

**Georgia State University**  
**ScholarWorks @ Georgia State University**

---

Psychology Theses

Department of Psychology

---

Spring 5-9-2015

# Identifying Predictors of Diagnostic Instability of Autism Spectrum Disorder and Global Developmental Delay In Toddlers

Danielle N. Abrams  
*Georgia State University*

Diana L. Robins  
*Drexel University*

Lauren B. Adamson  
*Georgia State University*

Christopher Henrich  
*Georgia State University*

Follow this and additional works at: [https://scholarworks.gsu.edu/psych\\_theses](https://scholarworks.gsu.edu/psych_theses)

---

## Recommended Citation

Abrams, Danielle N.; Robins, Diana L.; Adamson, Lauren B.; and Henrich, Christopher, "Identifying Predictors of Diagnostic Instability of Autism Spectrum Disorder and Global Developmental Delay In Toddlers." Thesis, Georgia State University, 2015. [https://scholarworks.gsu.edu/psych\\_theses/135](https://scholarworks.gsu.edu/psych_theses/135)

This Thesis is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

IDENTIFYING PREDICTORS OF DIAGNOSTIC INSTABILITY OF AUTISM SPECTRUM DISORDER AND GLOBAL  
DEVELOPMENTAL DELAY IN TODDLERS

by

DANIELLE N. ABRAMS

Under the direction of Committee Co-chairs:

Diana L. Robins, PhD and Lauren B. Adamson, PhD

**ABSTRACT**

Although Autism Spectrum Disorder (ASD) is considered to be a lifelong condition, some toddlers experience diagnostic instability over time. In particular, some toddlers' diagnosis changes between ASD and Global Developmental Delay (GDD). However, little is known about the subset of children who change diagnosis. In a total of 424 toddlers who either maintained or changed diagnosis, the current study identified predictors of change in diagnosis and severity in those who change from ASD to non-ASD (ASD-NON), ASD to GDD (ASD-GDD), non-ASD to ASD (NON-ASD), and GDD to ASD (GDD-ASD) between two years old and four years old. Initial ASD symptom severity and participation in intervention services were predictive of all transitions. Additionally, receptive language predicted ASD-NON transition and socioeconomic status predicted ASD-GDD transition. Implications for informing

prognosis of children, identifying targets of intervention, refining of screening and diagnostic measures, and measuring change in severity regardless of categorical change are discussed.

INDEX WORDS: Diagnostic stability, Early diagnosis, Child development, Developmental delay, Symptom severity, Intervention

IDENTIFYING PREDICTORS OF DIAGNOSTIC INSTABILITY OF AUTISM SPECTRUM DISORDER AND GLOBAL  
DEVELOPMENTAL DELAY IN TODDLERS

by

DANIELLE N. ABRAMS

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2015

Copyright by  
Danielle Nicole Abrams  
2015

IDENTIFYING PREDICTORS OF DIAGNOSTIC INSTABILITY OF AUTISM SPECTRUM DISORDER AND  
GLOBAL DEVELOPMENTAL DELAY IN TODDLERS

by

DANIELLE N. ABRAMS

Committee Co-Chairs: Diana L. Robins

Lauren B. Adamson

Committee: Christopher Henrich

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2015

**DEDICATION**

Thank you to Dane, for your endless support and constant willingness to talk stats with me.

Thank you also to my family for your unwavering encouragement.

## **ACKNOWLEDGEMENTS**

Thank you to Diana Robins, Lauren Adamson, and Chris Henrich for supporting me in completion of this project, and to all of the families who kindly gave their time for this research.



## TABLE OF CONTENTS

|   |                  |
|---|------------------|
| <b>ACKNOWLEDGEMENTS .....</b>   | <b>v</b>         |
| <b>LIST OF TABLES .....</b>   | <b>x</b>         |
| <b>LIST OF FIGURES.....</b>   | <b>xii</b>       |
| <b>1 INTRODUCTION .....</b>   | <b>1</b>         |
| <b>1.1 Autism Spectrum Disorder.....</b>  | <b>3</b>         |
| <b>1.2 Rates of Diagnostic Stability of ASD .....</b>                                   | <b>4</b>         |
| <b>1.3 Importance of Accurate Early Diagnosis of ASD and Developmental Delays .....</b> | <b>7</b>         |
| <b>1.4 Change in ASD Symptom Presentation Over Time .....</b>                           | <b>9</b>         |
| <b>1.5 Factors Contributing to Change in Symptom Presentation .....</b>                 | <b>11</b>        |
| <b>1.6 Predictors of Diagnostic Instability of ASD.....</b>                             | <b>12</b>        |
| <b><i>1.6.1 ASD symptom severity.....</i></b>   | <b><i>13</i></b> |
| <b><i>1.6.2 Communication ability.....</i></b>  | <b><i>14</i></b> |
| <b><i>1.6.3 Cognitive ability .....</i></b>   | <b><i>15</i></b> |
| <b><i>1.6.4 Demographic factors .....</i></b>   | <b><i>16</i></b> |
| <b><i>1.6.5 Intervention .....</i></b>  | <b><i>16</i></b> |
| <b><i>1.6.6 Other factors.....</i></b>  | <b><i>17</i></b> |
| <b>1.7 Global Developmental Delay .....</b>   | <b>17</b>        |
| <b>1.8 GDD Stability and Outcomes.....</b>  | <b>18</b>        |
| <b>1.9 ASD compared to GDD .....</b>  | <b>20</b>        |

|        |  |    |
|--------|--|----|
| 1.9.1  | <i>Social abilities</i>  | 21 |
| 1.9.2  | <i>Communication skills</i>  | 22 |
| 1.9.3  | <i>Motor and sensory behaviors</i>   | 22 |
| 1.9.4  | <i>Cognitive and adaptive skills</i>   | 23 |
| 1.9.5  | <i>Developmental milestone attainment</i>                                    | 23 |
| 1.9.6  | <i>Factors contributing to difficult differentiation between ASD and GDD</i> | 24 |
| 1.10   | <i>Current Study</i>   | 24 |
| 1.11   | <i>Hypotheses</i>  | 27 |
| 1.11.1 | <i>Hypothesis 1A: Diagnostic transition from ASD to non-ASD</i>              | 27 |
| 1.11.2 | <i>Hypothesis 1B: Diagnostic transition from ASD to GDD</i>                  | 27 |
| 1.11.3 | <i>Hypothesis 2A: Diagnostic transition from non-ASD to ASD</i>              | 28 |
| 1.11.4 | <i>Hypothesis 2B: Diagnostic transition from GDD to ASD</i>                  | 29 |
| 2      | <b>METHODS</b>   | 33 |
| 2.1    | <b>Participants</b>  | 33 |
| 2.2    | <b>Measures</b>  | 34 |
| 2.2.1  | <i>Autism Diagnostic Observation Schedule</i>                                | 34 |
| 2.2.2  | <i>Mullen Scales of Early Learning</i>                                       | 36 |
| 2.2.3  | <i>Vineland Adaptive Behavior Scales (Second Edition)</i>                    | 37 |
| 2.2.4  | <i>Demographics and History Forms</i>  | 38 |
| 2.2.5  | <i>Intervention Data</i>   | 38 |

|       |  |    |
|-------|--|----|
| 2.3   | Procedure .....  | 39 |
| 3     | RESULTS .....  | 44 |
| 3.1   | Predictors of diagnostic instability .....                           | 45 |
| 3.1.1 | <i>Hypothesis 1A: ASD to non-ASD</i> .....                           | 45 |
| 3.1.2 | <i>Hypothesis 1B: ASD to GDD</i> .....                               | 47 |
| 3.1.3 | <i>Hypothesis 2A: non-ASD to ASD</i> .....                           | 47 |
| 3.1.4 | <i>Hypothesis 2B: GDD to ASD</i> .....                               | 48 |
| 3.2   | Predictors of change in continuous ASD symptom severity .....        | 49 |
| 3.2.1 | <i>Hypothesis 1A: ASD to non-ASD</i> .....                           | 49 |
| 3.2.2 | <i>Hypothesis 1B: ASD to GDD</i> .....                               | 50 |
| 3.2.3 | <i>Hypothesis 2A: non-ASD to ASD</i> .....                           | 50 |
| 3.3   | Study Attrition .....  | 51 |
| 3.4   | Site Comparison .....  | 51 |
| 4     | DISCUSSION.....  | 67 |
| 4.1   | Diagnostic Stability.....  | 67 |
| 4.2   | Hypothesis 1A: ASD to non-ASD.....                                   | 69 |
| 4.3   | Hypothesis 1B: ASD to GDD .....                                      | 74 |
| 4.4   | Hypothesis 2A: non-ASD to ASD.....                                   | 80 |
| 4.5   | Hypothesis 2B: GDD to ASD .....                                      | 84 |
| 4.6   | Diagnostic change versus change in continuous symptom severity ..... | 85 |

|            |                                    |           |
|------------|------------------------------------|-----------|
| <b>4.7</b> | <b>DSM-IV-TR versus DSM-5.....</b> | <b>86</b> |
| <b>4.8</b> | <b>Limitations.....</b>            | <b>88</b> |
| <b>4.9</b> | <b>Implications.....</b>           | <b>90</b> |
|            | <b>REFERENCES.....</b>             | <b>94</b> |

## LIST OF TABLES

|  |    |
|--|----|
| Table 1. Methodological summary of studies examining diagnostic stability of ASD in toddlers .....   | 30 |
| Table 2. Summary of results of studies examining predictors of diagnostic instability of ASD.....  | 31 |
| Table 3. Sample characteristics by T1 and T2 diagnostic category.....  | 41 |
| Table 4. Number of participants and hours of services for those who received intervention between T1<br>and T2 based on diagnostic categories at both time points..... | 52 |
| Table 5. Correlations for continuous T1 variables in those with ASD at T1 and ASD or GDD at T2 (for<br>Hypothesis 1A) .....  | 53 |
| Table 6. Contingency coefficients between categorical T1 variables and diagnostic transition outcome.  | 53 |
| Table 7. Logistic regression results for final model of hypothesis 1A (predictors of transition from ASD to<br>non-ASD) .....  | 53 |
| Table 8. Correlations for continuous T1 variables in those with ASD at T1 and GDD at T2 (for Hypothesis<br>1B).....  | 54 |
| Table 9. Logistic regression results for final model of hypothesis 1B (ASD-GDD) .....  | 54 |
| Table 10. Correlations for continuous T1 variables in those without ASD at T1 and ASD at T2 (for<br>Hypothesis 2A) .....   | 55 |
| Table 11. Logistic regression results for final model of hypothesis 2A (NON-ASD).....  | 55 |
| Table 12. Linear regression results corresponding to final model of hypothesis 1A (predictors of ASD<br>symptom severity change in ASD-ASD and ASD-NON) .....          | 55 |
| Table 13. Linear regression results corresponding to final model of hypothesis 1B (predictors of ASD<br>symptom severity change in ASD-ASD and ASD-GDD) .....          | 56 |
| Table 14. Linear regression results using bootstrapping corresponding to final model of hypothesis 2A<br>(NON-NON and NON-ASD).....                                    | 56 |

Table 15. Descriptive statistics and group comparisons in those who remained in the study and those  
who dropped out ..... 56

## LIST OF FIGURES

|   |    |
|---|----|
| Figure 1. The percentage of studies that have found specific rates of diagnostic stability of ASD from toddlerhood .....  | 32 |
| Figure 2. Visual representations of sample sizes of each group of participants .....  | 43 |
| Figure 3. Mean number of intervention hours in six-month intervals by group membership in those with T1 ASD .....   | 57 |
| Figure 4. Mean number of intervention hours in six-month intervals by group membership in those with T1 ASD and T2 ASD or GDD.....  | 58 |
| Figure 5. Mean number of intervention hours in six-month intervals by group membership in those without T1 ASD .....  | 59 |
| Figure 6. Scatterplot of T1 and T2 ASD symptom severity in participants who maintained GDD diagnosis and who transitioned from GDD to ASD.....  | 60 |
| Figure 7. Scatterplot of T1 ASD symptom severity and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD .....                   | 61 |
| Figure 8. Scatterplot of T1 receptive language ability and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD .....             | 62 |
| Figure 9. Scatterplot of T1 expressive language ability and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD .....            | 63 |
| Figure 10. Scatterplot of T1 parent reported communication ability and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD ..... | 64 |
| Figure 11. Scatterplot of T1 parent reported socialization skills and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD.....   | 65 |

Figure 12. Scatterplot of T1 nonverbal problem-solving ability and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD ..... 66



## 1 INTRODUCTION

Autism Spectrum Disorder (ASD) has recently been gaining attention in the media due to rising prevalence rates. With ASD diagnosis becoming increasingly common, there has been a rising interest regarding the stability of early diagnosis. ASD is behaviorally defined, meaning that diagnosis depends on history or present symptoms of impaired social communication and restricted, repetitive, and stereotyped interests and behaviors. ASD is thought to be a lifelong condition, and diagnosis is generally stable over time. However, some toddlers diagnosed with ASD at a very young age no longer meet criteria when they are older, whereas some toddlers at risk for ASD do not meet criteria initially but show symptoms of ASD when they are older (Fein et al., 2013; Kleinman et al., 2008; Lord, 1995; Orinstein et al., 2014). For example, some children may transition between diagnoses of ASD and Global Developmental Delay (GDD) over time, and others may even show no delays at all later in life.

The reasons for this diagnostic instability are unclear, particularly whether children were misdiagnosed at their initial evaluations, received and responded well to rigorous intervention, or have strengths and weaknesses, such as in cognitive or language ability, which could predict improvement or worsening of symptoms over time. These skills or intervention may help to predict developmental outcomes in the child, when assessed at a young age. However, most studies have emphasized the relatively high rates of early diagnostic stability of ASD. Therefore, there is currently little research focusing on the predictors of diagnostic instability of ASD. Furthermore, studies examining diagnostic stability typically focus on the loss of ASD diagnosis over time, rather than newly meeting criteria or transitioning between ASD and GDD. Finally, stability of GDD over time has received little attention in the literature.

The current study fills these gaps in previous literature by clarifying predictors of diagnostic instability of ASD and GDD. Children who changed diagnosis between ASD, non-ASD, and more specifically, GDD, were identified from a large sample of toddlers who had screened positive for risk of

ASD. Various skills and demographic features were identified based on previous research and theoretical considerations as potential predictors of change between the diagnostic categories. Final models indicated which predictors increased the odds of diagnostic transition in the toddlers. Significant predictors were then also explored as predictors of change in a continuous measure of ASD symptom severity; this procedure identified predictors which were important for significant change in severity, in contrast to those that contributed to minimal change sufficient only to cross the diagnostic boundary.

Identifying these predictors will help efforts to recognize when children are likely to change diagnostic category over time. These results have significant implications for predicting prognosis of children with ASD, based on whether or not they exhibit the characteristics associated with positive outcomes. It will also help for planning appropriate targets of treatment and will help to identify individuals who should be re-assessed later despite not meeting criteria for ASD at a young age. For example, if strong language ability was found to predict change from ASD to non-ASD, then a child diagnosed with ASD and strong language abilities at two years old would likely have a more positive prognosis. Furthermore, language ability would also be identified as an important target for intervention, as language ability would be associated with positive outcomes. Information on predictors of diagnostic instability may also reduce the number of early missed diagnoses by identifying children who will need re-evaluations based on their initial characteristics. These findings apply to diagnostic instability of GDD as well. Certain characteristics differentiate those who change between an ASD and GDD diagnosis and those who maintain GDD diagnosis. These predictors will be important to identifying which children will need to be re-evaluated to determine appropriate services (i.e., targeting ASD symptoms or not), and identify whether they should be assessed for intellectual disability when they are older.

## 1.1 Autism Spectrum Disorder

ASD is a neurodevelopmental disorder characterized by impairments in social interaction and communication skills, as well as the presence of restricted, repetitive, and stereotyped behaviors (American Psychiatric Association [APA], 2013). ASD has received significant media attention in recent years due to its rising prevalence rates; it is currently reported to affect one in 68 children (Baio, 2014). It is thought to be a lifelong condition, though some toddlers diagnosed with ASD no longer meet criteria when they are older, whereas some toddlers at risk for ASD but who do not meet criteria initially show symptoms of ASD when they are older (Chawarska et al., 2007; Cox et al., 1999; Fein et al., 2013; Kleinman et al., 2008; Lord, 1995; Pellicano, 2012; Soke et al., 2011; Stone et al., 1999; Turner & Stone, 2007).

Although the Diagnostic and Statistical Manual (DSM) has recently been revised (DSM-5; APA, 2013), the diagnostic criteria used in the current study are based on the previous version of the DSM (DSM-IV-TR; APA, 2000). The DSM-IV-TR identified several subtypes within the ASD category. They include Autistic Disorder, Asperger's Disorder, and Pervasive Developmental Disorders- Not Otherwise Specified (PDD-NOS; thought to be a milder ASD). An individual meets criteria for Autistic Disorder according to the DSM-IV-TR if he or she experiences at least two symptoms regarding social interaction (i.e., impairment in the use of nonverbal behaviors, failure to develop peer relationships, lack of spontaneous seeking to share enjoyment or interests, lack of social or emotional reciprocity), at least two communication symptoms (i.e., delay in or lack of development of spoken language not compensated for by other means, impairment in conversational ability, stereotyped or idiosyncratic language, lack of varied and spontaneous make-believe or social imitative play), and at least one restricted, repetitive, and stereotyped pattern of behavior or interest (RRB; i.e., encompassing preoccupation with restricted patterns of interest, inflexible adherence to nonfunctional routines or rituals, stereotyped and repetitive motor mannerisms, persistent preoccupation with parts of objects).

Individuals in the current study who received a diagnosis of Autistic Disorder, PDD-NOS (at least one social symptom and one communication or RRB symptom), or Asperger's Disorder (at least two social symptoms and one RRB symptom, no language or cognitive delay) were included in the ASD category. Subtypes were combined into one ASD category because dividing the sample into subtypes would limit power for analyses, research has indicated that the subtypes may not be different and valid constructs (Happé, 2011), and new diagnostic conventions do not distinguish between these subtypes as the DSM-5 has eliminated them in favor of an overall "ASD" diagnosis (APA, 2013).

Beyond broad diagnostic criteria, specific impairments have been found to distinguish children with ASD from others at around age two in the research setting, based on assessment and screening measures. These findings provide support for the validity of early diagnosis, as they demonstrate that toddlers with ASD exhibit behaviors distinguishable from those of typically-developing children. These include social impairment, such as ignoring people, reduced peer interest, frequency and/or quality of eye contact (Mitchell, Cardy, & Zwaigenbaum, 2011), joint attention skills (Robins et al., 2001), less joint functional play, and less responsive smiling, responding to name, following pointing, and looking to read faces (Trillingsgaard, Sorensen, Nemec, & Jorgensen, 2005). In terms of communication skills, those with ASD tend to demonstrate less initiation of requests using verbal and non-verbal behavior (Trillingsgaard et al., 2005) and fewer gestures (Veness et al., 2012). Wetherby and colleagues (2007) summarized these findings by identifying five signs of ASD in children at age 3, including impairments in gaze shifting, gaze and point following, rate of communicating, joint attention, and gestures. Overall, findings support the validity of early diagnosis of ASD, as toddlers with ASD demonstrate identifiable impairments that distinguish them from typically developing peers.

## **1.2 Rates of Diagnostic Stability of ASD**

Several studies have examined rates of diagnostic stability in young children with ASD at an initial or later evaluation, with emphasis on the high percentage of the total sample who stayed within

the same diagnostic category over time. See Figure 1 for a summary of stability findings in toddler studies; see Table 1 (studies examining rates of stability) and Table 2 (studies examining predictors of instability) for methodological summaries of these studies. Most studies of diagnostic stability in toddlers found relatively high levels of stability over time, ranging from 80 to 85% (Cox et al., 1999; Gillberg et al., 1990; Kleinman et al., 2008; Pellicano, 2012; Soke et al., 2011; Stone et al., 1999; Sutera et al., 2007; van Daalen et al., 2009; Wiggins et al., 2012). However, Turner and Stone (2007) found the lowest rate of stability of any studies, in that only 63% of participants diagnosed with ASD at age two still met criteria for ASD at age four. Guthrie and colleagues (2013) reported 100% stability of ASD diagnoses given at an initial evaluation, though diagnosis was deferred for 17% of participants for whom diagnosis could not be confidently confirmed. Including the participants with deferred diagnosis in the ASD group at the initial evaluation (3 of whom were ultimately diagnosed with ASD and 10 of whom were not), stability was determined to be at 84%.

A few studies found rates of stability slightly higher than the typical range of 80 to 85 percent. In a landmark study by Lord (1995), 90% of individuals stayed within their diagnostic category based on clinical judgment, while those judged to have ASD based only on the ADI showed poor diagnostic stability; all seven were later diagnosed with cognitive disability or language impairment. This finding highlights the difficulty of differentiating between ASD and other developmental delays in toddlers based on parent report. Hedvall and colleagues (2013) found that 90% of two-and-a-half-year-olds diagnosed with ASD still met criteria at age four. Stability in three studies was even higher, ranging from 94 to 100 percent (Charman et al., 2005- 96%; Chawarska et al., 2007- 100%; Eaves & Ho, 2004- 94%).

Diagnostic stability has also been examined from toddlerhood to later childhood. A seven year follow up study by Turner, Stone, Pozdol, and Coonrod (2006) yielded high rates of diagnostic stability of ASD from age two to age nine (88%), whereas another study found 81% stability between age five and age eight (Pellicano et al., 2012). In an older sample, Helles and colleagues (2014) followed 100 males

with Asperger's Syndrome from around age 11 until around age 33, and found stability of PDD of 91 percent after 10 years and 76 percent (in a subset of 47 participants) after 20 years.

Only four of these studies included a group of children who transitioned specifically from non-ASD to ASD (Chawarska et al., 2007; Cox et al., 1999; Soke et al., 2011; van Daalen et al., 2009). In addition to this small number of studies, each individual study also included a very small number of children who made this transition in diagnostic category. Three of the studies had a sample of one or two in this group (out of a total N ranging from 31 to 131), whereas one study included nine of these children (out of 46; 80%; Cox et al., 1999). Other studies without any participants in this transition group either did not identify children who changed from non-ASD to ASD, or excluded them from analyses because they focused only on the ASD-NON transition. Due to the limited research available on children who transition from non-ASD to ASD, further exploration of the frequency of this occurrence as well as information about these children is warranted.

Many of these studies on diagnostic instability of ASD are limited in that most had a sample of fewer than 50 children in total. Such a small sample size decreases power for analyses and contributes to less precise variance estimates and larger confidence intervals. For these analyses in particular, small sample sizes could contribute to variability in stability rates between studies. Moreover, a few studies of ASD diagnostic stability did not use validated measures of direct child observation to assess ASD symptom presentation. This is a shortcoming because the measures used instead depended entirely on parental report (Charman et al., 2004). Although parent report is valuable, since an evaluation allows for child observation during a limited window of time, many parents are not aware of the symptoms of ASD and how they present in children. Therefore, the parents may give inaccurate responses to questions about their children's behavior (Risi et al., 2006; Trillingsgaard et al., 2005). Despite these limitations, generally between five and 20 percent of participants across studies meet criteria for ASD as toddlers but no longer meet criteria one or more years later, or gain an ASD diagnosis over time. However, most

of the studies examining diagnostic stability of ASD emphasize the high rates of stability, and do not further examine the subset of children with unstable diagnoses.

### **1.3 Importance of Accurate Early Diagnosis of ASD and Developmental Delays**

It is widely accepted that early intervention leads to better prognosis in children with ASD (Harris & Handleman, 2000; Lovaas, 1987; Warren et al., 2011). Studies have suggested that children who receive early intervention show greater gains in cognitive and adaptive functioning and a greater reduction in ASD symptoms than those who do not receive early services (Dawson, 2008; Warren et al., 2011). However, ASD diagnosis informs intervention targets, and most children do not receive intensive ASD-specific interventions before receiving a diagnosis. As a result, there is currently an emphasis on diagnosing children as early as the second year of life in order to facilitate entry into intensive ASD-specific intervention services. Screening measures have been developed and validated to identify children as young as 18 months old who may have ASD (Charman et al., 2007; Chlebowski, Robins, Barton, & Fein, 2013; Robins et al., 2014; Siegel, 2004). The American Academy of Pediatrics even recommends that children be screened for ASD between 18 and 24 months of age (Johnson & Meyers, 2007).

Unfortunately, diagnosis tends to occur much later than the suggested age range. The median age of initial ASD diagnosis is 53 months and varies based on specific subtype: Autistic Disorder at 48 months, PDD-NOS at 50 months, and Asperger's Disorder at 74 months (Baio, 2014). A contributing factor to delayed diagnosis is the doubt that early diagnosis is reliable and valid (Kleinman et al., 2008). The source of this doubt may include the concern that ASD diagnosis is sometimes unstable over time. Early ASD diagnoses have been considered to be unstable for various reasons, including difficulty differentiating between ASD and other developmental delays, difficulty identifying impairments before school age when children are seen interacting with peers, and inability to apply DSM-IV criteria to young children (Aitken, 1991; Kleinman et al., 2008; Siegel, Pliner, Eschler, & Elliott, 1988; Stone et al., 2003;

Wiggins et al., 2012b). For example, Stone and colleagues (1999) found that three DSM-IV symptoms were consistently considered “not applicable” to their two-year-olds, including failure to develop peer relationships (60%), impaired conversational skills (100%), and stereotyped language (80%). Also, Wiggins and colleagues (2012b) determined that a subgroup of toddlers did not exhibit the significant repetitive behaviors and abnormal sensory responses included in DSM-IV-TR criteria. Furthermore, children with ASD tend to have reduced compliance during cognitive testing, which may lead to confusion between an ASD and GDD diagnosis (Akshoomoff, 2006). These factors contribute to some diagnostic instability over time, such that some children may not be diagnosed at a young age but meet criteria as they get older.

The issue of early diagnosis and diagnostic instability also applies to the diagnosis of other developmental delays, particularly Global Developmental Delay (GDD). Early intervention has been found to improve target skills, such as motor skills and language ability, in children with developmental delays (Rydz, Shevell, Majnemer, & Oskoui, 2005). As a result, there has been an attempt to diagnose children as early as possible in order to make services available (Rydz et al., 2006). However, research has found that fewer than half of children with delays are identified before entering school (Glascoe, 2005). Authors have proposed that this delay in identification may be due to lack of concern by parents, or lack of adherence to widespread screening practices (Glascoe, 1994; Glascoe, 2005; Rydz et al., 2006). This finding indicates that more widespread screening through medical practice may be essential to improving outcomes for children with delays. On the other hand, diagnosing children earlier may mean that a small group of children who have received a diagnosis will no longer meet criteria when they are older. Further research on the diagnostic stability of developmental delays is necessary to clarify this risk.

Diagnostic stability is important for identifying impairments at a young age so that children can receive the appropriate interventions for their needs. There are potentially harmful outcomes to an



unstable diagnosis. In particular, a child who ends up with ASD or GDD may not be initially diagnosed, and would be deprived of timely intervention services that are known to be associated with a more positive prognosis. This outcome highlights the importance of striving for more reliable early diagnoses. It would also be beneficial to families to better understand how their child's strengths and weaknesses at an early age predict prognosis. The current study aims to elucidate these skills and characteristics, so that more accurate predictions of the ASD and GDD trajectories may be made at initial diagnosis, and so that characteristics consistent with phenotypic improvement may be targeted in intervention.

#### **1.4 Change in ASD Symptom Presentation Over Time**

Sufficient evidence has been provided to indicate that diagnostic instability is not simply the result of initial misdiagnosis, but that ASD symptom presentation regularly changes over time, sometimes contributing to diagnostic instability. Some research has been dedicated to the identification of developmental trajectories of children with ASD, particularly identifying which children are most likely to improve or decline in functioning over time. Lord, Luyster, Guthrie, and Pickles (2012) evaluated 78 children at risk for ASD every six months from 18 to 36 months of age. They identified four different groups of children based on their developmental trajectories. Twenty-one percent (N=16) were in the severe persistent group, whereas 40 percent (N=31) were in the persistently non-ASD group. However, 21 percent (N=16) were in the worsening group and 19 percent (N=15) were in the improving group. Although the improving group did not necessarily move off the spectrum altogether, this finding is consistent with the rates of diagnostic instability generally found in the literature; some studies have found that around 20 percent of children improve to no longer meet criteria for ASD after an initial early diagnosis. Similarly, the trajectories studied by Venker and colleagues (2014) found that 14 percent (N=28 of 129) belonged to the group that improved from age 2.5 to 5.5, and eight percent (N=10 of 129) belonged to the worsening group.

The trajectories identified by Lord and colleagues (2012) differed by changes in verbal and nonverbal mental age over time. In particular, the improving group showed faster increases in nonverbal mental age than the persistent group, and both the improving and non-ASD groups increased more quickly in verbal mental age than the persistent group. Notably, the improving and worsening groups did not differ on Autism Diagnostic Interview-Revised (ADI-R; Le Couteur, Lord, & Rutter, 2003) scores, though, as expected, the non-ASD group did have significantly lower ADI-R scores than the other groups. The trajectories also did not differ in terms of amount of intervention received. In the study by Venker and colleagues (2014), the group of children with the most stable, severe symptoms had the lowest baseline nonverbal cognitive ability and growth in receptive and expressive language skills. Furthermore, another study found that 10 percent of 6,975 individuals between the age of two and 14 years with severe autism, as identified through the California Department of Developmental Services database, experienced enough communication and social gains over time to be functioning at the same level as those with only mild impairments, and that those who were least severe improved most rapidly (Fountain, Winter, & Bearman, 2012). Overall, groups of children who improve, worsen, and remain consistent in their ASD symptom severity are common across samples and can be distinguished by other skills; this change in symptom severity contributes to diagnostic instability, and is not the result of inaccurate initial diagnosis.

In terms of general improvement of ASD symptoms over time, several studies identified changes in scores on ASD diagnostic measures over time which could ultimately contribute to diagnostic instability. Soke and colleagues (2011) found that in a sample of two- to four-year-olds evaluated and re-evaluated two years later, the mean total, social, and communication scores on the ADI-R decreased over time, indicating reduced impairment. However, restricted, repetitive, and stereotyped behavior (RRB) scores did not change over time. Similarly, Guthrie, Swineford, Nottke, and Wetherby (2013) also found that Autism Diagnostic Observation Schedule-Toddler Module (ADOS-T; Luyster et al., 2009) Social

Affect scores decreased over time for both children with ASD and children without ASD. However, RRB scores increased for children on the spectrum, supporting the theory that RRB may be less likely to be observed in very young toddlers and may appear later in development (Wiggins et al., 2012b). Also, a study of the stability of the 10-point ADOS Calibrated Severity Score (CSS; Gotham, Pickles, & Lord, 2009) determined after 12 to 24 months, 23.6 percent of participants improved by one or more point, and 12.4 percent worsened by one or more point (Shumway et al., 2012). In terms of language, changes over time in children with ASD have been found to be related to the emergence of atypical language patterns, such as echolalia and stereotyped speech (Chawarska, Klin, Paul, & Volkmar, 2007). The results of these studies indicate that ASD symptoms often fluctuate over time, and may contribute to diagnostic change with a large enough fluctuation. Therefore, diagnostic instability can result from symptom change over time, and not simply from confusion or misdiagnosis at initial evaluation.

### **1.5 Factors Contributing to Change in Symptom Presentation**

The theoretical basis for our exploration of diagnostic instability in the current study is a dynamic systems model. Smith and Thelen (2003) have described a dynamic system, in which development is influenced by various systems interacting in a complex environment. In other words, skills related to cognitive ability, communication ability, and social interaction influence each other throughout development. Early strengths in one of these areas may facilitate growth in other areas of development, resulting in diagnostic transition from ASD to non-ASD. Similarly, early impairments in certain domains may hinder progress in other areas of functioning, which could result in movement onto the autism spectrum later in childhood. For example, research has supported the finding that joint attention skills early in childhood impact social cognitive skills (e.g., Mundy, Sullivan, & Mastergeorge, 2009) and communication ability (e.g., Adamson et al., 2009) in children with ASD. Communication ability in children with ASD has also been found to support performance on social cognitive tasks

(Happé, 1995). These examples illustrate that interactions between various systems are likely to contribute to changes across various aspects of development.

The influence of child characteristics and environmental factors on child development seems to be important to change in ASD symptoms over time. For example, this seems to be true for skills such as language development, which requires input from others. A child attends to surrounding individuals, and is thus able to learn language. However, if children with ASD are less attuned to social stimuli, they may be less likely to attend to individuals who are speaking, which may contribute to a delay in language learning. In this case, it may be expected that a child with more social interest may develop stronger language skills or develop them sooner. At the same time, a child who has a language delay may be less likely to engage with others who are speaking, either out of a lack of understanding or due to frustration at ineffective communication. In this way, a language delay may limit social interactions, which may then further delay language development. Furthermore, a child with limited language abilities may appear to develop other cognitive abilities more slowly, because they do not have the linguistic capacity to internally represent or communicate cognitive reasoning. Therefore, young children may be supported by early strengths or impaired by early weaknesses that impact multiple domains of functioning, which may contribute to diagnostic instability.

## **1.6 Predictors of Diagnostic Instability of ASD**

A few studies have examined initial predictors of diagnostic instability of ASD, particularly the factors that can predict no longer meeting criteria for ASD at re-evaluation. See Table 2 for a summary of methodology and findings regarding predictors of instability of ASD diagnosis. The majority of studies examined predictors of positive outcomes in general, rather than those resulting directly in diagnostic instability. Overall, the findings have been variable.

### **1.6.1 ASD symptom severity**

Several studies have examined early ASD symptom severity as a possible predictor of experiencing diagnostic transition later in childhood. Findings are inconsistent, but low ASD symptom severity has generally been associated with positive outcomes and even diagnostic instability. A few studies have surprisingly not found symptom severity to be a strong indicator of diagnostic transition (Pellicano, 2012; Sutera et al., 2007). However, some studies have identified early differences in symptom severity as being indicative of later diagnostic transition. Stone and colleagues (1999) determined that in 37 children between age two and three, the six children who made diagnostic transitions of decreasing severity had two or three fewer DSM-IV symptoms endorsed at the initial evaluation. Also, the two children who showed an increase in severity had two to three more symptoms initially endorsed. Therefore, symptom number in this small sample was related to change over time. Furthermore, Lord (1995) found that at age two, only the social domain of the ADI and the Childhood Autism Rating Scale (CARS; Schopler et al., 1984) total score, which reflects general ASD symptom severity, differed between the children who maintained and those who no longer met criteria for their ASD diagnosis at age three. Finally, Wiggins and colleagues (2012b) found that children with mild ASD-related impairments were five times more likely than the children in the severe ASD group to receive a later non-ASD diagnosis, though those with moderate severity did not differ from those with severe impairments in terms of prediction of diagnostic change. Other studies support lower social impairment as being predictive of moving off the spectrum (Turner & Stone, 2007)

Lower ASD symptom severity has also been associated with general improvement over time (Fountain, Winter, & Bearman, 2012). For example, one study of 78 toddlers found that those with lower ASD symptom severity and those who were younger at the initial evaluation made greater cognitive gains after receiving intervention (Ben-Itzhak & Zachor, 2011). However, Charman and colleagues (2005) found that in 29 toddlers, ASD symptom severity at age three but not age two

predicted ASD symptom and adaptive behavior outcome by age seven. Specifically, higher symptom severity at age three correlated with higher symptom severity and lower adaptive skills at age seven. In general the findings are mixed, but most studies have found low ASD symptom severity at an early age as being predictive of general improvement, and even movement off of the autism spectrum.

### **1.6.2 Communication ability**

Communication abilities, including receptive language, expressive language, and use of gestures, have been associated with improvements in ASD presentation over time, but not with diagnostic instability of ASD. First, initial receptive and expressive language ability have been found to predict later outcomes. Luyster and colleagues (2007) found that at age two and three, receptive and expressive language scores predicted language, IQ, and ADOS/ADI-R composite scores at age nine. Similarly, another study of children in Italy found that low initial levels of impairment in communication, language comprehension, and gestures predicted improvements in ASD severity, but not diagnostic instability, after six months (Muratori & Narzisi, 2014). Paul, Chawarska, Cicchetti, and Volkmar (2008) found that the number of sounds and words produced, use of symbolic play schemes, and response to joint attention bids at age two predicted expressive language ability by age four. Expressive language ability at age two in Turner and colleagues' (2006) study predicted conversational ability at age nine, though the number of words at age two was not predictive of later language ability. Furthermore, those with high language scores at age two were more likely to be in the outcome group with the highest language and cognitive performance. Early verbal ability has also been found to be predictive of later adaptive skills (Ben-Itzhak & Zachor, 2011). Next, in terms of nonverbal communication skills in toddlers, initial nonverbal behaviors, such as imitation and joint attention, have been found to predict later language and social skills (Charman et al., 2005; Toth et al., 2006). Gestures at age two have also been found to predict age nine verbal IQ, expressive language, and adaptive skills (Luyster et al., 2007). Finally, weak expressive language abilities have been found to be associated with global cognitive deficits in

individuals with ASD (Hedvall et al., 2013). Although all of these early communication abilities have been associated with later abilities, it is notable that none of these studies have identified characteristics which predict diagnostic instability specifically.

### **1.6.3 Cognitive ability**

Intellectual ability has generally been found to predict later positive outcomes in children with ASD. In particular, higher cognitive level at age two has been found to predict later communication ability (Paul, Chawarska, Cicchetti, & Volkmar, 2008; Turner et al., 2006) and cognitive ability (Turner et al., 2006). Nonverbal IQ at age three, but not age two, was associated with the ADOS social score at age seven in a study by Charman and colleagues (2005). Similarly, but in terms of diagnostic stability specifically, Turner and Stone (2007) found that those with higher IQs were more likely to move off the autism spectrum.

Another study found that while rigorous intervention was the main contributing factor to improvement of ASD symptoms over time, response to therapy was much greater in individuals with higher nonverbal IQ (Mazurek, Kanne, & Miles, 2012). Similarly, Ben-Itzhak, Watson, and Zachor (2014) found that those with high IQ transferred their social-communication gains successfully from intervention to daily functioning. Therefore, higher intellectual ability seems to be related to greater receptiveness to intervention, and ultimately contributes to the reduction of ASD symptom severity over time. Similarly, those with intellectual disabilities have also been found to be less likely to make significant gains than other children with ASD, contributing to greater diagnostic stability in this group (Fountain, Winter, & Bearman, 2012). Another study found that the group that moved from ASD to non-ASD showed a greater increase in cognitive ability over time, though not necessarily higher initial cognitive scores (van Daalen et al., 2009). Interestingly, verbal IQ was found to be highest in a group of children with ASD that worsened the most and in a group that improved the most, though it increased the most for the improving group (Gotham, Pickles, and Lord, 2012). Overall, higher intellectual ability

seems to be associated with improvement in symptoms over time, and possibly with diagnostic transition from ASD to non-ASD.

#### **1.6.4 Demographic factors**

Very few studies have investigated diagnostic instability in different racial, ethnic, and socioeconomic groups. Despite this lack of attention in the literature, a few studies have identified predictors of outcomes for these groups. Fountain, Winter, and Bearman (2012) found that individuals with White, non-Hispanic, well-educated mothers were most likely to show improvement over time. However, these results were based on a large database of individuals enrolled in the California Department of Developmental Services, and therefore the data lack consistency in the method and basis of diagnosis (e.g., different measures used and differences in clinician judgment). Another study found that higher maternal education and older maternal age were associated with greater cognitive gains over a span of two years from toddlerhood (Ben-Itzhak & Zachor, 2011). Further research is necessary to establish relationships between diagnostic instability and demographic characteristics, in order to provide information for making diagnoses and providing services in individuals with ASD.

#### **1.6.5 Intervention**

Intervention has been studied as a factor influencing improvement and diagnostic change over time. Findings generally indicate that intervention predicts positive outcomes, though some studies have not found it to be a significant factor in influencing ASD symptoms. Turner and colleagues (2006) found that hours of speech-language therapy were predictive of later cognitive and language functioning in an ASD population from age two to nine. Also, Orinstein and colleagues (2014) established that children who met criteria at age two but no longer met criteria for any disorder between age eight and 21 had earlier parental concern, had earlier referral for intervention, had more intensive intervention, and were more likely to have had applied behavior analysis (ABA) therapy than those with high functioning ASD. Another study found that those who changed diagnostic category from ASD to non-ASD



differed only in that they were younger when they began intervention (Pellicano, 2012). A recent study of 39 children in Turkey found that children who met criteria for ASD at age 2.5 but not at five and who participated in early intervention services had the best outcomes in terms of high IQ and strong language ability (Mukkades et al., 2014). However, some studies have surprisingly not found intervention to be predictive of significant improvement in ASD symptoms or loss of ASD diagnosis over time (Lord, Luyster, Guthrie, & Pickles, 2012; Turner & Stone, 2007). Camarata (2014) explained that it is difficult to study intervention effects on ASD because diagnostic instability may occur independently of or in association with intervention.

#### **1.6.6 Other factors**

Other factors, including age of diagnosis and motor abilities, have been found to be related to improvements in ASD presentation over time. Sutera and colleagues (2007) identified a group of toddlers who met criteria for ASD at an initial evaluation but did not meet criteria for any other diagnosis by re-evaluation. They found very few differences between the groups at the initial evaluation; the children who transitioned differed from the ASD-persistent group only on Mullen Scales of Early Learning (MSEL; Mullen, 1995) motor scores. However, this finding has not been replicated in other studies of diagnostic stability. Furthermore, Turner and Stone (2007) found that children diagnosed before 30 months of age were more likely to lose their ASD diagnosis than those older than 30 months. Wiggins and colleagues (2012a) found the same age contribution to diagnostic instability in a large surveillance study of diagnosis by community professionals. Similarly, younger age of diagnosis has been found to be predictive of later higher cognitive and language functioning (Turner et al., 2006). Therefore, earlier diagnosis could contribute to developmental progress.

#### **1.7 Global Developmental Delay**

Global Developmental Delay (GDD) is another diagnosis given to children younger than age five with developmental delays. Although the DSM-5 definition of GDD does not identify specific criteria

required for diagnosis aside from failure to meet several developmental milestones (APA, 2013), other professionals have defined GDD as requiring a significant delay (i.e., at least two standard deviations below the expected mean performance) in two or more domains of functioning, including motor skills, speech and language, cognition, social/personal behavior, and activities of daily living in children under five years old (Shevell et al., 2003; Tirosh & Jaffe, 2011). GDD is estimated to affect about one to three percent of children under five years old (Shevell et al., 2010).

Children with GDD are often diagnosed with intellectual disability (ID) after the age of five, if their intellectual functions by this time fall two standard deviations below the population mean, and they have impairments in adaptive functioning (APA, 2013). However, not all children with GDD end up with ID, as GDD can be based on language, personal, and motor impairments, and not necessarily on cognitive impairment. For example, one study found that 73% of children with a GDD diagnosis aged two to five (based on impairments in two domains of functioning) earned IQ scores higher than 70 according to the Wechsler Preschool and Primary Scale of Intelligence, Third edition (WPPSI-III; Wechsler, 2002; Riou, Ghosh, Francoeur, & Shevell, 2009). The authors explain that although GDD is generally associated with cognitive impairment, the difficulty with reliably assessing very young children leads to the need to evaluate a variety of areas in addition to cognitive ability; this consideration leads to variability in the number of children with GDD who actually have cognitive impairment.

### **1.8 GDD Stability and Outcomes**

GDD diagnoses, which indicate delays in two or more domains, are considered to be relatively stable. However, very few studies have examined diagnostic stability of delays in young children diagnosed with GDD, and those that have found variable results. One study found that 43 out of 45 children (96%) who met criteria for GDD at age three had impairments in multiple domains that persisted through age seven (Shevell et al., 2005b), and another found that based on parent report, 60 out of 62 children (97%) continued to show impairments at age seven consistent with their GDD

diagnosis from age three (Webster, Majnemer, Platt, & Shevell, 2008). However, in a study of 394 children from age four to age six, 40 percent of those with mild developmental delays (but not official GDD diagnosis) achieved normal developmental status by age six, whereas more severe impairments remained stable over time, based on a Finnish neurodevelopmental screener assessing language, motor skills, and school readiness (Valtonen, Ahonen, Lyytinen, & Tolvanen, 2007). Overall, stability of GDD diagnosis before age five has not been examined. However, impairments associated with GDD diagnoses seem to be fairly stable past age five, though there is evidence that less severe impairments may be less stable.

Outcome research on children diagnosed with GDD is surprisingly limited. Most studies that have examined this issue were retrospective, in that it was noted that individuals with current intellectual disability were previously labeled as having GDD (Shevell et al., 2005b). Other studies have primarily focused on academic skills, identifying early issues such as motor and cognitive delays associated with later school difficulties, such as reading problems (Silva, McGee, & Williams, 1985). Generally, research has found that functional outcomes are often better than developmental outcomes, indicating that children and their families may be able to adapt to the child's limitations and enhance daily functioning (Shevell et al., 2005a; Shevell et al., 2005b).

In terms of predictors of outcomes in children with GDD, research has been limited. However, some variables have been identified as potential predictors of later outcomes. Some children with GDD have a known underlying etiology (e.g., genetic disorder), whereas others do not; a lack of known underlying etiology predicted poorer motor skills compared to those with known etiology, perhaps because the parents of children without these conditions were less likely to seek intervention (Shevell et al., 2005b). Furthermore, more severe initial delays were predictive of poorer communication and adaptive skills, as reported by parents on the Vineland Adaptive Behavior Scales (VABS; Sparrow, Cicchetti, & Balla, 1989). A recent study of children under five with GDD found that poorer outcomes

after two years were associated with intrauterine growth restriction during gestation, low socioeconomic status (SES), noncompliance to habilitation plan, and prematurity (Thomaidis et al., 2014). Finally, maternal employment was found to predict VABS communication scores, and paternal education level was found to predict VABS communication, socialization, and total scores. Shevell and colleagues (2005b) suggest that these parents with higher education may be more likely to advocate for and accommodate the needs of their children. Together, these limited findings suggest that children with GDD may be at risk for learning and adaptive difficulties as they grow older. Another study found that no single variable predicted outcomes from age four to age six, but that greater overall severity and different modalities affected at age four predicted a stable impaired outcome by age six (Valtonen, Ahonen, Lyytinen, & Tolvanen, 2007). Poor psychosocial health and high levels of parenting stress have also been associated with GDD diagnoses in school aged children, though the role that these stressors play in the progression of GDD is unclear (Webster, Majnemer, Platt, & Shevell, 2008). Overall, research has identified some predictors of specific scores and outcomes, but predictors of stability of GDD have not been explored.

Although GDD diagnoses are generally considered to be stable, it is notable that research on diagnostic stability of GDD in toddlers is apparently unavailable, and information about stability of delays associated with GDD beyond age five is limited. Research on factors which predict outcomes in these children is also quite limited. Given this limited literature, research on predictors of diagnostic instability of GDD specifically is unavailable. Further research is necessary to explore these predictors, in order to identify targets of intervention and characteristics of children which predict transition to and from other diagnoses, such as ASD.

### **1.9 ASD compared to GDD**

ASD and other developmental delays (DD), such as GDD, are often difficult to distinguish at a young age for various reasons. ASD and GDD are both very heterogeneous. Different children on the

spectrum may have extremely different symptom presentations; similarly, children with GDD can have impairments in any two of five domains, also contributing to a high level of heterogeneity. Furthermore, many symptoms of ASD and GDD overlap, and specific characteristics of each can be difficult to differentiate (Tirosh & Jaffe, 2011; Zwaigenbaum et al., 2009). As a result of this symptom diversity and overlap, it can sometimes be difficult to provide an accurate differential diagnosis for a young child, despite the presence of symptoms at a young age. For example, one study found that all seven children who were diagnosed with ASD based on the ADI at age two but who no longer met ASD criteria a year later, were diagnosed with cognitive disability or language impairment (Lord, 1995). This finding highlights the difficulty of differentiating between ASD and other developmental delays in toddlers. Research has identified characteristics which may help to distinguish between these conditions, though some results are still inconsistent.

### **1.9.1 Social abilities**

Major differences between toddlers with ASD and those with developmental delays have been identified in several domains, including socialization, communication, motor skills, and temperament. In the socialization domain at age two, Mitchell, Cardy, and Zwaigenbaum (2011) found that compared to children with non-ASD DD (such as those with language delay or GDD), those with ASD showed more limited responsiveness to others (e.g., less peer interest and sharing of enjoyment), limited use of gaze to modulate social interactions, reduced positive affect and facial expressions (though interestingly no differences in negative affect or social smiling), and decreased response to name. Ventola and colleagues (2007) found that more severe ratings on all of the items from the reciprocal social interaction domain of the ADOS algorithm differentiated the groups. In particular, items involving joint attention differentiated the ASD group from the group experiencing other delays. Trillingsgaard and colleagues (2005) found that the ASD and DD group of two- to four-year-olds differed on seven out of 15 measures of social interaction during a semi-structured interactive play measure, for which reliability

and validity have not yet been established. The measures on which groups differed included responsive social smile, joint attention, joint interactive play, and looking at faces for social information. Overall, several social symptoms have been found to distinguish between children ASD and those with DD.

### **1.9.2 Communication skills**

In the communication domain, comparisons between two-year-olds with ASD and DD have been variable when guided by standardized measures of cognitive, language, and motor ability (i.e., MSEL, Mitchell, Cardy, & Zwaigenbaum, 2011). Some studies have found no differences, some found weaker receptive language scores, and some have found both weaker receptive and expressive scores in those with ASD. In terms of identified differences, a DD group has been shown to attain higher receptive language scores on the MSEL than a group with ASD (Trillingsgaard et al., 2005). Those with ASD have also been found to use fewer words and early emerging communicative gestures (e.g., pointing, nodding) in general than those with DD (Mitchell, Cardy, & Zwaigenbaum, 2011). Children with ASD were also characterized by having greater expressive than receptive language abilities. Ventola and colleagues (2007) compared toddlers with ASD to toddlers with GDD or developmental language delay (DLD). Similar to the other studies, they found that groups differed on communication ability as reported by parents on the VABS and as tested using the MSEL. However, the groups did not differ on other interactive measures of communication ability. In particular, they found that only one algorithm item from the communication domain of the ADOS, a measure of pointing, differentiated between the ASD group and the DD group. Items related to communicative ability did not differentiate the groups, likely because both groups tend to experience delays in communication. This finding of lack of differences in communication ability has been supported elsewhere (Veness et al., 2012).

### **1.9.3 Motor and sensory behaviors**

Several characteristics regarding motor and sensory behaviors have been found to distinguish between children with ASD and those with other DD. In terms of motor behaviors, one study found that

based on videotapes of one-year-olds, children later diagnosed with ASD tended to engage in more repetitive motor actions and atypical sensory responses than those later diagnosed with intellectual disability (Osterling, Dawson, & Munson, 2002). Another study found that a group of children with ASD engaged in more repetitive movements and hand or body posturing than a group with developmental delays (Wetherby et al., 2004). Additionally, children with ASD have been found to have more sensory sensitivities than children with other developmental delays (Ventola et al., 2007; Wiggins, Robins, Bakeman, & Adamson, 2009).

#### **1.9.4 Cognitive and adaptive skills**

Children with ASD tend to score lower than those with GDD on measures of cognitive ability. In one study, the children with ASD scored lower than the group with other delays on measures of adaptive skills and nonverbal problem-solving skills (Mitchell, Cardy, & Zwaigenbaum, 2011). Consistent with this finding, a comparison of cognitive ability between a group of toddlers with ASD and a group with GDD or DLD determined that the ASD group had lower visual reception scores according to the MSEL (Ventola et al., 2007). This result has been found elsewhere as well (Trilingsgaard et al., 2005), indicating that children with ASD tend to have more impaired nonverbal cognitive ability than those with other developmental delays.

#### **1.9.5 Developmental milestone attainment**

In another study examining differences in developmental milestone attainment based on parent report between children with ASD and non-ASD atypical development (i.e., language delay, intellectual disability, spina bifida, premature birth, pre-natal exposure to substances, etc.), a greater percentage of toddlers in the ASD group were found to have had delays in language and motor milestones (Matson, Mahan, Kozlowski, & Shoemaker, 2010). In terms of age of acquisition of these milestones, those in the ASD group were also found to have acquired these milestones at a later age than those with non-ASD atypical development.

### **1.9.6 Factors contributing to difficult differentiation between ASD and GDD**

In spite of some reliable differences between children with ASD and those with other developmental delays, several studies indicate that these conditions can be difficult to distinguish. Veness and colleagues (2012) found that the only characteristic that clearly differentiated them at 24 months of age was the gesture score on the Communication and Symbolic Behavior Scale Developmental Profile (CSBS; Wetherby & Prizant, 2002). However, it is important to note that sample sizes in this study were small, so they may have lacked adequate power to detect differences. One study noted that these disorders are very difficult to differentiate in two-year-olds based solely on parents' retrospective reports, because the parents may not be aware of deficits or they may be more likely to remember specific instances of positive behaviors (e.g., a child smiling back at them) rather than the high frequency of negative behaviors (e.g., the child does not look at the parent's face; Trillingsgaard et al., 2005). Finally, children with ASD were found to exhibit more off-task behaviors and less engagement with the examiner than children with GDD during completion of the MSEL (Akshoomoff, 2006). Although this behavioral distinction is notable, it contributes to confusion between the diagnoses; non-compliance would result in underestimates of cognitive, language, and motor ability for children with ASD, which may result in confusion with GDD.

### **1.10 Current Study**

Existing literature on predictors of diagnostic stability of ASD and GDD is limited and equivocal. For both diagnoses, some research is available on the predictors of specific outcomes, such as language or cognitive ability. However, findings are variable in the case of both conditions, and they are especially limited in the case of GDD as there are very few toddler studies examining these predictors. Inconsistencies in assessment and diagnostic methods, as well as small sample sizes, may contribute to the variability in ASD and GDD outcome data. Some potentially important variables, such as race/ethnicity and SES, have also received little attention.



Furthermore, research on predictors of diagnostic instability specifically is quite limited. In the case of ASD, most research has focused on the high rates of reliability of diagnosis over time, not on the consistent finding that five to 20 percent of children have diagnoses that are not stable over time. The few studies that actually examined the group of children who changed diagnosis (e.g., Lord, 1994; Sutera et al., 2007; Turner & Stone, 2007) found variable results, had small sample sizes, or focused only on one outcome (i.e., transitioning from ASD to no diagnosis, versus gaining an ASD diagnosis over time). More attention should be devoted to characteristics of children who meet criteria for ASD later after not initially meeting criteria. Understanding of these predictors may inform identifying these cases at an earlier age to determine appropriate intervention, or refining screening and diagnostic measures to consider these characteristics. Furthermore, the ADOS Calibrated Severity Score (CSS) is a recent development that was unavailable for use in previous studies of stability (Gotham et al., 2009). The use of the ADOS CSS score may help to clarify the role of ASD symptom severity in the diagnostic stability of ASD.

Predictors of diagnostic instability of GDD have not been examined to this author's knowledge. This may be because this diagnosis is generally considered to be stable over time, though its stability has also not been well studied. Similarly, predictors of diagnostic transition from GDD to ASD and vice versa have not been identified in the literature. Therefore, the current study examined diagnostic instability of GDD in addition to ASD. Although some children changed between diagnoses of ASD, GDD, and developmental language delay (DLD), instability of DLD was not explicitly examined due to small sample sizes.

The current study addressed these limitations by examining diagnostic instability of ASD and GDD using a large sample of toddlers with initial diagnoses of ASD and non-ASD (including GDD). This allowed for the identification of predictors of no longer meeting criteria for ASD over time, newly meeting criteria for ASD after initially receiving no ASD diagnosis, and transitioning between an ASD and

GDD diagnosis. Also, using rigorous diagnostic methods involving both parent report and direct child interaction, the current study contributes to knowledge about predictors of diagnostic instability, such as initial symptom severity and language ability, because findings regarding their predictive ability have been inconsistent. The current study also includes exploratory analyses examining predictors of diagnostic transition between ASD and GDD specifically, knowledge of which has been missing in the literature. Finally, in addition to predictors of diagnostic instability, the role of each of these variables is considered in the change in ASD symptom severity between T1 and T2 regardless of diagnostic change.

The current study used an archival dataset from a large ASD screening study to identify the variables which, at a two-year-old evaluation (T1), predicted diagnostic instability at a four-year-old evaluation (T2). Predictors of diagnostic instability were examined in toddlers whose diagnostic categories changed from ASD to non-ASD (ASD-NON) and from non-ASD to ASD (NON-ASD). A subset of these children was also specifically examined, including those who changed from ASD to GDD (ASD-GDD) and those who changed from GDD to ASD (GDD-ASD). The potential predictors of diagnostic transition included cognitive ability, language ability (expressive language and receptive language), ASD symptom severity, demographic characteristics (race and SES), and weekly hours of intervention between T1 and T2. Also, the role of these predictors was examined in the continuous change in symptom severity between time points, in addition to the aforementioned categorical diagnostic transition. This analysis allowed for comparison between a fluctuation in a continuous measure of severity and a categorical change in diagnosis. The results of this study may be used to make early diagnosis more accurate and effective, guide prognostic planning, and help to inform which child skills may be important targets of intervention.

## **1.11 Hypotheses**

Analyses addressing each hypothesis first explored the role of predictors in categorical diagnostic change, and then examined whether these predictors were involved in change in a continuous measure of symptom severity.

### ***1.11.1 Hypothesis 1A: Diagnostic transition from ASD to non-ASD***

The predictors of diagnostic transition from ASD at a two-year-old evaluation (T1) to non-ASD at a four-year-old evaluation (T2) were examined (ASD-NON). Based on findings of previous research, high scores on cognitive and language measures at T1 were likely to predict no longer meeting criteria for ASD at T2, though it was thought that the inclusion of individuals with other diagnoses in the non-ASD group may alter this outcome. Although the literature has mixed results regarding the role of initial ASD symptom severity in diagnostic instability, it was presumed that those with lower ASD symptom severity at T1 would be more likely to no longer meet criteria for ASD by T2, because less change over time would be required to cross diagnostic boundaries, and the ADOS CSS was expected to provide a more reliable estimate of ASD symptom severity than other measures that had been used previously. Based on the limited literature available, it was expected that those from the majority race, individuals from a high SES, and those who have received more frequent intervention services would also be more likely to make this diagnostic transition. Furthermore, it was expected that the same predictors contributing to diagnostic change over time would contribute to change in a continuous measure of ASD symptom severity over time; this would indicate that these factors are contributing to significant change over time, rather than minimal change across a diagnostic boundary.

### ***1.11.2 Hypothesis 1B: Diagnostic transition from ASD to GDD***

The predictors of diagnostic transition from ASD at a two-year-old evaluation (T1) to GDD at a four-year-old evaluation (T2) were examined (ASD-GDD). Lower ASD symptom severity at T1 was expected to predict transition to GDD because it would require a smaller change over time to cross the

diagnostic boundary. It was expected that greater time spent in intervention services would likely predict diagnostic transition because it would contribute to the reduction of social impairments. The predictive ability of language level was unclear due to the variable findings regarding communication comparisons between individuals with ASD and GDD. Furthermore, the role of race and SES was unclear due to lack of previous research. Therefore, these analyses were exploratory. ASD symptom severity and intervention were also expected to predict change in ASD symptom severity between T1 and T2, indicating that they would contribute to a significant reduction in ASD symptoms over time regardless of categorical diagnostic transition.

### ***1.11.3 Hypothesis 2A: Diagnostic transition from non-ASD to ASD***

The predictors of diagnostic transition from non-ASD at a two-year-old evaluation (T1) to ASD at a four-year-old evaluation (T2) were examined (NON-ASD). Little research had been designated to determining predictors of gaining a diagnosis at T2 after not receiving an ASD diagnosis at T1. In these individuals, it was expected that lower cognitive and language ability at T1 would predict gaining an ASD diagnosis, because these abilities would be expected to be less developed than those who never met criteria for ASD. Although the NON-ASD group's T1 ASD symptoms were at subclinical levels, it was expected that higher subthreshold ASD symptom severity would likely predict transition from non-ASD to ASD. Finally, it was expected that individuals with less time spent in intervention services and those from lower SES background would be more likely to gain an ASD diagnosis after not initially meeting criteria for a diagnosis because these children would have had less access to enriching environments and educational resources. The role of race in this transition was unclear based on previous literature, so analysis of this predictor was exploratory. It was expected that the same variables from the categorical analysis would contribute to a significant increase in ASD symptom severity between T1 and T2, and not only the categorical change from non-ASD to ASD.

#### **1.11.4 Hypothesis 2B: Diagnostic transition from GDD to ASD**

The characteristics of toddlers who changed diagnostic category from GDD at a two-year-old evaluation (T1) to ASD at a four-year-old evaluation (T2) were examined qualitatively due to small sample size of this rare outcome (GDD-ASD). It was expected that ASD symptom severity of the children in the GDD-ASD group would differ from that of the GDD-GDD group, even though they would have had sub-threshold levels of ASD symptoms. However, it was unclear whether language ability would differ between this group and the GDD-GDD group, because comparisons of language ability between toddlers with GDD and those with ASD had been variable. Furthermore, the relationship of race and SES on the transition from GDD to ASD was unclear. The current study describes these characteristics in the GDD-ASD and GDD-GDD groups.

**Table 1. Methodological summary of studies examining diagnostic stability of ASD in toddlers**

| Study                   | T1 age    | T2 age          | N   | Sample                          | Measures used                 |                               |   |          | N of NON-ASD group | N of ASD-NON group | % Stable ASD diagnosis |
|-------------------------|-----------|-----------------|-----|---------------------------------|-------------------------------|-------------------------------|---|----------|--------------------|--------------------|------------------------|
|                         |           |                 |     |                                 | Cog.                          | Lang.                         | ASD sx                                  | Adaptive |                    |                    |                        |
| Charman et al., 2005    | 2         | 3, 7            | 26  | Screened using CHAT or referred | GMDS                          | CDI at 2/3; RDLS at 7         | ADI-R; ESCS at 2; ADOS-G at 7           | VABS     | 0                  | 1                  | 96%                    |
| Chawarska et al., 2007  | 14-25 mo. | 3               | 31  | Clinic-referred                 | MSEL                          | MSEL                          | ADOS-G; ADI-R                           | -        | 1                  | 0                  | 100%                   |
| Cox et al., 1999        | 20 mo.    | 42 mo.          | 46  | Screened using CHAT             | GSID or LIPS                  | RDLS                          | ADI-R, DSM-IV                           | -        | 9                  | 0                  | 80%                    |
| Eaves & Ho, 2004        | 2.5       | 4.5             | 49  | Clinic-referred                 | Bayley at 2.5; WPPSI-R at 4.5 | Bayley at 2.5; WPPSI-R at 4.5 | CARS; DSM-IV                            | VABS     | 0                  | 3                  | 94%                    |
| Gillberg et al., 1990   | 1-3       | 3-11 (mostly 5) | 28  | Clinic-referred                 | GSID                          | GSID                          | ABC; unnamed questionnaire; DSM-III-R   | -        | 0                  | 5                  | 82%                    |
| Guthrie et al., 2013    | 20 mo.    | 36 mo.          | 82  | Screened by FIRST WORDS project | MSEL                          | MSEL                          | ADOS; ESAC; RBS; SCQ if >36 mo.; DSM-IV | VABS     | 0                  | 14                 | 84%                    |
| Hedvall et al., 2013    | 2.5       | 4               | 195 | Clinic-referred                 | GSID, WPPSI-III if over 3     | PARIS Schedule                | PARIS Schedule, DISCO 10, ABC           | VABS-II  | 0                  | 21                 | 90%                    |
| Kleinman et al., 2008   | 2         | 4               | 77  | Screened using M-CHAT           | MSEL; Bayley; DAS             | MSEL; Bayley; DAS             | ADI-R; ADOS; CARS; DSM-IV               | VABS     | 0                  | 15                 | 80%                    |
| Soke et al., 2011       | 2-4       | 4-6             | 28  | Community-based                 | MSEL                          | MSEL                          | ADI-R; ADOS-G                           | -        | 2                  | 3                  | 82%                    |
| van Daalen et al., 2009 | 26 mo.    | 45 mo.          | 131 | Screened using ESAT             | MSEL; Bayley                  | MSEL; Bayley                  | SEEC; WADIC; ADOS-G; ADI-R; DSM-IV      | -        | 2                  | 7                  | 84%                    |

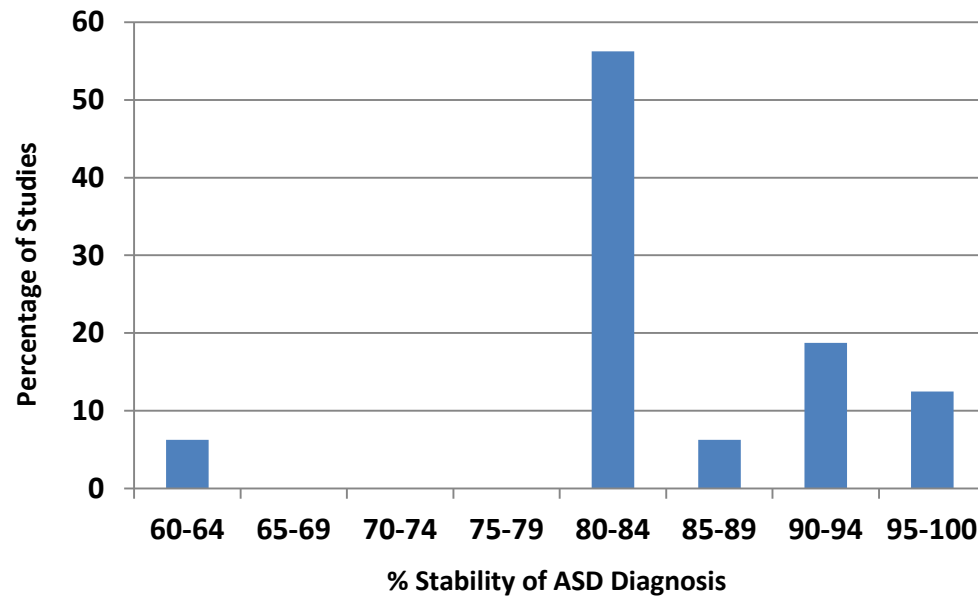
Note. Cog.=Cognitive; Lang.=Language; ASD sx=ASD symptoms. See Table 2 for Lord, 1995; Pellicano, 2012; Sutera et al., 2007; Turner & Stone, 2007; Wiggins et al., 2012. Charman et al., 2005 studied participants at three time points. ABC=Autism Behavior Checklist; ADI=Autism Diagnostic Interview; ADI-R=Autism Diagnostic Interview-Revised; ADOS=Autism Diagnostic Observation Schedule; ADOS-G=ADOS-General;

Bayley=Bayley Scales of Infant Development; CARS=Childhood Autism Rating Scale; CDI=MacArthur Communicative Development Inventory; DAS=Differential Abilities Scale; DISCO 10= Diagnostic Interview for Social and Communication disorders; ESAC=Early Screening for Autism and Communication Disorders; ESAT=Early Screening of Autistic Traits; ESCS=Early Social Communication Scales; GMDS=Griffiths Mental Development Scales; GSID=Griffiths Scale of Infant Development; LIPS=Leiter International Performance Scale; MSEL=Mullen Scales of Early Learning; PARIS Schedule= Paris Autism Research In Sib-pairs, includes DSM-IV-TR and language and behavioral ratings; RBS=Repetitive Behavior Scale; RDLS=Reynell Developmental Language Scales – III; SCQ=Social Communication Questionnaire; SEEC=Vineland Social-Emotional Early Childhood Scales; VABS=Vineland Adaptive Behavior Scales; WADIC=Wing Autistic Disorder Interview Checklist; WPPSI-R=Wechsler Preschool and Primary Scale of Intelligence- Revised.

**Table 2. Summary of results of studies examining predictors of diagnostic instability of ASD**

| Study                      | T1 age | T2 age | N   | Sample                        | Measures used                         | % Stable ASD diagnosis | Cognitive ability | Language ability | ASD symptom severity | Motor ability | SES | Race | Intervention |
|----------------------------|--------|--------|-----|-------------------------------|---------------------------------------|------------------------|-------------------|------------------|----------------------|---------------|-----|------|--------------|
| Stone et al., 1999         | 2      | 3      | 37  | Clinic-referred               | CARS, PL-ADOS, DSM-IV                 | 84%                    | -                 | -                | Y                    | -             | -   | -    | -            |
| Lord, 1995                 | 2      | 3      | 30  | Clinic-referred               | ADI, CARS                             | 90%                    | -                 | -                | Y                    | -             | -   | -    | -            |
| Pellicano, 2012            | 5      | 8      | 37  | Recruited                     | SCQ, ADOS-G                           | 81%                    | N                 | N                | N                    | -             | -   | -    | Y            |
| Sutera et al., 2007        | 2      | 4      | 90  | Screened using M-CHAT         | ADI-R, ADOS, CARS, VABS, MSEL         | 82%                    | N                 | N                | N                    | Y             | -   | -    | -            |
| Turner & Stone, 2007       | 2      | 4      | 48  | Clinic-referred               | ADOS-G at T1, ADI-R at T2, CARS, MSEL | 63%                    | Y                 | -                | Y                    | -             | N   | N    | N            |
| Orinstein et al., in press | N/A    | 8-21   | 58  | Clinic-referred and recruited | WAIS, VABS, ADOS, ADI-R               | N/A                    | -                 | -                | -                    | -             | -   | -    | Y            |
| Wiggins et al, 2012        | 2      | 4      | 136 | Screened using M-CHAT         | ADI-R, ADOS, VABS, MSEL               | 83%                    | -                 | -                | Y                    | -             | -   | -    | -            |

Note. Y indicates that a variable was found to be predictive of ASD-NON transition; N indicates that a variable was not found to be predictive of ASD-NON transition; other variables were not noted as having been examined in the given studies. Orinstein et al., 2014 is not longitudinal, but compares a group with high functioning ASD to a group with previous but no current ASD diagnosis. Therefore, it was not included in Figure 1. ADI=Autism Diagnostic Interview; ADI-R=Autism Diagnostic Interview-Revised; ADOS=Autism Diagnostic Observation Schedule; ADOS-G=ADOS-General; PL-ADOS=Pre-Linguistic ADOS; CARS=Childhood Autism Rating Scale; MSEL=Mullen Scales of Early Learning; SCQ=Social Communication Questionnaire; VABS=Vineland Adaptive Behavior Scales; WAIS=Weschler Adult Intelligence Scale.



Note. N=16 studies, which are listed in Tables 1 and 2.

**Figure 1.** The percentage of studies that have found specific rates of diagnostic stability of ASD from toddlerhood



## 2 METHODS

### 2.1 Participants

Participants were recruited from the metro-Atlanta area and Connecticut as part of a larger ongoing cross-validation study investigating an early screening measure for ASD, the Modified Checklist for Autism in Toddlers (M-CHAT or M-CHAT-Revised; Chlebowski et al., 2013; Robins, Fein, Barton, & Green, 2001; Robins et al., 2014). Participants were included if they screened positive on the M-CHAT(-R) and Follow-Up Interview, or if they were flagged by pediatricians or service providers for having possible ASD at around age two and were not found to be typically developing. Participants were excluded if they already had a diagnosis of ASD prior to screening, if they were younger than 16 months, if the family did not speak fluent English (or Spanish, when available), if the child was physically unable to complete the evaluation, or if they did not screen positive on the M-CHAT(-R) and Follow-Up (i.e., they were only flagged by a pediatrician) and were found to be typically developing.

The sample consisted of 424 toddlers who were evaluated at around two years old ( $M=26.31$  months,  $SD=4.46$ ) and then again at around four years old ( $M=49.88$  months,  $SD=7.19$ ; see Table 3). Of the 424 children in the sample, 318 were males (75.0%). The sample was racially/ethnically diverse, which maximizes generalizability of results. Specifically, 288 identified as Caucasian (67.9%), 63 as African American (14.8%), 38 as Hispanic/Latino (9.0%), 14 as Asian (3.3%), 18 as Bi/Multiracial (4.2%), and three as Other or unreported (.7%). In terms of caregiver's level of education, 134 (31.6%) attended or completed high school and/or vocational or technical school, 156 (36.8%) attended college or earned a college degree, and 98 (23.1%) attended graduate school or earned an advanced degree. Data on education level were unavailable for 36 participants (8.5%).

The specific diagnoses for each group over time within the category of non-ASD are presented in Table 3 and Figure 2. The sample included a mixture of children who, based on their T1 evaluation, had a diagnosis of ASD ( $N=270$ ), GDD ( $N=60$ ), Developmental Language Delay ( $N=34$ ), other diagnoses ( $N=6$ ),

no diagnosis (if they were not designated typically developing but did not meet criteria for a specific diagnosis, N=35), or typical development (N=19). By T2, there were 238 children with ASD, 58 with GDD, 13 with Developmental Language Disorder, one with Attention-Deficit/Hyperactivity Disorder (ADHD), 19 with other diagnoses, 78 without a diagnosis (but not designated as typically developing), and 17 considered to be typically developing. For the purpose of maintaining large sample sizes, sufficient power for analyses, and examination of transition between ASD and non-ASD, those with GDD, other diagnoses, and no diagnosis were combined into a non-ASD group (with the exception of analyses focusing on ASD-GDD or GDD-ASD transition).

The majority of children maintained their diagnostic categories over time (N=354, 83.5%); 219 met criteria for ASD at both T1 and T2, and 135 children did not meet criteria for ASD at both T1 and T2. However, several children changed diagnosis between T1 and T2 (see Figure 2 for diagrams of group sample sizes). Specifically, of the 270 children who met criteria for ASD at T1, some no longer met ASD criteria at T2 (ASD-NON; N=51; 18.9%), 17 of whom met criteria for GDD rather than ASD at T2 (ASD-GDD; 6.3%). Also, of the 154 who did not initially meet criteria for ASD, 19 children newly met criteria at T2 (NON-ASD; 12.3%), eight of whom changed from a GDD to ASD diagnosis (GDD-ASD). Also, of the 60 toddlers diagnosed with GDD at T1, 28 (46.7%) were not diagnosed with GDD at T2. Of these 28 participants who no longer met criteria for GDD at T2, eight were diagnosed with ASD, four were diagnosed with Developmental Language Delay (DLD), 13 did not receive any diagnosis, and three were classified in the category of "Other diagnosis."

## **2.2 Measures**

### **2.2.1 Autism Diagnostic Observation Schedule**

The Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1989) and ADOS-2 (Rutter et al., 2012) are semi-structured interactive play-based assessments of a child's communication, social interactions, play skills, imagination, and repetitive or stereotyped behaviors. Four different ADOS

modules are available based on the examinee's age and language ability. The same modules are available for ADOS-2, as well as a Toddler Module. Depending on their language ability, all participants were tested using either Module 1 (little or no speech) or Module 2 (phrase speech). The assessment took between 30 and 60 minutes to administer.

For both ADOS and ADOS-2, children receive a score for each item ranging from 0 (no abnormality noted) to 3 (abnormality clearly present) based on structured interactions with the examiner. The scores of specific items are transferred to an algorithm (with all 3's converted to 2's), where they are summed to yield overall scores. The ADOS algorithm produces a social relatedness score, communication score, a restricted/repetitive behaviors score, and a total score. In contrast, the ADOS-2 yields a social affect score, a restricted/repetitive behavior score, and a total score. Based on the total score, a child is considered to be in a mild, moderate, or severe concern for ASD range (with higher scores indicating higher risk).

The original validation sample for the ADOS consisted of 381 children, including children with autistic disorder (one half of sample), other ASD (one third), and non-spectrum disorders (one sixth). Participants varied in terms of race and sex (2:1 male to female ratio). The Module 1 sample consisted of 190 participants from 15 months to 11 years of age, and the Module 2 sample consisted of 111 participants from two to eight years of age (Lord et al., 2012). Inter-rater reliability was acceptable, with weighted kappas ranging from .61 to .92 per item, or .58 to .87 overall. Test-retest reliability was also adequate for test items (.57 to .84) and overall ratings (.58 to .92). For the algorithm, inter-rater reliability ranged from .75 to .96 and test retest reliability from .70 to .92 (Lord et al., 1989). The ADOS-2 extended validation sample consisted of 1,139 participants aged 14 months to 16 years, 912 (56%) of which had autistic disorder, 439 (27%) had other ASD, and 279 (17%) had non-ASD conditions. The sample was racially diverse, and sex ranged from 57 to 86% male, depending on the module. ADOS-2 algorithms were also replicated with 1,259 participants. ADOS-2 exhibited strong psychometric

properties, with most individual item inter-rater reliability kappas exceeding .60, 95 percent (Module 1) to 98 percent (Module 2) agreement of classification between raters, and overall intraclass correlation of .90 for algorithm test-retest reliability (Lord, Luyster, Gotham, & Guthrie, 2012; Lord et al., 2012).

An algorithm has recently been validated to create a Calibrated Severity Score (CSS), which allows for comparison over time and between modules (Gotham, Pickles, & Lord, 2009). This is important because the module appropriate for the child may change between T1 and T2, making direct comparisons difficult. The CSS was validated in a sample of 2,195 racially diverse children aged two to 16 years with ASD and non-ASD developmental delays (Gotham, Pickles, & Lord, 2009). The CSS has been found to be fairly stable over 12 to 24 months, with 62 to 64 percent of children staying within one point of their original score (range of possible scores: 0 to 10; Shumway et al., 2012). The CSS has also been used and validated in a large number of studies on child development and diagnostic stability since its development (e.g., Dawson et al., 2010; Gotham, Pickles, & Lord, 2012; Shumway et al., 2012). The CSS (referred to henceforth as ADOS CSS regardless of version and module) was used as a continuous measure of diagnostic symptom severity in the current study. It is important to note that the ADOS CSS was validated in individuals ranging from two to 16 years old; therefore, scores for children under two in the current study were determined using the same protocol as for two-year-olds.

### **2.2.2 Mullen Scales of Early Learning**

The Mullen Scales of Early Learning (MSEL; Mullen, 1995) is an interactive structured assessment of cognitive, language, and motor abilities in children from birth to 68 months of age. It is composed of five subscales: visual reception, receptive language, expressive language, fine motor, and gross motor skills. The MSEL is administered in about 45 to 85 minutes. Children are administered items at different start points based on their ages, and continue each subscale until a basal of three consecutive zero-rated items is achieved. Each item is rated on a scale from 0 (item not completed) to 1 through 5, depending on the item; the maximum possible score for each item varied.

Raw scores were converted to T scores using age-based norms for each subscale. The early learning composite (ELC) score included all subscales except gross motor, as per the MSEL manual. T scores were converted to a developmental quotient (DQ; age equivalent/chronological age x 100) to eliminate floor effects, because a large proportion of children had T scores of 20. The visual reception scale DQ was used as a continuous measure of nonverbal cognitive ability, and the receptive and expressive language scale DQs were used as continuous measures of language ability.

The normative sample was based on 1,849 children ranging from two days to 69 months old, with demographic distributions based on 1990 US Census data in terms of sex, race/ethnicity, community size, and father's occupation. There were between 84 and 156 children in each age group. The MSEL shows strong psychometric properties, including strong convergent validity with similar measures of cognitive ability in young children (Pearson's  $r$  of .53 to .70; Bishop, Guthrie, Coffing, & Lord, 2011), internal consistency (Chronbach's alpha of .75 to .91), test-retest reliability (.82 to .96), and inter-rater reliability (.91 to .99; Bradley-Johnson, 1997).

### **2.2.3 Vineland Adaptive Behavior Scales (Second Edition)**

The Vineland Adaptive Behavior Scales, Survey Interview Form (VABS; Sparrow, Balla, & Cicchetti, 1984) and VABS, Second Edition Survey Interview Form (VABS-II; Sparrow, Cicchetti, & Balla, 2005) are semi-structured parent interviews used to assess adaptive functioning for individuals from infancy through age 89 in the domains of communication, daily living, socialization, and motor skills, as well as an overall adaptive behavior composite score. These measures give insight into a child's typical daily behaviors as reported by the parent. They can help to determine whether behaviors observed during the assessment are typical of the child and can also give additional information about a child's behavior which may not be available through brief direct observation. The VABS and VABS-II each take between 45 and 60 minutes to administer. The two versions are similar, but the VABS-II was developed

in order to be more culturally sensitive, to incorporate recent research regarding developmental disabilities, and to better assess impairments in older individuals.

Each item is rated as a behavior *rarely or never* completed by the child (0), *sometimes* completed by the child (1), or *usually* completed by the child (2). Parents were administered items at different start points based on their child's age, and continued each subscale until basal was achieved. The measure yields standard scores using age-based norms for each domain and for an overall adaptive behavior composite (ABC). In this study, the Communication Scale standard score was used as a measure of parent-reported language ability, and the Socialization Scale standard score was used as a measure of social skill ability.

The standardization sample for the VABS-II consisted of 3,695 individuals ranging in age from birth to 90, with a larger concentration of individuals from birth to age five. The sample was distributed based on the 2001 US Census data in the areas of race/ethnicity, SES, geographic region, community size, and special education placement. The VABS-II has been shown to have high levels of internal consistency (in the mid- to high-90s), test-retest reliability (low .80s to mid .90s), concurrent validity (mostly .80s and .90s), and modest inter-rater reliability (mid .40s to mid .60s; Sparrow, Cicchetti, & Balla, 2005).

#### **2.2.4 Demographics and History Forms**

Demographics and participant history forms were collected from each participant in order to record background information such as date of birth, age, sex, race/ethnicity, parental education level (SES), presence of parental concerns, and known medical or developmental conditions.

#### **2.2.5 Intervention Data**

Intervention data were collected in terms of number of hours of intervention per week in six-month intervals between T1 and T2. These numbers of hours were summed across different

intervention types, and then averaged across the four time-intervals to yield the average number of intervention hours received per week.

### **2.3 Procedure**

Parents were recruited for the screening study during their child's well visit to the pediatrician at around age two, during participation in an early intervention program, or if they were referred by a psychologist or for having a sibling with ASD. Parents signed the consent form and completed the M-CHAT or M-CHAT-R, as well as the M-CHAT Follow-Up Interview if the child showed risk for developmental delays. If they continued to show risk for developmental delays after the Follow-Up Interview, or if a parent or pediatrician had a concern about developmental delay, children were invited for a free developmental evaluation at Georgia State University (GSU) or the University of Connecticut (UConn) as well as a follow-up evaluation at around age four. Participants' families were informed that they would be financially compensated and that they would receive feedback about their child's development.

The child and at least one caregiver came to GSU or UConn at T1 and informed consent was obtained. A team of graduate students and supervising clinicians completed child assessments and parent report measures, including the MSEL, VABS (-II), ADOS(-2), and either the Autism Diagnostic Interview, Revised (ADI-R; Le Couteur, Lord, & Rutter, 2003) or Toddler Autism Symptom Interview (a semi-structured interview developed by the research team to acquire parent report information on DSM-IV-TR symptoms of ASD in children under three years old). The graduate student and clinician then both completed the Childhood Autism Rating Scale (CARS, Schopler et al., 1980) or CARS, Second Edition, Standard Version (CARS2-ST; Schopler et al., 2010) together. Furthermore, the parent completed the Behavior Assessment System for Children, Second Edition (BASC-2; Reynolds & Kamphaus, 2004), which is not used in the current study. After completing these measures, the clinical team scored the measures and assigned a diagnosis to the child based on all available data, using DSM-

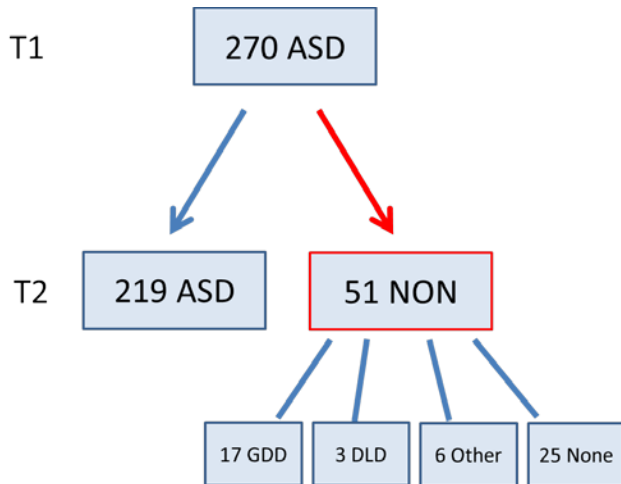
IV-TR symptom checklists. The clinical team then gave verbal feedback to the caregiver about the child's performance and diagnosis, and the family was compensated for their time. The caregiver received a detailed psychological report four to eight weeks later, describing the child's performance, as well as recommendations for intervention. At around age four, parents were again contacted for a re-evaluation (T2) consisting of the same diagnostic, adaptive, and cognitive measures. When the family came in for T2, the T1 procedure was repeated.



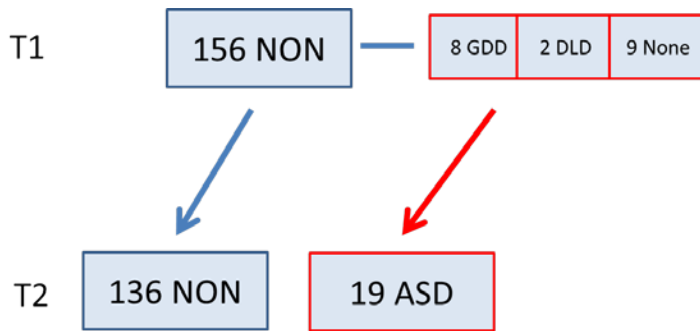
**Table 3. Sample characteristics by T1 and T2 diagnostic category**

| Characteristic      | ASD-ASD      | ASD-NON                | ASD-GDD <sup>A</sup> | NON-NON                | NON-ASD      | GDD-ASD <sup>A</sup> | All          |
|---------------------|--------------|------------------------|----------------------|------------------------|--------------|----------------------|--------------|
| Total N             | 219          | 51                     | 17                   | 135                    | 19           | 8                    | 424          |
| Child Age in Months |              |                        |                      |                        |              |                      |              |
| T1 M(SD)            | 26.46 (4.30) | 26.98 (5.22)           | 27.25 (5.58)         | 25.80 (4.36)           | 26.51 (4.69) | 25.86 (4.71)         | 26.31 (4.46) |
| T2 M(SD)            | 49.41 (6.60) | 50.75 (8.21)           | 49.51 (9.28)         | 50.20 (7.40)           | 50.59 (9.27) | 51.44 (10.72)        | 49.88 (7.19) |
| Child Sex           |              |                        |                      |                        |              |                      |              |
| Male                | 177 (80.8%)  | 35 (68.6%)             | 12 (70.6%)           | 92 (68.1%)             | 14 (73.7%)   | 7 (87.5%)            | 318 (75.0%)  |
| Female              | 42 (19.2%)   | 16 (31.4%)             | 5 (29.4%)            | 43 (31.9%)             | 5 (26.3%)    | 1 (12.5%)            | 106 (25.0%)  |
| Child Race          |              |                        |                      |                        |              |                      |              |
| White               | 150 (68.5%)  | 31 (60.8%)             | 5 (29.4%)            | 93 (68.9%)             | 14 (73.7%)   | 4 (50.0%)            | 288 (67.9%)  |
| Black               | 29 (13.2%)   | 10 (19.6%)             | 7 (41.2%)            | 21 (15.6%)             | 3 (15.8%)    | 3 (37.5%)            | 63 (14.8%)   |
| Other               | 40 (18.3%)   | 10 (19.6%)             | 5 (29.4%)            | 21 (15.6%)             | 2 (10.5%)    | 1 (12.5%)            | 73 (17.2%)   |
| Mother's Education  |              |                        |                      |                        |              |                      |              |
| High School         | 67 (30.6%)   | 21 (41.2%)             | 10 (58.8%)           | 43 (31.9%)             | 3 (15.8%)    | 1 (12.5%)            | 134 (31.6%)  |
| Beyond High School  | 134 (61.2%)  | 27 (52.9%)             | 6 (35.3%)            | 78 (57.8%)             | 15 (78.9%)   | 6 (75.0%)            | 254 (59.9%)  |
| Data unavailable    | 18 (8.2%)    | 3 (5.9%)               | 1 (5.9%)             | 14 (10.4%)             | 1 (5.3%)     | 1 (12.5%)            | 36 (8.5%)    |
| ADOS CSS            |              |                        |                      |                        |              |                      |              |
| T1 M(SD)            | 6.54 (2.04)  | 5.31 (2.28)            | 5.00 (1.93)          | 2.03 (1.48)            | 2.79 (1.69)  | 2.63 (1.77)          | 4.79 (2.79)  |
| T2 M(SD)            | 6.38 (2.17)  | 1.77 (1.16)            | 2.18 (1.24)          | 1.48 (.94)             | 5.21 (1.90)  | 5.50 (1.77)          | 4.21 (2.93)  |
| T1 diagnoses        |              |                        |                      |                        |              |                      |              |
| ASD                 | 219 (100%)   | 51 (100%)              | 17 (100%)            | 0                      | 0            | 0                    | 270 (63.7%)  |
| GDD                 | 0            | 0                      | 0                    | 52 (38.5%)             | 8 (42.1%)    | 9 (100%)             | 60 (14.2%)   |
| Language delay      | 0            | 0                      | 0                    | 32 (23.7%)             | 2 (10.5%)    | 0                    | 34 (8.0%)    |
| Other diagnosis     | 0            | 0                      | 0                    | 6 <sup>A</sup> (4.4%)  | 0            | 0                    | 6 (1.4%)     |
| No diagnosis        | 0            | 0                      | 0                    | 28 (20.7%)             | 7 (36.8%)    | 0                    | 35 (8.3%)    |
| Typical Dev.        | 0            | 0                      | 0                    | 17 (12.6%)             | 2 (10.5%)    | 0                    | 19 (4.5%)    |
| T2 diagnoses        |              |                        |                      |                        |              |                      |              |
| ASD                 | 219 (100%)   | 0                      | 0                    | 0                      | 19 (100%)    | 9 (100%)             | 238 (56.1%)  |
| GDD                 | 0            | 17 (33.3%)             | 17 (100%)            | 41 (30.4%)             | 0            | 0                    | 58 (13.7%)   |
| Language delay      | 0            | 3 (5.9%)               | 0                    | 10 (7.4%)              | 0            | 0                    | 13 (3.1%)    |
| ADHD                | 0            | 0                      | 0                    | 2 (1.5%)               | 0            | 0                    | 2 (.5%)      |
| Other diagnosis     | 0            | 6 <sup>B</sup> (11.8%) | 0                    | 12 <sup>C</sup> (8.9%) | 0            | 0                    | 18 (4.2%)    |
| No diagnosis        | 0            | 25 (49.0%)             | 0                    | 54 (40.0%)             | 0            | 0                    | 79 (18.6%)   |
| Typical Dev.        | 0            | 0                      | 0                    | 16 (11.9%)             | 0            | 0                    | 16 (3.8%)    |

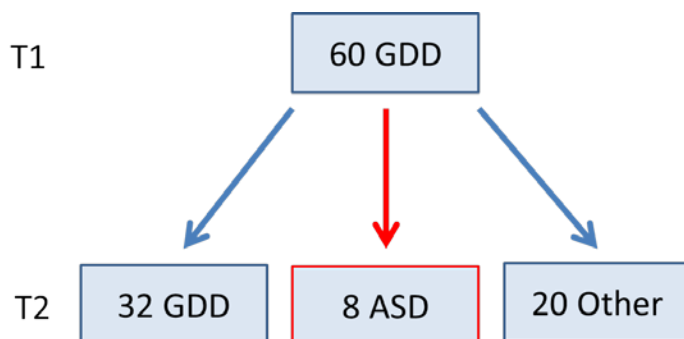
Note. ^These groups are subsets of the previous groups (e.g., ASD-NON group contains the ASD-GDD group). Other diagnoses include: ^1 Epilepsy, 1 Phonological Disorder; <sup>B</sup>1 Articulation Disorder; <sup>C</sup>2 Phonological Disorder, 1 Developmental Coordination Disorder, 1 Motor Delay; all others are unspecified. Typical Dev.= Typical Development. ADOS CSS= Autism Diagnostic Observation Schedule Calibrated Severity Scale



A. Sample sizes for ASD-ASD, ASD-NON, and ASD-GDD groups. Note. "NON" refers to non-ASD.



B. Sample sizes for NON-NON and NON-ASD groups, with specific T1 diagnoses for NON-ASD group specified. Note. "NON" refers to non-ASD.



C. Sample sizes for GDD-GDD and GDD-ASD groups. Note. Other group includes participants without GDD or ASD who were diagnosed with GDD at T1.

**Figure 2. Visual representations of sample sizes of each group of participants**

### 3 RESULTS

All analyses were completed using IBM SPSS Statistics, Version 20. Descriptive data obtained during each evaluation were used to test hypotheses 1A, 1B, and 2A using separate logistic regression models with the dichotomous outcome of diagnostic stability or diagnostic instability. These hypotheses were then each tested using linear regression models with the continuous outcome of the change in symptom severity between T1 and T2. The predictors in each linear regression model mirrored those used in the logistic regression models for ease of comparison between using the dichotomous outcome of diagnostic instability, and the corresponding change in continuous symptom severity regardless of categorical change.

Possible predictors in each model included sex; a continuous measure of nonverbal cognitive functioning (MSEL visual reception DQ); continuous measures of language ability (MSEL expressive language DQ, MSEL receptive language DQ, VABS(-II) communication scores); a continuous measure of ASD symptom severity (ADOS CSS); a continuous measure of parent-reported social interaction skills; a continuous measure of weekly intervention hours participated in between T1 and T2 (collapsed across intervention types; see Table 4); a dichotomous measure of SES (caregiver-reported education level; college or graduate education, or less than college education ); and a dichotomous measure of race (majority or minority). For the regression model variable selection, continuous predictors significantly related to the outcome were identified using point biserial correlations. Phi coefficients were used to identify categorical predictors related to the dichotomous outcome, and Fisher's exact test  $p$ -values were used when expected cell values were less than five. Final models included the predictors explaining variance in the outcome.

Power analyses for logistic regression were completed as described in Hsieh, Bloch, and Larsen (1998). For hypothesis 1A, given the sample size ( $N=270$ ), proportion of the sample who changed diagnosis from ASD to non-ASD (18.9%), and desired power of .8, the lowest detectable odds ratio (OR)

for one standard deviation of change in the predictor is 1.55. For hypothesis 1B, given the sample size (N=270), proportion who change diagnosis from ASD to GDD (6.3%), and power of .8, the lowest detectable OR is 2.02. For hypothesis 2A, given the sample size (N=154), proportion who change diagnosis from non-ASD to ASD (12.3%), and power of .8, the lowest detectable OR is 1.99. Therefore, current sample sizes yield adequate power for detecting moderate effect sizes similar to those found in previous literature (Fountain, Winter, & Bearman, 2012; Guthrie et al., 2013; Sutera et al., 2007). Logistic regression results with dichotomous diagnostic outcomes were first examined, followed by linear regression results with change in continuous symptom severity as outcomes.

### **3.1 Predictors of diagnostic instability**

Assumptions of logistic regression include linearity and independence of errors, neither of which were violated in the following logistic regression models.

#### **3.1.1 Hypothesis 1A: ASD to non-ASD**

A logistic regression model including relevant predictors was developed for hypothesis 1A. The sample included all participants with T1 ASD diagnosis, and the outcome was a diagnosis of ASD (ASD-ASD) or non-ASD (ASD-NON). Predictors were selected based on correlations (see Table 5) and contingency coefficients (see Table 6) between each variable and the outcome (T2 Non-ASD status).

Based on these correlations and contingency coefficients, the initial model included ADOS CSS (Pearson's  $r = -.225$ ,  $p < .001$ ), MSEL Expressive Language (Pearson's  $r = .147$ ,  $p = .018$ ), MSEL Receptive Language (Pearson's  $r = .207$ ,  $p = .001$ ), MSEL Visual Reception (Pearson's  $r = .189$ ,  $p = .002$ ), VABS(-II) Communication (Pearson's  $r = .145$ ,  $p = .017$ ), VABS(-II) Socialization (Pearson's  $r = .177$ ,  $p = .004$ ), and number of intervention hours (Pearson's  $r = -.141$ ,  $p = .023$ ). Two possible final models resulted; one included ADOS CSS, number of intervention hours, and MSEL receptive language, whereas the other model included ADOS CSS, number of intervention hours, and MSEL visual reception. MSEL receptive language and MSEL visual reception were significantly correlated (Pearson's  $r = .553$ ,  $p < .001$ ) and had

significant overlapping variance, and both models had similar odds ratios and  $p$ -values for all three variables. Therefore, the model using MSEL receptive language ( $p=.025$ ) was selected as the final model, as it had a lower  $p$ -value compared to MSEL visual reception ( $p=.032$ ). However, it is important to note that either variable was significant in the model without affecting the odds ratios and significance of the other variables.

Therefore, the final model included ADOS CSS, MSEL Receptive Language, and number of intervention hours (see Table 7). In the sample of 253 participants, ADOS CSS was predictive of the outcome such that higher T1 ADOS CSS led to lower odds of transitioning from ASD to non-ASD,  $OR=.816$ ,  $p=.014$ . Receptive language ability was predictive of diagnostic transition; higher T1 MSEL receptive language DQ score was associated with greater likelihood of transitioning,  $OR=1.016$ ,  $p=.025$ . Finally, hours of intervention received predicted diagnostic transition, as greater participation in intervention was associated with reduced odds of changing from ASD to non-ASD,  $OR=.949$ ,  $p=.032$ . The other cognitive and language measures were not predictive of diagnostic change,  $ps>.05$ .

Several mechanisms were explored which could correspond with the finding that more hours of participation in intervention were associated with smaller likelihood of transitioning from ASD to non-ASD. To test the hypothesis that those who participate in these services had more severe symptoms and were thus less likely to transition, the correlation between intervention hours and ADOS CSS was examined. Number of intervention hours and ASD symptom severity were not significantly correlated, Pearson's  $r=.075$ ,  $p=.228$ . Furthermore, maternal education did not predict number of intervention hours a participant engaged in,  $B=1.661$ ,  $p=.164$ . However, differences in participation in intervention between the ASD-ASD and ASD-NON groups were examined at each of four time points between evaluations (see Figure 3); they were found to be initially similar ( $t(65.23)=-.996$ ,  $p=.323$ ), but diverged over time and were significantly different as children approached T2,  $t(92.56)=-2.905$ ,  $p=.005$ . Change in intervention from two to 3.5 years also differed by group,  $B=-3.002$ ,  $p=.009$ .

### 3.1.2 Hypothesis 1B: ASD to GDD

For hypothesis 1B, the logistic regression model included all participants who had ASD at T1 and a T2 diagnosis of GDD (ASD-GDD) or ASD (ASD-ASD). Based on correlations (see Table 8) and contingency coefficients (see Table 6), the original model included ADOS CSS (Pearson's  $r=-.193$ ,  $p=.003$ ), VABS(-II) socialization skills (Pearson's  $r=.153$ ,  $p=.019$ ), intervention hours (Pearson's  $r=-.183$ ,  $p=.006$ ), maternal education ( $\Phi=-.159$ ,  $p=.019$ ), and race ( $\Phi=.213$ ,  $p=.001$ ) as predictors; the final model included all of these variables except VABS(-II) socialization skills (see Table 9). In the sample of 208 participants, ADOS CSS was a significant predictor, as higher ADOS CSS contributed to greater odds of diagnostic transition from ASD to GDD,  $OR=.658$ ,  $p=.005$ . Maternal education was a significant predictor; having a parent who attended college and/or graduate school led to lower odds of changing from ASD to GDD,  $OR=.291$ ,  $p=.034$ . Finally, number of hours of intervention was a significant predictor, as more hours of intervention led to smaller likelihood of changing diagnosis,  $OR=.851$ ,  $p=.039$ . Race was not a significant predictor, though its  $p$ -value approached significance; belonging to a minority race was associated with greater odds of transitioning from ASD to GDD,  $OR=3.342$ ,  $p=.051$ . Post-hoc analysis revealed that maternal education level was not significantly correlated with race ( $\Phi=-.081$ ,  $p=.230$ ). Also, group differences in change in participation in intervention over time (see Figure 4) were not significant,  $B=-2.85$ ,  $p=.159$ , though comparisons between ASD-ASD and ASD-GDD groups at each time point had moderate to high effect sizes (age two Cohen's  $d=-.54$ , age 2.5 Cohen's  $d=-.61$ , age three Cohen's  $d=-.78$ , age 3.5 Cohen's  $d=-.72$ ).

### 3.1.3 Hypothesis 2A: non-ASD to ASD

For hypothesis 2A, the logistic regression model included all participants who did not have ASD at T1, and the outcome was a diagnosis of ASD (NON-ASD) or non-ASD (NON-NON). Based on correlations (see Table 10) and contingency coefficients (see Table 6), ADOS CSS (Pearson's  $r=.165$ ,  $p=.041$ ) and hours of intervention (Pearson's  $r=.193$ ,  $p=.018$ ) were included in the model. Both variables

were significant predictors of transitioning from non-ASD to ASD, and were included in the final model (see Table 11). In particular, T1 ASD CSS led to greater odds of transitioning,  $OR=1.397$ ,  $p=.022$ , and more participation in intervention contributed to greater odds of transitioning from non-ASD to ASD,  $OR=1.072$ ,  $p=.030$ . At age two, using the Mann Whitney U test, the NON-NON and NON-ASD groups participated in a similarly low number of intervention hours ( $p=.898$ ), though their participation diverged by age 3.5 ( $p=.065$ , Cohen's  $d=-.51$ ; see Figure 5).

### **3.1.4 Hypothesis 2B: GDD to ASD**

Hypothesis 2B, comparing children who changed from a GDD to an ASD diagnosis to those who maintained GDD diagnosis, was examined qualitatively due to small sample sizes. See Table 3 for demographic characteristics of GDD-GDD and GDD-ASD groups. As expected, the GDD-GDD group generally had lower ADOS CSS scores at both T1 and T2 than the GDD-ASD group (see Figure 6). Similarly, ADOS CSS in the GDD-GDD group tended to stay consistent or decrease over time, in contrast with the majority of the GDD-ASD group who experienced an increase in ADOS CSS (see Figure 7). In terms of T1 language, communication, and social skills (see Figures 8-11), the GDD-ASD group appeared to have fewer individuals with high receptive language ability (see Figure 8). In contrast, there were no apparent differences between groups in T1 expressive language ability (see Figure 9) or on a T1 measure of parent reported communication ability (see Figure 10). Finally, a measure of nonverbal cognitive ability also did not appear to distinguish between the GDD-ASD and GDD-GDD groups (see Figure 12).

Post-hoc analyses were used to explore possible differences between the group that maintained GDD diagnosis ( $N=32$ ) and the group that changed from GDD to non-GDD ( $N=28$ ), as the diagnosis was only 53.3 percent stable. Those who maintained GDD diagnosis did not differ from those who no longer met GDD criteria in terms of T1 nonverbal problem-solving ability (MSEL Visual Reception;  $t(50.64)=1.013$ ,  $p=.316$ ), expressive language ability (MSEL Expressive Language;  $t(50.35)=-.173$ ,  $p=.864$ ),



receptive language ability (MSEL Receptive Language;  $t(51.17) = .005$ ,  $p = .996$ ), or fine motor ability (MSEL Fine Motor;  $t(53.58) = 1.886$ ,  $p = .065$ , Cohen's  $d = .498$ ).

### **3.2 Predictors of change in continuous ASD symptom severity**

Linear regression models with change in ASD symptom severity as the outcome were completed, using the same predictors as in the previous logistic regression models. Assumptions for linear regression include linearity, homoschedasticity, normality, and independence of observations. Normality was assessed using Q-Q plots and histograms of standardized residuals, whereas linearity and homoschedasticity were assessed using partial residual plots. Only the regression corresponding to hypothesis 2A did not meet the assumption of normality; bootstrapping methods were implemented as described below. All other analyses met all assumptions. All analyses included a measure of the difference between T1 and T2 ASD symptom severity (T1 ADOS CSS minus T2 ADOS CSS) as a continuous outcome.

#### **3.2.1 Hypothesis 1A: ASD to non-ASD**

Similar to hypothesis 1A, a linear regression model was developed with T1 ADOS CSS, T1 MSEL receptive language DQ, and hours of intervention services included as predictors (see Table 12). The sample included all participants with an ASD diagnosis at T1, including those who maintained and those who did not maintain ASD diagnosis. Consistent with the logistic regression results, ADOS CSS was predictive of the outcome, such that a one-point decrease in T1 symptom severity was associated with a .66 unit decrease in symptom severity between T1 and T2,  $B = .661$ ,  $p < .001$ . Also, a one standard deviation increase in MSEL receptive language DQ was indicative of a .21 unit decrease in ADOS CSS over time,  $B = .025$ ,  $\beta = .211$ ,  $p = .001$ . However, number of hours of intervention services was not a significant predictor of change in ASD symptom severity at T2,  $B = -.034$ ,  $p = .075$ .

### **3.2.2 Hypothesis 1B: ASD to GDD**

To correspond with hypothesis 1B, a linear regression was used including all ASD-ASD participants and all ASD-GDD participants. The model included all variables from the final logistic regression model, including T1 ADOS CSS, number of intervention hours, race, and maternal education (see Table 13). Only T1 ADOS CSS was a significant predictor of change in ASD symptom severity, as for every one-point increase in T1 severity there was a .57 point decrease in severity over time,  $B=.566$ ,  $p=.005$ . The remaining variables were not predictive of ASD symptom severity change,  $ps>.05$ .

### **3.2.3 Hypothesis 2A: non-ASD to ASD**

The linear regression model corresponding to hypothesis 2A included individuals without ASD at T1. T1 ADOS CSS and participation in ASD-specific intervention were included as predictors. There was evidence of non-normal residuals in this model, which would impact standard errors, confidence intervals, and  $p$ -values. However, linearity, homoscedasticity, and independence of observations were not violated, so point estimates remained valid. Therefore, a bootstrapping method was implemented; 1000 bootstrap samples were used to estimate the standard errors, confidence intervals, and  $p$ -values of the regression model (see Table 14). T1 ADOS CSS approached significant as a predictor of change in ASD symptom severity, as everyone one-point increase in T1 ADOS CSS was associated with a .59 unit decrease in severity over time,  $B=.040$ ,  $p=.051$ . In other words, those with higher T1 symptom severity increased less in symptom severity between T1 and T2. A one-sample  $t$ -test was used to compare change in severity in the NON-ASD and NON-NON groups. The mean increase in severity in the NON-ASD group was 2.42 ADOS CSS points ( $SD=2.22$ ),  $t(18)=-4.756$ ,  $p<.001$ . In contrast, the mean decrease in severity in the NON-NON group was 0.54 ADOS CSS points,  $t(128)=-3.960$ ,  $p<.001$ . Furthermore, participation in ASD-specific intervention was not associated with change in ASD symptom severity over time,  $p>.05$ .

### 3.3 Study Attrition

In addition to the 424 participants who were evaluated at both time points, 281 participants (39.9%) were evaluated at T1 and did not return for re-evaluation. Differences between those who returned and those who did not were examined across variables (see Table 15). Overall, the groups differed in their T1 ADOS CSS ( $t(594)=-4.277$ ,  $p<.001$ ; higher for those who returned for re-evaluation), VABS(-II) social score ( $t(534)= 2.155$ ,  $p=.032$ ; higher in those who did not return), and race ( $\Phi=-.199$ ,  $p<.001$ ; minorities less likely to return). Altogether, those with greater ASD symptom severity and weaker social skills were more likely to return for a re-evaluation, whereas children of minority status were less likely to return. Maternal education level did not differ between those who returned and those who did not, though it is important to note that 168 participants (23.8%) of the 705 were missing maternal education data. Also, those with an ASD diagnosis were more likely to return for a re-evaluation ( $\Phi=.193$ ,  $p<.001$ ), whereas those with GDD were less likely than other children to return for re-evaluation ( $\Phi=-.195$ ,  $p<.001$ ).

### 3.4 Site Comparison

Differences in diagnostic instability between the two study sites were examined using logistic regression, with the site as the predictor and the outcome as diagnostic transition or stability. Site did not predict transition from ASD to non-ASD ( $B=-.378$ ,  $p=.236$ ), non-ASD to ASD ( $B=-.212$ ,  $p=.670$ ), or GDD to ASD ( $B=-.636$ ,  $p=.529$ ). Site did predict transition from ASD to GDD, as six of the 17 who transitioned were assessed at UConn (2.2% of the 274 UConn participants) and 11 of 17 were assessed at GSU (7.3% of the 150 GSU participants),  $B=-1.341$ ,  $p=.011$ . The sites did not differ in attrition rates,  $\Phi=.007$ ,  $p=.842$ .

**Table 4. Number of participants and hours of services for those who received intervention between T1 and T2 based on diagnostic categories at both time points**

|   | <b>ASD-ASD</b>          | <b>ASD-NON</b>        | <b>ASD-GDD<sup>^</sup></b> | <b>NON-NON</b>         | <b>NON-ASD</b>        | <b>GDD-ASD<sup>^</sup></b> | <b>All</b>             |
|---|-------------------------|-----------------------|----------------------------|------------------------|-----------------------|----------------------------|------------------------|
| Total Sample N  | 219                     | 51                    | 17                         | 135                    | 19                    | 8                          | 424                    |
| Any intervention N <sup>+</sup>                                   | 156 (71.2%)             | 31 (60.8%)            | 8 (47.1%)                  | 80 (59.3%)             | 14 (73.7%)            | 6 (75.0%)                  | 281 (66.3%)            |
| ASD-specific  | 61 (27.9%)              | 9 (17.6%)             | 2 (11.8%)                  | 7 (5.2%)               | 4 (21.1%)             | 2 (25.0%)                  | 81 (19.1%)             |
| State Early Intervention  | 109 (49.8%)             | 17 (33.3%)            | 4 (23.5%)                  | 42 (31.1%)             | 8 (42.1%)             | 4 (50.0%)                  | 176 (41.5%)            |
| Occupational Therapy  | 104 (47.5%)             | 17 (33.3%)            | 6 (35.3%)                  | 43 (31.9%)             | 9 (47.4%)             | 3 (37.5%)                  | 173 (40.8%)            |
| Speech Therapy  | 120 (54.8%)             | 22 (43.1%)            | 6 (35.3%)                  | 63 (46.7%)             | 13 (68.4%)            | 6 (75.0%)                  | 218 (51.4%)            |
| Physical Therapy  | 37 (16.9%)              | 5 (9.8%)              | 2 (11.8%)                  | 15 (11.1%)             | 3 (15.8%)             | 2 (25.0%)                  | 60 (14.2%)             |
| Social Skills   | 4 (1.8%)                | 1 (2.0%)              | 0 (0.0%)                   | 0 (0.0%)               | 0 (0.0%)              | 0 (0.0%)                   | 5 (1.2%)               |
| Other Intervention  | 32 (14.6%)              | 2 (3.9%)              | 0 (0.0%)                   | 18 (13.3%)             | 4 (21.1%)             | 2 (25.0%)                  | 56 (13.2%)             |
| Special Education   | 74 (33.8%)              | 13 (25.5%)            | 2 (11.8%)                  | 21 (15.6%)             | 6 (31.6%)             | 3 (37.5%)                  | 114 (26.9%)            |
| Mean hours of intervention<br>M(SD)                               | 8.117 (8.805)<br>N=210  | 5.048 (7.226)<br>N=50 | 2.105 (3.465)<br>N=17      | 2.792 (5.461)<br>N=131 | 6.476 (9.276)<br>N=17 | 7.272 (11.037)<br>N=7      | 5.963 (8.047)<br>N=408 |
| Mean hours of intervention<br>(only if participated) <sup>#</sup> | 11.517 (8.415)<br>N=148 | 8.143 (7.698)<br>N=31 | 4.474 (3.915)<br>N=8       | 4.630 (6.410)<br>N=79  | 8.469 (9.820)<br>N=13 | 10.181 (12.071)<br>N=5     | 8.977 (8.393)<br>N=271 |

Note. <sup>^</sup> These groups are subsets of the previous groups (e.g., ASD-NON group contains the ASD-GDD group). <sup>+</sup> Any intervention N does not equal the sum of subtype intervention Ns, as some children may be involved in multiple types of intervention. <sup>#</sup> Mean hours of intervention in those who participated in intervention are reported; those who participated in zero hours were only included in previous variable (Mean hours of intervention). N may not equal Any intervention N because some parents indicated that they participated in intervention, but did not report number of hours.

**Table 5. Correlations for continuous T1 variables in those with ASD at T1 and ASD or GDD at T2 (for Hypothesis 1A)**

| Predictor          | T2 non-ASD | MSEL Exl | MSEL Rel | MSEL VR | ADOS CSS | VABS Soc | VABS Comm | Age     |
|--------------------|------------|----------|----------|---------|----------|----------|-----------|---------|
| T2 non-ASD         | 1          |          |          |         |          |          |           |         |
| MSEL Exl           | .147*      | 1        |          |         |          |          |           |         |
| MSEL Rel           | .207**     | .681**   | 1        |         |          |          |           |         |
| MSEL VR            | .189**     | .543**   | .553**   | 1       |          |          |           |         |
| ADOS CSS           | -.225**    | -.242**  | -.374**  | -.275** | 1        |          |           |         |
| VABS Soc           | .177**     | .368**   | .382**   | .324**  | -.287*   | 1        |           |         |
| VABS Comm          | -.145*     | .637**   | .610**   | .447**  | -.277**  | .705**   | 1         |         |
| Age                | .046       | .009     | -.031    | -.131*  | .137*    | -.217**  | -.005     | 1       |
| Intervention Hours | -.141*     | .026     | .022     | .015    | .075     | -.197**  | -.116     | -.216** |

Note. \* $p < .05$ , \*\* $p < .01$ . MSEL=Mullen Scales of Early Learning, Exl=Expressive Language, Rel=Receptive language, VR=Visual reception; ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score; CARS2-ST=Childhood Autism Rating Scale, Second Edition, Standard Form; VABS=Vineland Adaptive Behavior Scales, Soc=Socialization domain score, Comm=Communication domain score.

**Table 6. Contingency coefficients between categorical T1 variables and diagnostic transition outcome**

| Predictor                       | T2 non-ASD | T2 GDD             | T2 ASD <sup>^</sup> |
|---------------------------------|------------|--------------------|---------------------|
| Sex                             | -.116      | -.066 <sup>#</sup> | .039                |
| Maternal Education <sup>+</sup> | -.086      | -.159*             | .138                |
| Race/Ethnicity                  | .064       | .213**             | -.034               |

Note. \* $p < .05$ , \*\* $p < .01$ . Phi coefficient presented unless otherwise indicated. <sup>+</sup>Values presented are tau-b coefficients. <sup>#</sup> $p$ -value based on Fisher's exact test. <sup>^</sup>Sample includes all participants with T1 ASD diagnosis, unless T2 ASD measured as outcome (includes T1 non-ASD participants).

**Table 7. Logistic regression results for final model of hypothesis 1A (predictors of transition from ASD to non-ASD)**

| Predictor               | Exp(B) | Wald  | $p$  | 95% CI         |
|-------------------------|--------|-------|------|----------------|
| ADOS CSS                | .816*  | 6.047 | .014 | [.694, .960]   |
| MSEL Receptive Language | 1.016* | 5.008 | .025 | [1.002, 1.029] |
| Intervention Hours      | .949*  | 4.574 | .032 | [.905, .996]   |

Note. \* $p < .05$ . ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score; MSEL=Mullen Scales of Early Learning Receptive Language

**Table 8. Correlations for continuous T1 variables in those with ASD at T1 and GDD at T2 (for Hypothesis 1B)**

| Predictor          | T2 GDD  | MSEL Exl | MSEL Rel | MSEL VR | ADOS CSS | VABS Soc | VABS Comm | Age     |
|--------------------|---------|----------|----------|---------|----------|----------|-----------|---------|
| T2 GDD             | 1       |          |          |         |          |          |           |         |
| MSEL Exl           | .009    | 1        |          |         |          |          |           |         |
| MSEL Rel           | .094    | .661**   | 1        |         |          |          |           |         |
| MSEL VR            | .018    | .542**   | .538**   | 1       |          |          |           |         |
| ADOS CSS           | -.193** | -.240**  | -.373**  | -.248** | 1        |          |           |         |
| VABS Soc           | .153*   | .373**   | .370**   | .297**  | -.314**  | 1        |           |         |
| VABS Comm          | .028    | .658**   | .615**   | .413**  | -.299**  | .690**   | 1         |         |
| Age                | .047    | -.063    | -.056    | -.180** | .143*    | -.227**  | -.031     | 1       |
| Intervention Hours | -.183** | .069     | .076     | .069    | .059     | -.205**  | -.102     | -.232** |

Note. \* $p < .05$ , \*\* $p < .01$ . MSEL=Mullen Scales of Early Learning, Exl=Expressive Language, Rel=Receptive language, VR=Visual reception; ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score; CARS2-ST=Childhood Autism Rating Scale, Second Edition, Standard Form; VABS=Vineland Adaptive Behavior Scales, Soc=Socialization domain score, Comm=Communication domain score.

**Table 9. Logistic regression results for final model of hypothesis 1B (ASD-GDD)**

| Predictor          | Exp(B) | Wald  | $p$  | 95% CI         |
|--------------------|--------|-------|------|----------------|
| ADOS CSS           | .658** | 7.831 | .005 | [.491, .882]   |
| Intervention Hours | .851*  | 4.261 | .039 | [.729, .992]   |
| Race/Ethnicity     | 3.342  | 3.794 | .051 | [.993, 11.254] |
| Maternal Education | .291*  | 4.487 | .034 | [.093, .912]   |

Note. \* $p < .05$ , \*\* $p < .01$ . ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score

**Table 10. Correlations for continuous T1 variables in those without ASD at T1 and ASD at T2 (for Hypothesis 2A)**

| Predictor          | T2 ASD | MSEL Exl | MSEL Rel | MSEL VR | ADOS CSS | VABS Soc | VABS Comm | Age  |
|--------------------|--------|----------|----------|---------|----------|----------|-----------|------|
| T2 ASD             | 1      |          |          |         |          |          |           |      |
| MSEL Exl           | -.031  | 1        |          |         |          |          |           |      |
| MSEL Rel           | -.039  | .515**   | 1        |         |          |          |           |      |
| MSEL VR            | -.128  | .585**   | .731**   | 1       |          |          |           |      |
| ADOS CSS           | .165*  | .028     | -.237**  | -.204*  | 1        |          |           |      |
| VABS Soc           | .074   | .430**   | .419**   | .370**  | -.014    | 1        |           |      |
| VABS Comm          | .061   | .603**   | .579**   | .511**  | -.075    | .635**   | 1         |      |
| Age                | .054   | .067     | -.116    | -.148   | .159*    | -.183*   | -.127     | 1    |
| Intervention Hours | .193*  | -.214**  | -.261**  | -.232** | .034     | -.231**  | -.280**   | .096 |

Note. \*  $p < .05$ , \*\*  $p < .01$ . MSEL=Mullen Scales of Early Learning, Exl=Expressive Language, Rel=Receptive language, VR=Visual reception; ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score; CARS2-ST=Childhood Autism Rating Scale, Second Edition, Standard Form; VABS=Vineland Adaptive Behavior Scales, Soc=Socialization domain score, Comm=Communication domain score.

**Table 11. Logistic regression results for final model of hypothesis 2A (NON-ASD)**

| Predictor          | Exp(B) | Wald  | $p$  | 95% CI         |
|--------------------|--------|-------|------|----------------|
| ADOS CSS           | 1.397* | 5.281 | .022 | [1.050, 1.858] |
| Intervention Hours | 1.072* | 4.702 | .030 | [1.007, 1.141] |

Note. \*  $p < .05$ , \*\*  $p < .01$ . ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score; ASD intervention=exposure to ASD-specific intervention.

**Table 12. Linear regression results corresponding to final model of hypothesis 1A (predictors of ASD symptom severity change in ASD-ASD and ASD-NON)**

| Predictor               | B      | Standard Error | $\beta$ | $t$    | $p$   |
|-------------------------|--------|----------------|---------|--------|-------|
| ADOS CSS                | .661** | .081           | .509    | 8.183  | <.001 |
| MSEL Receptive Language | .025** | .007           | .211    | 3.414  | .001  |
| Intervention Hours      | -.034  | .019           | -.103   | -1.786 | .075  |

Note. \*\*  $p < .01$ . ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score; MSEL=Mullen Scales of Early Learning Receptive Language

**Table 13. Linear regression results corresponding to final model of hypothesis 1B (predictors of ASD symptom severity change in ASD-ASD and ASD-GDD)**

| Predictor          | B      | Standard Error | $\beta$ | $t$    | $p$  |
|--------------------|--------|----------------|---------|--------|------|
| ADOS CSS           | .566** | .078           | .471    | 7.286  | .005 |
| Intervention Hours | -.032  | .019           | -.112   | -1.653 | .100 |
| Race               | -.238  | .356           | -.045   | -.667  | .506 |
| Maternal Education | .029   | .341           | .006    | .086   | .932 |

Note. \*\* $p < .01$ . ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score

**Table 14. Linear regression results using bootstrapping corresponding to final model of hypothesis 2A (NON-NON and NON-ASD)**

| Predictor          | B    | Standard Error | $p$  | 95% CI        |
|--------------------|------|----------------|------|---------------|
| ADOS CSS           | .040 | .020           | .051 | [.004, .083]  |
| Intervention Hours | .010 | .007           | .132 | [-.002, .024] |

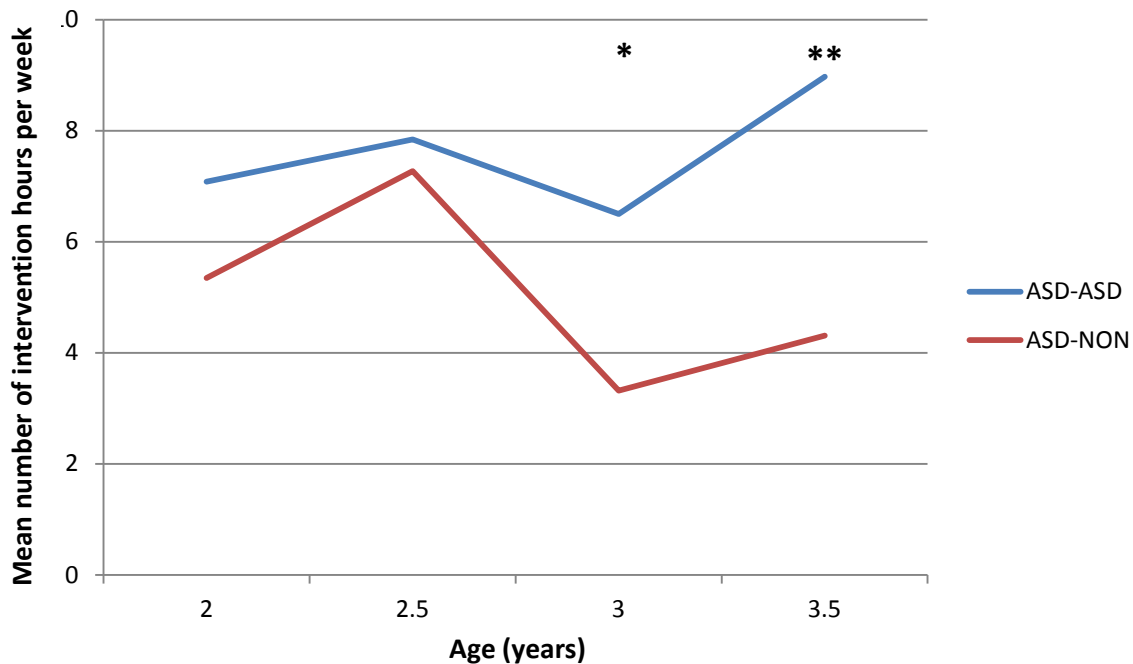
Note. ADOS CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score.

**Table 15. Descriptive statistics and group comparisons in those who remained in the study and those who dropped out**

| T1 Variable        | T1 only: M(SD) | T1 and T2: M(SD) | $t$ or $\phi^+$ | $p$   |
|--------------------|----------------|------------------|-----------------|-------|
| Age                | 26.632(5.986)  | 26.308(4.459)    | .777            | .438  |
| MSEL ExI           | 54.562(22.666) | 58.009(24.024)   | -1.906          | .057  |
| MSEL Rel           | 57.468(26.263) | 57.484(29.089)   | -.007           | .994  |
| MSEL VR            | 72.296(24.759) | 72.805(23.524)   | -.270           | .788  |
| ADOS CSS           | 3.872(2.832)   | 4.788(2.794)     | -4.227**        | <.001 |
| VABS Comm          | 76.315(14.057) | 75.209(13.483)   | 1.036           | .301  |
| VABS Soc           | 79.907(13.446) | 77.786(11.610)   | 2.155*          | .032  |
| Race               | -              | -                | -.199**         | <.001 |
| Maternal Education | -              | -                | .001 (N=537)    | .974  |
| ASD dx             | -              | -                | .193**          | <.001 |
| GDD dx             | -              | -                | -.195**         | <.001 |
| No dx              | -              | -                | -.010           | .798  |
| Other dx           | -              | -                | -.045           | .230  |

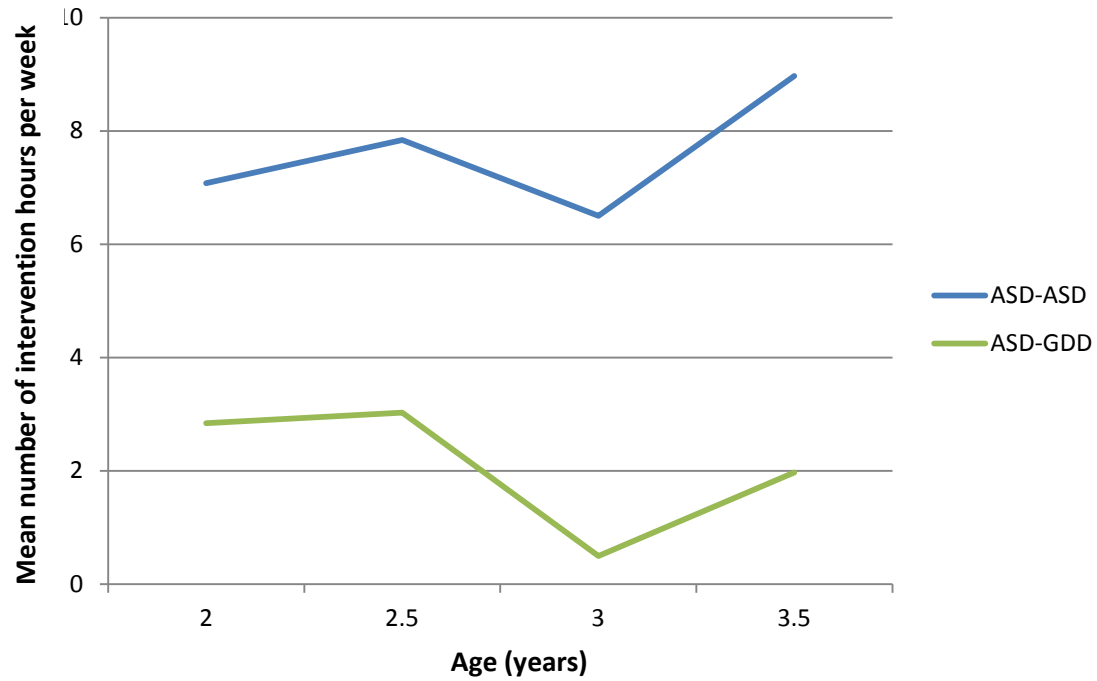
Note. \* $p < .05$ , \*\* $p < .001$ . <sup>+</sup>Phi coefficient used for categorical variables;  $t$  used for continuous variables.





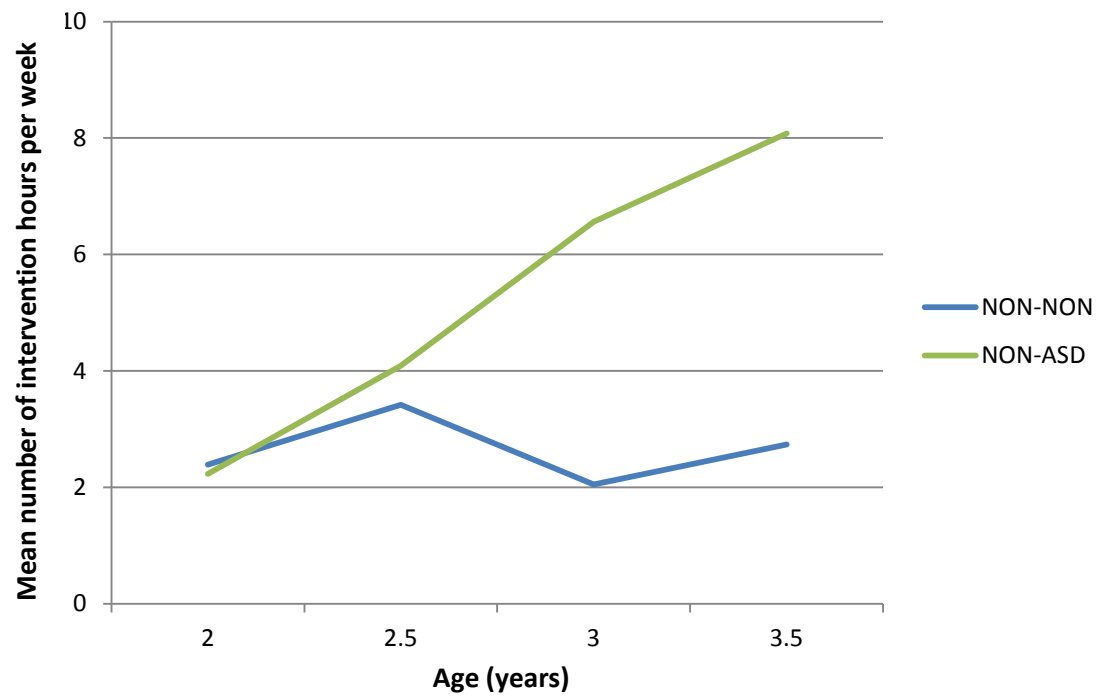
*Note. Comparisons at each time point: Age 2:  $t(65.23)=-.996$ ,  $p=.323$ ; Age 2.5:  $t(56.54)=-.269$ ,  $p=.789$ ; Age 3:  $t(90.84)=-2.324$ ,  $p=.022$ ; Age 3.5:  $t(92.56)=-2.905$ ,  $p=.005$ . Change in intervention from 2 to 3.5 years differs by group,  $B=-3.002$ ,  $p=.009$ .*

**Figure 3. Mean number of intervention hours in six-month intervals by group membership in those with T1 ASD**



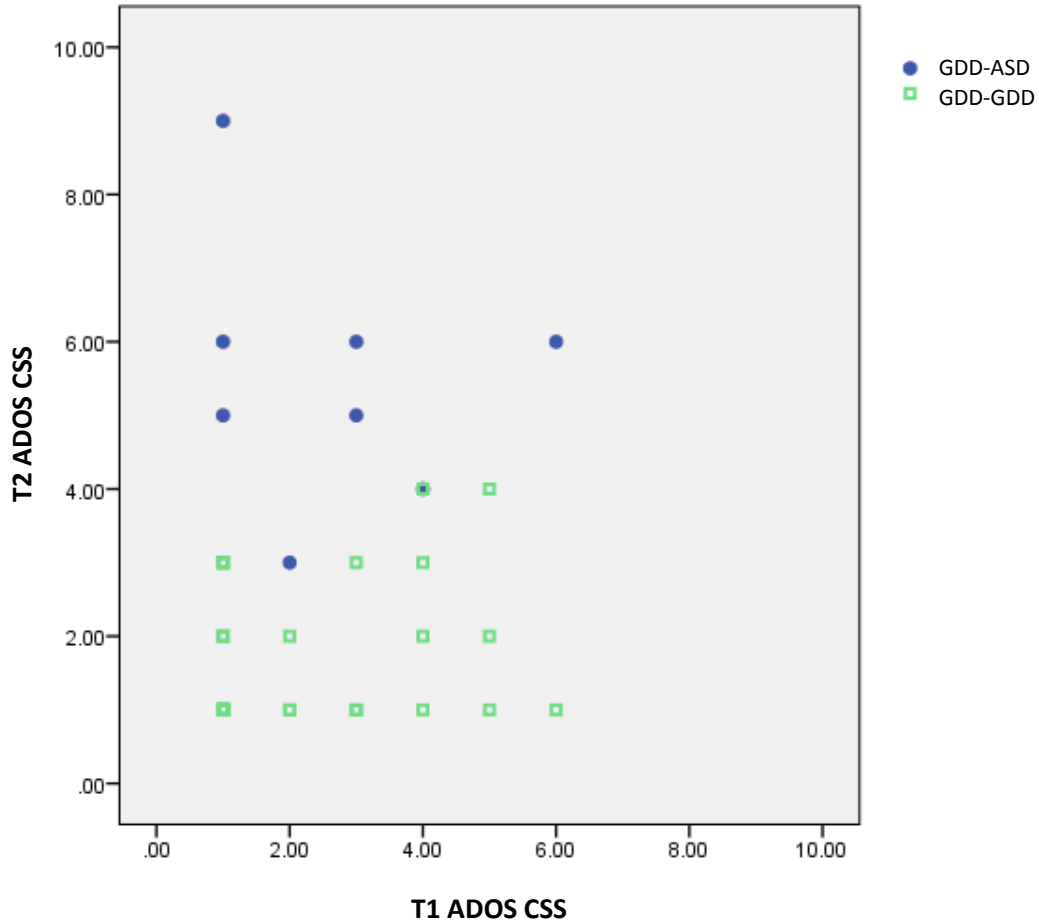
*Note. Change in intervention from 2 to 3.5 years not significantly different (small N),  $B=-2.85$ ,  $p=.159$ .*

**Figure 4. Mean number of intervention hours in six-month intervals by group membership in those with T1 ASD and T2 ASD or GDD**



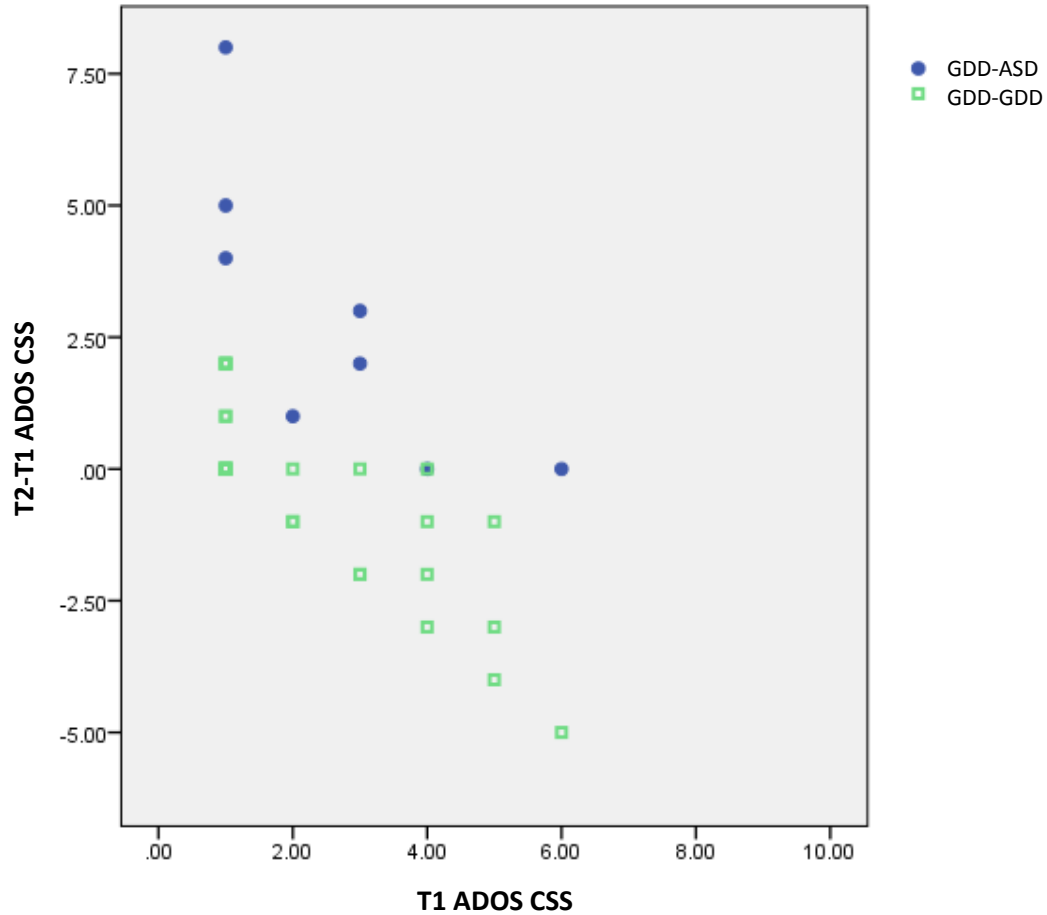
*Note. Change in intervention from 2 to 3.5 years not significantly different (small N),  $B=4.22$ ,  $p=.101$*

**Figure 5. Mean number of intervention hours in six-month intervals by group membership in those without T1 ASD**



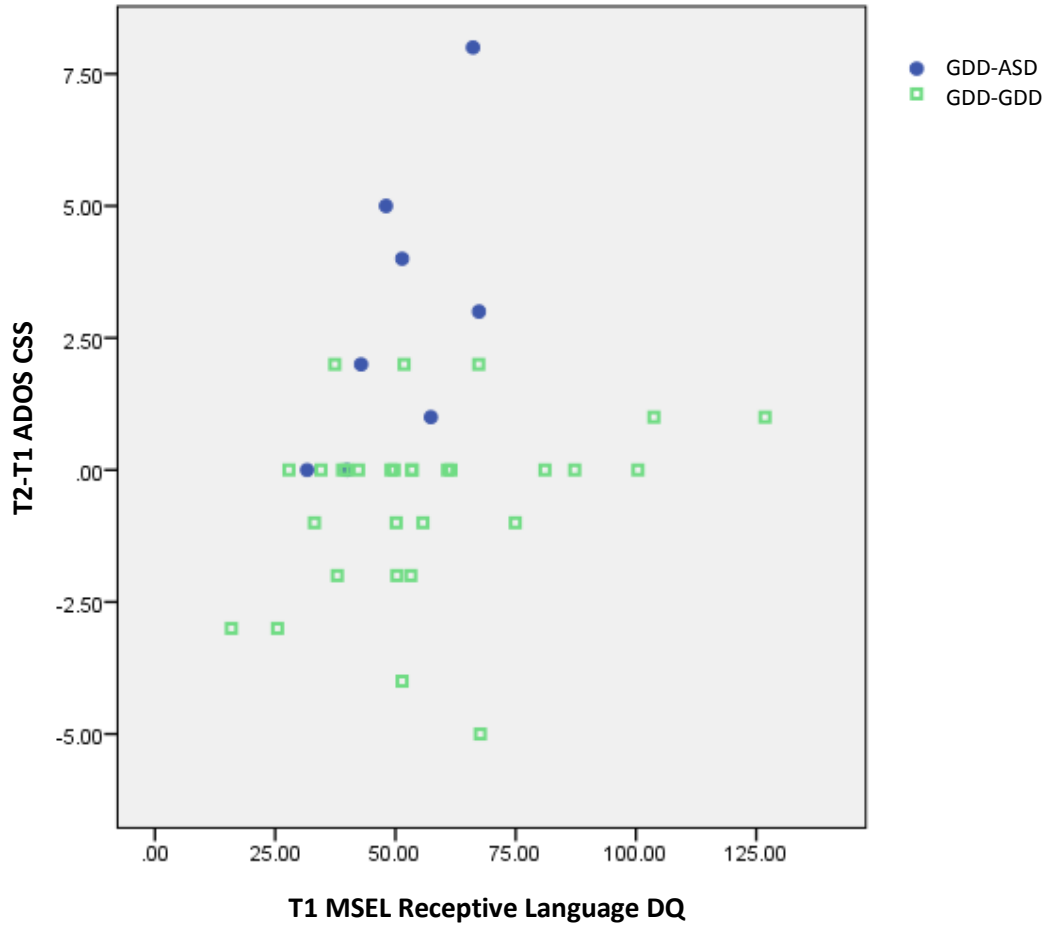
*Note. ADOS CSS= Autism Diagnostic Observation Schedule Calibrated Severity Score.*

**Figure 6. Scatterplot of T1 and T2 ASD symptom severity in participants who maintained GDD diagnosis and who transitioned from GDD to ASD**



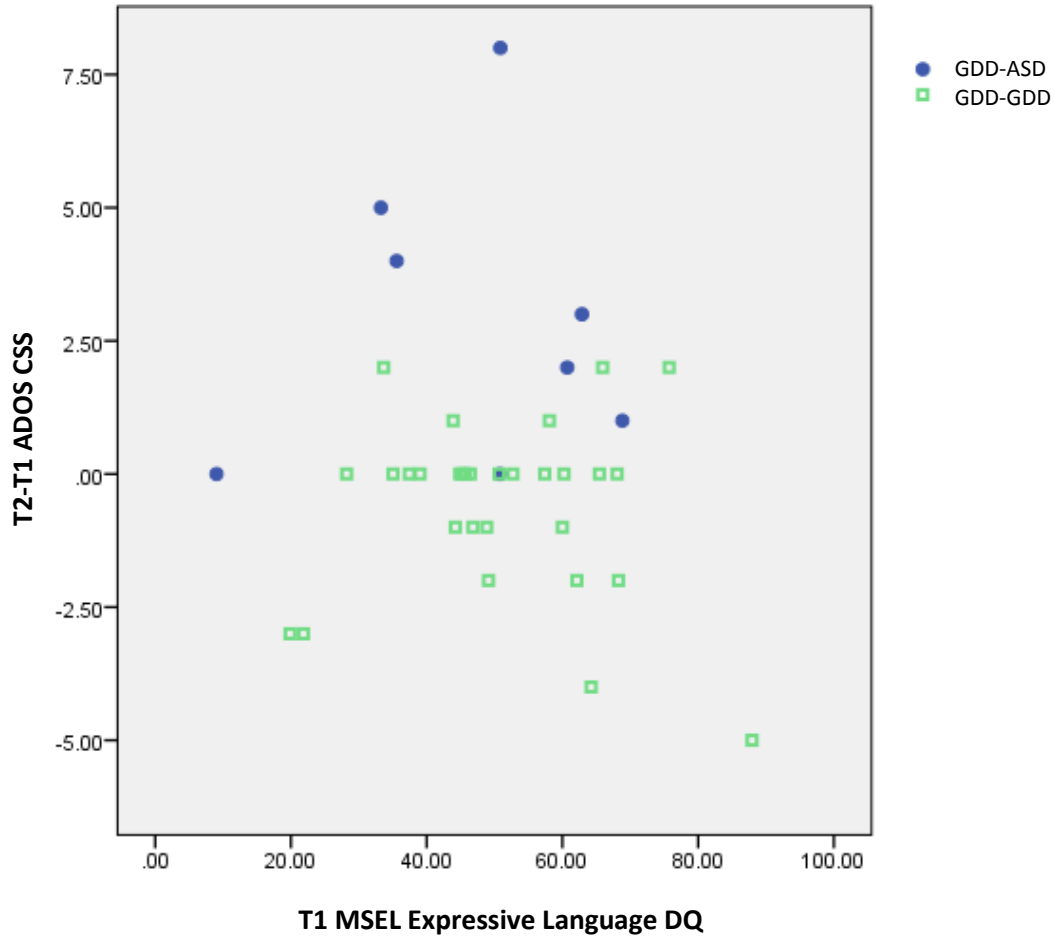
*Note. Increasing values on the Y axis indicate an increase in ASD symptom severity between T1 and T2. ADOS CSS= Autism Diagnostic Observation Schedule Calibrated Severity Score.*

**Figure 7. Scatterplot of T1 ASD symptom severity and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD**



*Note. Increasing values on the Y axis indicate an increase in ASD symptom severity between T1 and T2. MSEL=Mullen Scales of Early Learning, DQ= Developmental Quotient.*

**Figure 8. Scatterplot of T1 receptive language ability and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD**

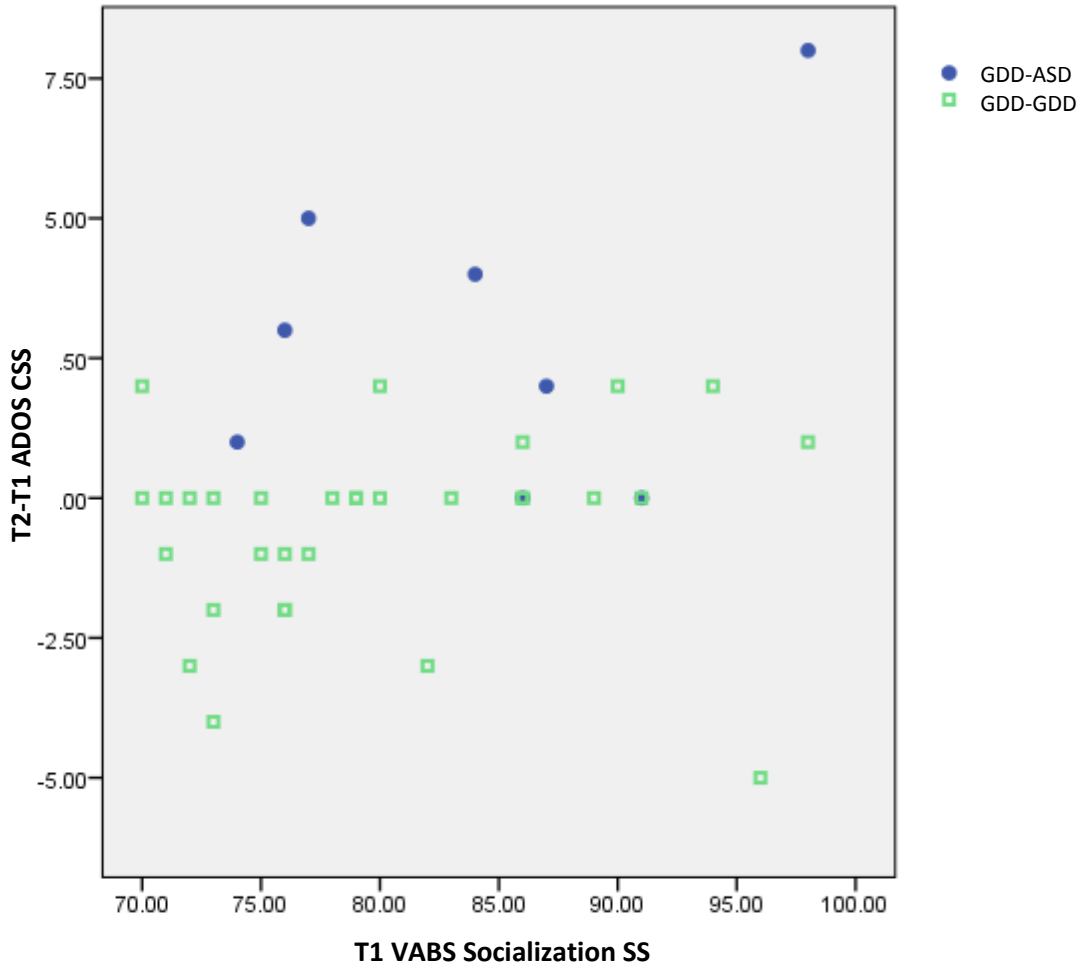


*Note. Increasing values on the Y axis indicate an increase in ASD symptom severity between T1 and T2. MSEL=Mullen Scales of Early Learning, DQ= Developmental Quotient.*

**Figure 9. Scatterplot of T1 expressive language ability and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD**

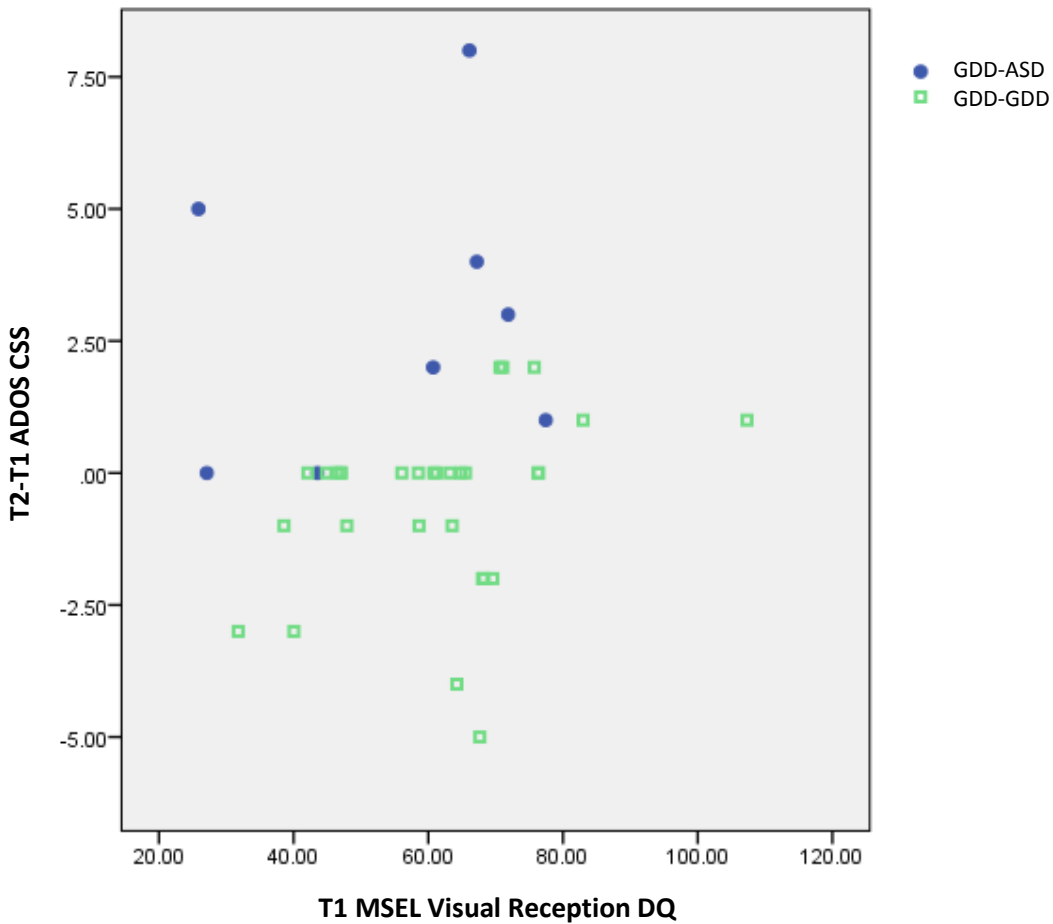






*Note. Increasing values on the Y axis indicate an increase in ASD symptom severity between T1 and T2. VABS=Vineland Adaptive Behavior Scales, SS= Standard Score*

**Figure 11. Scatterplot of T1 parent reported socialization skills and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD**



*Note. Increasing values on the Y axis indicate an increase in ASD symptom severity between T1 and T2. MSEL=Mullen Scales of Early Learning, DQ= Developmental Quotient.*

**Figure 12. Scatterplot of T1 nonverbal problem-solving ability and change in ASD symptom severity over time in participants who maintained GDD diagnosis and who transitioned from GDD to ASD**

## 4 DISCUSSION

### 4.1 Diagnostic Stability

The current study examined predictors of diagnostic instability across four groups of participants: ASD-NON, ASD-GDD, NON-ASD, and GDD-ASD. Stability of initial diagnostic category (ASD or non-ASD) was 83.5 percent; stability of ASD in those with a T1 diagnosis was 81.1 percent; stability of T1 non-ASD status was 87.7 percent. Overall, rates of diagnostic stability of ASD were comparable to rates found in previous studies. In particular, the majority of previous studies of diagnostic stability of ASD (56 percent, see Figure 1) found stability rates between 80 and 84 percent (e.g., Pellicano 2012; Stone et al., 1999). The current finding lends support to the consideration that about 15 to 20 percent of toddlers diagnosed with ASD no longer meet criteria when they are older. This result emphasizes the need for exploration of factors predicting this diagnostic instability in the current study.

Furthermore, research has not focused on examining the rates or predictors of diagnostic transition between ASD and GDD; the current study is the first to examine rates of transition between these conditions. The stability rate of T1 GDD diagnosis was fairly low, at 53.3 percent. This stability rate was lower than expected given research on stability of children with GDD beyond age five. Previous studies have found that 97 to 98 percent of children with GDD at around age three still demonstrated significant impairments at age seven (Shevell et al., 2005b; Webster, Majnemer, Platt, & Shevell, 2008). This contrast with the current study may be related to earlier assessment of children in the present study (at around age two) in contrast with age three assessment in the previous studies. Differences in measures used may also contribute to this inconsistency. For example, Shevell and colleagues (2005b) used the Battelle Developmental Inventory (Newborg, 1984), a developmental assessment that had been normed nearly 20 years beforehand and which has produced documented discrepancies in test-retest scoring (Boyd, 1989).

Furthermore, differences between the group that maintained GDD diagnosis and the group that did not maintain GDD diagnosis may help to explain this low stability rate. One study demonstrated instability rates that were more consistent with results of the current study, in that 40 percent of children with mild delays at age four were assessed to be developmentally on track by age six (Valtonen, Ahonen, Lyytinen, & Tolvanen, 2007). In the context of the current study, this finding would indicate that the children who maintained GDD diagnosis likely had more severe impairments across multiple domains than those who no longer met GDD criteria. However, post-hoc analyses indicated that those who maintained GDD diagnosis in the current sample did not differ from those who no longer met GDD criteria in terms of T1 nonverbal problem-solving ability, expressive language ability, or receptive language ability. Interestingly, the GDD-NON group had marginally higher T1 fine motor ability than the GDD-GDD group ( $p=.065$ , Cohen's  $d=.498$ ). Therefore, although many of the children who changed diagnosis from GDD to non-GDD may have had initial motor delays, they demonstrated a relative strength compared to those who maintained GDD diagnosis. This strength was apparent at T2 as well, when the GDD-NON group had even higher scores ( $M=80.672$ ,  $SD=27.658$ ) than the GDD-GDD group ( $M=59.012$ ,  $SD=17.383$ ). Therefore, these children were also less likely to have motor delays at T2, thus eliminating one delay that would qualify them for GDD diagnosis. This finding helps to explain why a large proportion of toddlers diagnosed with GDD at T1 did not meet criteria at T2.

Finally, another explanation for the high rates of instability of T1 GDD diagnosis is the idea that some children simply have delays in skill acquisition and will eventually reach a typical developmental level, whereas some children experience deviance from the typical developmental trajectory. The group that would be considered delayed would likely meet criteria for GDD at age two, though they may make sufficient gains by age four, resulting in diagnostic instability. However, the children who deviate from the typical trajectory would likely still meet criteria for GDD at T2. This group may also be more likely to

meet criteria for intellectual disability when they are older, though this would be dependent on one of their delays being cognitive impairment (Riou, Ghosh, Francoeur, & Shevell, 2009; Shevell, 2008).

Overall, these findings indicate that gains in fine motor skills may account for some children transitioning from GDD to non-GDD. It is also important to note that many of the children who changed diagnosis were still diagnosed with other conditions at T2, including DLD. Therefore, it is possible that children who transitioned from GDD to DLD made gains in other domains, such as fine motor skills or nonverbal problem-solving ability, but maintained significant language delays. Children who transitioned from GDD to ASD are discussed below.

#### **4.2 Hypothesis 1A: ASD to non-ASD**

Analysis of the first hypothesis revealed that when accounting for other variables in the model, low ASD symptom severity, high receptive language ability, and fewer hours of intervention services predicted transition between ASD at age two and non-ASD at age four.

The finding that lower ASD symptom severity predicts transition to non-ASD is consistent with some previous literature (Lord, 1995; Stone et al., 1999; Turner & Stone, 2007; Wiggins et al., 2012b). Notably, all of these studies examined an age group similar to the current study, in that children were evaluated initially at around at age two and then again at age three or four. Similarly, the diagnostic measures that each study used were similar to those in the current study (e.g., ADOS, ADI(-R), and CARS). However, the current results are inconsistent with two prior studies, which found that ASD symptom severity was not associated with diagnostic transition (Pellicano, 2012; Suter et al., 2007). Pellicano (2012) found that only intervention predicted diagnostic transition. However, it is notable that the sample and methods differed significantly from the present study. In particular, the age of participants was older (five at T1 and eight at T2), the sample size was much smaller (N=37), and the methods differed. Specifically, the authors relied on the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) and ADOS-G to determine the children's diagnostic status. Although the

specificity of the SCQ improves in combination with the ADOS, the use of the ADOS with the ADI-R as in the present study has demonstrated the best diagnostic discriminatory ability (Corsello et al., 2007). Therefore, the rigorous method and larger sample size in the current study demonstrate an improvement upon previous research. Sutera and colleagues (2007) also did not find ASD symptom severity to predict diagnostic instability. Interestingly, the sample used in that paper was a subset (N=90) of the sample in the current study. Therefore, it is possible that the lack of findings in Sutera's study may be attributed to a lack of power due to small sample size.

Beyond ASD symptom severity, high T1 receptive language ability was associated with a decrease in ASD symptom severity and transition from ASD to non-ASD over time. The finding regarding diagnostic instability has not been established in previous research; the only two studies that examined language ability as a predictor did not find it to be significant, and these were the same two studies that did not find ASD symptom severity to be a predictor (Pellicano, 2012; Sutera et al., 2012). Therefore, it is possible that the limited findings in these two studies are related to limited power from small sample size, and methods that differ from the current study.

However, strong initial receptive language ability has been found to be associated with positive outcomes later in childhood, such as stronger language ability, IQ, and lower ASD symptom severity (Luyster et al., 2007; Muratori & Narzisi, 2014). The present study is consistent with these findings in that receptive language ability predicted a decrease in continuous ASD symptom severity over time as well as diagnostic instability. Furthermore, strong language ability has been shown to support the development of social understanding in children, particularly theory of mind development (Astington & Jenkins, 1999) and performance on false belief tasks (Happé, 1995). As theory of mind skills are a central deficit in children with ASD, and language skills seem to contribute to their development, it is not surprising that strong receptive language skills support a reduction in ASD symptom severity over time. These findings indicate a significant role for receptive language skills in diagnostic instability of ASD.

Both receptive language ability and ASD symptom severity predicted change in a continuous measure of symptom severity. Early strengths in terms of receptive language ability and minimal ASD symptom severity likely contribute to the development of other skills over time. Developmental theorists assert that development is influenced by various systems interacting in a complex environment (Smith & Thelen, 2003). Therefore, abilities such as language, cognitive, and social skills interact with each other throughout development, with strengths supporting skill development and weaknesses contributing to further delays. Research has demonstrated that early social engagement influences various aspects of learning and development (Mundy, Sullivan, & Mastergeorge, 2009; Ochsner & Lieberman, 2001). Furthermore, it is not surprising that toddlers with relatively stronger abilities, such as social engagement and communication skills, are more likely to make gains in these and other domains over time. Thus, children with ASD whose symptoms are less severe are more likely to experience a decrease in symptom severity and diagnostic change than those with more severe initial symptoms. Early language skills have also been shown to be associated with the development of social and cognitive skills (Astington & Jenkins, 1999; Happé, 1995; Harris, de Rosnay, & Pons, 2005). Therefore, early language strengths support other aspects of development, which contribute to improvement in ASD symptom severity and movement off the autism spectrum. Overall, dynamic systems interacting throughout development help to explain why early strengths support developmental gains.

Furthermore, less participation in early intervention services between T1 and T2 increased the odds of no longer meet criteria for ASD at T2. This result is contradictory to hypotheses, as early intervention is generally associated with positive developmental outcomes in children with ASD (Harris & Handleman, 2000; Lovaas, 1987; Warren et al., 2011). The result was initially thought to be explained by a higher level of ASD symptom severity in the group receiving more intervention services, as those with more severe symptoms may be more likely to be referred to and enroll in services. However, this mechanism was tested, and no significant relationship between ASD symptom severity and hours of

participation in intervention was found. Furthermore, there was no relationship between SES and participation in intervention.

However, there seems to be an issue of reverse causality contributing to the unexpected intervention finding, since the intervention data span a two-year time period, rather than the specific time point assessed by the other study measures (see Figure 3). Parents of children who demonstrated improvements after receiving an ASD diagnosis were more likely to decrease intervention hours over time. In contrast, parents of those who maintained clinically significant weaknesses were more likely to increase their children's participation in services. Therefore, participation in intervention between age two and age four may be more related to parents' perception of change in severity during this time period, rather than initial symptom severity. As the intervention variable used in the logistic regression models was an overall measure of participation, it did not capture this change in participation over time. This helps to explain why greater participation in intervention was associated with lower odds of changing from ASD to non-ASD; it is not that intervention prevented diagnostic change, but that the variable did not capture changes in participation over time associated with change in the child's perceived severity.

Another possible factor contributing to the unexpected intervention finding is that the intervention measure in the present study has some limitations. First, some parents indicated that their children participated in intervention, but did not indicate the number of hours. As a result, data from 15 participants (3.5%) who indicated that they participated in intervention but did not indicate number of hours were excluded from analyses. An additional 54 participants (12.7%) had mean intervention hours based on partial data (i.e., they reported hours for at least one, but not all, of four six-month time periods). Furthermore, information on intensity of intervention and quality of services is unknown, as this information would be subjective and difficult to quantify and measure. However, varying quality of services could certainly contribute to differences in developmental trajectories. Additionally, all types of



intervention were combined into one variable to analyze the effect of intervention in general, to maximize power for analyses, and to minimize the number of potential variables. However, it is possible that different intervention types may differentially affect developmental trajectories, which would not be conveyed using the current variable. Another downside to the current intervention data is that the information was retrospectively reported by parents, whose reports may not always be accurate. Finally, a lack of reporting of intervention was interpreted as a lack of intervention due to the manner in which data were collected, whereas in reality a parent may have forgotten to report one or more types of intervention. Overall, these weaknesses in the intervention data available may be related to the finding that less participation in intervention predicted transition from ASD to non-ASD.

The explanation that the finding may be specific to the data or related to change in severity and participation in services over time is supported by the result that the same intervention variable was not associated with change in a continuous measure of ASD symptom severity over time. Therefore, the number of intervention hours does not seem to be related to a reduction in symptom severity. Importantly, these results do not indicate that intervention has negative consequences for children. Intervention was not related to an increase in symptom severity, and the previously mentioned explanations have likely contributed to the unexpected finding.

Despite these weaknesses, intervention information was still included in analyses as intervention may be a significant contributor to diagnostic instability of ASD, and not including it as a predictor could neglect an important factor in developmental change. This idea is supported by previous research finding that early intervention contributes to developmental gains (Mukkades et al., 2014; Turner et al., 2006) and even diagnostic instability of ASD (Orinstein et al., 2014; Pellicano, 2012). Despite these findings, other studies have also found intervention to contribute little to change in ASD symptom severity, which is consistent with the results of the current study (Lord, Luyster, Guthrie, & Pickles, 2012; Turner & Stone, 2007). This is not to say that early intervention is ineffective, but that

measuring intervention in a meaningful way is difficult. As described above, there are many facets of intervention which may impact its effectiveness and which may be difficult to quantify and measure. Furthermore, diagnostic instability may occur regardless of intervention due to developmental changes, which may also impact predictive power of intervention data (Camarata, 2014).

Notably, an alternative final model was identified, which was similar to the current final model but with cognitive ability in place of receptive language ability,  $B=1.019$ ,  $p=.032$ . This alternative model was consistent with hypotheses, which were based on findings that cognitive strengths support positive outcomes (e.g., Ben-Itzhak, Watson, & Zachor, 2014; Charman et al., 2005), that children with ASD and intellectual disabilities are more likely to maintain ASD diagnosis (Fountain, Winter, & Bearman, 2012), and that cognitive ability has been found to predict diagnostic instability of ASD (Turner & Stone, 2007). Therefore, this alternative model lends support to previous research finding that cognitive ability was associated with positive outcomes and diagnostic instability of ASD.

Contrary to hypotheses, race and SES did not predict transition from ASD to non-ASD. Race and SES were expected to predict transition from ASD to non-ASD due to the possibility that minority race and low SES groups would have less access to services, and would thus make fewer developmental gains over time. This hypothesis was based on previous findings that racial minorities and those from lower SES background experience disparities in diagnosis and access to services (Mandell et al., 2009; Thomas et al., 2011). The role of race and SES in diagnostic instability is discussed further below.

#### **4.3 Hypothesis 1B: ASD to GDD**

Analysis of hypothesis 1B revealed that initial ASD symptom severity, maternal education level, and number of intervention hours were significant predictors of diagnostic transition between ASD and GDD; also, race approached significance as a predictor. However, only T1 ASD symptom severity was a predictor of change in severity over time regardless of diagnostic transition.

Like transition from ASD to non-ASD in hypothesis 1A, T1 ASD symptom severity was predictive of transition off of the autism spectrum, even when children ended up with a T2 diagnosis of GDD. This is likely because those with lower T1 severity needed to make fewer gains in order to cross the diagnostic boundary from ASD to GDD, though low initial severity was also predictive of a significant decrease in severity over time. The children who transitioned from ASD to GDD likely had cognitive, language, and/or motor delays, as these delays are often associated with children with both ASD and GDD (Mitchell, Cardy, & Zwaigenbaum, 2011). Typically, social-communicative characteristics— such as responsiveness to others, facial expressions including social smile, and joint attention—discriminate between children with ASD and those with GDD, in that those with ASD are often impaired in these skills (Mitchell, Cardy, & Zwaigenbaum, 2011; Trillingsgaard et al., 2005; Ventola et al., 2007). Similarly, children with ASD have been found to engage in more repetitive motor mannerisms and atypical sensory responses than those with other delays (Osterling, Dawson, & Munson, 2002; Ventola et al., 2007; Wetherby et al., 2004; Wiggins, Robins, Bakeman, & Adamson, 2009). These social impairments and repetitive or sensory behaviors are core symptoms of ASD, not of GDD, and would be expected to decrease in children who transition from ASD to GDD. Relatedly, these characteristics would be expected to manifest in lower initial frequency and severity in children with ASD who will ultimately meet criteria for GDD instead of ASD, in comparison to those who maintain ASD diagnosis. This assumption is consistent with the current findings, that ASD symptom severity at T1 predicts diagnostic change from ASD to GDD. Similarly, a large surveillance study found that children with ASD with specific developmental delays, such as language or motor delays, were more likely to transition from ASD to non-ASD over time (Wiggins et al., 2012a).

In contrast, other difficulties common to both ASD and GDD would not be expected to predict transition from ASD to GDD. Studies of language ability between the two diagnostic groups have been variable, and do not find reliable differences (Veness et al., 2012; Ventola et al., 2007). Therefore,

language ability would not be expected to predict transition between these diagnoses, and indeed it was not associated. Furthermore, studies of cognitive ability have indicated that children with ASD often score lower than children with other delays on nonverbal problem-solving tasks (Mitchell, Cardy, & Zwaigenbaum, 2011; Trillingsgaard et al., 2005; Ventola et al., 2007). However, children with GDD often do have cognitive impairment (Riou, Ghosh, Francoeur, & Shevell, 2009), so higher initial cognitive ability would not necessarily be expected to predict transition from ASD to GDD. This may explain why cognitive ability was not predictive of diagnostic transition in the current study. Although previous research has not been designated to identifying predictors of diagnostic transition from ASD to GDD, results are consistent with impairments expected of each diagnostic group.

Characteristics of the child, as well as family-clinician dynamics at T1, can lead to difficulty differentiating the diagnoses of ASD and GDD and could potentially contribute to diagnostic instability. Mandell and colleagues (2009) found that the presence of global intellectual disability can make ASD harder to identify. The most common early symptoms of ASD, including delayed speech, poor response to others, and hyperactivity may be difficult to differentiate from GDD (Mandell et al., 2002), as they are common to both disorders. Also, the rising prevalence of ASD may contribute to increased attention to ASD symptoms at a young age, and decreased attention to evidence instead confirming GDD (Tirosh & Jaffe, 2013). All of these factors may contribute to difficulty with differential diagnosis at T1 and thus a higher likelihood of diagnostic instability between ASD and GDD.

Maternal education level was also found to be a significant predictor of diagnostic transition from ASD to GDD (but not ASD to non-ASD). In particular, having a parent who was not educated past high school was associated with increased odds of changing from ASD to GDD over time. In other words, these children with T1 ASD from lower SES backgrounds are more likely than children with ASD from high SES backgrounds to be perceived as having fewer ASD symptoms at T2 than T1 but having other

remaining delays at T2. This finding could be related to differences in experiences and developmental trajectories for children from different SES.

Research has demonstrated that SES can have a significant impact on development, and could thus contribute to different developmental trajectories. In particular, SES has been shown to have a strong relationship with cognitive ability, with those from low SES have been found to demonstrate lower IQ and academic achievement when they reach school age (Bradley & Corwyn, 2002; Hackman & Farah, 2009). Relatedly, children with ASD with more educated mothers were found to make greater cognitive gains in toddlerhood in one study (Ben- Itzhak & Zachor, 2011). Furthermore, children diagnosed with ASD have been consistently found to follow several potential trajectories, such as worsening and improving trajectories (e.g., Lord et al., 2012; Venker et al., 2014). Therefore, it is possible that the children from low SES who changed from ASD to GDD were on a trajectory of decreasing symptoms of ASD as discussed previously (i.e., social/communication symptoms and possibly RRB), though they may have been more likely to maintain cognitive delays and make fewer cognitive gains, based on previous research regarding SES and cognitive functioning. As cognitive impairment is often (but not always) associated with GDD (Riou, Ghosh, Francoeur, & Shevell, 2009), it is possible that SES impacted cognitive development, contributing to change from ASD to GDD in those from low SES.

Research has also demonstrated that other aspects of SES may contribute to identification of ASD symptoms at T1 but not T2. For example, children from low SES may have more limited exposure to enriching materials and experiences (Bradley & Corwyn, 2002), which may contribute to change from ASD to GDD in children from low SES. In particular, limited exposure to a variety of educational and imaginative toys may manifest in less developed play skills at T1, thus indicating a sign of ASD. Also, access to these enriching experiences and materials has been found to mediate the relationship between SES and behavior problems in children (Corwyn & Bradley, 2001). These increased behavioral problems may be interpreted as inflexibility in the context of an evaluation for ASD. Furthermore,

families from low SES backgrounds have been reported to experience greater levels of life stress and lower neighborhood quality (Hackman & Farah, 2009). These factors may contribute to reduced opportunity for social interactions between children; stress may interfere with time and ability to arrange social interactions for children, and low neighborhood quality may also limit peer socialization within the neighborhood. This limited social interaction at a young age may contribute to ASD diagnosis at T1, as it would affect reported peer relationships and, potentially, interactions during the evaluation. However, participants will have entered school by the time of their re-evaluation at T2. As a result, they will have had increased opportunity for social interaction with peers, and presumably more exposure to stimulating environments and resources within the school environment. Therefore, increased experience may result in reduced ASD symptom endorsement and thus a lack of ASD diagnosis by T2. However, as previously discussed, other delays may remain, as long-term cognitive effects have been reported in low SES populations. These combined factors may contribute to transition from ASD to GDD in those from low SES, as limited initial exposure to enriching environments and social interaction are more likely in the low SES group.

It is also notable that race approached significance as a predictor of transition from ASD to GDD. Previous research has generally supported the presence of disparities in access to services as a contributor to delayed ASD diagnosis and reduced improvement over time. For example, studies have found that that Black children are likely to be diagnosed at a later age than White children (Mandell et al., 2002; Mandell et al., 2009), and that Hispanic children are less likely to receive a diagnosis at all (Mandell et al., 2009). Also, one study found that individuals with White, non-Hispanic, well-educated mothers were most likely to show improvement over time (Fountain, Winter, & Bearman, 2012). Research often attributes underdiagnosis of minorities to delayed access to services and referral bias, in that minority children are less likely to seek or be referred for ASD evaluation (Begeer et al., 2008; Mandell et al., 2002). However, the current study minimizes these disparities, as children across sites

with varying racial and SES backgrounds are screened and evaluated within a standardized research protocol, resulting in a large number of minority participants and variability in SES. Therefore, it is possible that the method of widespread screening as recruitment, as well as the availability of free evaluations in the current study, minimizes disparities that often contribute to delayed diagnosis in minority children from low SES backgrounds. Consistent with this hypothesis, Herlihy and colleagues (2014) found that in a sample overlapping with the current study's sample, the standardized screening and evaluation procedure used in both studies minimized racial disparities in age of diagnosis and access to services.

In the current study, the finding that minority children who received an initial ASD diagnosis were more likely to cross the diagnostic boundary from ASD to GDD approached significance. This may be attributed to a decrease in ASD symptoms, such as social impairments and repetitive and sensory behaviors, but maintenance of cognitive, language, and/or motor delays contributing to their T2 diagnosis of GDD. Therefore, minorities may have been less likely to make non-ASD-specific developmental gains over time, as has been found in previous studies.

In terms of the contribution of race and SES in transition of ASD to GDD, it was initially thought that race and SES may be confounded, as often occurs in the literature (Kaufman, Cooper, & McGee, 1997). However, the two variables were not significantly correlated, and SES was significant in the logistic regression model when accounting for race (and race was nearly significant when accounting for SES). It is notable that site of evaluation predicted differences in transition from ASD to GDD, as 2.2 percent (N=6) of UConn participants compared to 7.3 percent (N=11) of GSU participants made this transition. The GSU sample was also more diverse, as 73 (48.7%) were of majority racial status and 77 (51.3%) were minorities, compared to 217 (78.6%) majority race participants and 59 (21.4%) minorities at UConn. Therefore, race may contribute to different rates of diagnostic change from ASD to GDD between sites.

Therefore, in addition to previously discussed explanations, it is possible that cultural differences between the racial groups may be impacting the assessment of minority children. One possibility is bias in the assessment of individuals of racial minorities. Research has found that raters often give biased accounts on assessments of children of minorities, though sometimes in a positive direction (Chang & Sue, 2003; Garb, 2006; Lethermon et al., 1986). Also, clinicians may interact with families with a minority racial status in different ways, affecting assessment results (Garb, 2006). Overall, it is extremely important for examiners to be aware of cultural differences in the assessment of children, and exert effort to provide unbiased assessments. However, it is also important to note that the number of participants transitioning from ASD to GDD at each site was low, so the apparent relationship between race and transition may not persist with larger sample sizes.

Finally, intervention was predictive of transition from ASD to GDD in the opposite direction as expected, as greater participating in intervention was associated with lower odds of changing diagnosis. This finding again seems to be related to change in participation in intervention over time (see Figure 4). The group ASD-GDD group participated in less initial intervention than the ASD-ASD group, and they continued to diverge over time. The overall measure of intervention hours does not take this change into account, which likely contributes to the unexpected intervention finding.

#### **4.4 Hypothesis 2A: non-ASD to ASD**

This study examined predictors of diagnostic instability between non-ASD and ASD. In particular, 19 participants who did not meet criteria for ASD at T1 did meet criteria at T2, which is a much larger sample than has been found in previous research. However, it is notable that the overall sample of participants in the current study (N=424) and those without ASD at T1 (N=154) is also significantly larger than in previous studies, which had total sample sizes ranging from a total of 31 to 131 participants with and without ASD at T1. A few studies noted the number of participants who made the transition from non-ASD to ASD, but the participants were not extensively studied due to small sample sizes of one or



two (Chawarska et al., 2007; Soke et al., 2011; van Daalen et al., 2009). In one study with nine participants newly meeting criteria for ASD at a second evaluation based on the ADI-R, authors were unable to provide a possible explanation for the transition, except that these participants initially more closely represented those who would end up with a final diagnosis of language delay than those with ASD (Cox et al., 1999). This statement implies that these participants' language delays were more apparent than social concerns at their initial evaluation, though data are not presented to support this hypothesis. Overall, occasional participants have been found to gain an ASD diagnosis over time, but factors contributing to this change have not been explored.

In the present study, initial (sub-threshold) ASD symptom severity and number of intervention hours predicted newly meeting criteria for ASD at T2. The finding regarding ASD symptom severity mirrors the finding of the ASD-NON analysis; those without ASD who had higher subthreshold symptom severity were more likely to later meet criteria for ASD, as they required less change over time to change diagnosis. Also, having greater impairment in social interaction skills, or other behaviors such as inflexibility and repetitive behaviors, are likely to hinder further development. Specifically, social interaction skills influence various aspects of development (Ochsner & Lieberman, 2001), so decreased engagement could have a cumulative effect to result in more significant delays over time. Furthermore, characteristics such as inflexibility and focus on repetitive interests may not be as pronounced at a young age (Wiggins et al., 2012b), but may become more apparent as demands increase with school participation and peer interaction. Therefore, not only would these symptoms be newly observable at a later evaluation, but they would likely interfere with other activities (e.g., attention at school, peer interactions) to further impact development.

Similarly, some have documented the failure of DSM-IV-TR criteria to fully apply to young children, resulting in many symptomatic children not meeting official ASD criteria (Stone et al., 1999). In particular, two-year-old children may not have opportunity to demonstrate symptoms related to

developing peer relationships, having impaired conversational ability, and using repetitive language (Stone et al., 1999). However, children who would end up meeting ASD criteria would be more likely to exhibit impairments in these domains by the time of their re-evaluations at age four, as they have had more opportunity for peer interaction, and their language skills are more likely to have developed to the point that these symptoms would be apparent. Anecdotally, parents in the current study often have difficulty reporting the quality of their toddlers' interactions with other children, as many do not have much opportunity to interact with peers at a young age. Because the examiners do not have the opportunity to observe the child with peers during their evaluations, the relevant DSM-IV-TR symptom is often not endorsed, though the symptom may in reality be a concern for the participant. Notably, of the 20 participants who changed from non-ASD to ASD in the current study, marked impairment in conversational ability was marked as a "not applicable" symptom for seven participants (and only two had the symptom endorsed), two had stereotyped language marked as "not applicable" (and only three had it endorsed), and no other symptoms were marked as "not applicable" at T1. Therefore, these symptoms were rarely endorsed at T1, which may reflect difficulty applying these criteria to toddlers. By T2, only three participants had the conversation symptom marked as "not applicable" (and 13 had the symptom endorsed), and no participants had stereotyped language marked as "not applicable" (and 12 had the symptom endorsed). These numbers present a decrease in the inapplicability of these symptoms, as well as an increase in the number of participants for whom the symptoms were endorsed. Therefore, the shift in applicability of these symptoms may have contributed to their diagnostic instability. These factors could contribute to participants with high subthreshold symptom severity newly meeting criteria for ASD later in childhood.

Initial ASD symptom severity also approached significance as a predictor of continuous change in severity over time, as higher initial severity in those without ASD at T1 was associated with more of a decrease in severity over time. This finding may seem counterintuitive, as the 19 participants who

changed diagnosis from non-ASD to ASD experienced an increase in symptoms. Notably, the mean increase in severity for the NON-ASD group was 2.42 ADOS CSS points (out of 10) compared to a .54-point decrease for the NON-NON group. However, the linear regression finding is likely related to regression toward the mean, as participants with extreme scores at T1 are less likely to have extreme scores at T2. Furthermore, intervention did not predict change in ASD symptom severity, which was consistent with previous analyses.

As in previous analyses, intervention predicted transition from non-ASD to ASD in the opposite direction of expectations. In particular, greater participation in intervention predicted transition from non-ASD to ASD. The potential explanations for this finding mirror those from analysis of hypothesis 1A, in which less intervention predicted transition from ASD to non-ASD. In sum, this finding is likely attributed to change in participation in intervention between the two evaluations (see Figure 5), as the NON-ASD group increased their intervention over time; however, aspects of severity not captured by the ADOS CSS or characteristics of the intervention variable itself may also contribute to this finding. Future research should clarify the role of intervention in children who newly meet criteria for ASD after not initially meeting criteria.

One question that arises in the consideration of children who are not diagnosed with ASD initially but who later receive a diagnosis is whether any of these children experienced clinically significant developmental regression, or whether development of expected skills may have slowed over time. Developmental regression refers to a marked loss in skills that were previously present (Shattuck et al., 2009) and which seems to be specific to ASD (Lord, Shulman, & DiLavore, 2004). It was found that only one of the NON-ASD children experienced developmental regression, according to parent report on the ADI-R. In particular, the parent reported that the child demonstrated a definite loss in social engagement and responsiveness, and that this change was temporally associated with a non-meningeal and non-encephalitic illness. Interestingly, various ratings of the child's ASD symptom severity

demonstrated low severity levels at both T1 and T2, though five DSM-IV-TR symptoms were endorsed at T2 compared to zero at T1. These findings indicate that the child may have demonstrated a high enough quantity of symptoms at T2 to receive a diagnosis, but that the severity of the symptoms was not high. This case example supports the notion that continuous symptom severity is important in consideration of developmental progression, and that a categorical diagnosis may not best represent a child's strengths and weaknesses.

It is important to note that the use of the ADI-R may not be the most reliable method of determining the frequency of developmental regression. In the present study, 28 of 219 children (12.8%) in the ASD-ASD group were reported to have experienced developmental regression according to parent report on the ADI-R. This proportion of children who experienced regression is lower than levels reported in other studies. A meta-analysis of regression in ASD including studies resulting in a total of 29,035 participants found that regression rates ranged from 25 to 39 percent, depending on the domain affected (e.g., language, social; Barger, Campbell, & McDonough, 2013). However, it is notable that most data (as in the present study) come from retrospective parent report, and many parents have difficulty separating delayed skills from actual developmental regression (Shattuck et al., 2009). Therefore it is possible that previous estimates of regression frequency are inaccurate, or that the proportion of children who were reported to have regressed in the current study may not be fully representative of the sample.

#### **4.5 Hypothesis 2B: GDD to ASD**

The present study examined a small group (N=8) of participants who met criteria for GDD at T1, but newly met criteria for ASD at T2. Children who made similar diagnostic transitions have not been examined in prior studies. One participant from Chawarska and colleagues' (2007) study made the transition from developmental delay to ASD, though factors contributing to this transition were not discussed. Due to a lack of previous research on these participants, they were explored in the current

study. However, small sample size (N=8) resulted in limited power for analyses, so T1 variables which may contribute to diagnostic instability were examined qualitatively.

The GDD-ASD group had higher T1 and T2 ASD symptom severity than the GDD-GDD group, as expected. This result is consistent with findings of the other hypotheses in this study, in that initial symptom severity is indicative of later diagnostic transition. Overall, the GDD-GDD group did not appear to differ from the GDD-ASD group on measures of expressive language and cognitive ability, which is consistent with the finding that individuals with both diagnoses tend to exhibit delays in these domains (Tirosh & Jaffe, 2011; Zwaigenbaum et al., 2009). However, the GDD-ASD group seemed to have less variability in receptive language ability, with scores concentrated on the lower end of the spectrum rather than spread throughout the full range as did the GDD-GDD group. This finding is not surprising given the high prevalence of language delays in children with ASD (Trillingsgaard et al., 2005). It is also consistent with some comparisons of receptive language ability between children with ASD and those with GDD, in which those with GDD tended to have higher scores (Trillingsgaard et al., 2005). However, many studies have not found reliable communication differences between children with ASD and those with GDD, which supports the lack of difference between groups in expressive language ability. Further research should explore which of these factors will predict diagnostic transition between GDD and ASD using larger sample sizes. The current study suggests that ASD symptom severity and communication ability may be promising predictors of diagnostic instability.

#### **4.6 Diagnostic change versus change in continuous symptom severity**

In addition to predictors of diagnostic instability, there are important implications for predictors of change in a continuous measure of symptom severity. In particular, some of the variables found to predict diagnostic instability were not predictive of change in ASD symptom severity. This indicates that although they were associated with transition across a diagnostic boundary, they were not associated with significant change in severity. Although a change in severity may seem like the more important

measure of developmental gains and delays, current diagnostic conventions focus on categorical distinctions between receiving a diagnosis or not. Diagnosis is based on endorsement in a minimum number of symptoms, meaning that a child with a few severe symptoms may not meet criteria, whereas a child with multiple mild symptoms will meet criteria. The finding that several variables predicted diagnostic transition but not change in symptom severity supports the notion that continuous ratings of symptom severity may be more useful indicators of progress and developmental delay than diagnostic status, in contrast with current diagnostic conventions. Therefore, it may be beneficial to move toward a more continuous method of indicating impairment. In a study with a sample overlapping with the current sample, Wiggins and colleagues (2012b) found that the rate and intensity of symptoms better distinguished ASD subgroups than number of symptoms, further supporting a dimensional view of ASD. Although the DSM-5 sought to develop a more continuous rating of severity for individuals with ASD, this rating is still categorical (i.e., mild, moderate, severe), and the diagnostic criteria are arguably more stringent (i.e., requiring a child to meet all three criteria in the social communication domain, requiring restricted or stereotyped interest or behavior symptoms). Overall, changes in diagnosis do not necessarily reflect true change in symptom severity, which indicates that categorical distinctions are not the most valuable measure of change over time.

#### **4.7 DSM-IV-TR versus DSM-5**

As previously mentioned, the diagnostic criteria for ASD were recently revised. There has been significant debate over whether these revisions represent an improvement over previous criteria, or whether they present a regression because some individuals who would have met DSM-IV-TR criteria will not meet DSM-5 criteria for ASD. In general, the DSM-5 criteria were designed to respond to criticism of DSM-IV-TR criteria. For example, research has cited that the DSM-IV-TR subtypes of ASD (e.g., Autistic Disorder, PDD-NOS, Asperger's Disorder) are not different and valid constructs; they are not all well-defined and do not consistently differ in characteristics such as developmental level or IQ,

etiology, or response to treatment (Happé, 2011). Instead, the DSM-5 includes modifiers indicating severity level of ASD to replace the previous subtypes (Happé, 2011). This change in itself may improve the validity of ASD diagnosis by guiding intervention with a more dimensional measure of severity.

Another change to ASD criteria in the DSM-5 included more stringent criteria which will minimize the issue of “overdiagnosis” that has been a focus in the media. However, these changes also present some negative effects. The more stringent criteria (i.e., requiring *all* three symptoms in the social-communication domain rather than a subset, as well as two out of four symptoms related to restricted, repetitive, and stereotyped behaviors) may result in a large subset of children who would have met ASD criteria on the DSM-IV-TR to not meet criteria on the DSM-5. Several studies have examined the proportion of children who would have met DSM-IV-TR criteria but not DSM-5 criteria. Some have found that most participants with DSM-IV-TR diagnoses would be identified by the DSM-5 (e.g., 91% in Huerta et al., 2012). However, many studies have found that a significant proportion of individuals, especially those who would have been diagnosed with a PDD aside from Autistic Disorder, no longer meet criteria according to the DSM-5 (e.g., 23.4% in Gibbs et al., 2012; 39.4% in McPartland, Reichow, & Volkmar, 2012). This finding stands in research on toddlers specifically, as Matson and colleagues (2012) found that 47.8 percent of toddlers with ASD based on DSM-IV-TR criteria did not meet criteria on DSM-5. Unfortunately, this means that many children who would benefit from ASD-related services will have more difficulty obtaining them, as they do not have a clinical diagnosis (Matson et al., 2012). Barton and colleagues (2013) even sought to propose revised criteria to minimize the impact of the criteria change on individuals who do not meet stringent criteria. Overall, the DSM-5 criteria have been shown to reduce the number of individuals diagnosed with ASD compared to DSM-IV-TR criteria.

In the case of the current study, this criteria change also means that a portion of the discussed participants who received an ASD diagnosis based on the DSM-IV-TR would likely not receive a diagnosis

with the DSM-5. Therefore, the rates of diagnostic instability of ASD may differ when using DSM-5 criteria. However, little research has examined stability of DSM-5 diagnosis, as the revision was fairly recent. It is notable that toddlers who meet criteria for ASD based on the DSM-5 often have more severe symptomatology than those who met on the DSM-IV-TR (Matson et al., 2012). The current study revealed that higher initial symptom severity results in a more stable diagnosis, which indicates that DSM-5 diagnosis in toddlers may be more stable. However, it is also possible that these children do not actually have more severe symptoms, but a greater number of symptoms.

Furthermore, the DSM-5 criteria for ASD include a modification that symptoms may be endorsed based on current or historical behaviors (APA, 2013), in contrast with current behaviors only in the DSM-IV-TR. This inclusion of history would in itself guarantee diagnostic stability of initial ASD diagnosis, as those children would always have a history of symptoms. However, the idea of measuring stability of symptoms and meeting symptom thresholds over time will still be important to study. The current study supports the finding that ASD symptoms are not always persistent, and that even children who have symptoms severe enough to meet ASD criteria make enough developmental progress to no longer meet criteria later. This change in symptom severity also may call for a change in the academic setting and services that the child receives. Therefore, it would not be helpful to assume that all children with an initial diagnosis will always meet criteria for ASD, as this could result in involvement in outdated and unhelpful services, as well as the ongoing stress of being labeled with a disorder that no longer applies. As a result, the inclusion of historical symptoms in the DSM-5 criteria for ASD should not indicate that all diagnoses are stable, but that ASD diagnostic and symptom stability should still be examined and considered in prognosis and treatment of children.

#### **4.8 Limitations**

The current study has several limitations. First, attrition bias may affect the results, as there were differences between the group who returned for a re-evaluation and those who did not return. In



particular, those with greater T1 ASD severity, weaker social skills, ASD diagnosis, and non-minority racial status were more likely to return for a re-evaluation than other participants. It is not surprising that parents of children with an ASD diagnosis and those with more severe symptoms (and weaker social skills) would have greater interest in returning for a re-evaluation in order to track their children's developmental progress and reassess the need for services. However, this result also implies that the study was unable to acquire T2 data from a large number of participants who did not initially have ASD or who had low symptom severity. This may bias results as more data are missing from the T1 non-ASD group than from the ASD group, and it is known from the results of this and other studies that even non-ASD status can be unstable. Furthermore, families of minority racial status were less likely to return for a re-evaluation. Individuals from minority races may be less likely to return for a re-evaluation due to several factors, such as mistrust of the medical community. Also, individuals of minority racial status may appraise diagnoses differently due to cultural differences, for example believing that developmental disabilities are punishment for sin, or that it is not appropriate to label a child with a diagnosis (Dyches et al., 2004). Due to differences in attrition rates, information on diagnostic stability may be more applicable to non-minority status children for whom more information is available.

Furthermore, due to the large scale of this study, data were collected by multiple examiners at multiple sites. Although all examiners were trained on measures, individual variability in scoring, interpretation, and diagnosis may exist. As previously discussed, site predicted differences in transition from ASD to GDD. However, the lack of diagnostic instability differences in the other transition groups (e.g., ASD-NON, NON-ASD) contributes to confidence that widespread biases between sites are not significant. Additionally, the large sample size was a benefit to the study, although the sample sizes of certain subgroups (e.g., NON-ASD) were still relatively small.

Measurement issues may also contribute to study limitations. As previously stated, the intervention data available were not ideal as they did not include information on intensity, quality, or

type of services, and they were based on retrospective parent report. Future research should examine the influence of these aspects of intervention on diagnostic stability. Also, although the ADOS CSS was selected due to its validation for use across time, age, and language ability, this may also be a weakness of the measure. In particular, the measure is standardized across age and language groups, which may result in a loss of variance. Also, as previously mentioned, regression toward the mean may impact comparisons between severity at T1 and T2. Finally, it is possible that changes in perception by clinicians assigning diagnosis may contribute to different severity ratings between T1 and T2 and ultimately to diagnostic stability rates. Although reliability of diagnosis by clinicians is an important consideration in this study of diagnostic instability, findings in previous literature lend support to the findings in the current study. In particular, previous studies have consistently found that children's strengths and weaknesses change throughout development, and that around 20 percent of participants with ASD change diagnosis over time. Nonetheless, it is important to acknowledge that some changes in results of child assessment over time may be associated with change in the clinician's perception of the child.

Despite these limitations, the current study contributes to the literature in various ways. It is a rare longitudinal study of a large sample of toddlers. The sample was also racially/ethnically diverse, contributing to generalizability of results. Diagnoses were based on the integration of observation and parent report, based off of reliable and valid assessments. Furthermore, the participant recruitment method was beneficial because it resulted in multiple diagnostic groups and avoided selection biases resulting from other forms of recruitment. The study also had a direct benefit to participants, as they were given full evaluations, clinical reports, and recommendations.

#### **4.9 Implications**

The results of the current study have significant implications for early diagnosis and treatment of ASD and GDD. First, the study replicated previous findings regarding rates of early diagnostic instability of ASD of nearly 20 percent. Although some may cite this finding as support that early

diagnosis is unreliable and that diagnosis should be avoided until a later age, the identification of predictors of this instability indicates that there is opportunity for early diagnosis to be refined. For example, clinicians may relate these results to a child's individual profile, which can provide guidance for intervention and prognosis. Furthermore, diagnostic instability of ASD does not necessarily imply that early diagnosis is inaccurate, but that children with ASD may have skills that support development and which may be strengthened in certain situations, such as targeted interventions and increased peer interaction in school. The case of participation in intervention contributing to diagnostic instability is a positive outcome, rather than evidence that early diagnosis is not valuable.

Relatedly, the variables found to be significant predictors of diagnostic instability of ASD or GDD may be specifically targeted in interventions, as they appear to be associated with developmental gains. In particular, ASD symptom severity and receptive language ability were associated with decreases in severity over time and loss of ASD diagnosis. This indicates that, as expected, interventions that target ASD symptoms may help a child to make gains to no longer meet criteria over time. Similarly, receptive language ability in particular may be another area to focus on, as strengths in this domain seem to relate to overall developmental gains. Although a child with a language delay would likely be referred for language support regardless of ASD diagnosis, the current finding suggests that perhaps more children with language weaknesses and ASD but without a clinical language delay would benefit from this type of intervention as well.

Variables predicting diagnostic instability will be important to consider when determining a child's prognosis. For example, severity of a child's ASD symptoms and receptive language ability have been shown to be related to instability of ASD diagnosis. Therefore, a child's strengths and weaknesses in these domains will be important for relaying prognostic information to parents. This information could be a source of comfort to a parent receiving a diagnosis for their child who has relatively mild symptoms and strong receptive abilities. On the other end of the spectrum, it would also provide some

indication for future preparation, such as planning for longer term services in children with more severe symptoms and limited communication ability. Regardless, this information would help to identify children who will benefit from repeated evaluations; if they have a high likelihood of having an unstable diagnosis, then it will be important to have them re-evaluated frequently in order to track developmental progress and identify appropriate services. Thus, receptive language may be an important target for diagnostic assessment measures; because receptive language ability is important in determining odds of diagnostic instability, clinicians making diagnoses would benefit from integrating information from language measures when determining a child's diagnosis.

Furthermore, this study highlights the importance of demographic factors when assessing toddlers for developmental delays. Discrepancies in the prevalence of ASD, age at initial diagnosis, and access to health professionals between races have been documented in surveillance studies (e.g., Mandell et al., 2002; Mandell et al., 2009; Mandell, Novak, & Zubritzky, 2005). However, previous research has not indicated that SES or race are associated with diagnostic transition from ASD to GDD. The finding that SES and possibly race may predict diagnostic instability calls into question various factors which may contribute to this finding, such as a lack of resources or information on child development for families before an initial diagnosis, and possible parent-child-clinician dynamics in which cultural differences impact assessment and diagnosis. Therefore, education related to environment and behaviors that support child development for parents from lower SES backgrounds is recommended, even before a child is diagnosed with a developmental delay. Also, education for clinicians on cultural competency is indicated in order to reduce assessment bias. Clinicians should also consider the cultural sensitivity of the measures that they use in determining a diagnosis, as they may not all be appropriate for use with children with different racial and ethnic backgrounds. Despite these possible biases contributing to diagnostic reassignment in children from low SES, it is also reassuring

that children are able to make up for some delays, to an extent, after supports such as school are in place.

Although the findings regarding intervention and diagnostic instability are unexpected in the current study, this is still an important direction for research. The current study indicated that participation in intervention has a role in influencing child development. However, it will be important to better quantify frequency, intensity, quality, and type of intervention in order to be able to measure benefits more directly.

Overall, results of the current study support the finding that ASD and GDD diagnoses may be unstable in toddlers, and that initial child characteristics and participation in intervention can predict this diagnostic instability. Therefore, these predictors can be used to inform early diagnostic assessment, prognosis, and intervention. The study further supports consideration of ASD impairment on a continuum, as predictors of change in ASD symptom severity did not consistently correspond with predictors of change in diagnostic category. Therefore, measures of change in severity over time may be more useful in evaluating developmental progress than diagnostic transition itself. Regardless, it is important to reassess developmental progress throughout childhood, as a diagnostic label may not continue to be relevant.

## REFERENCES

- Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism and Developmental Disorders, 39*(1), 84-96.
- Aitken, K. (1991). Diagnostic Issues in Autism—Are We Measuring the Emperor for Another Suit of Clothes?. *Developmental Medicine & Child Neurology, 33*(11), 1015-1020.
- Akshoomoff, N. (2006). Use of the Mullen Scales of Early Learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychology, 12*(4-5), 269-277.
- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders, Text Revision* (4th ed.). Washington, DC: American Psychiatric Publishing.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: American Psychiatric Publishing.
- Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*(5), 1311.
- Bachevalier, J., & Loveland, K. A. (2006). The orbitofrontal–amygdala circuit and self-regulation of social–emotional behavior in autism. *Neuroscience & Biobehavioral Reviews, 30*(1), 97-117.
- Baio, J. (2014). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Report: Surveillance Summaries* (Washington, DC: 2002), 63, 1.

- Barton, M. L., Robins, D. L., Jashar, D., Brennan, L., & Fein, D. (2013). Sensitivity and specificity of proposed DSM-5 criteria for autism spectrum disorder in toddlers. *Journal of Autism and Developmental Disorders, 43*(5), 1184-1195.
- Barger, B. D., Campbell, J. M., & McDonough, J. D. (2013). Prevalence and onset of regression within autism spectrum disorders: a meta-analytic review. *Journal of Autism and Developmental Disorders, 43*(4), 817-828.
- Bayley, N. *Bayley Scales of Infant Development-II*. (1994). San Antonio: Psychological Corporation.
- 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.
- Ben-Itzhak, E., & Zachor, D. A. (2011). Who benefits from early intervention in autism spectrum disorders?. *Research in Autism Spectrum Disorders, 5*(1), 345-350.
- Ben-Itzhak, E., Watson, L. R., & Zachor, D. A. (2014). Cognitive Ability is Associated with Different Outcome Trajectories in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 1*-9.
- Boyd, R. D. (1989). What a difference a day makes: Age-related discontinuities and the Battelle Developmental Inventory. *Journal of Early Intervention, 13*(2), 114-119.
- Bradley-Johnson, S. (1997). Mullen Scales of Early Learning. *Psychology in the Schools, 34*(4), 379-382.
- Camarata, S. (2014). Early identification and early intervention in autism spectrum disorders: Accurate and effective?. *International Journal of Speech-Language Pathology, 16*(1), 1-10.
- Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, i*-174.

- Casasola, M. (2008). The development of infants' spatial categories. *Current Directions in Psychological Science, 17*(1), 21-25.
- Chang, D. F., & Sue, S. (2003). The effects of race and problem type on teachers' assessments of student behavior. *Journal of Consulting and Clinical Psychology, 71*(2), 235-242.
- Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., & Pickles, A. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *The British Journal of Psychiatry, 191*(6), 554-559.
- Charman, T., Howlin, P., Berry, B., & Prince, E. (2004). Measuring developmental progress of children with autism spectrum disorder on school entry using parent report. *Autism, 8*(1), 89-100.
- Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J. A., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry, 46*(5), 500-513.
- Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry, 48*(2), 128-138.
- Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics, 131*(4), e1121-e1127.
- Chomsky, N. (1966). *Cartesian Linguistics: Chapter in the History of Rationalist Thought*. Harper and Row.
- Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Leventhal, B. L., & Lord, C. (2007). Between a ROC and a hard place: decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, 48*(9), 932-940.



- Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., ... & Wheelwright, S. (1999). Autism spectrum disorders at 20 and 42 months of age: Stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry*, 40(5), 719-732.
- Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, 20(03), 775-803.
- Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., ... & Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics*, 125(1), e17-e23.
- Dyches, T. T., Wilder, L. K., Sudweeks, R. R., Obiakor, F. E., & Algozzine, B. (2004). Multicultural issues in autism. *Journal of Autism and Developmental Disorders*, 34(2), 211-222.
- Eaves, L. C., & Ho, H. H. (2004). The very early identification of autism: Outcome to age 41/2–5. *Journal of Autism and Developmental Disorders*, 34(4), 367-378.
- Elsabbagh, M. (2014). Risk and resilience in the developing brain. *The Psychologist*, 27(10), 754-757.
- Elsabbagh, M., Mercure, E., Hudry, K. et al. (2012). Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Current Biology*, 22(4), 338–342.
- Fein, D., Barton, M., Eigsti, I. M., Kelley, E., Naigles, L., Schultz, R. T., ... & Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, 54(2), 195-205.
- Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics*, 129(5), e1112-e1120.
- Garb, H. N. (1997). Race bias, social class bias, and gender bias in clinical judgment. *Clinical Psychology: Science and Practice*, 4(2), 99-120.

- Gibbs, V., Aldridge, F., Chandler, F., Witzlsperger, E., & Smith, K. (2012). Brief report: an exploratory study comparing diagnostic outcomes for autism spectrum disorders under DSM-IV-TR with the proposed DSM-5 revision. *Journal of Autism and Developmental Disorders, 42*(8), 1750-1756.
- Gillberg, C. (1990). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 31*(1), 99-119.
- Glascoe, F. P. (1994). It's Not What It Seems The Relationship Between Parents' Concerns and Children With Global Delays. *Clinical Pediatrics, 33*(5), 292-296.
- Glascoe, F. P. (2005). Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(3), 173-179.
- Goldin-Meadow, S., Levine, S.C., Hedges, L.V., Huttenlocher, J., Raudenbush, SW., & Small, S.L. (2014). New Evidence About Language and Cognitive Development Based on a Longitudinal Study: Hypotheses for Intervention. *American Psychologist, 69*(6), 588-599.
- Gopnik, A. (1996). The post-Piaget era. *Psychological Science, 221*-225.
- Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(5), 693-705.
- Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics, 130*(5), e1278-e1284.
- Gotham, K., Risi, S., Dawson, G., Tager-Flusberg, H., Joseph, R., Carter, A., ... & Lord, C. (2008). A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(6), 642-651.
- Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry, 54*(5), 582-590.

- Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences, 13*(2), 65-73.
- Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843-855.
- Happé, F. (2011). Criteria, categories, and continua: autism and related disorders in DSM-5. *Journal of the American Academy of Child & Adolescent Psychiatry, 50*(6), 540-542.
- Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science, 14*(2), 69-73.
- Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: A four-to six-year follow-up. *Journal of Autism and Developmental Disorders, 30*(2), 137-142.
- Hedvall, Å., Westerlund, J., Fernell, E., Holm, A., Gillberg, C., & Billstedt, E. (2014). Autism and developmental profiles in preschoolers: stability and change over time. *Acta Paediatrica, 103*(2), 174-181.
- Helles, A., Gillberg, C. I., Gillberg, C., & Billstedt, E. (2014). Asperger syndrome in males over two decades: stability and predictors of diagnosis. *Journal of Child Psychology and Psychiatry.*
- Helt, M., Kelley, E., Kinsbourne, M., Pandey, J., Boorstein, H., Herbert, M., & Fein, D. (2008). Can children with autism recover? If so, how?. *Neuropsychology Review, 18*(4), 339-366.
- Herlihy, L. E., Brooks, B., Dumont-Mathieu, T., Barton, M. L., Fein, D., Chen, C. M., & Robins, D. L. (2014). Standardized screening facilitates timely diagnosis of autism spectrum disorders in a diverse sample of low-risk toddlers. *Journal of Developmental & Behavioral Pediatrics, 35*(2), 85-92.
- Holmboe, K., Rijdsdijk, F. V., Hallett, V., Happé, F., Plomin, R., & Ronald, A. (2014). Strong genetic influences on the stability of autistic traits in childhood. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*(2), 221-230.

- Hsieh, F. Y., Bloch, D. A., & Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, *17*(14), 1623-1634.
- Huerta, M., Bishop, S. L., Duncan, A., Hus, V., & Lord, C. (2014). Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. *American Journal of Psychiatry*, *169*(10), 1056-1064.
- Jencks, C. & Phillips, M., eds. (1998). Racial bias in testing. *The Black-White test score gap*. The Bookings Institution, Washington, D.C.
- Johnson, C. P., & Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics*, *120*(5), 1183-1215.
- Kaufman, J. S., Cooper, R. S., & McGee, D. L. (1997). Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. *Epidemiology*, 621-628.
- Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S., ... & Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(4), 606-615.
- Le Couteur, A., Lord, C., & Rutter, M. (2003). *The Autism Diagnostic Interview—Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services.
- Lethermon, V. R., Williamson, D. A., Moody, S. C., & Wozniak, P. (1986). Racial bias in behavioral assessment of children's social skills. *Journal of Psychopathology and Behavioral Assessment*, *8*(4), 329-337
- Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*(4), 1130-1144.
- Lord, C. (1995). Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry*, *36*(8), 1365-1382.

- Lord, C., Luyster, R., Guthrie, W., & Pickles, A. (2012). Patterns of developmental trajectories in toddlers with autism spectrum disorder. *Journal of Consulting and Clinical Psychology, 80*(3), 477.
- Lord, C., Luyster, R.J., Gotham, K., & Guthrie, W. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module*. Torrance, CA: Western Psychological Services.
- Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., & Bishop, S.L. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1-4*. Torrance, CA: Western Psychological Services.
- Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism Diagnostic Observation Schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19*, 185–212.
- Lord, C., Shulman, C., & DiLavore, P. (2004). Regression and word loss in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry, 45*(5), 936-955.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*(1), 3.
- Luyster, R., Gotham, K., Guthrie, W., Coffing, 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*(9), 1305-1320.
- Luyster, R., Qiu, S., Lopez, K., & Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur–Bates Communicative Development Inventory. *Journal of Speech, Language, and Hearing Research, 50*(3), 667-681.

- Mandell, D. S., Listerud, J., Levy, S. E., & Pinto-Martin, J. A. (2002). Race differences in the age at diagnosis among Medicaid-eligible children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(12), 1447-1453.
- Mandell, D. S., Novak, M. M., & Zubritsky, C. D. (2005). Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics, 116*(6), 1480-1486.
- Mandell, D. S., Wiggins, L. D., Carpenter, L. A., Daniels, J., DiGuseppi, C., Durkin, M. S., ... & Kirby, R. S. (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health, 99*(3), 493.
- Matson, J. L., Kozlowski, A. M., Hattier, M. A., Horovitz, M., & Sipes, M. (2012). DSM-IV vs DSM-5 diagnostic criteria for toddlers with autism. *Developmental Neurorehabilitation, 15*(3), 185-190.
- Mazurek, M. O., Kanne, S. M., & Miles, J. H. (2012). Predicting improvement in social-communication symptoms of autism spectrum disorders using retrospective treatment data. *Research in Autism Spectrum Disorders, 6*(1), 535-545.
- McPartland, J. C., Reichow, B., & Volkmar, F. R. (2012). Sensitivity and specificity of proposed DSM-5 diagnostic criteria for autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(4), 368-383.
- Mitchell, S., Cardy, J. O., & Zwaigenbaum, L. (2011). Differentiating autism spectrum disorder from other developmental delays in the first two years of life. *Developmental Disabilities Research Reviews, 17*(2), 130-140.
- Mukaddes, N. M., Tutkunkardas, M. D., Sari, O., Aydin, A., & Kozanoglu, P. (2014). Characteristics of Children Who Lost the Diagnosis of Autism: A Sample from Istanbul, Turkey. *Autism Research and Treatment, 2014*.
- Mullen, E. M. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service. *Inc.*(21).

- Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research, 2*(1), 2-21.
- Muratori, F., & Narzisi, A. (2014). Exploratory study describing 6 month outcomes for young children with autism who receive treatment as usual in Italy. *Neuropsychiatric Disease and Treatment, 10*, 577-586.
- Newborg, J. (1984). *Battelle developmental inventory*. Allen, TX: DLM Teaching Resources.
- Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist, 56*(9), 717.
- Orinstein, A.J., Helt, M., Troyb, E., Tyson, K.E., Barton, M.L., Eigsti, I., Naigles, L., & Fein, D.A. (2014). Intervention for Optimal Outcome in Children and Adolescents with a History of Autism. *Journal of Developmental and Behavioral Pediatrics*.
- Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and psychopathology, 14*(02), 239-251.
- Paul, R., Chawarska, K., Cicchetti, D., & Volkmar, F. (2008). Language outcomes of toddlers with autism spectrum disorders: a two year follow-up. *Autism Research, 1*(2), 97-107.
- Pellicano, E. (2012). Do autistic symptoms persist across time? Evidence of substantial change in symptomatology over a 3-year period in cognitively able children with autism. *American Journal on Intellectual and Developmental Disabilities, 117*(2), 156-166.
- Piaget, J. (1973). *The child and reality: Problems of genetic psychology*. NY: Grossman Publishers.
- Pinker, S. (1994). *The language instinct: The new science of language and mind* (Vol. 7529). Penguin UK.
- Reynolds, C. R., & Kamphaus, R. W. (2004). *BASC-2: Behavior Assessment System for Children (2nd ed.)*. Circle Pines, MN: AGS.

- Riou, E. M., Ghosh, S., Francoeur, E., & Shevell, M. I. (2009). Global developmental delay and its relationship to cognitive skills. *Developmental Medicine & Child Neurology*, *51*(8), 600-606.
- Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., ... & Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(9), 1094-1103.
- Robins, D. L., Casagrande, K., Barton, M., Chen, C. M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised With Follow-up (M-CHAT-R/F). *Pediatrics*, *133*(1), 37-45.
- Robins, D. L., Fein, D., Barton, M. L., & 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*(2), 131-144.
- Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*(5916), 951-953.
- Rutter, M., Bailey, A., & Lord, C. *Social Communication Questionnaire (SCQ)*. (2003). Los Angeles, CA: Western Psychological Services.
- Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule: ADOS-2*. Los Angeles, CA: Western Psychological Services.
- Rydz, D., Shevell, M. I., Majnemer, A., & Oskoui, M. (2005). Topical review: developmental screening. *Journal of Child Neurology*, *20*(1), 4-21.
- Rydz, D., Srour, M., Oskoui, M., Marget, N., Shiller, M., Birnbaum, R., ... & Shevell, M. I. (2006). Screening for developmental delay in the setting of a community pediatric clinic: a prospective assessment of parent-report questionnaires. *Pediatrics*, *118*(4), e1178-e1186.



- Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders, 10*(1), 91-103.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *The Childhood Autism Rating Scale—Second Edition Manual*. Los Angeles, CA: Western Psychological Services.
- 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 & Adolescent Psychiatry, 48*(5), 474-483
- Shevell, M. I. (2010). Present conceptualization of early childhood neurodevelopmental disabilities. *Journal of Child Neurology, 25*(1), 120-126.
- Shevell, M., Majnemer, A., Platt, R. W., Webster, R., & Birnbaum, R. (2005a). Developmental and functional outcomes in children with global developmental delay or developmental language impairment. *Developmental Medicine & Child Neurology, 47*(10), 678-683.
- Shevell, M., Majnemer, A., Platt, R. W., Webster, R., & Birnbaum, R. (2005b). Developmental and functional outcomes at school age of preschool children with global developmental delay. *Journal of Child Neurology, 20*(8), 648-654.
- Shumway, S., Farmer, C., Thurm, A., Joseph, L., Black, D., & Golden, C. (2012). The ADOS calibrated severity score: relationship to phenotypic variables and stability over time. *Autism Research, 5*(4), 267-276.
- Siegel, B., Pliner, C., Eschler, J., & Elliott, G. R. (1988). How children with autism are diagnosed: Difficulties in identification of children with multiple developmental delays. *Journal of Developmental & Behavioral Pediatrics, 9*(4), 199-204.

- Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., & Frith, U. (2008). Levels of emotional awareness and autism: an fMRI study. *Social Neuroscience, 3*(2), 97-112.
- Silva, P. A., McGee, R., & Williams, S. (1985). Some characteristics of 9-year-old boys with general reading backwardness or specific reading retardation. *Journal of Child Psychology and Psychiatry, 26*(3), 407-421.
- Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*(8), 343-348.
- Soke, G. N., Philofsky, A., Diguseppi, C., Lezotte, D., Rogers, S., & Hepburn, S. (2011). Longitudinal changes in Scores on the Autism Diagnostic Interview—Revised (ADI-R) in pre-school children with autism Implications for diagnostic classification and symptom stability. *Autism, 15*(5), 545-562.
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales: Interview edition, survey form manual*. Circle Pines, MN: American Guidance Service.
- Sparrow, S., Balla, D., & Cicchetti, D. V. (2005). *Vineland Adaptive Behavior Scales, Second Edition*. Circle Pines, MN: American Guidance Services.
- Stone, W. L., Lee, E. B., Ashford, L., Brissie, J., Hepburn, S. L., Coonrod, E. E., & Weiss, B. H. (1999). Can autism be diagnosed accurately in children under 3 years?. *Journal of Child Psychology and Psychiatry, 40*(2), 219-226.
- Sutera, S., Pandey, J., Esser, E. L., Rosenthal, M. A., Wilson, L. B., Barton, M., ... & Fein, D. (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(1), 98-107.
- Thomaidis, L., Zantopoulos, G. Z., Fouzas, S., Mantagou, L., Bakoula, C., & Konstantopoulos, A. (2014). Predictors of severity and outcome of global developmental delay without definitive etiologic yield: a prospective observational study. *BMC pediatrics, 14*(1), 40.

- Thomas, P., Zahorodny, W., Peng, B., Kim, S., Jani, N., Halperin, W., & Brimacombe, M. (2011). The association of autism diagnosis with socioeconomic status. *Autism, 0*(0), 1-13.
- Tirosh, E., & Jaffe, M. (2011). Global developmental delay and mental retardation—a pediatric perspective. *Developmental Disabilities Research Reviews, 17*(2), 85-92.
- Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology, 509-529*.
- Toth, K., Munson, J., Meltzoff, A. N., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. *Journal of Autism and Developmental Disorders, 36*(8), 993-1005.
- Trillingsgaard, A., Sørensen, E. U., Němec, G., & Jørgensen, M. (2005). What distinguishes autism spectrum disorders from other developmental disorders before the age of four years?. *European Child & Adolescent Psychiatry, 14*(2), 65-72.
- Turner, L. M., & Stone, W. L. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry, 48*(8), 793-802.
- Turner, L. M., Stone, W. L., Pozdol, S. L., & Coonrod, E. E. (2006). Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism, 10*(3), 243-265.
- Valtonen, R., Ahonen, T., Lyytinen, P., & Tolvanen, A. (2007). Screening for developmental risks at 4 years of age: Predicting development two years later. *Nordic Psychology, 59*(2), 95.
- Van Daalen, E., Kemner, C., Dietz, C., Swinkels, S. H., Buitelaar, J. K., & 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 & Adolescent Psychiatry, 18*(11), 663-674.
- Vaughan, C. A. (2011). Test Review: E. Schopler, ME Van Bourgondien, GJ Wellman, & SR Love Childhood Autism Rating Scale. *Journal of Psychoeducational Assessment, 29*(5), 489-493.

- Veness, C., Prior, M., Bavin, E., Eadie, P., Cini, E., & Reilly, S. (2012). Early indicators of autism spectrum disorders at 12 and 24 months of age: A prospective, longitudinal comparative study. *Autism, 16*(2), 163-177.
- Venker, C. E., Ray-Subramanian, C. E., Bolt, D. M., & Weismer, S. E. (2014). Trajectories of Autism Severity in Early Childhood. *Journal of autism and developmental disorders, 44*(3), 546-563.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children, 34*-41.
- Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-VanderWeele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics, 127*(5), e1303-e1311.
- Webb, S. J., Jones, E. J., Kelly, J., & Dawson, G. (2014). The motivation for very early intervention for infants at high risk for autism spectrum disorders. *International Journal of Speech-Language Pathology, 16*(1), 36-42.
- Webster, R. I., Majnemer, A., Platt, R. W., & Shevell, M. I. (2008). Child health and parental stress in school-age children with a preschool diagnosis of developmental delay. *Journal of Child Neurology, 23*(1), 32-38.
- Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III)*. Sydney, NSW: Pearson.
- Wetherby, A. M., & Prizant, B. M. (2002). *Communication and Symbolic Behavior Scales: Developmental Profile*. Baltimore, MD: Paul H Brookes Publishing.
- Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders, 37*(5), 960-975.

- Wetherby, A. M., 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*(5), 473-493.
- Wiggins, L. D., Baio, J., Schieve, L., Lee, L. C., Nicholas, J., & Rice, C. E. (2012a). Retention of autism spectrum diagnoses by community professionals: Findings from the Autism and Developmental Disabilities Monitoring Network, 2000 and 2006. *Journal of Developmental & Behavioral Pediatrics, 33*(5), 387-395.
- Wiggins, L. D., Robins, D. L., Adamson, L. B., Bakeman, R., & Henrich, C. C. (2012b). Support for a dimensional view of autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders, 42*(2), 191-200.
- Wiggins, L. D., Robins, D. L., Bakeman, R., & Adamson, L. B. (2009). Brief report: sensory abnormalities as distinguishing symptoms of autism spectrum disorders in young children. *Journal of Autism and Developmental Disorders, 39*(7), 1087-1091.
- Zwaigenbaum, L., Bryson, S., Lord, C., Rogers, S., Carter, A., Carver, L., ... & Yirmiya, N. (2009). Clinical assessment and management of toddlers with suspected autism spectrum disorder: insights from studies of high-risk infants. *Pediatrics, 123*(5), 1383-1391.