behaviors were coded, suggesting that the decision to refer may include a complex set of factors.

The subtle differences between the ASD group showing Typical responses 89% of the time and the TDC group showing Typical responses 98% of the time can certainly make an accurate clinical judgment problematic, as was seen in the rate of inaccurate rater judgments in the study (25% across all groups).

Brief Observations May Not Provide Enough Information to Detect ASD Symptoms

For 39% of cases, a 10-minute observation period was too short for even an ASD expert to detect enough signs of ASD to decide to refer. Raters in this study were autism specialists, licensed psychologists with similar training and years of professional experience in autism assessment and early childhood development. They were able to carefully consider each 10-second interval, watching each three times to code separately for Responding/Response to Name, Sounds/Play, and Initiating. Raters also had the ability to rewind and rewatch any segment for clarity. Despite this high level of expertise and a focused, repeated observation, the symptomatic presentation of toddlers in the sample did not provide sufficient information for accurate rater judgments of need for referral more than 61% of the time.

Training efforts to build ASD diagnostic expertise in health care providers are commendable (Warren, Stone & Humberd, 2009), but may not be the most efficient or effective way to increase referral rates and accuracy of clinical judgments. Results of this study suggest that even high levels of expertise and experience with focused and repeated observation of a 10-minute behavioral sample, a substantial number of children needing referral will be missed if clinical judgment alone is the only screening method used.

Time Constraints in Pediatric Practice

Health care providers are the professionals with the most consistent and comprehensive access to toddlers for early identification of ASD, yet they are extremely limited by schedule demands in the amount of time allotted for each patient visit. Although the uptake of standardized screening for autism is increasing in pediatric primary care settings, more than half of providers surveyed still rely on clinical judgment alone rather than standardized developmental screening to guide decisions about referrals to Early Intervention, special education systems or other specialists for early identification of ASD (Radecki et al., 2011). The current study results suggest that the rate and ratio of behaviors that might be observed during a brief interaction with a child would not yield sufficient information to inform a consistently correct clinical judgment. Since extending the duration of observation for toddler health care visits is not very feasible, finding effective methods for gathering other information (including parent report) will help health care providers determine the need for more extensive screening or evaluation.

The current study did not show that an increase of 20-30 minutes spent in the evaluation environment produced significantly more informative behavioral observations. No significant differences in behavior were found between Time 1 and Time 2, but rater judgment was found to be significantly affected by age at Time 1. Younger children ($M =$ 20.21 months, $SD =$ 4.01

months) were more likely to be in the incorrect rater judgment group at Time 1, but at Time 2, no differences in age between the incorrect and correct group was found. The data are not specific as to what factors related to age may have caused more diagnostic uncertainty at Time 1 that resolved by Time 2. If a longer "warming up" period was needed for younger children, their atypical behaviors would have been found to be significantly different between Time 1 and Time 2, and it was not.

Strengths of the Current Study

Representative Nature of Community-Based Sample

Results of the study come from a community-based sample with diversity of race, ethnicity, language (English and Spanish) and SES, including children with Medicaid or CHIP (government subsidized) health insurance. Participants in the two clinical groups were recruited from a single suburban catchment area through screening of over 80% of patients seen in a large pediatric practice. Additional TDC participants were recruited from the same catchment area (including the same pediatric practice). The range of severity of ASD symptoms seen in the sample is therefore likely to be representative of the range that would be seen in many pediatric practices. Differences in symptomatic presentation seen within the sample are thus not as vulnerable to selection bias as they might be if participants had been recruited from higher risk or clinical populations.

Cognitive and Adaptive Profiles

The three groups were not significantly different in demographic make-up, but were significantly different in cognitive development and adaptive skills. The cognitive profile found in the ASD group is similar to that reported in studies of early autism in high-risk samples (Ray-Subramanian, Huai, & Weismer, 2011).

Coder Expertise

The high level of expertise of raters maximized the likelihood of accurate coding of observed behaviors and accuracy of rater judgment. The resulting accuracy of rater judgments (63%) may thus represent a best-case scenario in terms of expertise, suggesting that even highly skilled clinicians are not able to consistently identify early signs of ASD from a brief observation.

Focused Observation

Observations were made from video, which offered unlimited opportunity for review of clips to decrease the likelihood of missing behaviors or passing behavior off as unclear because of distraction in the moment. The raters' attention could thus be controlled for. The opportunity for repeated views of the videos for specific behaviors represents a best-case scenario for observing any atypical behaviors. A clear difference between study conditions and pediatric exams is that the observers were focused solely on watching the child, and watched the videos three times, rather than trying to interact with the child or ask the parents questions during the observation. Furthermore, the observers were looking specifically for ASD related behaviors, rather than trying to screen for the full range of physical and developmental issues included in a well child exam. These factors may limit generalization of findings, but this may perhaps be offset by the higher likelihood of seeing atypical behaviors during ADOS activities under close observation, as opposed to observations made during a broad based physical exam and discussion with parents.

Limitations

Power

One of the study aims was to determine if children with autism show a "novelty" effect where they appear more interactive and responsive in a novel environment, and then become less interactive as they become familiar with the people and objects in the room. These data suggested a possible trend in this regard, and post-hoc power analysis revealed that the loss of videos in the ASD group due to technical difficulties resulted in this analysis being significantly underpowered. Because the sample was drawn from a specific community of 1,000 children, augmentation of the ASD sample with clinically referred children was not done, as it may have distorted the representative nature of the sample. Future studies with larger groups may be able to answer the question of novelty effect more conclusively.

Reliability and Low Frequency Behaviors

Given the very low occurrence of Atypical behaviors, the kappa statistic may not accurately reflect the degree of agreement between raters. Kappa is known to be problematic in the case of very low rates of occurrence (Cunningham, 2009; Ludbrook, 2002, Viera & Garrett, 2005), which was the case with these data.

Atypical behaviors were originally the variable of greatest interest, but at the very low rates of occurrence in these behavioral samples (2%), agreement on atypical behaviors is not very informative regarding overall reliability between raters. Positive agreement of .35 may indicate that not enough codes were recorded for a reliable picture of how much each coder would agree given more opportunities for judgment. It may also indicate how difficult it is for raters to qualitatively judge a behavior as atypical under conditions of a brief observation in a novel environment. There may not have been enough atypical behavior occurring in a brief sample to consistently determine reliability. A third possibility is that the subtle early signs described in this very young age group (Zweigenbaum et al., 2009) are difficult to identify with reliability for any set of observers.

Typical behaviors occurred much more frequently at a rate of 42%, with agreement between raters on 97% of codes. Raters also agreed 84% of the time that a behavior had occurred and that nothing occurred in 87% of the time. The overall exact agreement measure of .82 may be the best general description of reliability across different behaviors and raters.

Setting

Study conditions were different from a pediatric exam. The child was engaged in the ADOS with lots of toys, rather than a physical exam in a pediatric exam room. In a pediatric exam, the provider might try to elicit specific behaviors within the brief exam (e.g., response to name, response to joint attention), whereas in this study the observers watched a 10-minute period of a longer examination, and thus were not able to witness all the social presses that were administered.

Additional Factors Contributing to Clinical Judgment

Clinical judgment of early signs of ASD includes more than just a brief observation. Medical and developmental history and information about environmental factors are also important components of clinical judgment that were not provided to raters in the study. Specific evaluation for cognitive abilities or presses for behavior that were not included in our behavioral samples may also be part of a pediatric health care appointment that could provide more information for clinical judgment to practitioners. The results of this study provide specific information about rates of typical and atypical behavior with a focus on novel situations and evaluators. The study was not designed to include all possible sources of information for clinical judgment. It is not known what factors may influence clinical judgment for an ASD referral (Allison, Auyeung, & Baron-Cohen, 2012). There may be many factors that influence the decision to refer a child for evaluation that have not yet been discovered or defined. Data regarding what types of behavior may be observed within the constraints of a brief medical appointment may help to identify some of these factors.

Areas of Future Study

Recruitment of a larger sample of children with ASD would have increased power to determine if a novelty effect exists. Due to the relatively low base rate of ASD, large participant pools would be required to obtain a larger number of participants with ASD. The current sample was recruited through daily screening over a period of 6 months in a large and busy pediatric clinic. Increasing the number of recruited participants with ASD will necessitate a longer screening partnership with a pediatric practice or recruitment from multiple sites.

Another direction for future study is taking a more detailed look at the behaviors that are seen under novel conditions. Standardizing behavioral presses across observations would provide information about behaviors that can be elicited in novel situations as well as behaviors that are observed under more naturalized conditions. It may also help to further identify consistent differences in symptomatic presentation in early ASD compared to SSD and TDC groups. However, there is still no guarantee that a brief set of presses would yield sensitive results, as the press for response to name does not seem to.

Replication of the study in pediatric settings under conditions more representative of a well child exam is yet another direction for future study. The intention of this study was to provide data to promote efforts to study symptomatic presentation in a pediatric setting directly. Further research within community pediatric practice settings is ideal for generalization of findings, but may be difficult to achieve. These results speak to the need for more research on actual practices in real pediatric health care settings.

A final area of future study is examination of the determining factors in a health care provider's decision to refer a child for an ASD evaluation or not. This study has identified one factor that may influence decision making for referrals by showing the predominance of typical behaviors compared to atypical behaviors in the symptomatic presentation of ASDs in brief observations. Other factors such as medical and developmental history, co-morbid conditions, opportunities for discussion, clinician awareness of early signs of ASD, experience with ASD patients, and clinician awareness and opinions regarding early intervention options may all contribute to the accuracy of clinical judgment for referral.

Conclusions

Identification of early signs of ASD is the most critical element in efforts to improve the outcome of children with ASD, which is a rapidly increasing population. Pediatric health care providers are the professionals with the most comprehensive access to children in the toddler age range when signs of ASD first become apparent, somewhere between the 18- and 36-month well child visits. In brief or limited observations of toddlers with early signs of ASD, the low base rate of atypical behavior and the high ratio of typical behaviors to atypical behaviors may make it difficult to determine if a child is truly symptomatic and needs further evaluation.

Current research indicates more than half of pediatric care providers use clinical judgment alone to determine risk of developmental delays such as ASD in toddlers. In this study, experts trained in ASD diagnosis were given a brief 10-minute opportunity for observation to inform their judgment of need for referral. Under these conditions, they were not able to detect sufficient problems in 39% of children with early signs of ASD. Raters' judgments of need for referral were more accurate in typically developing and speech delayed populations, which are much larger groups than ASD populations in community pediatric practice. This suggests that accuracy in detecting the lack of developmental problems in the majority of toddlers through brief observations alone may not necessarily predict accuracy in identification of developmental problems in children with early signs of ASD.

Symptomatic presentation of ASDs in this study was captured in a sample that is representative of the general population in the United States in many ways. Within the sample, symptomatic presentation of ASD was complex in that much higher rates of typical behavior were seen compared to atypical behavior rates. Even high levels of early childhood and ASD expertise and experience were not sufficient to consistently form accurate clinical judgments based on these brief symptomatic presentations. Ongoing research continues to establish the sensitivity, specificity, and positive predictive value of standardized ASD screening tools. Use of standardized screening tools, as prescribed by AAP guidelines, can provide health care professionals with more information about the risk of ASD in a child than clinical judgment of symptomatic presentation alone. Steps taken by pediatric health care providers to increase accurate referrals can have the benefit of earlier referral and entrance into intervention programs and improved outcomes for children with ASDs.

APPENDIX A

“LEARN THE SIGNS. ACT EARLY.” DEVELOPMENTAL MILESTONES AND WARNING SIGNS

The following checklists are adapted from CARING FOR YOUR BABY AND YOUNG CHILD, BIRTH TO AGE 5, edited by Steven Shelov and Tanya Reimer Altmann © 1991, 1993, 1996, 2004, 2009 by the American Academy of Pediatrics and BRIGHT FUTURES; GUIDELINE FOR HEALTH SUPERVISION OF INFANTS, CHILDREN, AND ADOLESCENTS, Third Edition, edited by Joseph Hagan, Jr., Judith S. Shaw, and Paula M. Duncan, 2006, Elk Grove Village, IL: American Academy of Pediatrics

Your Baby at 2 Months,
Your Baby at 4 Months,
Your Baby at 6 Months,
Your Child at 1 Year,
Your Child at 18 Months (1 ½ Years),
Your Child at 2 Years,
Your Child at 3 Years,
Your Child at 4 Years,
Your Child at 5 Years.

The checklists have been modified in font size and formatting to fit within thesis guidelines. For the original checklists, please visit http://www.cdc.gov/ncbddd/actearly/milestones/index.html

Your Baby at 2 Months

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by the end of 2 months. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next.

What most babies do at this age:

Social/Emotional

☐ Begins to smile at people
☐ Can briefly calm himself (may bring hands to mouth and suck on hand)
☐ Tries to look at parent

Language/Communication

☐ Coos, makes gurgling sounds
☐ Turns head toward sounds

Cognitive (learning, thinking, problem-solving)

☐ Pays attention to faces
☐ Begins to follow things with eyes and recognize people at a distance
☐ Begins to act bored (cries, fussy) if activity doesn't change

Movement/Physical Development

☐ Can hold head up and begins to push up when lying on tummy
☐ Makes smoother movements with arms and legs

Act early by talking to your child's doctor if your child:

☐ Doesn't respond to loud sounds
☐ Doesn't watch things as they move
☐ Doesn't smile at people
☐ Doesn't bring hands to mouth
☐ Can't hold head up when pushing up when on tummy

Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

Your Baby at 4 Months

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by the end of 4 months. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next.

What most babies do at this age:

Social/Emotional

- ☐ Smiles spontaneously, especially at people
- ☐ Likes to play with people and might cry when playing stops
- ☐ Copies some movements and facial expressions, like smiling or frowning

Language/Communication

- ☐ Begins to babble
- ☐ Babbles with expression and copies sounds he hears
- ☐ Cries in different ways to show hunger, pain, or being tired

Cognitive (learning, thinking, problem-solving)

- ☐ Lets you know if she is happy or sad
- ☐ Responds to affection
- ☐ Reaches for toy with one hand
- ☐ Uses hands and eyes together, such as seeing a toy and reaching for it
- ☐ Follows moving things with eyes from side to side
- ☐ Watches faces closely
- ☐ Recognizes familiar people and things at a distance

Movement/Physical Development

- ☐ Holds head steady, unsupported
- ☐ Pushes down on legs when feet are on a hard surface
- ☐ May be able to roll over from tummy to back
- ☐ Can hold a toy and shake it and swing at dangling toys
- ☐ Brings hands to mouth
- ☐ When lying on stomach, pushes up to elbows

Act early by talking to your child's doctor if your child:

- ☐ Doesn't watch things as they move
- ☐ Doesn't smile at people
- ☐ Can't hold head steady
- ☐ Doesn't coo or make sounds
- ☐ Doesn't bring things to mouth
- ☐ Doesn't push down with legs when feet are placed on a hard surface
- ☐ Has trouble moving one or both eyes in all directions

Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

www.cdc.gov/actearly | 1-800-CDC-INFO

Learn the Signs. Act Early.

Your Baby at 6 Months

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by the end of 6 months. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next.

What most babies do at this age:

Social/Emotional

- ☐ Knows familiar faces and begins to know if someone is a stranger
- ☐ Likes to play with others, especially parents
- ☐ Responds to other people's emotions and often seems happy
- ☐ Likes to look at self in a mirror

Language/Communication

- ☐ Responds to sounds by making sounds
- ☐ Strings vowels together when babbling ("ah," "eh," "oh") and likes taking turns with parent while making sounds
- ☐ Responds to own name
- ☐ Makes sounds to show joy and displeasure
- ☐ Begins to say consonant sounds (jabbering with "m," "b")

Cognitive (learning, thinking, problem-solving)

- ☐ Looks around at things nearby
- ☐ Brings things to mouth
- ☐ Shows curiosity about things and tries to get things that are out of reach
- ☐ Passes things from one hand to the other

Movement/Physical Development

- ☐ Rolls over in both directions (front to back, back to front)
- ☐ Begins to sit without support
- ☐ When standing, supports weight on legs and might bounce
- ☐ Rocks back and forth, sometimes crawling backward before moving forward

Act early by talking to your child's doctor if your child:

- ☐ Doesn't try to get things that are in reach
- ☐ Shows no affection for caregivers
- ☐ Doesn't respond to sounds around him
- ☐ Has difficulty getting things to mouth
- ☐ Doesn't make vowel sounds ("ah", "eh", "oh")
- ☐ Doesn't roll over in either direction
- ☐ Doesn't laugh or make squealing sounds
- ☐ Seems very stiff, with tight muscles
- ☐ Seems very floppy, like a rag doll.

Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

www.cdc.gov/actearly | 1-800-CDC-INFO

Learn the Signs. Act Early.

Your Baby at 9 Months

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by the end of 9 months. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next.

What most babies do at this age:

Social/Emotional

☐ May be afraid of strangers
☐ May be clingy with familiar adults
☐ Has favorite toys

Language/Communication

☐ Understands "no"
☐ Makes a lot of different sounds like "mamamama" and "bababababa"
☐ Copies sounds and gestures of others
☐ Uses fingers to point at things

Cognitive (learning, thinking, problem-solving)

☐ Watches the path of something as it falls
☐ Looks for things he sees you hide
☐ Plays peek-a-boo
☐ Puts things in her mouth
☐ Moves things smoothly from one hand to the other
☐ Picks up things like cereal o's between thumb and index finger

Movement/Physical Development

☐ Stands, holding on
☐ Can get into sitting position
☐ Sits without support
☐ Pulls to stand
☐ Crawls

Act early by talking to your child's doctor if your child:

☐ Doesn't bear weight on legs with support
☐ Doesn't sit with help
☐ Doesn't babble ("mama", "baba", "dada")
☐ Doesn't play any games involving back-and-forth play
☐ Doesn't respond to own name
☐ Doesn't seem to recognize familiar people
☐ Doesn't look where you point
☐ Doesn't transfer toys from one hand to the other

Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

www.cdc.gov/actearly | 1-800-CDC-INFO

Learn the Signs. Act Early.

Your Child at 1 Year

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by his first birthday . Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next. **What most children do at this age:**

Social/Emotional

☐ Is shy or nervous with strangers
☐ Cries when mom or dad leaves
☐ Has favorite things and people
☐ Shows fear in some situations
☐ Hands you a book when he wants to hear a story
☐ Repeats sounds or actions to get attention
☐ Puts out arm or leg to help with dressing
☐ Plays games such as "peek-a-boo" and "pat-a-cake"

Language/Communication

☐ Responds to simple spoken requests
☐ Uses simple gestures, like shaking head "no" or waving "bye-bye"
☐ Makes sounds with changes in tone (sounds like speech)
☐ Says "mama" and "dada" and exclamations like "uh-oh!"
☐ Tries to say words you say

Cognitive (learning, thinking, problem-solving)

☐ Explores things in different ways, like shaking, banging, throwing
☐ Finds hidden things easily
☐ Looks at the right picture or thing when it's named
☐ Copies gestures
☐ Uses things correctly; drinks from cup, brushes hair
☐ Bangs two things together
☐ Puts things in a container, takes things out of a container
☐ Lets things go without help
☐ Pokes with index (pointer) finger
☐ Follows simple directions like "pick up the toy"

Movement/Physical Development

☐ Gets to a sitting position without help
☐ Pulls up to stand, "cruises" (walks) holding on to furniture
☐ May take a few steps without holding on
☐ May stand alone

Act early by talking to your child's doctor if your child:

☐ Doesn't crawl
☐ Can't stand when supported
☐ Doesn't search for things that she sees you hide
☐ Doesn't say single words like "mama" or "dada"
☐ Doesn't learn gestures like waving or shaking head
☐ Doesn't point to things
☐ Loses skills he once hadTell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

www.cdc.gov/actearly | 1-800-CDC-INFO

Learn the Signs. Act Early.

Your Child at 18 Months (1½ Years)

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by 18 months. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next. **What most children do at this age:**

Social/Emotional

☐ Likes to hand things to others as play
☐ May have temper tantrums
☐ May be afraid of strangers
☐ Shows affection to familiar people
☐ Plays simple pretend, such as feeding a doll
☐ May cling to caregivers in new situations
☐ Points to show others something interesting
☐ Explores alone but with parent close by

Language/Communication

☐ Says several single words
☐ Says and shakes head "no"
☐ Points to show someone what he wants

Cognitive (learning, thinking, problem-solving)

☐ Knows what ordinary things are for; for example, telephone, brush, spoon
☐ Points to get the attention of others
☐ Shows interest in a doll or stuffed animal by pretending to feed
☐ Points to one body part
☐ Scribbles on his own
☐ Can follow 1-step verbal commands without any gestures; for example, sits when you say "sit down"

Movement/Physical Development

☐ Walks alone
☐ May walk up steps and run
☐ Pulls toys while walking
☐ Can help undress herself
☐ Drinks from a cup
☐ Eats with a spoon

Act early by talking to your child's doctor if your child:

☐ Doesn't point to show things to others
☐ Can't walk
☐ Doesn't know what familiar things are for
☐ Doesn't copy others
☐ Doesn't gain new words
☐ Doesn't have at least 6 words
☐ Doesn't notice or mind when a caregiver leaves or returns Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

Your Child at 2 Years

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by his second birthday. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next. **What most children do at this age:**

Social/Emotional

☐ Copies others, especially adults and older children
☐ Gets excited when with other children
☐ Shows more and more independence
☐ Shows defiant behavior (does what he was told not to)
☐ Plays mainly beside other children, but is beginning to include other children, such as in chase games

Language/Communication

☐ Points to things or pictures when they are named
☐ Knows names of familiar people and body parts
☐ Says sentences with 2 to 4 words
☐ Follows simple instructions
☐ Repeats words overheard in conversation
☐ Points to things in a book

Cognitive (learning, thinking, problem-solving)

☐ Finds things when hidden under two or three covers
☐ Begins to sort shapes and colors
☐ Completes sentences and rhymes in familiar books
☐ Plays simple make-believe games
☐ Builds towers of 4 or more blocks
☐ Might use one hand more than the other
☐ Follows two-step instructions such as "Pick up your shoes and put them in the closet."
☐ Names items in a picture book such as a cat, bird, dog

Movement/Physical Development

☐ Stands on tiptoe
☐ Kicks a ball
☐ Begins to run
☐ Climbs onto and down from furniture without help
☐ Walks up and down stairs holding on
☐ Throws ball overhand
☐ Makes or copies straight lines and circles

Act early by talking to your child's doctor if your child:

☐ Doesn't use 2-word phrases (for example, "drink milk")
☐ Doesn't know what to do with common things, like a brush, phone, fork, spoon
☐ Doesn't copy actions and words
☐ Doesn't follow simple instructions
☐ Doesn't walk steadily
☐ Loses skills she once had

Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

www.cdc.gov/actearly | 1-800-CDC-INFO

Learn the Signs. Act Early.

Your Child at 3 Years

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by his third birthday. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next.

What most children do at this age:

Social/Emotional

- ☐ Copies adults and friends
- ☐ Shows affection for friends without prompting
- ☐ Takes turns in games
- ☐ Shows concern for crying friend
- ☐ Understands the idea of "mine" and "his" or "hers"
- ☐ Shows a wide range of emotions
- ☐ Separates easily from mom and dad
- ☐ May get upset with major changes in routine
- ☐ Dresses and undresses self

Language/Communication

- ☐ Follows instructions with 2 or 3 steps
- ☐ Can name most familiar things
- ☐ Understands words like "in," "on," and "under"
- ☐ Says first name, age, and sex
- ☐ Names a friend
- ☐ Says "I," "me," "we," and "you" and some plurals
- ☐ Talks well enough for strangers to understand
- ☐ Carries on a conversation using 2 to 3 sentences

Cognitive (learning, thinking, problem-solving)

- ☐ Can work toys with buttons, levers, and moving parts
- ☐ Plays make-believe with dolls, animals, and people
- ☐ Does puzzles with 3 or 4 pieces
- ☐ Understands what "two" means
- ☐ Copies a circle with pencil or crayon
- ☐ Turns book pages one at a time
- ☐ Builds towers of more than 6 blocks
- ☐ Screws and unscrews jar lids or turns door handle

Movement/Physical Development

- ☐ Climbs well
- ☐ Runs easily
- ☐ Pedals a tricycle (3-wheel bike)
- ☐ Walks up and down stairs, one foot on each step

Act early by talking to your child's doctor if your child:

- ☐ Falls down a lot or has trouble with stairs
- ☐ Drools or has very unclear speech
- ☐ Can't work simple toys (such as peg boards, simple puzzles, turning handle)
- ☐ Doesn't speak in sentences
- ☐ Doesn't understand simple instructions
- ☐ Doesn't play pretend or make-believe
- ☐ Doesn't want to play with other children or with toys
- ☐ Doesn't make eye contact

Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

www.cdc.gov/actearly | 1-800-CDC-INFO

Learn the Signs. Act Early.

Your Child at 4 Years

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by his fourth birthday. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next. **What most children do at this age:**

Social/Emotional

☐ Enjoys doing new things
☐ Plays "Mom" and "Dad"
☐ Is more creative with make-believe play
☐ Prefers play with other children than by himself
☐ Cooperates with other children
☐ Can't tell what's real and what's make-believe
☐ Talks about what she likes and what she is interested in

Language/Communication

☐ Knows some basic rules of grammar, such as correctly using "he" and "she"
☐ Sings song or says a poem from memory such as the "Itsy Bitsy Spider" or the "Wheels on the Bus"
☐ Tells stories
☐ Can say first and last name

Cognitive (learning, thinking, problem-solving)

☐ Names some colors and some numbers
☐ Understands the idea of counting
☐ Starts to understand time
☐ Remembers parts of a story
☐ Understands the idea of "same" and "different"
☐ Draws a person with 2 to 4 body parts
☐ Uses scissors
☐ Starts to copy some capital letters
☐ Plays board or card games
☐ Tells what he thinks will happen next in a book

Movement/Physical Development

☐ Hops and stands on one foot up to 2 seconds
☐ Catches a bounced ball most of the time
☐ Pours, cuts with supervision, and mashes own food

Act early by talking to your child's doctor if your child:

☐ Can't jump in place
☐ Has trouble scribbling
☐ Shows no interest in interactive games or make-believe
☐ Ignores or doesn't respond to people outside the family
☐ Resists dressing, sleeping, and using the toilet
☐ Can't retell a favorite story
☐ Doesn't follow 3-part commands
☐ Doesn't understand "same" and "different"
☐ Confuses "me" with"you"
☐ Speaks unclearly
☐ Loses skills he once had

Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

www.cdc.gov/actearly | 1-800-CDC-INFO

CDC

Learn the Signs. Act Early.

Your Child at 5 Years

How your child plays, learns, speaks, and acts offers important clues about your child's development. Developmental milestones are things most children can do by a certain age. Check the milestones our child has reached by his fifth birthday. Take this with you and talk with you child's doctor at every visit about the milestones your child has reached and what to expect next.

What most children do at this age:

Social/Emotional

- ☐ Wants to please friends
- ☐ Wants to be like friends
- ☐ More likely to agree with rules
- ☐ Likes to sing, dance, and act
- ☐ Shows concern and sympathy for others
- ☐ Is aware of gender
- ☐ Can tell what's real and make-believe
- ☐ Shows more independence
- ☐ Is sometimes demanding and sometimes very cooperative

Language/Communication

- ☐ Speaks very clearly
- ☐ Tells a simple story using full sentences
- ☐ Uses future tense
- ☐ Says name and address

Cognitive (learning, thinking, problem-solving)

- ☐ Counts 10 or more things
- ☐ Can draw a person with 6+ body parts
- ☐ Can print some letters or numbers
- ☐ Copies triangle, other geometric shapes
- ☐ Knows about things used every day, like money and food

Movement/Physical Development

- ☐ Stands on one foot for 10+ seconds
- ☐ Hops; may be able to skip
- ☐ Can do a somersault
- ☐ Uses a fork, spoon, table knife
- ☐ Can use the toilet on her own
- ☐ Swings and climbs

Act early by talking to your child's doctor if your child:

- ☐ Doesn't show a wide range of emotions
- ☐ Shows extreme behavior (unusually fearful, aggressive, shy or sad)
- ☐ Unusually withdrawn, not active
- ☐ Is easily distracted, has trouble focusing on one activity for more than 5 minutes
- ☐ Doesn't respond to people, or responds only superficially
- ☐ Can't tell what's real or make-believe
- ☐ Doesn't play a variety of games and activities
- ☐ Can't give first and last name
- ☐ Doesn't use plurals or past tense
- ☐ Doesn't talk about daily activities or experiences
- ☐ Doesn't draw pictures
- ☐ Can't brush teeth, wash and dry hands, or get undressed without help
- ☐ Loses skills he once had

Tell your child's doctor or nurse if you notice any of these signs of possible developmental delay for this age, and talk with someone in your community who is familiar with services for young children in your area, such as your state's public early intervention program. For more information, go to www.cdc.gov/concerned or call 1-800-CDC-INFO.

www.cdc.gov/actearly | 1-800-CDC-INFO

Learn the Signs. Act Early.

APPENDIX B

RED FLAGS

First Signs ® Red Flags https:www.firstsigns.org.concerns/flags/htm

Red Flags

The following red flags may indicate a child is at risk for an autism spectrum disorder, and is in need of an immediate evaluation.

In clinical terms, there are a few "absolute indicators," often referred to as "red flags," that indicate that a child should be evaluated. For a parent, these are the "red flags" that you child should be screened to ensure that he/she is on the right developmental path.

Red flags of Autism Spectrum Disorders:

If your baby shows two or more of these signs, please ask your pediatric healthcare provider for an immediate evaluation.

Impairment in Social Interaction:

- Lack of appropriate eye gaze
- Lack of warm, joyful expressions
- Lack of sharing interest or enjoyment
- Lack of response to name

Impairment in Communication:

- Lack of showing gestures
- Lack of coordination of nonverbal communication
- Unusual prosody (little variation in pitch, odd intonation, irregular rhythm, unusual voice quality)

Repetitive Behaviors & Restricted Interests:

- Repetitive movements with objects
- Repetitive movements or posturing of body, arms, hands, or fingers

Wetherby, A., Woods, J. Allen, L., Cleary, J., Dickinson, H. & Lord, C. (2004) Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders*, 34, 473-493.

Based on research at the Florida State University FIRST WORDS ® Project.

"*Most Mommies and Daddies tell me "I thought there was a problem at 14 or 15 months . . . and they told me let's wait and see because sometimes some kids grow out of it." Well, that's not a good answer. We've got to make the distinction between less important problems, where we can wait and see, and core problems, which involve a lack of reciprocity and a lack of getting to know your world. For these core problems, we have to act on it yesterday. We can't wait nine months, we can't wait two months."* (Stanley I. Greenspan, MD, Child Psychiatrist)

APPENDIX C

CODING SHEET

Video ID# ______________________ Coder ________________________ Date ____________________

Responding					**Response to Name**				**Sounds/ Words**				**Play with objects and people,** **Stereotypy**				**Initiating**				**No opportunity**	**Notes**
response to context cues, response to joint attention expression of shared enjoyment									prosody, echolalia, lack of vocalization with consonant, pitch, volume								gaze, shared enjoyment or interest, gaze/sound/ expression/ gesture for Joint Attn, shows, points				Examiner not engaged	
T = = behavior not consistent with ASD **A** = = behavior consistent with ASD **U** = = not clear, unknown **blank** = none observed																						
0:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
0:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
0:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
0:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
0:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
1:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
1:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
1:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
1:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
1:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
1:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
2:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
2:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
2:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
2:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
2:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
2:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
3:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
3:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
3:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
3:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
3:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
3:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
4:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
4:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
4:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
4:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
4:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
4:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
5:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			

Responding					**Response to Name**				**Sounds/ Words**				**Play with objects and people, Stereotypy**				**Initiating**				**No opportunity**	**Notes**
response to context cues, response to joint attention expression of shared enjoyment									prosody, echolalia, lack of vocalization with consonant, pitch, volume								gaze, shared enjoyment or interest, gaze/sound/ expression/ gesture for Joint Attn, shows, points				Examiner not engaged	
5:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
5:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
5:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
5:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
5:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
6:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
6:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
6:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
6:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
6:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
6:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
7:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
7:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
7:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
7:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
7:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
7:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
8:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
8:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
8:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
8:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
8:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
8:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
9:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
9:10	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
9:20	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
9:30	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
9:40	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
9:50	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			
10:00	T	A	U		T	A	U		T	A	U		T	A	U		T	A	U			

Based on this observation alone, would you refer this child for an ASD evaluation?

YES/NO

APPENDIX D

ATYPICAL BEHAVIOR CODE SUMMARY

Atypical Behavior Code Summary

Grp	#	Responding		Resp. To Name		Sounds		Play		Initiating	
		Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
ASD	1	4	8	2	5	14	0	0	1	0	1
ASD	2	0	1	2	1	0	6	0	3	0	0
ASD	3	0	0	0	4	0	0	0	0	0	0
ASD	4	0	0	2	1	0	0	0	0	0	0
ASD	5	1	6	0	5	0	1	0	17	0	1
ASD	6	0	8	0	5	0	6	0	0	0	0
ASD	7	9	2	6	2	5	3	25	0	0	0
ASD	8	25	14	5	7	4	11	9	0	1	1
ASD	9	1	18	1	1	8	5	0	23	6	4
ASD	10	2	0	2	0	0	0	0	0	0	0
ASD	11	2[a]	2	1[a]	0	1[a]	0	0[a]	0	0[a]	0
ASD	12	0	0	0	0	0	1	0	0	0	0
ASD	13	1	3	2	3	0	1	0	0	0	0
ASD	14	4	2[a]	4	1[a]	14	15[a]	11	1[a]	4	2[a]
SSD	1	1	0	0	2	0	0	0	0	0	0
SSD	2	0	7	0	1	0	0	0	0	0	0
SSD	3	0	0	1	0	0	0	0	0	0	0
SSD	4	0	0	0	0	0	0	0	0	0	0
SSD	5	1	1	0	1	0	0	0	0	0	4
SSD	6	8	0	0	1	11	0	0	0	3	1
SSD	7	1	0	1	0	0	0	0	0	0	0
SSD	8	0	1	0	1	0	0	1	1	0	0
SSD	9	0	0	0	0	0	0	0	0	0	0
SSD	10	0	0	0	0	0	0	0	0	0	0
SSD	11	0	2	0	3	1	0	0	0	2	0
SSD	12	0	0	0	0	0	0	0	0	0	0
SSD	13	0	2	0	0	0	0	0	0	0	0
SSD	14	0	0	0	0	0	0	0	0	0	0
TDC	1	0	0	0	0	0	0	0	0	0	0
TDC	2	1	0	3	0	0	0	0	0	0	0
TDC	3	1	1	1	0	0	1	1	3	0	2
TDC	4	0	0	0	2	0	0	0	0	2	0
TDC	5	0	0	0	0	0	0	0	0	0	0
TDC	6	1	0	0	0	0	0	0	0	0	0
TDC	7	0	0	0	0	0	0	0	0	0	0
TDC	8	2	0	0	0	0	0	1	0	0	0

Grp	#	Responding		Resp. To Name		Sounds		Play		Initiating	
		Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
TDC	9	8	4	1	4	2	2	10	0	1	0
TDC	10	0	0	1	0	2	0	0	0	0	0
TDC	11	0	0	1	0	34	0	0	0	0	0
TDC	12	0	0	0	0	0	0	0	0	0	0
TDC	13	0	2	1	2	0	0	0	0	0	0
TDC	14	0	0	0	0	0	0	0	0	0	0
Total ASD		49	64	27	35	46	49	11	45	45	9
Total SSD		11	13	2	9	12	0	1	1	5	5
Total TDC		13	7	8	8	38	3	12	3	3	2
Totals		73	84	37	52	96	52	58	49	19	16

[a] excluded from analysis of overall atypical behavior, responding, play, and initiating due to excessive time off-camera.

APPENDIX E

TYPICAL BEHAVIOR CODE SUMMARY

Typical Behavior Code Summary

		Responding		Response to Name		Sounds		Play		Initiating	
Grp	#	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
ASD	1	17	26	1	6	5	0	56	51	3	6
ASD	2	45	34	3	1	9	20	42	31	37	26
ASD	3	41	24	2	5	1	1	60	56	22	13
ASD	4	33	44	2	2	43	35	59	38	13	28
ASD	5	29	29	1	7	3	9	50	22[c]	13	14
ASD	6	55	35	0	5	13	6	8	2	1	6
ASD	7	9	47	0	1	23	41	34[c]	53	2	31
ASD	8	9	16	1	2	4	7	46	46	5	0
ASD	9	32	27	0	1	16	23	44	11[c]	2	5
ASD	10	33	44	8	4	12	16	59	49	13	16
ASD	11	1[a]	43	0[a]	1	3[a]	25	20[a]	43	5[a]	14
ASD	12	48	42	7	0	29	44	48	29	14	10
ASD	13	38	31	3	4	51	48	59	42	17	22
ASD	14	35	17[a]	5	5[a]	22	22[a]	48	29[a]	14	12[a]
SSD	1	52	51	4	1	3	0	58	40	35	35
SSD	2	28	48	0	5	0	12	56	47	22	31
SSD	3	46	34	3	1	0	1	43	59	4	28
SSD	4	53	46	3	1	7	13	44	44	29	34
SSD	5	41	36	0	3	13	28	60	34	6	18
SSD	6	48	54	0	1	23	16	47	56	30	37
SSD	7	41	43	4	4	8	22	28	57	35	25
SSD	8	41	35	2	5	0	5	44	55	22	26
SSD	9	45	54	0	1	2	19	37	43	12	22
SSD	10	44	47	2	1	13	6	25	29	15	21
SSD	11	32	37	6	2	58	44	58	54	16	30
SSD	12	37	28	0	2	0	25	35	56	11	20
SSD	13	50	56	0	1	0	9	50	47	30	27
SSD	14	42	51	3	2	12	22	53	60	11	42
TDC	1	43	58	2	1	3	3	47	49	12	29
TDC	2	50	59	2	0	33	38	58	52	27	11
TDC	3	45	46	3	1	13	10	51	43	23	30
TDC	4	55	43	1	3	6	19	60	60	20	26
TDC	5	54	40	3	1	4	12	58	15	38	14
TDC	6	39	51	1	4	11	30	49	22	11	29
TDC	7	42	44	2	4	3	23	51	45	28	37
TDC	8	40	47	5	3	38	31	46	26	24	29
TDC	9	36	38	0	0	6	15	41	24	12	17
TDC	10	42	58	6	0	12	24	48	42	14	44
TDC	11	40	55	0	2	10	34	58	51	30	36
TDC	12	37	49	2	1	30	52	60	56	34	38

		Responding		Response to Name		Sounds		Play		Initiating	
Grp	#	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
TDC	13	36	43	6	3	31	18	60	40	26	22
TDC	14	31	39	4	4	5	19	54	50	13	6
Total ASD		425	459	33	44	234	297	633	502	161	203
Total SSD		600	620	27	30	139	222	638	681	278	396
Total TDC		590	670	37	27	205	328	741	575	312	368
Totals		1615	1749	97	101	578	847	2012	1758	751	967

[a] excluded from analysis of overall atypical behavior, responding, play, and initiating due to excessive time off-camera.

APPENDIX F

RATES AND RATIOS OF CODED BEHAVIORS – ATYPICAL

Rates and Ratios of Coded Behaviors – Atypical

Behavior	ASD				SSD				TDC			
	Total Count	% Total Codes	% Codable Behaviors	Ratio A:T	Total Count	% Total Codes	% Codable Behaviors	Ratio A:T	Total Count	% Total Codes	% Codable Behaviors	Ratio A:T
Responding	113	6.73	11.33	.13	24	1.43	1.93	.02	20	1.19	1.56	.02
Response to Name	62	3.69	44.60	.81	11	.65	16.18	.19	16	0.95	20.00	.25
Sounds and Words	95	5.65	15.18	.18	12	.71	3.22	.03	41	2.44	7.14	.08
Play	90	5.36	7.35	.08	2	.11	0.15	.00*	15	0.89	1.13	.01
Initiating	20	1.19	5.21	.06	10	.59	1.46	.02	5	0.30	0.73	.01
Total	380	4.52	11.3	.13	59	.70	1.6	.02	98	1.17	2.5	.03

*ratio = .002

APPENDIX G

RATES AND RATIOS OF CODED BEHAVIORS – TYPICAL

Rates and Ratios of Coded Behaviors- Typical

	ASD				SSD				TDC			
Behavior	Total Count	% Total	% Occur-rence	Ratio T:A	Total	% Total	% Occur -rence	Ratio T:A	Total	% Total	% Occur-rence	Ratio T:A
Responding	884	17.54	88.67	7.82	1,220	24.21	98.07	50.83	1,260	25.00	98.44	63.00
Response to Name	77	1.53	55.40	1.24	57	1.13	83.82	5.18	64	1.27	80.00	4.00
Sounds and Words	531	10.54	84.82	5.59	361	7.16	96.78	30.08	533	10.58	92.86	13.00
Play	1,135	22.52	92.65	12.61	1,319	26.17	99.85	659.50	1,316	26.11	98.87	87.73
Initiating	364	7.22	94.79	18.20	674	13.37	98.54	67.40	680	12.49	99.27	136.00
Total	2,991	35.6	88.72	7.87	3,631	43.22	98.40	61.54	3,853	45.83	97.54	39.72

REFERENCES

Ad Council (Producer) (2011). *Autism awareness.* Available from http://www.adcouncil.org/Impact/Case-Studies-Best-Practices/Autism-Awareness

Allison, C., Baron-Cohen, S., Wheelwright, S., Charman, T., Richler, J., Pasco, G., & Brayne, C. (2008). The Q-CHAT (Quantitative Checklist for Autism in Toddlers): A normally distributed quantitative measure of autistic traits at 18-24 months of age: Preliminary report. *Journal of Autism and Developmental Disorders 38,* 1414-1425. doi: 10.1007/s10803-007-0509-7

Allison,C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "red flags" for autism screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child and Adolescent Psychiatry 51* (2), 202-212. doi: 10.1016/j.jaac.2011.11.003

Al-Qabandi, M., Gorter, J.M., & Rosenbaum, P. (2011) Early Autism Detection: Are We Ready for Routine Screening? *Pediatrics 128* (1), e211-e217. doi: 10.1542/peds.2010-1881

American Academy of Pediatrics. (2003). Periodic Survey of Fellows #53: Identification of Children <36 months at risk for developmental problems and referral to early identification programs. Retrieved January 18, 2010 from http://www.aap.org/research/periodicsurvey/ps53exs.htm.

American Academy of Pediatrics, Rose, S. R., American Thyroid Association, Brown, R. S., & Lawson Wilkins Pediatric Endocrine Society (2006). Update of newborn screening and therapy for congenital hypothyroidism. *Pediatrics 117* 2290-2303 doi: 10.1542/peds.2006-0915

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edition, text revision). Washington, DC: Author.

Autism Speaks, Inc. (Producer) (2012). *Learn the signs--Ad Council campaign.* Available from http://www.autismspeaks.org/what-autism/learn-signs/ad-campaign.

Barbaro, J., Ridgway, L., & Dissanayake, C. (2011) Developmental surveillance of infants and toddlers by maternal and child health nurses in an Australian community-based setting: Promoting the early identification of autism spectrum disorders. *Journal of Pediatric Nursing 26,* 334-347.

Baron-Cohen, S., Wheelwright, S., Cox, A., Baird, G. Charman, T, Swettentham, J., Drew, A., & Doehring, P. (2010). Early identification of autism by the Checklist for Autism in Toddlers (CHAT). *Journal of the Royal Society for Medicine 93,* 521-535.

Barnette, J. J., & McLean, J. E. (2006). Confidence intervals of common effect sizes: What are they good for? San Francisco, CA: Annual Meeting of the American Educational Research Association.

Baird, G., Charman, T., Cox, A., Baron-Cohen, S., Swettenham, J., Wheelwright, S., & Drew, A. (2001). Screening and surveillance for autism and pervasive developmental disorders. *Archives of Disease in Childhood 84*, 468-475. doi:10.1136/adc.84.6.468

Bayley, N. (2005). *Bayley Scales of Infant and Toddler Development®, Third Edition.* San Antonio, TX: Pearson.

Begeer, S., El Bouk, S., Boussaid, W., Terwogt, M.M., & Koot, H.M.(2009) Underdiagnosis and referral bias of autism in ethnic minorities. *Journal of Autism and Developmental Disorders 39* (1), 142-148.

Bhasin, T.K., & Schendel, D. (2007). Sociodemographic risk factors for autism in a US metropolitan area. *Journal of Autism and Developmental Disorders 37*, 667-677.

Binns, H. J., LeBailly, S. A., Fingar, A. R., & Saunders, S. (1999). Evaluation of risk assessment questions used to target blood lead screening in Illinois. *Pediatrics, 103*(1), 100-106.

Byrt, T. (1996). How good is that agreement? *Epidemiology 7* (5), 561.

Caglayan, A.O. (2010). Genetic causes of syndromic and non-syndromic autism. *Dev Med Child Neurol.* [Epub ahead of print] PMID: 20059518

Centers for Disease Control and Prevention (2007). Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002. *Mortality and Morbidity Weekly Report, 56* (SS-1) 12-28.

Centers for Disease Control and Prevention (2009). Centers for Disease Control and Prevention: Prevalence of Autism Spectrum Disorders — Autism and Developmental Disabilities Monitoring Network, United States, 2006. Surveillance Summaries. *Mortality and Morbidity Weekly Report, 58* (No. SS-10).

Centers for Disease Control and Prevention (2010). Autism spectrum disorders (ASDs): Treatments. Retrieved January 23, 2011 from http://www.cdc.gov/ncbddd/autism/treatment.html

Centers for Disease Control and Prevention (2010). Learn the Signs. Act Early. Retrieved January 24, 2012 from http://www.cdc.gov/ncbddd/actearly/

Centers for Disease Control and Prevention (2012). Prevalence of autism spectrum disorders – Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *Mortality and Morbidity Weekly Report, 61* (SS - 3), 1-19.

Chawarska, K., Paul, R., Klin, A., Hannigen, S., Dichtel, L.E., & Volkmar, F. (2007). Parental recognition of developmental problems in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37* (1), 62-72.

Christensen, L, Hutman, T., Rozga, A, Young, G.S., Ozonoff, S., Rogers, S.J., . . . Sigman, M. (2010). Play and developmental outcomes in infant siblings of children with autism. *Journal of Autism and Developmental Disorders, 40*, 946-957. doi 10.1007/s10803-010-0941-y

Cohen, D., Pichard, N., Tordjman, S., Baumann, C., Burglen, L. Excoffier, E., Lazar, G., . . . Heron, D. (2005). Specific genetic disorders and autism: Clinical contribution towards their identification. *Journal of Autism and Developmental Disorders 35* (1), 103-116. doi: 10.1007/s10803-004-1038-2

Combating Autism Reauthorization Act of 2011, Public Law 112-32, 112th Congress, 125 Stat. 361 (2011).

Comeau, A.M. et al. (2004) Population-based newborn screening for genetic disorders when multiple mutation DNA testing is incorporated: A cystic fibrosis newborn screening model demonstrating increased sensitivity but more carrier detections. *Pediatrics 113*, 1573-1581

Cuccaro, M.L., Wright, H.H., Rownd, C.V., Abramson, R. K., Waller, J., & Fender, D. (1996). Brief report: Professional perceptions of children with developmental difficulties: The influence of race and socioeconomic status. *Journal of Autism and Developmental Disorders. 26*(4):461-469.

Cunningham, M. (2009). More than just kappa coefficient: A program to fully characterize inter-rater reliability between two raters. *SAS Global Forum, Statistics and Data Analysis.* Retrieved from *support.sas.com/resources/papers/proceedings09/242-2009.pdf.*

Daniel, K.L., Prue, C., Taylor, M.K., Thomas, J., & Scales, M. (2009). "Learn the signs. Act early.": A campaign to help every child reach his or her full potential. *Public Health 123*(Suppl 1), e11-e16.

Dawson, G., Meltzoff, A.N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient naturally occurring social stimuli. *Journal of Autism and Developmental Disorders 28* (6), 479-485.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., et al. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology 40* (2), 271-283.

Dosreis, S., Weiner, C. L., Johnson, L., & Newschaffer, C.J., (2006). Autism spectrum disorder screening and management practices among general pediatric providers. *Developmental and Behavioral Pediatrics 27* (2), S88-S94.

Dover, C., & LeCouteur, A. L. (2007). How to diagnose autism. *Archives of Disease in Childhood 92* 540-545.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice 40*(5), 532-538.

Filipek, P.A., Accardo, P.J., Baranek, G.T., Cook, E.H., Dawson, G., Gordeon, B., Gravel, J.S., Johnson, C.P., . . . Volkmar, F.R. (1999). The screening and diagnosis of autism spectrum disorders. *Journal of Autism and Developmental Disorders 29*(6), 439-484. doi: 10.1023/A:1021943802493

First Signs, Inc. (2012). Red flags. Retrieved January 24, 2012 from https://www.firstsigns.org/concerns/flags.htm

Ganz, M.L. (2007). The lifetime distribution of the incremental societal costs of autism. *Archives of Pediatric & Adolescent Medicine 161,* 343-349.

Gomot, M., Belmonte, M.K., Bullmore, E.T., Bernard, F.A., & Baron-Cohen, S. (2008). Brain hyper-reactivity to auditory novel targets in children with high-functioning autism. *Brain 131*, 2479-2488.

Grosse, S.D. et al. (2004) Newborn screening for cystic fibrosis: Evaluation of benefits and risks and recommendations for state newborn screening programs. *Mortality and Morbidity Weekly Report 53* (RR13);1-36. Retrieved February 12, 2012 from http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5313a1.htm

Grove, W.M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology 61* (10), 1233-1243. doi: 10.1002/jclp.20179

Gupta, V. B., Hyman, S. l., Johnson, C. P., Bryant, J., Byers, B, Kallen, R. et al. (2007). Identifying children with autism early? *Pediatrics 119* (1) 152-153.

Halfon, J., Stevens, G.D., Larson, K., & Olson, L.M. (2011). Duration of a well child visit: Association with content, family-centeredness, and satisfaction. *Pediatrics 128* (4), 657-664. doi: 10.1542/peds.2011-0586

Harrop, A., & Daniels, M. (1986). Methods of time sampling: A reappraisal of momentary time sampling and partial interval recording. *Journal of Applied Behavior Analysis 19*, 73-77.

Harrop, A., Daniels, M., & Foulkes, C. (1990). The use of momentary time sampling and partial interval recording in behavioral research. *Behavioural Psychotherapy 18*, 121-127.

Horner, R., Carr, E., Strain, P., Todd, A., & Reed, H. (2002). Problem behavior interventions for young children with autism: A research synthesis. *Journal of Autism and Developmental Disorders, 32*(5), 423-446.

Interagency Autism Coordinating Committee (Department of Health and Human Services). (2010). The 2010 strategic plan for autism spectrum disorder research. Retrieved March 6, 2010 from http://iacc.hhs.gov/strategic-plan/.

Jacobson, J.W., Mulick, J.A., & Green, G. (1998). Autism and ABA.*Behavioral Interventions*, 13, 201-226.

Johnson, C. P., Myers, S. M., & Council on Children with Disabilities, (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120* (5) 1183-1215.

Keehn, B. & Joseph, R.M. (2008). Impaired prioritization of novel onset stimuli in autism spectrum disorder. *Journal of Child Psychology and Psychiatry 49* (12), 1296-1303.

Keren, R., Helfand, M., Homer, C., McPhillips, H., & Lieu, T. A. (2002). Projected cost-effectiveness of statewide universal newborn hearing screening. *Pediatrics, 110*(5), 855-864.

Kim, Y.S., Leventhal, B.L., Koh, Y.,J., Fombonne, E., Laska, E.,, Lim, F.C., . . . Grinker, R. R. (2011) Prevalence of autism spectrum disorders in a total population sample. *The American Journal of Psychiatry168*(9), 904-912, doi: 10.1176/appi.ajp.2011.10101532

Knapp M., Romeo R., & Beecham, J. (2009). Economic cost of autism in the UK. *Autism 13,* 317-336.

Kogan, M.D., Blumberg, S.J., Scheive, L.A., Boyle, C.A., Perrin, J.M., Ghandour, R.M., et al. (2009). Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics 124* (5), 1395-1403.

Landry, R. & Bryson, S.E. (2004) Impaired disengagement of attention in young children with autism. *Journal of Child Psychology and Psychiatry 45* (6), 1115-1122.

Lantz, C.A., & Nebenzahl, E. (1996). Behavior and interpretation of the κ statistic: Resolution of two paradoxes. *Journal of Clinical Epidemiology 49* (4), 431-434.

LeCouteur, A., Lord, C., & Rutter, M. (2003). *The Autism Diagnostic Interview-Revised.* Los Angeles: Western Psychological Services.

Limperopoulos, C. (2009). Autism spectrum disorders in survivors of extreme prematurity. *Clinical Perinatology 36* , 791-805. doi:10.1016/j.clp.2009.07.010

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurn, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry 63* 694-701.

Lord, C., Rutter, M, DiLavore, P. S., & Risi, S., (1999) *Autism Diagnostic Observation Schedule (ADOS).* Los Angeles: Western Psychological Services.

Lord, C., Rutter, M., DiLavore, P. S., Risi, S., (in press) *Autism Diagnostic Observation Schedule Toddler Research Module (ADOS-T), Prepublication Research Version.* Los Angeles: Western Psychological Services.

Ludbrook, J. (2002). Statistical techniques for comparing measurers and methods of measurement: A critical review. *Clinical and Experimental Pharmacology and Physiology 29*, 527–536.

Luyster, R., Gotham, K., Guthrie, W., Doffing, M., Petrak, R., Pierce, K. . . . Lord, C. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders 39*, 1305-1320. doi: 10.1007/s10803-009-0746-z

Mandell, D.S., Ittenbach, R.F., Levy, S.E., & Pinto-Martin, J.A. (2007). Disparities in diagnoses received prior to a diagnosis of autism spectrum disorder. *Journal of Autism and Developmental Disorders 37,* 1795-1802.

Mandell, D.S., Morales, K.H., Xie, M., Lawer, L., Stahmer, A.C., & Marcus, S.C. (2010). Age of diagnosis in Medicaid-enrolled children with autism, 2001-2004. *Psychiatric Services. 61*(8), 822-829. doi: 10.1176/appi.ps.61.8.822

Mandell, D.S., Wiggins, L.D., Yeargin-Allsopp, M., Carpenter, L.A., Arnstein, L., Daniels, J., et al. (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health, 99* (3), 493-498.

McEachin, J., Smith, T., & Lovaas, I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal of Mental Retardation, 97,* 359-372.

Meany-Daboul, M.G., Roscoe, E. M., Bourret, J.C., & Ahearn, W.H. (2007). A comparison of momentary time sampling and partial interval recording for evaluating functional relations. *Journal of Applied Behavioral Analysis 40* (3), 501-514.

Miller, J.S., Gabrielsen, T., Villalobos, M., Alleman, R., Wahmhoff, N., Carbone, P.S., & Segura, B. (2011). The Each Child Study: Systematic screening for autism spectrum disorders in a pediatric setting. *Pediatrics 127,* 866-871. doi: 10.1542/peds.2010-0136

Mullen, E. M. (1995). *Mullen Scales of Early Learning.* Bloomington, MN: Pearson

National Dissemination Center for Children with Disabilities. (2011). Overview of early intervention. Retrieved January 24, 2011 from http://nichcy.org/babies/overview

Nadel, S., & Poss, J. E. (2007). Early detection of autism spectrum disorders: Screening between 12 and 24 months of age. *Journal of the American Academy of Nurse Practitioners 19* 408-417.

Nadig A.S., Ozonoff S., Young G.S., Rozga A., Sigman M., & Rogers, S.J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of Pediatrics & Adolescent Medicine 161* (4), 378-83.

National Autism Center (2009). *National Autism Center's National Standards Report.* Randolph, MA: National Autism Center

Nordahl, C.W., Lange, N., Li, D.D., Barnett, L.A., Lee, A., Buonocore, M.H., Simon, T.J., Rogers, S.R., Ozonoff, S., & Amaral, D.G. (2011). Brain enlargement is associated with regression in preschool-age boys with autism spectrum disorders. *Proceedings of the National Academy of Sciences 108* (50), 20195-20200. doi: 10.1073/pnas.1107560108

Oosterling, I.J., Wensing, M., Swinkels, S.H., van der Gaag, R.J., Visser, J.C., Woudenberg, T., Minderaa, R. . . . Buitelaar, J.K. (2010). Advancing early detection of autism spectrum disorder by applying an integrated two-stage screening approach. *Journal of Child Psychology and Psychiatry 51* (3), 250-258. doi:10.1111/j.1469-7610.2009.02150.x

Ozonoff, S., Iosif, A., Baguio, F., Cook, I.C., Hill, M.M., Hutman, T., Rogers, S. . . . Young, G.S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry 49* (3), 256-266.

Ozonoff, S., Iosif, A., Young, G.S., Hepburn, S., Thompson, M., Colombi, C., Cook, I.C., . . . Rogers, S.J. (2011). Onset patterns in autism: Correspondence between home video and parent report. *Journal of the American Academy of Child & Adolescent Psychiatry 50* (8), 796-806. doi: 10.1016/j.jaac.2011.03.012

Pandey, J. (2008). The modified checklist for autism in toddlers (M-CHAT) sibling study: Are younger siblings representative of the general ASD population*? Dissertation Abstracts International: Section B: The Sciences and Engineering, Vol 68*(8-B), pp. 5586

Papanikolaou, K., Paliokosta, E., Houliaras, G., Vgenopoulou, S., Giouroukou, E., Pehlivanidis, A., et al. (2009). Using the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule-Generic for the diagnosis of autism spectrum disorders in a Greek sample with a wide range of intellectual abilities. *Journal of Autism and Developmental Disorders 39*, 414-420.

Peacock, G., & Lin, S.C. (2012). Enhancing early identification and coordination of intervention services for young children with autism spectrum disorders: Report from the Act Early Regional Summit Project. *Disability Health Journal 5* (1), 55-59.

Pinborough-Zimmerman, J., Bakian, A., Fombonne, E., Bilder, D., Taylor, J., & McMahon, W. (2011). Changes in the administrative prevalence of autism spectrum disorders: Contribution of special education and health, 2002-2008. *Journal of Autism and Developmental Disabilities.* doi: 10.1007/s10803-011-1265-2

Pinto-Martin, J. A., Dunkle, M., Earls, M., Fliedner, D., & Landes, C. (2005). Developmental stages of developmental screening: Steps to implementation of a successful program. *American Journal of Public Health 95* (11) 1928-1932.

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998) Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *Journal of Child Psychology and Psychiatry 39* (5), 765-775.

Radecki, L., Sand-Loud, N., O'Connor, K. G., Sharp, S., & Olson, L.M. (2011). Trends in the use of standardized tools for developmental screening in early childhood: 2002 – 2009. *Pediatrics 128* (1), 14-19. doi: 10.1542/peds.2010-2180

Ray-Subramianian, C.E., Huai, N., & Weismer, S.E. (2011). Brief report: Adaptive behavior and cognitive skills for toddlers on the autism spectrum. *Journal of Autism and Developmental Disabilities 41*(5), 679-684.

Repp, A.C., Deitz, D.E.D., Boles, S.M., Deitz, S.M., & Repp, C.F. (1976). Technical article: Differences among common methods for calculating interobserver agreement. *Journal of Applied Behavior Analysis 9*, 109-113.

Robins, D. L., & Dumont-Mathieu, T. M. (2006) Early screening for autism spectrum disorders: Update on the Modified Checklist for Autism in Toddlers and other measures. *Developmental and Behavioral Pediatrics* 27 (2) S111-S119.

Robins, D. L., Fein, D., Barton, M., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*, 131-151.

Robins, D., (2008). Screening for autism spectrum disorders in primary care settings. *Autism 12*(5), 537 – 556.

Rogers, S. J. (1996). Brief Report: Early Intervention in Autism. *Journal of Autism and Development Disorders 26* (2), 243-24.

Rogers, S.J. (2009). What are infant siblings teaching us about autism in infancy? *Autism Research 1,* 1-13.

Rogers, S.J., & Vismara, L.A. (2008). The Early Start Denver Model: A case study of innovative practice. *Journal of Early Intervention 31*(1), 93-115.

Runco, M.A.., Charlop, M.H., & Screibman, L. (1986). The occurrence of autistic children's self-stimulation as a function of familiar versus unfamiliar stimulus conditions. *Journal of Autism and Developmental Disabilities 16* (1), 31-44.

Schertz, H.H., Baker, C., Hurwitz, S., & Benner, L. (2011). Principles of early intervention reflected in toddler research in autism spectrum disorders. *Topics in Early Childhood Special Education 31*(4) 4-21. doi: 10.1177/0271121410382460

Schreibman, L. (2000). Intensive behavioral/psychoeducational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disabilities 30*, 373-378.

Schwartz, I.S., & Sandall, S.R. (2010). Is autism the disability that breaks Part C? A commentary on "Infants and toddlers with autism spectrum disorder; Early identification and early intervention," by Boyd, Odom, Humphreys, and Sam. *Journal of Early Intervention 32* (2), 105-109. doi: 10.1177/1053815110366698

Sell, N.K., Giarelli, E., Blum, N., Hanlon, A.L., & Levy, S.E. (2012). A comparison of Autism Spectrum Disorder DSM-IV criteria and associated features among African American and white children in Philadelphia County. *Disability and Health Journal 5* (1), 9-17. doi:10.1016/j.dhjo.2011.08.002

Shattuck, P.T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D.S., Wiggins, L. . . . Cuniff, C. (2009). Timing of identification among children with an autism spectrum disorder: Findings from a population based surveillance study. *Journal of the American Academy of Child and Adolescent Psychiatry 48* (5), 474-483. doi: 10.1097/CHI.0b013e31819b3848

Shumway, S., Thurm, A. Swedo, S.A., Deprey, L., Barnett, L.A., Amaral, D.G., Rogers, S. & Ozonoff, S. (2011). Brief report: Symptom onset patterns and functional outcomes in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders 41,* 1727 – 1732.

Sokhadze, E., Baruth, J., Tasman, A., Sears, L. Mathai, G., El-Baz, A., & Casanova, M.F. (2009). Event-related potential study of novelty processing abnormalities in autism. *Applied PsychophysiologicalBiofeedback 34*, 37–51 doi: 10.1007/s10484-009-9074-5

Smedley, B.D., Stith, A.Y., & Nelson, A.R. (Eds.), Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care, Board of Health Sciences Policy, & National Institute of Medicine of the National Academies (2002). *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* Washington, DC: The National Academies Press.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. (2005). *Vineland Adaptive Behavior Scales, Second Edition, Survey Forms Manual.* Minneapolis, MN: Pearson, Inc.

Stahmer, A.C., Shreibman, L., & Cunningham, A.B. (2010). Toward a technology of treatment individualization for young children with autism spectrum disorders. *Brain Research 1380,* 229-239. doi:10.1016/j.brainres.2010.09.043

Stone, W.S., Coonrod, E.E., & Ousley, O.Y. (2000). Brief report: Screening tool for autism in two-year-olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders, 30* (6), 607-612.

Sumi, S., Taniai, H., Miyachi, T., & Tanemura, M. (2006). Sibling risk of pervasive developmental disorder estimated by means of an epidemiologic survey in Nagoya, Japan. *Journal of Human Genetics. 51* (6), 518- 522.

Szatmari, P., Jones, M., Zwaigenbaum, L., & MacLean, J. (1998). Genetic epidemiology of autism spectrum disorders: Issues in matching. *Journal of Autism and Developmental Disorders, 28,* 351-368.

Uebersax, J. (2009). Raw agreement indices. Retrieved March 29, 2012 from from http://www.john-uebersax.com/stat/raw.htm#binspe

Utah Baby Watch Early Intervention Program (2008). Annual report certification of the interagency coordinating council under Part C of the Individuals with Disabilities Education Act (IDEA). Retrieved March 9, 2009 from http://www. utahbabywatch.org/PDF/apr0708.pdf.

Utah Baby Watch Early Intervention Program (2010). Individuals with Disabilities Education Act Part C of the (IDEA). FFY 2009-2010 Annual Performance Report. Retrieved January 23, 2011 from http://www.utahbabywatch.org/publicinformation/apr.htm

United States Department of Education. (2004). Building the legacy: IDEA 2004. Retrieved January 23, 2012 at http://idea.ed.gov/.

Viera, A.J. & Garrett, J.M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine 37* (5), 360-363.

Vismara, L.A., Columbi, C., & Rogers, S.J. (2009) Can one hour per week of therapy lead to lasting changes in young children with autism? *Autism 13*(1), 93-115.

Vivanti, G., Nadig, A., Ozonoff, S., & Rogers, S.J. (2008) What do children with autism attend to during imitation tasks? *Journal of Experimental Child Psychology 101*, 186-205.

van Daalen, E., Kemner, C., Dietz, C., Swinkels, S., Buitelaar, J., & van Engeland, H. (2009). Inter-rater reliability and stability of diagnoses of autism spectrum disorder in children identified through screening at a very young age. *European Child and Adolescent Psychiatry 18* (11), 663-674. doi: 10.1007/s00787-009-0025-8

Warren, Z., McPheeters, M.L., Sathe, N., Foss-Feig, J., Glasser, A., & Veenstra-VanderWeele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics 127*, e1303-e1311. doi: 10.1542/peds.2011-0426

Warren, Z., Stone, W., & Humberd, Q. (2009). A training model for the diagnosis of autism in community pediatric practice. *Journal of Developmental and Behavioral Pediatrics, 30,* 442-446.

Watkins M.W., & Pacheco, M. (2000). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education 10* (4), 205-212.

Wetherby, A., & Prizant, B. (2002) *Communication and symbolic behavior scales developmental profile—first normed edition.* Baltimore, MD: Paul H. Brookes.

Wetherby, A., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life*. Journal of Autism and Developmental Disorders,* 34, 473-493.

Wiggins, L., Baio, J., & Rice, C. (2006). Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *Journal of Developmental and Behavioral Pediatrics, 27* (2) S79-S87.

Wolff, J.J., Gu, H., Gerig, G., Elison, J.T., Styner, M., Gouttard, S., Botteron, K., . . . Piven, J. (2012). Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. *American Journal of Psychiatry,* published online in advance of publication February 17, 2012. *AiA,* 1-12. doi: 10.1176/appi.ajp.2011.11091447

Wong, V., Hui, L.S., Lee, W., Leung, L.J., Ho, P.P., Lau, W.C., . . . Chung, B. (2004). A modified screening tool for autism (Checklist for Autism in Toddlers [CHAT-23] for Chinese children. *Pediatrics 114* (2), e166-e176.

Yirmiya, N., & Ozonoff, S. (2007). The very early autism phenotype. *Journal of Autism and Developmental Disorders 37*, 1-11.

Yirmiya, N., & Charman, T. (2010). The prodrome of autism: Early behavioral and biological signs, regression, peri-and post-natal development and genetics. *Journal of Child Psychology and Psychiatry, 51* (4), 432-458. doi: doi:10.1111/j.1469-7610.2010.02214.x

Young, G.S., Merin, N., Rogers, S.J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants ad infants at risk for autism. *Developmental Science 12* (5), 798-814. doi: 10.1111/j.1467-7687.2009.00833.x

Zeiger, V.M. (2009) Screening for autism spectrum disorders: Pediatric practices eight years after publication of practice guidelines. *Dissertation Abstracts International: Section B; The Sciences and Engineering, 69* (7-B), 4463.

Zwaigenbaum, L., Bryson, S., Lord, C., Rogers, S., Carter, A., Carver, L. et al. (2009) Clinical assessment and management of toddlers with suspected autism spectrum disorder: Insights from studies of high-risk infants. *Pediatrics, 123*, 1383-1391