3-28-2014

The Diagnostic Stability of Developmental Delay and Developmental Language Disorder in Infants and Toddlers

Alexander J. Hinnebusch
University of Connecticut - Storrs, alexander.hinnebusch@uconn.edu

The Diagnostic Stability of Developmental Delay and Developmental Language Disorder in Infants and Toddlers

Alexander J. Hinnebusch

B.S., The College of William and Mary, 2010

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts at the

University of Connecticut

2014

APPROVAL PAGE

Master of Arts Thesis

The Diagnostic Stability of Developmental Delay and Developmental Language Disorder in Infants and Toddlers

Presented by

Alexander J. Hinnebusch, B.S.

Major Advisor_____

Deborah Fein, Ph.D.

Associate Advisor_____

Marianne Barton, Ph.D.

Associate Advisor_____

Chi-Ming Chen, Ph.D.

Associate Advisor_____

Marie Coppola, Ph.D.

University of Connecticut

2014

ACKNOWLEDGEMENTS

Committee

Deborah Fein, Ph.D., Marianne Barton, Ph.D., Chi-Ming Chen, Ph.D., Marie Coppola, Ph.D.

Early Detection Team

Deborah Fein, Ph.D., Thyde Dumont-Mathieu, MD, Sarah Hodgson, Ph.D., Jamie Kleinman, Ph.D. Lauren Herlihy, Kelley Knoch, Eva Troyb, Laura Brennan, Lauren Haisley, Dasal Jashar, Kate Bradbury, Emily Moulten, Lauren Miller, Cara Cordeux, Courtney Manning

Undergraduate Research Assistants

Participating Families, Pediatricians, and Early Intervention Providers

TABLE OF CONTENTS

ABSTRACT

A significant number of children are diagnosed with Developmental Delay (DD) and Developmental Language Disorder (DLD) each year in the US. There are inconsistent and ambiguous diagnostic definitions for these disorders. In 1994, The Individuals with Disabilities Education Act (IDEA) led to state-specific diagnostic criteria for DD and DLD. This has produced widely varying diagnostic definitions for DD and DLD, which has complicated national prevalence estimation and the analysis of diagnostic stability. In the current study, 37 children received a diagnosis of DD and 21 received a diagnosis of DLD at an initial evaluation as part of their participation in a larger study investigating the early detection of Autism Spectrum Disorders (ASDs), and were seen for a follow-up evaluation. The DD group was significantly more likely to retain this diagnosis at follow-up than the DLD group. The DD and DLD group made significant gains on developmental measures between evaluations, and the DLD group made significantly greater gains in language skills than children in the DD group. Children in the DD group who were more delayed at their initial evaluation were more likely to retain their diagnosis at follow-up than less delayed children, and this difference approached significance. Males from the DD group retained their diagnosis more often than males in the DLD group; the sample size was not large enough to analyze diagnostic trends in females. Maternal education and family income did not have an effect on diagnostic stability in the DD and DLD groups. A small number of children from both the DD and DLD group received a diagnosis of an Autism Spectrum Disorder at follow-up, but this did not differ significantly between the groups, and did not vary by gender of the participant. These findings emphasize the need for more clear and consistent diagnostic

criteria for DD and DLD to allow for clearer measurement of prevalence and analysis of diagnostic stability. This study highlights the need for rigorous research investigating the type and intensity of intervention that would be effective in treating children with DD and DLD at varying levels of impairment.

Title: The Diagnostic Stability of Developmental Delay and Developmental Language Disorder in Infants and Toddlers

Introduction

Developmental Delay and Developmental Language Disorder – Description, Prevalence, and Identification

A developmental delay is a significant and ongoing delay in one or more processes of a child's development. Developmental Delay (DD) is a diagnostic label given to children who have not attained developmental milestones in multiple domains of development at a rate expected for their chronological age, as compared to typically developing, similarly-aged peers. Developmental Language Disorder (DLD) is a more specific diagnostic category, referring to the delay in a child's spoken language skills and/or ability to comprehend language spoken to them. Research in child development has established what skills children typically develop by particular ages, and contemporary research in pediatrics and developmental psychology continues to expand upon this knowledge base to better define what is considered typical infant and child development. The National Institutes of Health and The Centers for Disease Control and Prevention have made this information accessible to the general public by developing a general timeline detailing when children typically meet particular developmental milestones (CDC, 2011), and pediatricians often assess a child's skill acquisition during well-child visits during the first few years of life to identify any delay in development or potential disability.

The broad definition of a child with a disability was established under an amendment of The Individuals with Disabilities Education Act in 2004 (IDEA), as a

child, **"**who is experiencing developmental delays as defined by the State and as measured by appropriate diagnostic instruments and procedures in one or more of the following areas: Physical development, cognitive development, communication development, social or emotional development, or adaptive development" (National Dissemination Center for Children with Disabilities, 2012).

There is no part of IDEA that formally describes the exact diagnostic criteria for Developmental Delay (DD), nor an entry for DD in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), the classification system conventionally used in the United States (US), to describe mental and developmental disorders and around which research efforts in the United States are organized.

Under IDEA, a speech or language impairment is defined as, "a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, that adversely affects a child's educational performance." Alternatively, specific language impairment (SLI) is a construct often used in speech and language research, with a few diagnostic analogs. There are entries in the DSM-IV for expressive language disorder (315.31) and mixed expressive-receptive language disorder (315.32), but not receptive language disorder (Appendix A). A diagnosis of expressive language disorder is assigned to children whose scores on a measure of expressive language development are substantially lower than their scores on measures of nonverbal intellectual ability and receptive language development. A diagnosis of mixed expressive-receptive language disorder is assigned to children that receive substantially lower scores on both receptive and expressive language measures than a measure of nonverbal intellectual capacity. The diagnostic criteria for these diagnoses are relatively

nonspecific, as they do not identify a specific threshold (i.e., a score that is a certain number of standard deviations below the mean score for the population) at which a child should score below to qualify for a diagnosis. In the DSM-V, these diagnoses have been collapsed into a single diagnostic category, language disorders (Appendix B) (American Psychiatric Association, 2013). Under DSM-IV, children who demonstrate developmental delay by age 6, constituted by having an intelligence quotient (IQ) lower than two standard deviations below the mean for the population (70) and the presence of delayed adaptive skills in at least two domains, receive a diagnosis of mental retardation (American Psychiatric Association, 2000). In the DSM-5, this diagnosis has been renamed intellectual disability, but maintains the core diagnostic criteria outlined in the DSM-IV (Appendix B).

The International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list used in the United States primarily for billing and insurance purposes, but widely used in other parts of the world, includes codes for DD and DLD in its most recent iteration, ICD-9, for Expressive Language Disorder (315.31), Mixed Receptive-Expressive Language Disorder (315.32), Mixed Developmental Disorder (315.5), and Mental Retardation mild (317), moderate (318), severe (318.1), profound (318.2), and not otherwise specified (319) (Centers for Medicare and Medicaid Services, 2013). These ICD codes, however, are not as thoroughly integrated into clinical research and medical practice in the United States as the diagnoses described in the DSM.

The recently published DSM-5 includes a diagnosis of Global Developmental Delay (315.8), which is assigned to children under five years old whose intellectual

functioning cannot be systematically assessed. This diagnostic category will eventually be integrated into research and clinical practice, but the majority of current, and all prior, research has been conducted without an official DD definition, and nonspecific definitions for DLD subtypes, in the DSM-IV TR.

Due to the lack of a nationally established definition and specific diagnostic thresholds for DD and DLD, the criteria for these diagnoses have been specified at the state level. Since the initial passing of IDEA in 1990, states have crafted their own diagnostic definitions of DD and DLD, expanding upon IDEA's definition of a child with a disability by specifying the level of delay that indicates developmental impairment and identifying the diagnostic instruments that can be used to determine eligibility for services (Shackelford, 2006). As a result, the criteria for DD and DLD in the United States varies widely between states and is instrumental in determining a child's ability to receive intervention services. For example, Connecticut's Birth to Three System provides developmental evaluations and early intervention services for infants and toddlers from birth to 36 months of age in the state of Connecticut. Connecticut's Birth to Three defines DD as "a delay of at least 2 standard deviations below the mean in a single developmental area, or 1.5 standard deviations below the mean in two or more of the following areas: cognitive development, physical development, hearing, motor and health development, communication development, social or emotional development, adaptive skills development" (Connecticut Birth to Three System, 2010). A Developmental Language Disorder (DLD) is defined by a delay of at least 2 standard deviations on either expressive or receptive language skills, or 1.5 standard deviations below on both expressive and receptive language skills. An expressive language delay can be present

independent of a receptive language delay, but they often co-occur and are referred to as a mixed expressive and receptive DLD. A child who qualifies for a DLD diagnosis should not demonstrate clinically significant delays in other domains of functioning (Nelson, Nygren, Walker, & Panoscha, 2006). DD is often considered a more severe diagnosis than DLD, because of the presence of delays in multiple domains of functioning, though this is not suggesting that DLDs cannot be severe in nature and involve a high degree of impairment. The difference between states can be illustrated by comparing Connecticut's DD and DLD definitions to those implemented in New York, which does not differentiate the two disorders, but rather broadly defines DD as, "a 12 month delay in one or more functional areas or a 33% delay in one functional area, or a 25% delay in two areas, or a score of 2 SD below the mean in one functional area or a score of at least 1.5 SD below the mean in 2 areas" (New York State Department of Health Bureau of Early Intervention, 2005).

The ability of the individual state to define DD and DLD has led to a heterogeneous classification system. In the current, state-based system of identification, it is possible that a child diagnosed with a DLD in Connecticut could be identified differently in New York, which could potentially impact a child's ability to receive services based on his or her residency. The varying definitions between states also complicate national efforts to collect data, estimate prevalence, and assess the stability of a particular diagnosis across development.

As a result of this heterogeneous classification system, estimates of the prevalence of DD and DLD among the general population are difficult to ascertain. Prevalence studies do not employ specific, consistent diagnostic criteria to allow for differentiation

among DD, DLD, and other developmental disorders. The 1994-1995 National Health Interview Survey on Disability estimated the prevalence of all types of developmental delays within a nationally representative sample of infants and children between the ages of four months and five years to be between 3.3-4.4% (Simpson, Colpe, & Greenspan, 2003). Rosenberg, Zhang, and Robinson (2008) estimated the prevalence of all types of developmental delay in a sample of children who participated in The Early Childhood Longitudinal Study (ECLS-B) at 24 months of age in the United States to be 13%; only 10% of those children were receiving intervention services. A study in 2011 analyzing the prevalence of all developmental disabilities in US children, using data on children aged 3 to 17 years from the 1997-2008 National Health Interview Surveys, found that the overall rate of any developmental disability had increased from 12.84% to 15.04% (Boyle, Boulet, Schieve, Cohen, Blumber, Yeargin-Allsopp, Visser, & Kogan, 2011). However, the categories of developmental disability used to classify the data in these studies are broad and include diagnostic categories that may in fact represent more than one developmental disability (i.e. learning disability, intellectual disability, and a category labeled "other developmental delay"), which makes it difficult to determine the specific prevalence of DD and DLD.

Between 2006 and 2008, the estimates of the prevalence of learning disability was 7.24%, intellectual disability .67%, and other developmental delay 4.24%, suggesting that the rate of developmental delay could be as high as the sum of each of these diagnostic categories, or that some of these children could qualify for more than one diagnosis, depending on the diagnostic criteria (Boyle et al., 2011). A 2004 meta-analysis of interventions for children with primary developmental speech and language

delays/disorders estimated that 6-8% of preschool children have a speech or language difficulty (Law, Garrett, & Nye, 2004). Bishop and Leonard (2000) cited a national survey conducted as part of a study funded by the National Institutes of Health that reported approximately 7% of children in the United States as meeting criteria for specific language impairment (SLI), which is a construct commonly referenced and applied in speech and language research, but with no current diagnostic counterpart. The Disability Status Report, generated by Cornell University, provides estimates of disability for different age groups at the state and national level. The report is based on data collected through a U.S. Census Bureau survey, the American Community Survey (ACS), but is primarily focused on the working-age population and in children under 4, focused on disability affecting vision and hearing. The 2011 Annual Disability Status Report found that 0.8% of children in the United States under 4 years of age had both a visual and hearing disability, 0.5% had a visual disability, and 0.6% had a hearing disability (Erickson, Lee, & von Schrader, 2012). Interestingly, the overall rate of disability for children between 5 and 15 years is 5.1%, and the highest prevalence rate was for "cognitive disability," at 3.9%, but information is not collected regarding DD and DLD, which are precursors to intellectual disability.

However, there is criticism regarding this method of relying on surveys to estimate prevalence, particularly that the style of questioning in population surveys does not reference any diagnostic criteria for defining developmental disability, and instead relies on subjective parent report about the growth and development of their child. Currently, practices for determining the rate of DD and DLD in the national population do not integrate state-specific diagnostic criteria used in assessment and qualification for

intervention services, and state-level agencies that confirm these diagnoses and provide services to affected children are minimally involved in prevalence estimation efforts.

The lack of a formal DSM-IV diagnostic category, and non-specific DLD diagnostic criteria, along with the differing definitions of DD and DLD by state, can account for a portion of these analytic issues. In addition, the high rates of cognitive disability in children between 5 and 15, and absence of DD and DLD rates in children under 4 suggests that the lack of national diagnostic criteria may result in poor, or neglected, prevalence estimation of children affected nationally by DD and DLD. There remains a pressing need for prevalence studies that employ well-defined diagnostic criteria for DD and DLD, and differentiate these diagnostic categories from other developmental disorders. A more consistent classification system across states would allow for the more accurate assessment of DD and DLD rates across the United States, which would thereby increase the ability to assess the diagnostic stability and developmental progress across time.

Early Identification and Intervention

Early identification of infants and young children with developmental delays is considered critical, as research has demonstrated that appropriate intervention and treatment can minimize the potential for more serious problems and can improve the opportunity for identified children to function more successfully within home, school, and public settings (Simpson, Colpe, & Greenspan, 2003). In the United States, the Congressional Public Law 99-457 in 1986 first encouraged individual states to identify developmentally delayed infants and young children and to organize comprehensive programs of early intervention services (Meisels, 1989). In 1990, The Education for All

Handicapped Children Act was renamed the Individuals with Disabilities Education Act (IDEA, PL 101-476), and expanded the role and involvement of the national government in promoting state-based detection and intervention programs. This legislation was based primarily on five landmark studies that assessed the effectiveness of intervention within groups of children at high risk for intellectual disability, a diagnosis assigned at 6 years of age to children that were previously showing delays and most likely had a diagnosis of DD or DLD. These studies include the Perry Preschool Project, the Milwaukee Project, the Abecedarian Project, Project CARE, and the Infant Health and Development Program (Ramey & Ramey, 1999). Ramey and Ramey, in their review of these studies, claim that these findings suggest that the rates of intellectual disability and special education can be reduced by 50% or greater by providing early intervention services to at-risk children. They also claim that these findings imply that direct, intensive, individualized intervention is most likely to result in the greatest benefit and alter early experiences for high-risk children. IDEA was designed to ensure that states provided early intervention and special education services to qualifying children in the United States, and the findings from the aforementioned studies form the foundation for current intervention practices for any child with a developmental disability.

In 1997, legislation was passed that effectively reorganized IDEA, and formally organized an infant-toddler component, entitled Part C (Bailey, Aytch, Odom, Symons, & Wolery, 1999). Though participation is voluntary, currently all states in the US participate in Part C and receive funding from the federal government, indicating that the country as a whole has adopted a policy that provides a statewide system of early intervention services for all children with disabilities. In 2002, Part C served 2.2% of

infants and toddlers (Rosenberg, Zhang, & Robinson, 2008). States have considerable flexibility in determining standards for eligibility for intervention, which, as previously discussed, allows each state to individually define the criteria for developmental delay and also to decide what type and intensity of intervention services should be provided to children with particular diagnoses (Bailey, 1999).

Support for intervention in cases of children with DD and DLD is based in longitudinal studies that track the progress of children that received intervention services across time, and newer studies specifically assessing the short-term effects of intervention. There is a rich literature demonstrating the efficacy of therapeutic and academic intervention for preschool and elementary-aged children, but fewer studies focusing on children between birth and pre-formal education age. Though it is an unstated assumption that intervention is beneficial for children who are developmentally delayed or intellectually disabled, there is little empirical data supporting the type, quantity, or style of intervention appropriate for children with DD.

Majnemer (1998) outlined the benefits of early intervention broadly for any child with a developmental disability, citing evidence that intervention improved scores on developmental outcome measures, as well as improved parent-child interactions and created a positive environment for the family. Evidence also suggests that structured programs that continue throughout childhood appear to have positive, long-lasting effects on development (Ramey & Ramey, 2004). Resnick and colleagues (1987) conducted a prospective longitudinal study assessing the effects of a developmental intervention program on the outcome of low birth weight infants, finding that infants who received the intervention were significantly less likely to have developmental delay and scored

significantly higher on measures of IQ and physical abilities. A meta-analysis conducted by Shonkoff and Hauser-Cram (1986) assessed 31 studies that evaluated the effectiveness of early intervention in children below 3 years of age with varying developmental disability, finding that early well-structured intervention programs had a positive impact on developmental progress. However, there is a lack of recent studies assessing the quality and effect of early intervention, or the effect of different levels of intensity, on children with a DD diagnosis specifically.

There is a wider breadth of studies focusing on the effectiveness of treatment and diagnostic outcome in children with DLD. In a practitioner review, Whitehurst and Fischel (1994) assessed research that suggested that children with a language delay that persists through age 4-5 and is accompanied by more general developmental delays, have a poorer long-term outcome. Law, Garret, and Nye (2004) performed a meta-analysis that included 13 studies conducted within a 25-year period assessing the effectiveness of intervention programs on linguistic development. The results of this study suggest that speech and language therapy may be effective in treating phonological or expressive language impairment, but there was mixed evidence supporting its effect on expressive syntax impairment, and little evidence supporting its effectiveness in treating receptive language impairment. The review identified longer duration of intervention, defined as greater than 8 weeks, as an influential factor in determining a favorable clinical outcome. Despite having a greater number of studies assessing intervention's effect on children with DLD, there is a need for more targeted research assessing the effect of intensity and duration of intervention and its effect on diagnostic stability.

For the sake of comparison, Connecticut's Birth to Three recommends that children under the age of three with an Autism Spectrum Disorder (ASD) receive 15-20 hours per week of services. There is a rich body of literature that suggests that between 25-40 hours is an appropriate level of intervention for children diagnosed with an ASD (Howard, Sparkman, Cohen, Green, Stanislaw, 2005). Additionally, studies focusing on the development of interventions for increasingly positive outcomes in young children with autism have gained recent attention (Dawson et al., 2010; Green, Brennan, & Fein, 2002; McGee, Morrier, & Daly, 1999).

The current literature would benefit from empirical studies, such as those recently conducted with samples of children with ASD, that examine the role intervention plays in developmental growth of a child with DD and DLD across time. Currently, the recommendations for intervention services are determined by Early Intervention providers by state and couched in intervention studies conducted nearly 30 years ago. However, this process of review is made difficult by the heterogeneous system of state-specific classification of DD and DLD, and the lack of a nationalized system of assessing the prevalence of DD and DLD. It is also difficult to assess intervention services as children receive them today, because of a lack of control over the quality, quantity, and actual receipt of services.

Outcome and Diagnostic Stability

Few published studies have examined the stability of a DD or DLD diagnosis in infants and toddlers across time. Shevell, Majnemer, Platt, Webster, and Birnbaum (2004) tracked the developmental trajectory of children diagnosed with Global Developmental Delay (GDD), a broad diagnostic category applied to children who are

exhibiting delays in more than one area of development, at age 3.5 years and reassessed at 7 years of age. Of the 48 children included in their analyses, at follow-up 74% were impaired in two or more domains of the Battelle Developmental Inventory, and 48% were impaired in at least two domains on the Vineland Adaptive Behavior Scales. Children who had received a DLD diagnosis continued to present with impairments not only in their communication skills, but also impairment in other developmental and functional skill domains. Silva (1980) first reported on the diagnostic stability of language delay in a sample of 937 children who presented first at age 3, and were later reevaluated at age 5. The findings suggested that specific language delay, either receptive or expressive, was not highly stable across time, and children tended to improve and no longer meet diagnostic criteria after 2 years. However, mixed language delay was highly stable across time, though 84% of those children were determined to be intellectually disabled.

There is a paucity of prospective outcome data for young children diagnosed with DD or DLD, and most long-term outcome studies have examined academic outcomes, such as difficulties with mathematics, handwriting, and reading skills (Shapiro, Palmer, Antell, Bilker, Ross, & Capute, 1990; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). Other studies have assessed cognitive, social, and behavioral outcomes common to children with early language impairments, finding more negative outcomes for affected children (Trower & Nico, 1996). For example, studies have demonstrated that children with DD or early language delays can later experience behavioral difficulties, such as adult delinquency and aggression, difficulties with establishing friendships, acquiring numeracy, and can express attention deficit difficulties, internalizing problems and psychiatric disorder, and dyslexia and other related reading disabilities (Bishop &

Leonard, 2000; Brownlie, Beitchman, Escobar, Young, Atkinson, Johnson, Wilson, & Douglas, 2004; Snowling, Bishop, & Stothard, 2000). Even children with an SLI at age 4 who had responded to treatment by age 5½ showed reemerging language difficulties at age 8, including oral language impairment, reading difficulties, and verbal deficits (Bishop & Adams, 1990).

There is a demand for studies that measure the effect of quantity of intervention on development and diagnosis, or those that can serve as an empirical foundation for justifying a certain degree and type of services for children with DD or DLD. At present, statewide organizations such as Birth to Three do not have strict guidelines regarding the type, quantity, or quality of services that a child with a DD or DLD should receive, and services are typically recommended by the individual assessment professional who evaluates the child. These recommendations can be quite varied, and Birth to Three designs an Individual Family Service Plan on a case-by-case basis (Connecticut Birth to Three System, 2011).

Diagnostic Outcome - Gender, Socioeconomic Status, and Maternal Education

The 2011-2012 National Survey of Children's Health revealed that males and females are equally likely to receive developmental screening, with 29.9% of males screened and 28.7% of females screened between the ages of 10 months and 5 years of age. Data collected through the 2011-2012 National Survey of Children's Health show that despite equitable screening, parents more frequently report concerns regarding their male children. 13% of parents of males, and 9% of parents of females, expressed two or more concerns with their child's development; 17% of parents of males, and 14% of parents of females, reported one concern (Child Trends Data Bank, 2013). A 2007 review

of preschool records in Florida for all children who received a diagnosis of a developmental delay reveals that males comprised 71.30% of children who were identified as DD and 67.99% of those who were speech delayed (Delgado, Vagi, & Scott, 2007). Similar gender ratios have been established by research for other developmental delays; a ratio of 4 to 1 male to female has been established in research on ASD's (Fombonne, 1999). There has been no research to date assessing the potential role gender may play in the diagnostic stability of DD and DLD over time.

Research has shown that socioeconomic status and maternal education impact a child's likelihood of having a developmental delay, affects their response to intervention, and influences their long-term outcome. Findings from the landmark studies that formed the basis for IDEA in 1990 suggested that maternal education and family income have an impact on the effectiveness of intervention and child outcome. In their analysis of records regarding children enrolled in preschools across Florida to assess early risk factors for developmental delay, Delgado, Vagi, and Scott (2007), found that low maternal education, along with prematurity, posed the most severe population-level risk. Studies assessing the development of a child's vocabulary have found a positive relationship between family socioeconomic status (SES) and the rate of vocabulary development (Arriaga, Fenton, Cronan, & Pethick, 1998; Hoff, 2003; Morrisset, Barnard, Greenberg, Booth, & Spieker, 1990). A widely cited study conducted by Hart and Risley (1992) demonstrated a significant difference in the number of words heard by children from families with different socioeconomic backgrounds in their first three years, with children from families with higher socioeconomic status hearing significantly more words. Their results also demonstrated that the larger the size of a child's vocabulary at age three, the

higher their performance on language-based measures when the child was ten (Hart & Risley, 1995).

Other studies assessing the impact of family SES on infant development have found that there is a lack of correlation between the two until 18 to 24 months, at which point a higher family SES is correlated with a higher level of measured infant development (Golden & Birns, 1976; McCall, 1979; McCall, 1981). Several studies described by Sonnander and Claesson (1999) demonstrated that measures of SES are reliable predictors of performance once a child is of preschool age. Noble, Norman and Farah (2005) reviewed studies linking SES to a variety of outcomes including IQ, achievement test scores, grade retention, and functional literacy. They concluded that SES predicts about 17-20% of the variance in IQ scores of school-age children (Gottfried, Gottfried, Bathurst, Guerin, & Parramore, 2003). Sameroff and colleagues (1987) explored more specific risk factors associated with low SES, such as low maternal education, poor maternal mental health, and incidence of stressful life events, and explained 51% of the variance in a sample of 215 school-age children using a mixed risk factor model of socioeconomic status variables.

A sample of 101 children from a larger study assessing the effectiveness of a developmental delay screening instrument, the Parental Assessment Screening, demonstrated that low SES and low maternal education are strongly associated with longer-term outcomes, like school achievement problems at 8 and 14 years of age (Sonnander & Claesson, 1999). More recent neurocognitive research has found that SES differences were associated with inconsistent and impaired performance in the language and executive function domains, which could have implications for intervention practices

that concurrently target executive function and language ability (Noble, Norman, & Farah, 2005). Hoff (2003) demonstrated that the properties of the language used by mothers from families with a higher SES were correlated with accelerated language development in toddlers. Though SES and maternal education have been established as risk factors for developmental delay, there have not been any investigations to date into the role these factors may play in the diagnostic stability of DD and DLD.

Developmental Delay and Autism Spectrum Disorders

Children with Autism Spectrum Disorders (ASDs) often present with developmental delays in one or more domains of functioning, such as language, cognition, and motor skills. To meet criteria for an ASD, a child must present with deficits in social skills and communication abilities, as well as exhibit stereotyped and/or repetitive behaviors (American Psychiatric Association, 2013). These additional impairments differentiate children with an ASD from those with a DD or DLD, though there is a considerable degree of diagnostic overlap between the groups.

Research studies focusing on deficits in children with ASDs will typically compare these children to those with diagnoses of DD or DLD, as well as typically developing children. This research often demonstrates similarities between the delays children with DD or DLD show and what a child with an ASD might show. For example, research has demonstrated comparable levels of motor skill impairment in children with DD and ASD, while being unable to point to distinct motor impairments that could be used to differentiate the groups (Green, Baird, Barnett, Henderson, Huber, & Henderson, 2002; Miyahara, Tsujii, Hori, Nakanishi, Kageyama, & Sugiyama, 1997; Provost, Lopez, & Heimerl, 2006). Though the level of language delay can be similar among children

with an ASD and DD or DLD, children with ASD are more likely to exhibit noticeable impairment in social communication skills and the pragmatic use of language (Eisenmajer, Prior, Leekam, Wing, Ong, Gould, & Welham, 1998; Howlin, Mawhood, & Rutter, 2000; Loveland & Landry, 1986). The similarities in impairment in particular domains can make it difficult to accurately diagnose a child. While a substantial body of literature has evaluated the diagnostic stability of ASDs and the change over time from an ASD diagnosis to DD and DLD (Cox et. al, 1999; Kleinman et. al, 2008; Lord et. al, 2000), there are no recent studies that have investigated the number of children who receive an initial diagnosis of DD or DLD, and later go on to meet criteria for an ASD.

Current Study

The current study seeks to expand the limited body of literature regarding the diagnostic stability of DD and DLD between an initial evaluation at 2 years of age and follow-up evaluation at 4 years of age, and to assess the change in developmental skill level over time within these two groups. This study also seeks to assess the role of impairment severity at first diagnosis on diagnostic stability in both DD and DLD. In addition, this study aims to assess the effect of gender, maternal education, and family income on diagnostic stability. Finally, this study aims to compare the DD and DLD groups on the proportion of children who receive a diagnosis of ASD at a follow-up evaluation. There were several hypotheses for the current study:

1. We predicted that an initial diagnosis of DD, involving delay in multiple domains of development, would demonstrate a higher degree of diagnostic stability than that of DLD. Consistent with this prediction, we predicted that children who received an initial diagnosis of DD would demonstrate less developmental growth across time in all domains of development time than children with a diagnosis of DLD.
2. We predicted that severity of impairment at initial diagnosis would impact diagnostic stability in both DD and DLD groups. Specifically, those children who

were more severely delayed at initial diagnosis would be more likely to retain that diagnosis at follow-up.

3. We predicted that female participants would be less likely than male participants to retain their initial diagnosis of DD and DLD at follow-up.
4. Consistent with the literature, we predicted that lower maternal education and family income would be correlated with a higher retention of both DD and DLD diagnoses at follow-up.
5. We predicted that children who received an initial diagnosis of DD would be more likely than those children that received an initial diagnosis of DLD to receive a diagnosis of ASD at follow-up. We also predicted that males who received an initial diagnosis of DD or DLD would be more likely than females to receive an ASD diagnosis at follow-up.

Methods

Participants

Participants were selected from the Early Detection study, an ongoing study that is assessing the sensitivity and specificity of an ASD-specific screening questionnaire for toddlers, the Modified Checklist for Autism in Toddlers (M-CHAT- Robins et al., 2001), as well as a second form of the questionnaire, the M-CHAT-Revised. Children were enrolled in the study via two primary referral sources: a pediatrician and an Early Intervention service provider. These children were enrolled through screening at an 18- or 24-month well-child visit with a pediatrician, or through screening by an Early Intervention staff member. The majority of study participants were residents of Connecticut, Massachusetts, and/or Rhode Island at the time of their participation, representing mostly rural and suburban living situations, with less urban representation. Children were screened for major motor and sensory impairments prior to the initial evaluation, and were excluded from the study due to the interference that these impairments would present in the administration and interpretation of the standardized measures used to assess cognitive, language, and adaptive skills. Data included in the

current study represent the subsection of the total sample collected for the Early Detection study that received an initial diagnosis of DD or DLD.

Children included in this study received either a DD or DLD diagnosis at their initial evaluation based on diagnostic criteria specified by the clinicians and researchers of the Early Detection Study (Appendix C). Diagnoses were based upon a child's performance on the Mullen Scales of Early Learning (Mullen, 1994) and the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) (see below for description).

Children that received a DD diagnosis at the time of their evaluation demonstrated a delay of at least 1.5 standard deviations on at least one of the following non-language subscales: Visual Reception and Fine Motor from the Mullen, and Motor Skills from the Vineland. These children also demonstrated a delay of at least 1.5 standard deviations on at least one of the following language subscales: Expressive and Receptive Language from the Mullen, and Communication from the Vineland. Additionally, at least one of the scores from the above categories must have been on the Mullen.

To receive a diagnosis of DLD, a child must have demonstrated a delay of at least 1.5 standard deviations on at least two of the following language subscales: Expressive and Receptive Language from the Mullen, and Communication from the Vineland. Alternatively, a child could have demonstrated a delay of at least two standard deviations on only one of the aforementioned language subscales. Additionally, these children must not have demonstrated any delays greater than 1.5 standard deviations on the non-language subscales tested by the Mullen and Vineland (Visual Reception and Fine Motor on the Mullen, and Motor Skills on the Vineland). All children that met these diagnostic

criteria were diagnosed with DLD, but were further specified within the study based on the particular type of language delay. Children that exhibited delays in both expressive and receptive language skills were diagnosed with mixed language disorder, children with a delay in just expressive language were diagnosed with expressive language disorder, and children with a delay in just receptive language were diagnosed with receptive language disorder. For the purposes of this study, all children that received a DLD diagnosis, regardless of type, were included as one group for analyses unless otherwise specified.

To date, 98 children received an initial diagnosis of DD, and 43 received a diagnosis of DLD through the study. Of those 98 children who received a DD diagnosis at their initial evaluation, 37 received a follow-up evaluation. Of the remaining 61, 14 had parents that refused a follow-up evaluation, the study was unable to contact 32, and 15 were not yet old enough to receive a follow-up evaluation.

Of those 43 children who received a DLD diagnosis at their initial evaluation, 21 received a follow-up evaluation. Of the remaining 22, 2 had parents that refused a follow-up evaluation, the study was unable to contact 11, and 9 were not yet old enough to receive a follow-up evaluation.

In summary, the sample includes 37 children that received a diagnosis of DD at their first evaluation and received a second evaluation and 21 children that received a diagnosis of DLD at their first evaluation and received a follow-up evaluation.

Demographic information for the children included in this study can be found in Figure 1. The mean age of the children diagnosed with DD at their initial evaluation was 27.53 months (SD= 4.73 months) with a range of 18.86 months to 35.45 months. The

mean age of the children diagnosed with DLD at their initial evaluation was 25.48 months (SD= 4.74) with a range of 17.02 months to 32.46 months. The groups were not significantly different in age, $t(57) = 1.565$, $p = .123$.

Of the DD participants, 31 were male (83.7%) and 6 were female (16.3%); Of the DLD participants, 13 were male (62%) and 8 were female (38%). The difference in gender between the DD and DLD group approached significance, $x^2 (1, N = 58) = 3.5$, $p = .0614$.

Child ethnicity was collected through parent report using the following categories: White/European American, Hispanic/Latino- not Puerto Rican, Puerto Rican, African American, Caribbean or Caribbean American, Asian or Asian American, Native Hawaiian or Pacific Islander, Native American Indian, or Other. Of the DD participants, ethnicity information was available for 35 children. 27 (77%) of these children were White/European American, 5 children were African American (14.3%), 2 (5.7%) children were Hispanic/Latino- not Puerto Rican, and 1 child (3%) was biracial, Hispanic/Latino- not Puerto Rican and African American. Ethnicity information was available for all 21 DLD participants; 18 (85.7%) children were White/European American, 2 (9.5%) children were African American, and 1 (4.8%) child was biracial, West Indian and Caribbean. Chi-square tests revealed no significant differences in ethnicity between the DD and DLD group, $x^2 (3, N = 56) = .980$, $p = .914$, however this result violated the assumptions of the chi-square test because six of the cells (75%) had an expected count that was less than five. Fisher's exact test revealed that the DD and DLD groups were not significantly different in ethnicity, ($p = .914$).

Maternal education was self-reported in the following categories: some high school, high school diploma/GED, some college, vocational or technical degree, Associate's Degree, Bachelor's Degree, Master's degree, and Graduate or Professional Degree (M.D., J.D., or Ph.D.). Maternal education information was available for 50 of the participants, in both the DD (n = 31) and DLD (n = 19) groups. The modal and median level of maternal education for the entire sample was college degree. Chi-square tests did not reveal significant differences between the maternal education of the DD and DLD groups, x^2 (5, N = 50) = .696, p = .755, though this result violated the assumptions of the chi-square test as seven of the cells (58.3%) had an expected count that was less than five. For analyses, maternal education was recoded as an ordinal variable, with a number (1-6) assigned to each tier in ascending order.

Family income was determined through self-report by indication of annual household income. Annual household income was stratified in $10,000 intervals, ranging from between less than $10,000 to greater than $100,000 (i.e., $10,000-20,000, $20,000-30,000, etc.). Family income information was available for 45 participants, in both the DD (n = 27) and DLD (n = 18) groups. Parents of the entire sample represented the full range of yearly incomes, and median annual income for the entire sample was about $65,000, with chi-square tests revealing no significant differences between the DD and DLD groups, x^2 (10, N = 45) = 9.156, p = .517, though the assumptions of this test were violated, as 21 of the cells (95.5%) had expected counts that were less than five. For analyses, family income was coded as an ordinal variable, with a number (1-11) assigned to each income tier in ascending order.

Measures

ASD screening and diagnostic measures and developmental level

As part of a standardized battery, children received measures assessing developmental level and adaptive skills. As the study is designed to detect ASD, study personnel also administered measures to assess ASD symptomatology, including the Autism Diagnostic Observation Schedule, the Childhood Autism Rating Scale, and the Autism Diagnostic Interview, Revised (Lord, Risi, Lambrecht, Cook, Leventhal & DiLavore et al., 2000; Schopler, Reichler, & Renner, 1988; Rutter, Le Couteur, &Lord, 2003). The Modified Checklist for Autism in Toddlers is used for screening for ASD, and has demonstrated excellent psychometric properties (Kleinman, Robins, Ventola, Pandey, 2008; Robins, Fein, Barton, Green, 2001).

Modified Checklist for Autism in Toddlers (M-CHAT) (Robins et al., 2001)

The Modified Checklist for Autism in Toddlers (M-CHAT) is a tool that screens for behaviors in children consistent with those observed in children with ASD. The M-CHAT is a 23-item questionnaire in which parents respond with either a "yes" or "no" answer to questions regarding their child's behavior (Robins et al., 2001). The measure was developed from the Checklist for Autism in Toddlers that identifies children aged 18 months who are at risk for autism (CHAT- Baron-Cohen, Allen, & Gillberg, 1992; Baron-Cohen, Cox, & Baird, 1996). Of the questionnaire's 23 items, four are reverse-scored, in which for a typically developing child a parent would most likely answer "no," such that response bias is reduced (e.g., "Does your child ever seem oversensitive to noise?"). A positive screen on the M-CHAT is considered to be a child failing three out of 23 total items, or two out of six "critical items." If a child screens positive on the M-CHAT, their caregivers receive a follow-up phone screening, in which failed items are

assessed in more detail. If a child continues to fail the M-CHAT after phone screening, they qualify for a free initial developmental evaluation, and a subsequent follow-up evaluation two years afterward. The M-CHAT's internal reliability was demonstrated to be adequate for the 23-item checklist, as well as six "critical items" ($\alpha = .85$ and $\alpha = .83$ respectively), in both the original study sample (Robins et al., 2001) and in an additional study ($\alpha = .85$ and $\alpha = .83$, Kleinman et al., 2008). The majority of children included in the sample for this study were screened using the M-CHAT (36 DD, 20 DLD).

Modified Checklist for Autism in Toddlers (M-CHAT-R) (Robins et al., 2014)

The M-CHAT-R is the current measure used in the Early Detection Study, composed of 20 yes/no parent-report items that were reworded to improve comprehension. Additionally, the order of items was revised to counteract a tendency of parents to endorse "yes" for all items, examples were provided to increase the clarity of items, and three low-performing M-CHAT items were removed. As with the M-CHAT, children who screened positive (failing two of seven "best 7" items, or any three items) on the M-CHAT-R were given a follow-up phone interview. Children that continued to screen positive on the M-CHAT-R on the phone interview were offered free diagnostic evaluations. Published findings show that the M-CHAT-R is an effective screening tool when used in a low-risk, pediatric sample (with a cut-off of two failed items, sensitivity = .94, specificity = .83). A small number of the children (2 DD, 1 DLD) included in the sample for this study were screened using the M-CHAT-R.

Mullen Scales of Early Learning

The Mullen Scales of Early Learning (Mullen, 1994) is a standardized test of cognitive ability, intended to evaluate children between birth and age 68 months. Of its five

subtests, Gross Motor, Visual Reception, Fine Motor, Expressive Language, and Receptive Language, all but the Gross Motor scale were administered in this study. The Early Learning Composite (ELC) is a score that is considered an overall estimate of a child's developmental age, and is generated by summing a child's performance across all four domains administered in this study. In each subtest, T-scores, percentile ranks, and age equivalents are produced, which reflect the child's current level of development in comparison to same-aged peers. The Mullen was normed on a nationally representative sample of 1,849 children (48.7% female, 51.3% male). It is a frequently used measure of developmental level and cognitive functioning in both typically developing children and children with developmental delays, and has demonstrated good reliability and validity. The Mullen demonstrates very satisfactory internal consistency of .75 to .83. The test re-test reliability of the Mullen is .84 for younger children, and .76 for older children (Mullen, 1994).

For analyses, a median split was used to divide the DD sample; children with an initial diagnosis of DD were coded as "more delayed" if their ELC standard score was 54 and below, and "less delayed" if their ELC standard score was 55 and above. Children with an initial diagnosis of DLD were classified by their performance on the language subtests of the Mullen as having either receptive language delay (T-score of 30 or less on Mullen Receptive Language subscale, and T-score of 36 or higher on Mullen Expressive Language subscale), expressive language delay (T-score of 30 or less on Mullen Expressive Language subscale, and T-score of 36 or higher on Mullen Receptive Language subscale), or mixed expressive-receptive language delay (T-score of 35 or less on both Mullen Expressive and Receptive language subscales).

Vineland Adaptive Behavior Scales- Interview Edition

The Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) is a standardized parent report interview that assesses a child's adaptive skills, including domains of Communication, Daily Living, Socialization, and Motor Skills. The measure yields domain scores, standard scores for individual subscales, and an overall Adaptive Behavior Composite (ABC), used to compare a child's skills to same-aged peers. The Vineland has established reliability and validity (Sparrow, Balla, &Cicchetti, 1984) and it is frequently used with varied clinical populations. The Vineland is commonly used and considered a valid instrument when assessing children with developmental delays and ASD in both research and clinical applications (Klin, Carter, & Sparrow, 1997). Standard scores for the Communication, Socialization, and Motor domains were included in analyses for the current study. Domain standard scores range from 20-160, with higher scores indicating better functioning or skill level attained. For the range of ages included in the Early Detection sample, the Vineland demonstrates high internal consistency for its adaptive behavioral composite (.90) and domain scores (.80-.90). Test-retest reliability for the subdomains was adequate (ICC of .85 and higher), and inter-rater reliability for the adaptive composite score (.87) and domain scores (.75) were acceptable (Sparrow, Cicchetti, Balla 2005).

Intervention Information

Intervention information was collected through a parent-completed history form at the follow-up evaluation. Parents indicated any intervention services that their child received over time in increments of 6 months. On the form the parent indicated what type of services were being received, in what setting, and the quantity of intervention (hours

per week). All participants included in analyses received some form of intervention between their initial and follow-up evaluations.

Procedures

When the child was between 16 and 30 months, their parent completed the M-CHAT or the M-CHAT-R through either the child's Early Intervention provider or pediatrician's office as described above (Robins et al., 2001). Parents whose children failed the M-CHAT or M-CHAT-R received a follow-up interview over the phone; if the child continued to fail upon follow-up, the family was offered a free developmental evaluation at the University of Connecticut conducted by a licensed psychologist or a developmental pediatrician, and a graduate student. Participants lacking transportation were provided a free taxi service from their homes to the study. Study staff traveled to conduct evaluations at participating pediatric offices in two large towns with a high proportion of low SES families. The diagnosis of DD or DLD was made based on meeting cut-off scores on the Mullen and Vineland derived by the clinicians on the Early Detection study (Appendix C). This first evaluation will be referred to as the initial evaluation, or Time 1.

Children became eligible to receive a follow-up evaluation, or Time 2, when they were 42 months or older, and were invited back to the University of Connecticut. This follow-up evaluation included the same measures assessing developmental and adaptive skills as at Time 1, and a diagnosis was made based on meeting cut-off scores on the Mullen and the Vineland.

All of the children in this study, in both the DD and DLD groups, failed the M-CHAT or M-CHAT-R, as well as the phone interview, and received an initial and follow-up evaluation.

Sample Size

The overall sample size for the current study (N= 58) provides sufficient power (power= .80, alpha= .05) to detect a large effect (Cohen's d > .8, r >.5) but it is not quite sufficient to detect a medium effect (Cohen's d >.35, r > .25) (see Cohen, 1988).

Results:

Diagnostic Stability of DD and DLD

A chi-square analysis was conducted to compare possible differences in the diagnostic stability of the DD (N = 37) and DLD (N = 21) groups, when considering every possible diagnostic outcome. For this analysis, children were grouped into five categories based on their Time 1 compared to their Time 2 diagnosis: retain initial diagnosis, reverse diagnosis, no diagnosis, other diagnosis, and autism spectrum disorder at Time 2. Participants in the "retain initial diagnosis" group received the same diagnosis at Time 2 that they qualified for at Time 1. Participants in the "reverse diagnosis" group received a diagnosis of DD at Time 1, and a diagnosis of DLD at Time 2 or a diagnosis of DLD at Time 1, and a diagnosis of DD at Time 2. Participants in the "no diagnosis" group did not receive a formal diagnosis at Time 2, though within the study these children are not considered typically developing, because at one point in their development they demonstrated clinically significant delays. Participants in the "other diagnosis" group received a different diagnosis at Time 2 (e.g., ADHD). Participants in the "ASD" received an autism spectrum disorder diagnosis at Time 2. Outcome data for

each group can be found in Table 2. This initial chi-square analysis revealed that there was an overall difference in the diagnostic stability between Time 1 diagnosis of DD and DLD, x^2 (4, N = 58) = 12.334, p = .015, Phi = .461. The assumptions of this chi-square analysis were violated, as six of the cells had expected counts that were less than five.

Specifically, significant differences were found between the DD (n = 37) and DLD (n = 21) groups with regard to the "retain initial diagnosis" and receive "no diagnosis" groups. Nineteen (51.4%) children received DD at Time 1 and Time 2, while 2 (9.5%) children received DLD at Time 1 and Time 2, x^2 (1, N = 58) = 10.15, p = .0014, Phi = 0.4183 (Table 3). Nine children (24.3%) that received DD at Time 1 received no diagnosis at Time 2, and 12 children (57.1%) that received DLD at Time 1 received no diagnosis at Time 2, x^2 (1, N = 58) = 6.25, p = .0124, Phi = 0.3283 (Table 4). Chi-square tests could not be conducted to assess the specific difference between DD and DLD groups regarding the "reverse diagnosis," "other diagnosis," and "ASD" groups due to small sample sizes. However, qualitatively, these outcomes did not appear to differ widely by group. One (2.7%) child from the DD group, and two (9.5%) from the DLD reversed diagnosis at Time 2. Three (8.1%) children from the DD group, and three (14.3%) children from the DLD group received an "other diagnosis" at Time 2. Five (13.5%) children from the DD group and two (9.5%) from the DLD group received ASD at Time 2.

To better assess the effect size of the significant differences found between DD and DLD, a follow-up chi-square analysis was conducted and included only children that either retained their Time 1 diagnosis at Time 2 or that received no diagnosis at Time 2, removing all other diagnostic outcomes. Essentially, this analysis was assessing

differences in the number of children from the DD and DLD group that improved, and received no diagnosis at Time 2, and the number that maintained their diagnosis and still demonstrated delay. Twenty-five children from the DD group, and 14 from the DLD group, were included in this analysis. Of the children that received a DD diagnosis at Time 1, 19 (67.9%) retained their initial diagnosis, while nine (32.1%) received no diagnosis at follow-up. Of the children that received a DLD diagnosis at Time 1, two (14.3%) retained their diagnosis, while 12 (85.7%) received no diagnosis. Chi-square analysis revealed that children that received a diagnosis of DD at Time 1 retained their diagnosis at a significantly higher rate at Time 2 than children that received a diagnosis of DLD at Time 1, $x^2(1, N = 42) = 10.714$, $p = .001$, Phi = .505 (Table 5).

Developmental Progress in Language and Visual Reception by Diagnosis (DD, DLD):

Independent samples t-tests were conducted to determine if there were differences between the DD and DLD groups in the amount of developmental gain made between the two evaluations in language and visual reception abilities. Data was available for 29 children from the DD group, and 15 children from the DLD group. Time 1 age equivalent scores on subtests of the Mullen (Receptive Language, Expressive Language, and Visual Reception) were subtracted from Time 2 age equivalent scores, the difference of which reflects the amount of developmental progress in months made between Time 1 and Time 2. A positive difference indicates that a child made developmental gains within an area between evaluations, a difference of zero would indicate no developmental gains, and a negative difference would indicate a lower developmental level at Time 2 compared to Time 1. The number of months between the Time 1 and Time 2 evaluations was calculated for each child.

The ratio estimate used in these analyses to assess developmental progress between evaluations (mental age divided by chronological age) was based upon a ratio used in a research study evaluating developmental progress in children with ASDs and common to outcome literature (Sallows & Graupner, 2005). For each child, the difference between Time 1 and Time 2 age equivalent scores from the Visual Reception subtest was then divided by the number of months between Time 1 and Time 2 for each specific child; this quotient represents the change in mental age, or the proportion of expected developmental gain over the actual time elapsed between evaluations. For example, this ratio would equal 1 if a child demonstrated 2 years of developmental gain as measured by the Mullen, and 2 chronological years had elapsed between evaluations. A similar quotient was generated for the language subtests, by first averaging the Receptive and Expressive Language change in age equivalent subtest scores, and then dividing that by the amount of time in months that had passed. A quotient greater than 1 indicates that the child made developmental progress in months greater than the amount of chronological time that had passed, a quotient equal to 1 indicates that a child made the same amount of developmental progress in months as the number of actual months that had passed, and a quotient less than 1 indicates that a child made less developmental progress in months than the actual number of months that had passed.

The mean mental growth rate on Visual Reception for children from the DD group was .82 (SD = .51), and for combined Expressive and Receptive Language it was .92 (SD = .56). The mean mental growth rate on Visual Reception for children from the DLD group was 1.06 (SD = .25), and for combined Language it was 1.09 (SD = .29). Levene's Test for Equality of Variances was significant for both Language ($F = 6.836$, p

= .012) and Visual Reception (F = 6.899, p = .012) indicating that equal variances for this analysis are not assumed; the statistics were interpreted accordingly. For Language, the DLD group made significantly higher developmental progress between Time 1 and Time 2, with a higher mean mental growth rate than children from the DD group, $t(41.629) = 2.223$, $p < .032$, Cohen's D = .38. The groups did not differ significantly, however, in mental age growth rate in Visual Reception $t(41.291) = 1.171$, $p < .248$, Cohen's D = .58 (Table 6).

Severity of Delay at Initial Evaluation and Diagnostic Stability: DD

A chi-square analysis was conducted to determine the impact that severity of delay at Time 1 has on diagnostic stability of a diagnosis of DD. Twenty-three participants were divided into two groups using the Early Learning Composite (ELC) standard score. Any child that received an ELC standard score greater than or equal to 55 at Time 1 was coded "less delayed," and any child that received an ELC standard score less than or equal to 54 was coded "more delayed." Eleven children (47.8%) were coded "less delayed," and of these 11, six (54.5%) retained their original diagnosis, while five (45.5%) received no diagnosis at Time 2. Twelve children (52.2%) were coded "more delayed," and of these 12, 11 (91.7%) retained their original diagnosis, while one (8.3%) received no diagnosis at Time 2. The assumptions of a chi-square analysis were violated, as two of the cells (50%) had expected counts less than five. A Fisher's exact test was utilized, and revealed a trend approaching statistical significance (two-sided, $p = .069$).

Severity of Delay at Initial Evaluation and Diagnostic Stability: DLD

A chi-square analysis was planned to determine the impact that severity of delay at Time 1 has on the diagnostic stability of a diagnosis of DLD, but three of the cells

(75%) had expected counts less than five, and a Fisher's exact test was employed. The DLD sample was coded according to the type of language disorder they were diagnosed with at Time 1. Children that met diagnostic criteria for a receptive or expressive language disorder were considered less delayed and grouped for analyses, and children that met diagnostic criteria for a mixed expressive/receptive language disorder were considered more delayed. Data were available for 15 children, and missing at one of the two time-points for six children. Eleven (73.3%) children were diagnosed with either a receptive or expressive language disorder and four (26.7%) children were diagnosed with mixed receptive/expressive language disorder. Of the 11 children diagnosed with a receptive or expressive language disorder at Time 1, three children (27.3%) retained their diagnosis, while eight children received no diagnosis (72.7%) at Time 2. Of the four children with a mixed, expressive and receptive language disorder, at time 2, one (25%) child retained the diagnosis, while three (75%) children received no diagnosis. The diagnostic outcomes of the expressive or receptive language disorder group and the mixed expressive and receptive language disorder group were not significantly different (two-sided, $p = 1.00$).

Chi-square test Investigating Diagnostic Stability of DD, DLD within Gender

A series of chi-square analyses were conducted to assess the impact of gender on the diagnostic stability of the DD and DLD groups. Analyses were performed separately by gender. All Time 2 diagnostic outcomes grouped by Time 1 diagnosis and gender can be found in Table 7.

Forty-four male participants were included in the first chi-square analysis, 31 of whom had received an initial diagnosis of DD, and 13 of whom had received an initial

diagnosis of DLD. Participants were coded for this analysis as"1" for those participants that retained their initial diagnosis at Time 2, and "2" for those participants that received any other diagnosis, or no diagnosis, at Time 2. Seventeen (54.8%) males who received a diagnosis of DD at Time 1 retained their diagnosis at Time 2, and 14 (45.2%) received another diagnosis, or no diagnosis, at Time 2. None of the males who were diagnosed with DLD at Time 1 retained their diagnosis at Time 2, while 13 (100%) received another diagnosis or no diagnosis at Time 2. Chi-square analysis revealed a significant difference between the DD and DLD groups, with males who received a diagnosis of DD at Time 1 retaining their diagnosis at Time 2 at a much higher rate than males diagnosed with DLD at Time 1, $x^2(1, N = 44) = 11.618$ $p = .001$, Phi = .514 (Table 8).

The assumptions of a chi-square analysis were violated due to three cells (75%) having expected counts of less than five, and a Fisher's exact test was conducted. 14 female participants were included in this analysis, six of whom had received an initial diagnosis of DD, while eight had received an initial diagnosis of DLD, at Time 1. Participants were coded for this analysis as"1" for those participants that retained their initial diagnosis at Time 2 and "2" for those participants that received any other diagnosis, or no diagnosis, at Time 2. Two (33.3%) females who received a diagnosis of DD at Time 1 retained their diagnosis at Time 2, and four (66.6%) females who received a diagnosis of DD at Time 1 received any other diagnosis or no diagnosis at Time 2. Two (25%) females who received a diagnosis of DLD at Time 1 retained their diagnosis at Time 2, and six (75%) females who received a diagnosis of DLD at Time 2 received any other diagnosis, or no diagnosis, at Time 2. Fisher's exact test, revealed no significant differences between the DD and DLD groups (two-sided, $p = 1.00$).

Chi-square test for Diagnostic Stability of DD, DLD between Genders

A chi-square test was conducted to compare how gender impacts the diagnostic stability of the DD and the DLD groups. Forty-four Males and 14 females from the DD and DLD groups were coded 1 if they retained their Time 1 diagnosis at Time 2, or as 2 if they received any other diagnostic outcome, including reverse diagnosis, no diagnosis, other diagnosis, or ASD, at Time 2. Seventeen (38.6%) males from the DD and DLD groups retained their diagnosis at Time 2, and 27 (61.4%) received another diagnostic outcome. Four (28.6%) females retained their diagnosis at Time 2, and 10 (71.4%) received another diagnostic outcome. This analysis revealed no significant difference between genders on diagnostic stability with regard to retaining initial diagnosis or receiving any other diagnostic outcome, $x^2(1, N = 58) = .466$ $p = .49$, Phi = .09 (Table 9).

Follow-up analyses were planned to more specifically assess how gender impacts the diagnostic stability of the DD and DLD groups, with regard to those that retain their diagnosis versus those that improve, and receive no diagnosis at Time 2. Due to insufficient sample sizes for a chi-square analysis, two of the cells (50%) had counts less than five, Fisher's exact test was employed. Males and females from the DD and DLD groups were coded as 1 if they retained their Time 1 diagnosis at Time 2, or as 2 if they received no diagnosis at Time 2. Participants in the DD and DLD groups that received an ASD or an "other" diagnosis were not included in this analysis. Of the 26 males that received a diagnosis of DD at Time 1, 17 (65.4%) retained that diagnosis, and nine (44.6%) received no diagnosis at Time 2. Of the six females that received a diagnosis of DD at Time 1, two (33%) retained that diagnosis, and four (66%) received no diagnosis at Time 2. Fisher's exact test revealed no significant differences in diagnostic outcome in

the DD group by gender (p = .194). Of the 12 males who received a diagnosis of DLD at Time 1, zero retained that diagnosis and 12 (100%) received no diagnosis at Time 2. Of the seven females who received a diagnosis of DLD at Time 1, two (28.6%) retained that diagnosis at Time 2, and five (71.4%) received no diagnosis. A Fisher's exact test revealed no significant differences in diagnostic stability by gender in the DLD group (p = 0.123).

Maternal Education and Family Income/SES Correlation and Impact on Diagnostic Stability (DD, DLD)

A logistic regression was conducted to investigate the influence of maternal education and family income, as indicators of SES, on diagnostic stability of both DD and DLD groups. Maternal education and family income/SES data was available for 45 participants from the total sample. Parametric correlation revealed that maternal education and yearly family income were highly correlated, $r^2(43) = .474$, $p < .001$ in this sample (Table 10).

Thirty-four participants from the DD and DLD groups that retained their diagnosis or received no diagnosis at Time 2 had maternal education and family income data available. Maternal education was entered as a factor into a logistic regression model, which did not improve the model's fit, and was not a significant predictor of diagnostic outcome in the DD or DLD groups, $x^2(5, N = 34) = 5.797$, p = .326. A separate logistic regression was conducted and income was entered as a factor into the model, which did not improve the model's fit, and was not a significant predictor of diagnostic outcome in the DD or DLD groups, $x^2(9, N = 34) = 6.049$, p = .735. A final logistic regression was conducted that first entered maternal education, $x^2(5, N = 34) = 5.610$ p =

.346, and then income, $x^2(9, N = 34) = 10.019$, $p = .349$, into the model, neither of which contributed to the prediction of diagnostic stability in the DD or DLD groups.

Chi-square test Comparing Rate of ASD Diagnosis at Time 2 by Initial Diagnosis (DD, DLD)

Due to insufficient sample size for a chi-square analysis, two of the cells (50%) had expected frequencies that were less than five, Fisher's exact test was utilized to determine if the likelihood of receiving an ASD diagnosis at Time 2 differed as a function of an initial diagnosis of DD or DLD. Of the 37 children who received a DD diagnosis at Time 1, five (13.5%) received a Time 2 diagnosis of an ASD, while 32 (86.5%) retained the initial diagnosis, reversed diagnoses, received no diagnosis, or received another diagnosis. Of the 21 children who received a DLD diagnosis at Time 1, two (9.5%) received a diagnosis of an ASD at Time 2, while 19 (90.5%) retained the diagnosis, reversed diagnoses, received no diagnosis, or received another diagnosis. Fisher's exact test revealed no significant differences between the DD and DLD groups with both the DD and DLD groups demonstrating a low rate of ASD diagnosis at Time 2 (two-tailed, $p = 1.00$).

Chi-square test Comparing ASD diagnosis at Time 2 by Gender

Due to insufficient sample size for a chi-square analysis, one of the cells (25%) had an expected frequency that was less than five and equaled one, a Fisher's exact test was conducted to assess the impact gender had on the likelihood that a child who received either a DD or DLD diagnosis at Time 1 would receive a diagnosis of an ASD at Time 2. Forty-four males received a diagnosis of either DD or DLD at Time 1; six (13.6%) of these males received an ASD diagnosis at Time 2, and 38 (86.4%) received

any other diagnostic outcome. Thirteen females received a diagnosis of either DD or DLD at Time 1, and one (7.6%) received an ASD diagnosis at Time 2, with 12 (92.4%) receiving any other diagnostic outcome. Due to insufficient sample sizes, a Fisher's Exact Test was conducted, and revealed no significant difference between males and females in the likelihood to receive an ASD diagnosis at Time 2 (two-tailed, $p = 1.00$).

Discussion:

The goal of the current study was to examine the diagnostic stability of Developmental Delay and Developmental Language Disorder, and explore the effect of different participant-centered variables on that stability. The children included in this study participated in an ongoing study conducted at the University of Connecticut, which is assessing the use of the M-CHAT, and its revised version (M-CHAT-R), as an ASD screening instrument for children between the ages of 16 to 30 months. Though this study was specifically investigating the M-CHAT and M-CHAT-R for their ability to effectively screen for an ASD, many of the children detected by these screening instruments presented with developmental delays at the time of their initial evaluations and qualified for a diagnosis of DD or DLD.

Summary of Results

When including all possible diagnostic outcomes at follow-up, we found that children who received a diagnosis of DD at their initial evaluation were significantly more likely to retain that diagnosis than children who received an initial diagnosis of DLD. These findings support our hypothesis and are consistent with the conceptualization of DD as a condition that involves more global, impactful delays as compared to DLD (American Psychiatric Association, 2000; Nelson, Nygren, Walker, &

Panoscha, 2006). The high retention of the DD diagnosis between initial and follow-up evaluations (51.4%), is in direct contrast to the trend observed in children with DLD, as a high number (57.7%) of these children receive no diagnosis whatsoever at follow-up. When including only the outcomes of "retain diagnosis" and "no diagnosis" outcomes and comparing the DD and DLD groups, this difference became more apparent, as more children from the DD group retained their original diagnosis (67.9%), as opposed to receiving no diagnosis, than the DLD group (14.3%).

Our analyses demonstrated that while all children improved on all four developmental subdomains of the Mullen (Receptive Language, Expressive Language, Fine Motor, and Visual Reception) between their initial and follow-up evaluations, those children with DLD made significantly more developmental progress in the Language subdomains during the time between evaluations than children with DD. Descriptively, the DLD group made more than 2 years of developmental gains in both Visual Reception (a developmental rate of 1.06) and Language (1.09) skills in the 2 years between evaluations. The DD groups made less than two years of developmental gains in the time between evaluations in both Visual Reception (.82) and Language (.92) skills. The difference between the DD and DLD groups was not statistically significant, however, with regard to the growth rate on the Visual Reception subtests of the Mullen. Qualitatively, the gap between the performance of the children with DD (.82) and DLD (1.06) is sizeable. Additionally, the effect sizes (Cohen's D = .38 for Language, = .58 for Visual Reception) for these analyses suggest that the smaller sample size may have resulted in this difference being non-significant; a significant difference may have been detected with a larger sample size. These findings partially supported our hypothesis, in

which we had predicted that children with DLD would make significantly greater gains in all areas of development compared to children with DD. These findings are consistent with prior studies that demonstrated that children with GDD show persistent, global delays across time (Shevell, Majnemer, Platt, Webster, & Birnbaum, 2004), and that children with language delay improve across all developmental domains across time (Silva, 1980).

The severity of delay at initial evaluation in the DD group had an impact on diagnostic stability that was approaching statistical significance; children who were coded as more impaired based on their ELC score at their initial evaluation appeared more likely to retain their diagnosis at follow-up (91.7% of the more delayed DD group retained their initial diagnosis, while only 54.5% of the less delayed retained their diagnosis). However, the diagnostic stability in children who received a DLD diagnosis did not appear to be affected by the severity of their language impairment. The finding approaching statistical significance for the DD group was consistent with our hypothesis, however, the sample sizes for these particular analyses were small, and may not have been large enough to detect a real effect.

Next we investigated the impact of gender on diagnostic stability, regarding both the diagnostic stability within and between genders. Our results suggest that males who receive a diagnosis of DD at their initial evaluation tend to retain this diagnosis significantly more often than males with DLD, as compared to receiving any other diagnostic outcome (54.8% of males with DD retain, 0% of DLD retain). This finding was consistent with our hypothesis and offers further support for the conceptualization of DD as more severe in presentation and effect. This difference was not observed in female

participants when comparing the DD and DLD groups (33.3% DD retain, 25% DLD retain). There were no significant differences in diagnostic stability when comparing genders, which was contrary to our prediction that males would retain their original diagnosis more often than females. Again, however, the sample size in these analyses was particularly small, and may not have been robust enough to detect an effect. Though this sample does not capture all of the children that received evaluations through the study, the ratio of males to females in this specific sample (44:14) was consistent with prior research and surveys that demonstrate that males are more often affected by developmental delays (Child Trends Data Bank, 2013; Delgado, Vagi, & Scott, 2007).

Research has conventionally paired maternal education and family income as indicators of overall familial socioeconomic status, and previous research has established both as significant risk factors for developmental delay in children (Delgado, Vagi, & Scott, 2007; Sameroff et al., 1987; Sonnander & Claesson, 1999), and has measurable effects on a child's language abilities (Arriaga, Fenton, Cronan, & Pethick, 1998; Hart & Risley, 1992; Hoff, 2003; Morrisset, Barnard, Greenberg, Booth, & Spieker, 1990). Within our sample, maternal education and family income were highly correlated with one another. However, contrary to our hypothesis, these results did not demonstrate a relationship between retaining initial diagnosis of DD and DLD and either maternal education or income. The sample size for this analysis (n = 37) may have impacted these results.

Finally, we compared the proportion of children from the DD and DLD groups who received a diagnosis of an ASD at their follow-up evaluation. There were no significant differences between the DD and DLD groups, as both had a small number of

children that received an ASD diagnosis at follow-up (DD – 13.5%, DLD – 9.5%). This was contrary to our prediction that those children who presented with more severe, global delays as observed in DD would be more likely to receive an ASD diagnosis at follow-up. We also found that participant gender did not increase the likelihood that a child from both the DD and DLD groups would receive a diagnosis of an ASD at their follow-up appointment. This was also contrary to our prediction that males at follow-up would have an increased likelihood of being diagnosed with an ASD, due to the increased number of males that are affected by ASDs (Fombonne, 1999). Sample sizes for theses analyses were small, however, and these results should be interpreted conservatively.

Implications

The results of this study demonstrate the difference between the diagnostic stability of DD and DLD diagnoses between two time points (approximately 2 years). A majority of children with DD retain this diagnosis (57.1%) at their follow-up evaluation compared to receiving no diagnosis (24.3%). This finding emphasizes the severity of impairment in DD, especially when compared to less globally impairing conditions, like DLD. In this study, a majority of children with DLD received no diagnosis (57.1%) at follow-up, with only a few children continuing to meet diagnostic criteria (9.5%). This high level of diagnostic maintenance in the DD group could suggest that the impairments accompanying this delay are less likely to resolve naturally across time or in response to the implementation of current standards of intervention practices (all of the children in the sample received some form of intervention services after their first evaluation). These children may also have more serious underlying impairment (e.g., intellectual disability) that might prevent them from making developmental gains after identification, or impair

their ability to benefit fully from intervention services as a higher functioning child might.

These results could also suggest that the children that continue to meet diagnostic criteria for DD at follow-up may need a higher intensity of intervention services than are currently provided. Additionally, a certain portion of these children may eventually be diagnosed as intellectually disabled at age 6; if so, their current scores (approximately two standard deviations below the mean), would be expected to be relatively consistent across all stages of development and indicate lifelong impairment. The high number of children in the DLD group that receive no diagnosis at follow-up could indicate that these children and their families are responding well to identification and subsequent provision of intervention services. This is not to downplay the significance of DLD as a disorder, or indicate that these children have "recovered," as evidence from research has demonstrated that children who experience SLI's, even when treated, are at a higher risk for a variety of disorders in later development (Bishop & Leonard, 2000; Brownlie, Beitchman, Escobar, Young, Atkinson, Johnson, Wilson, & Douglas, 2004; Snowling, Bishop, & Stothard, 2000).

Children from the DLD group also demonstrated a higher growth in language skills across time when compared to the DD group. This could be explained by the fact that children with DLD present with a specific impairment in a single domain of development, which is identified through evaluation and targeted by speech and language specific interventions, and are able to make more progress as a result of this less global impairment. On the other hand, children with DD are globally impaired and will be receiving a variety of interventions, and/or have an underlying intellectual disability that

delays speech acquisition even with the most robust intervention. However, while the difference between the DD and DLD groups in visual reception skills at time two was not significant, the growth rates between the Language and Visual Reception subscales were comparable qualitatively. A larger sample size may have allowed for a more clear assessment of the differences between the groups by subscale, as qualitatively the DLD group made more progress between evaluations on the Visual Reception subscale as well as the Language subscales. If, however, these findings were replicated in a larger sample size, it could suggest that perhaps intellectual disability is not exerting as profound an influence on these groups, as some children from both groups are making improvements. It could also be that the children with DD benefit from the language that they do learn, which allows cognitive progress in areas such as visual reception, but speech acquisition itself remains slow.

The finding that children with DD who are more impaired at initial evaluation are more likely to retain their diagnosis than children with DD who are less impaired was as predicted. These children could be likely to go on to receive a diagnosis of intellectual disability at a later date. They could have also made progress across the time between evaluations, but due to a floor effect (they are starting at a lower point developmentally), still meet diagnostic criteria for DD. However, on average children with DD were making less than 2 years developmental progress in both language and visual reception in the 2 years between evaluations. Since the sample size for the DLD group was small, and no effect of initial severity on diagnostic stability was found, this would be something to investigate in future studies with a larger number of participants. It is plausible that a

similar effect of initial severity would be observed if these analyses were repeated with a larger group of children with DLD.

Significant differences in the rate of diagnosis retention between males with DD and DLD were observed, with a majority (54.8%) of males retaining their DD diagnosis, but no males with DLD retaining their diagnosis. This could suggest that within males there may be a difference in the manifestation and course of developmental delays, depending on their type and severity. However, it was not possible to assess the difference in retention rate by diagnosis within females, as the sample size was too small to run chi-square analyses. If there had been a similar trend of higher DD diagnosis retention within females, this would not suggest a gender difference, but only further support the trend of higher DD diagnosis retention observed when genders were combined. Comparisons between genders were trending towards significance; no males from the DLD group retained this diagnosis at follow-up and 12 received no diagnosis, while 2 (28.6%) females received DLD at follow-up and 5 (72.4) received no diagnosis. Qualitatively it is interesting to observe that no males retained a diagnosis of DLD, however sample sizes were too small to draw any conclusions.

Maternal education and income were found to be highly correlated as expected, but did not influence diagnostic stability. It was predicted that low maternal education and income would predict maintenance of initial diagnosis, but these results do not support this hypothesis, or findings of previous research. Despite having variability in familial incomes within the sample, the sample size (n = 37) was not as robust as it could have been, which may have influenced these results and prevented the possible detection of a trend. Alternatively, within our study, these children are receiving a diagnosis at

approximately age 2, and then receiving intervention services of some degree and type; in previous studies assessing the impact of income and maternal education and other indicators of overall SES, these children did not receive early identification and intervention. This identification and intervention would likely increase the resources and education available to parents that otherwise would not have access to it, and may explain why no effect was present within this sample.

Though our data did not suggest a higher likelihood of ASD diagnosis at Time 2 based on initial diagnosis or gender, a small portion of our sample did receive ASD diagnoses at follow-up; DD - 5 (13.5%), DLD – 2 (9.5%). The sample sizes may have been too small to detect the typical gender ratio of 4 males to 1 female, but the elevated rate of ASD diagnosis within this population may indicate increased risk for ASDs with other early developmental disability compared to children with normal development at age 2. Another possibility is that these patients were true misses and received a diagnosis other than an ASD at Time 1 when they may have had subclinical ASD symptomatology and did not meet the criteria for an ASD.

It is vital to note that there could also be a variety of other patient- or family-based, or environmental, factors that were not measured in our study that could affect a child's diagnostic stability between time points. The abovementioned theories and explanations for the observed effects in this study could be better assessed through rigorous, controlled study on the effect of intervention (differing intensity, type) and its relationship to diagnostic stability in children with DD and DLD, across a variety of initial levels of development. These results are also based upon our specific diagnostic

criteria for DD and DLD, which would make generalization to other definitions difficult due to widely differing sets of diagnostic criteria.

Limitations

There were a number of limitations that must be considered when interpreting the findings of this study. The sample size (N=58) can be considered small, and made more sophisticated statistical analysis a challenge. This study was not designed to recruit participants with specific diagnoses, such as DD and DLD, but is a larger study designed to assess the effectiveness of an autism-specific screener, not a DD and/or DLD specific screener. Additionally, all of these children screened positive on the M-CHAT or M-CHAT-R, an autism-specific screening instrument, therefore this sample might not be representative or typical of the larger DD and DLD populations. The current sample size allowed for the detection of large effect sizes, but a lack of available data for every participant at both initial and follow-up, and missing data in some cases, resulted in using analyses that correct for small sample sizes, which may have reduced the ability to detect an effect in these cases. In several analyses, small sample size resulted in lower power; these analyses should be revisited in future studies with larger sample sizes. A portion of children who were seen for a Time 1 evaluation declined a Time 2 evaluation or were lost to follow-up. It is possible that the parents of these children were not concerned with their child's development at the time a follow-up evaluation was offered. This could have biased the data, as parents of a child who continued to demonstrate delay may have been more likely to return for a follow-up evaluation. Alternatively, some children that were lost to follow-up may have been more representative of the lower end of the maternal education and socioeconomic status scales, and were not able to return to an evaluation

due to financial circumstances, despite best efforts to provide travel arrangements to such families.

Another limitation to consider is the specific time frame in which this study is assessing diagnostic stability. The data included is from two time points, at two and four years of age; there is no data available beyond the follow-up evaluation. It is impossible to predict long-term outcomes for these children, which often experience difficulties that present later in development despite earlier response to intervention and initial symptom improvement. One potential impact of this is that we do not know which children will go on to receive a diagnosis of intellectual disability (ID) when they are 6 years old, when such a diagnosis is considered stable (American Psychological Association, 2013). This information could provide an additional dimension of specificity to our analyses, and would likely be a robust predictor of diagnostic stability within the sample. Our attempt to stratify the DD and DLD samples by initial severity embodies this assumption, that children who are more severely affected will likely retain their diagnoses at follow-up, and across time more generally, and be susceptible to other disorders such as dyslexia.

An additional limitation in this study was the inability to truly assess the impact and role of intervention on the diagnostic outcome for these participants. It was indicated on history forms at follow-up that each participant received some type of intervention, but specific information regarding the type of services, as well the intensity and frequency of services, was not always available or detailed enough to truly evaluate the impact and quality of services on a child's progress over time. Therefore, it is difficult to conclude from our results that a child's improvement across time was directly related to the services they received, or the increased parental understanding of their child's

deficits, or any other factor that may contribute to a child's progress, or lack thereof, across time. This study was not designed to evaluate the impact of intervention, but the findings emphasize the need for such a study to assess the type, intensity, and quality of services appropriate for each diagnosis, DD and DLD. Such studies have been conducted in children with ASD, but there are fewer research efforts investigating the true effects of current intervention practices in DD and DLD, despite the vital importance of identification and provision of intervention services at young ages.

The reliance upon a single measure of development is also a limitation of the study. Evaluations provided by the study are completed within a 3-hour time frame, which allows for an assessment of a child's skill-level and provision of a diagnosis, but does not necessarily allow for the most comprehensive skill-level measurement. The Mullen is not the most inclusive and thorough measure of language and visual skills, and there are a number of alternative assessments that require more time to administer and provide a more specific assessment of both developmental level (Bayley Scales of Infant and Toddler Development, Stanford-Binet Intelligence Skills) and language skills (Clinical Evaluation of Language Fundamentals), although the developmental level of many of the children at age two would be too low for these measures. The Mullen may not be sensitive enough to pick up on more subtle language impairment at Time 2, which could have resulted in children with DLD scoring within the normal range, but continuing to demonstrate atypical mastery and understanding of language for a child their age.

Future Directions

Recent trends within intervention organizations nationally are troubling. Within Connecticut alone, families and children are required to bear more of the financial burden

of intervention service provision than previously (Johnson, Oliff, & Williams, 2011), and less national funding is committed annually to services such as Head Start (Lu, 2013) that specifically target children at high risk due to low familial SES, a large risk factor for developmental delay and disability. Currently, each state controls the nature and extent of services available to children with developmental delays, and the services can profoundly impact a child's developmental progress. States also set the diagnostic threshold for delays, determine the nature and extent of services provided, and design and implement these services. Families have few options if their state does not provide adequate services, and often must rely on their own income to provide supplemental services that may be necessary to promote development in children with more severe delays. Conceptually, a dedication to earlier intervention creates a larger financial strain in the short-term, but over time, a child that reaches a higher level of functioning, or recovers from delay, will require less financial support as a result (Shonkoff & Hauser-Cram, 1987; Meisels & Shonkoff, 2000).

The lack of standardized diagnostic definitions shared among states will continue to impede proper measurement and assessment of current intervention practices. Under IDEA, states determine the diagnostic criteria of DD and DLD as well as the intervention services provided to children that receive these diagnoses. The quality and extent of intervention services, as well as broad decisions regarding inclusionary criteria, are influenced by a multitude of state-level factors (e.g., funding, awareness, organization of services). States could potentially begin relying on DSM-V or ICD-9 diagnosis as a metric when determining diagnostic criteria, which would more readily allow state-by-state prevalence estimation, assessment of intervention practices, and implementation of

widespread standards of practice. A national epidemiological study to determine prevalence rates and diagnostic trajectories of DD and DLD would be ideal, but likely complicated by inconsistent diagnostic criteria among states and unreliable or unavailable records of diagnosis, intervention, and educational outcomes for individual children. However, an epidemiological study conducted within a single state with clearly defined diagnostic criteria for DD and DLD, a well developed and organized intervention system, and consistent record keeping would allow for a more specific measurement of the rates and progression of DD and DLD.

A possibility for a follow-up to this specific study would be to contact the families of the participants to assess each child's current diagnosis, and compare that diagnosis to initial diagnosis. It would seem logical to suggest that those children who would go on to receive a diagnosis of ID at 6 years of age would almost certainly be included among those children with DD and DLD that maintained their diagnosis or were diagnosed with other disorders at age 4.

In conclusion, the current study found that developmental delay has a significantly higher stability across time than developmental language disorder. This was one of few studies conducted in the past 30 years to assess diagnostic stability for DD and DLD. This study also demonstrated that the severity of delay in children who are identified as DD can be used as a predictor of their later diagnostic stability. Additionally, males with DD tended to retain their diagnosis at a higher rate than males with DLD. However, small sample sizes prevented comparison to females, which makes it difficult to conclude whether there is in fact a gender difference. It is likely that a larger sample size of females would have produced a similar observed diagnostic stability in DD, and

further supported the effect of diagnosis. Additionally, a small portion of children who initially receive a diagnosis of DD and DLD will be later diagnosed with an ASD, indicating that the diagnostic overlap between these disorders can be difficult to distinguish at a young age, and further research to elucidate the differences between these groups at a young age could be helpful. Finally, though our results indicated that maternal education and SES do not have an impact on diagnostic stability across time, future efforts to assess this question utilizing a larger sample size would allow researchers to more clearly understand the influence of these factors that have previously been identified as risk factors for delay. This study demonstrates the need for more consistent diagnostic criteria across states to allow for more accurate measures prevalence and progress across time. It also highlights the need for research to assess the effect of intervention on these children, in order to establish an empirical basis for the type and intensity of services typically recommended for children who present with delay.

References

American Psychiatric Association (Ed.). (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. American Psychiatric Pub.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-V* Washington, DC: American Psychiatric Association.

Arriaga, R. I., Fenton, L., Cronan, T., & Pethick, S. J. (1998).Scores on the MacArthur Communicative Development Inventory of children from low-and middle-income families. *Applied Psycholinguistics*, *19*, 209-224.

Bailey, D.B., Aytch, L.S., Odom, S.L., Symons, F., Wolery, M. (1999). Early Intervention as we know it. *Mental Retardation and Developmental Disabilities Research Reviews, 5,* 11-20.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *British Journal of Psychiatry*, *161*, 839-843.

Baron-Cohen, S., Cox, A., & Baird, G. (1996). Psychological markers in the detection of autism in infancy in a large population. *British Journal of Psychiatry*, *168*, 158-163.

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry*, *31*(7), 1027-1050.

Bishop, D. V., & Leonard, L. B. (2000). *Speech and language impairments in children: causes, characteristics, intervention and outcome*. New York, NY: Psychology Press.

Boyle, C.A., Boulet, S., Schieve, R.A., Cohen, S.J., Blumberg, M., Visser, S., Kogan, M.D. (2011). Trends in the Prevalence of Developmental Disabilities in US Children. *Pediatrics,127*(6), 2010-2022.

Brownlie, E. B., Beitchman, J. H., Escobar, M., Young, A., Atkinson, L., Johnson, C., Wilson, B., & Douglas, L. (2004). Early language impairment and young adult delinquent and aggressive behavior. *Journal of Abnormal Child Psychology*, *32*(4), 453-467.

Centers for Disease Control and Prevention (July, 2012). *Developmental Milestones*. Retrieved from http://www.cdc.gov/ncbddd/actearly/milestones/

Child Trends Data Bank (July, 2013). *Screening and Risk for Developmental Delay.*

Retrieved from http://www.childtrends.org/?indicators=Screening-and-risk-for-developmental-delay

Clark, M.K., Kamhi, A.G. (2010). *Language Disorders (Child Language Disorders)*. In: JH Stone, M Blouin, editors. International Encyclopedia of Rehabilitation. Retrieved from http://cirrie.buffalo.edu/encyclopedia/en/article/31/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences. (2nd ed.).* Hillsdale, NJ: Lawrence Earlbaum Associates.

Connecticut Birth to Three System (2010). *Evaluation to Determine Eligibility.* Retrieved from http://www.birth23.org/providers/CurrentProcedures/EVALuation.doc

Connecticut Birth to Three System (2011). *Individualized Family Service Plan Handbook*. Retrieved at www.birth23.org/providers/IFSPhandbook.pdf

Centers for Medicare and Medicaid Services (2013).*ICD-9 Code Lookup*. Retrieved from http://www.cms.gov/medicare-coverage-database/staticpages/icd-9-code-lookup.aspx

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., Drew, A., & 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., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., Varley, J. (2010). Randomized, controlled trail of an intervention for toddlers with autism: the early start Denver model. *Pediatrics, 125* (1), 17-23.

Delgado, C. E., Vagi, S. J., & Scott, K. G. (2007).Identification of early risk factors for developmental delay. *Exceptionality*, *15*(2), 119-136.

Eisenmajer, R., Prior, M., Leekam, S., Wing, L., Ong, B., Gould, J., & Welham, M. (1998). Delayed language onset as a predictor of clinical symptoms in pervasive developmental disorders. *Journal of autism and developmental disorders*, *28*(6), 527-533.

Erickson, W., Lee, C., & von Schrader, S. (2012). *2011 Disability Status Report: United States.* Ithaca, NY: Cornell University Employment and Disability Institute (EDI).

Fombonne, E. (1999). The epidemiology of autism: a review. *Psychological medicine*, *29*(4), 769-786.

Green, D., Baird, G., Barnett, A. L., Henderson, L., Huber, J., & Henderson, S. E. (2002). The severity and nature of motor impairment in Asperger's syndrome: a

comparison with specific developmental disorder of motor function. *Journal of child psychology and psychiatry*, *43*(5), 655-668.

Green, G., Brennan, L.C., Fein, D. (2002). Intensive behavioral treatment for a toddler at high risk for autism. *Behavioral Modification, 26*(1), 69-102.

Golden, M., &Birns, B. (1976).Social class and infant intelligence. Origins of intelligence. New York: Plenum, 299-351.

Gottfried, A. W., Gottfried, A. E., Bathurst, K., Guerin, D. W., & Parramore, M. M. (2003). Socioeconomic status in children's development and family environment: Infancy through adolescence. *Socioeconomic status, parenting, and child development, 287*, 189-207.

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology, 28(6)*, 1096.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes Publishing.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74(5),* 1368-1378.

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005).A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in developmental disabilities*, *26*(4), 359-383.

Howlin, P., Mawhood, L., & Rutter, M. (2000). Autism and Developmental Receptive Language Disorder—a Follow-up Comparison in Early Adult Life. II: Social, Behavioural, and Psychiatric Outcomes. *Journal of Child Psychology and Psychiatry*, *41*(5), 561-578.

Johnson, N., Oliff, P., & Williams, E. (2011). An update on state budget cuts. *Center on Budget and Policy Priorities*, 1-16.

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S., Green, J., Dumont-Mathieu, T., Robins, D., & Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. Journal of autism and developmental disorders, 38(4), 606-615.

Kleinman, J., Robins, D., Ventola, P., Pandey, J. et al. (2008). The Modified Checklist for Autism in Toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*, 827-839.

Klin, A., Carter, A., & Sparrow, S. S. (1997). Psychological assessment. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of Autism and Developmental Disorders* (2nd ed., pp. 418-427). New York: Wiley.

Law, J., Garrett, Z., Nye, C. (2004). The Efficacy of Treatment for Children with Developmental Speech and Language Delay/Disorder: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research, 47,* 924-943.

Lord, C., Risi, S., Lambrecht, L., Cook Jr., E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*,*30*(3), 205-223.

Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of autism and developmental disorders*, *16*(3), 335-349.

Lu, Adrienne. (2013). Head Start hit with worst cuts in its history. *USA Today*. Retrieved from http://www.usatoday.com/story/news/nation/2013/08/19/stateline-head-start/2671309/

Majnemer, A. (1998, March). Benefits of early intervention for children with developmental disabilities. In *Seminars in Pediatric Neurology* (Vol. 5, No. 1, pp. 62-69). WB Saunders.

McCall, R. B. (1979). The development of intellectual functioning in infancy and the prediction of later IQ. *Handbook of infant development*, 707-741.

McCall, R. B. (1981). Early predictors of later IQ: The search continues. *Intelligence*, *5*(2), 141-147.

McGee, G.G., Morrier, M.J., Teresa, D. (1999). An incidental teaching approach to early intervention for toddlers with autism. *Research and Practice for Persons with Severe Disabilities, 24*(3), 133-146.

Meisels, S. J., & Provence, S. (1989). *Screening and Assessment: Guidelines for Identifying Young Disabled and Developmentally Vulnerable Children and Their Families*. National Center for Clinical Infant Programs, 733 15th St., NW, Suite 912, Washington, DC 20005.

Meisels, S. J., & Shonkoff, J. P. (2000). Early childhood intervention: A continuing evolution. *Handbook of early childhood intervention*, *2*, 3-31.

Miyahara, M., Tsujii, M., Hori, M., Nakanishi, K., Kageyama, H., & Sugiyama, T. (1997). Brief report: motor incoordination in children with Asperger syndrome and learning disabilities. *Journal of autism and developmental disorders*, *27*(5), 595-603.

Morisset, C. E., Barnard, K. E., Greenberg, M. T., Booth, C. L., & Spieker, S. J. (1990). Environmental influences on early language development: The context of social risk. *Development and Psychopathology*, *2*(2), 127-149.

Mullen, E. M. (1994). *The Mullen Scales of Early Development*. Circle Pines, MN: American Guidance Services.

New York State Department of Health Bureau of Early Intervention (2005).*Standards and Procedures for Evaluations, Evaluation Reimbursement, and Eligibility Requirements and Determinations Under the Early Intervention Program.* Retrieved from http://www.health.ny.gov/community/infants_children/ early_intervention/ memoranda/2005-02/docs/memorandum_2005-02.pdf

Nelson, H.D., Nygren, P., Walker, M., Panoscha, R. (2006).*Screening for Speech and Language Delay in Preschool Children*. Retrieved from www.ahrq.gov/ downloads/pub/prevent/pdfser/**speech**syn.pdf

National Dissemination Center for Children with Disabilities (2012) *Developmental Delay*. Retrieved athttp://nichcy.org/disability/specific/dd

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science*, *8*(1), 74-87.

Provost, B., Lopez, B. R., & Heimerl, S. (2007). A comparison of motor delays in young children: Autism spectrum disorder, developmental delay, and developmental concerns. *Journal of autism and developmental disorders*, *37*(2), 321-328.

Ramey, C. T., & Ramey, S. L. (1994). Which children benefit the most from early intervention? *Pediatrics*, *94*(6), 1064-1066.

Ramey, S.L., and Ramey, C.T. (1999).Early experience and early intervention for children “at risk” for developmental delay and mental retardation. *Mental Retardation and Developmental Disabilities Research Reviews, 5,* 1-10.

Resnick, M. B., Eyler, F. D., Nelson, R. M., Eitzman, D. V., & Bucciarelli, R. L. (1987). Developmental intervention for low birth weight infants: improved early developmental outcome. *Pediatrics*, *80*(1), 68-74.

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.

Rosenberg, S. A., Zhang, D., & Robinson, C. C. (2008). Prevalence of developmental delays and participation in early intervention services for young children. *Pediatrics*, *121*(6), e1503-e1509.

Rutter, M., Le Couteur, A., & Lord, C. (2003).Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*.

Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *Journal Information*, *110*(6).

Sameroff, A. J., Seifer, R., Barocas, R., Zax, M., & Greenspan, S. (1987). Intelligence quotient scores of 4-year-old children: social-environmental risk factors. *Pediatrics*, *79*(3), 343-350.

Silva, P. A. (1980). The prevalence, stability and significance of developmental language delay in preschool children. *Developmental Medicine & Child Neurology*, *22*(6), 768-777.

Shackelford, J. (2000). *State and jurisdictional eligibility definitions for infants and toddlers with disabilities under IDEA*. National Early Childhood Technical Assistance System. Issue 21. Retrieved fromhttp://www.nectac.org/~pdfs/pubs/nnotes_21.pdf

Shapiro, B. K., Palmer, F. B., Antell, S., Bilker, S., Ross, A., & Capute, A. J. (1990). Precursors of reading delay: neurodevelopmental milestones. *Pediatrics*, *85*(3), 416-420.

Shevell, M., Majnemer, A., Platt, R. W., Webster, R., & Birnbaum, R. (2005).Developmental and functional outcomes at school age of preschool children with global developmental delay. *Journal of child neurology*, *20*(8), 648-654.

Shonkoff, J. P., & Hauser-Cram, P. (1987). Early intervention for disabled infants and their families: a quantitative analysis. *Pediatrics*, *80*(5), 650-658.

Simpson, G.A., Colpe, L., Greenspan, S. (2003). Measuring functional developmental delay in infants and young children: prevalence rates from the NHIS-D. *Pediatric and Perinatal Epidemiology*, *17*, 68–80.

Snowling, M., Bishop, D. V. M., & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry*, *41*(5), 587-600.

Sonnander, K., & Claesson, M. (1999).Predictors of developmental delay at 18 months and later school achievement problems. *Developmental Medicine & Child Neurology*, *41*(3), 195-202.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984).The Vineland Adaptive Behavior Scales-Interview Edition. Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Cicchetti, D. V., & Balla, D.A., (2005).The Vineland Adaptive Behavior Scales-II, Second Edition. Circle Pines, MN: American Guidance Service, Inc. 109-117.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language and Hearing Research*, *41*(2), 407.

Trower, D. R., & Nicol, A.R. (1996). Life-span Intellectual Development of People with Mental Retardation. *Developmental Medicine & Child Neurology*, *38*(7), 645-650.

Whitehurst, G. J., & Fischel, J. E. (1994). Practitioner Review: Early Developmental language Delay: What. If Anything. Should the Clinician Do About It? *Journal of Child Psychology and Psychiatry*, *35*(4), 613-648.

Table 1. Participant Demographics

	DD Mean Age (months), SD, and range (n=37)	DLD Mean Age (months), SD, and range (n=21)	Total Mean Age (months), SD, and range (n=58)	Difference Between Groups
Time 1	27.53 (4.73) 18.82 – 35.38	25.48 (4.74) 16.98 – 32.39	26.78 (4.79) 16.98-35.38.	$t(57)= 1.565$, $p = .123$
Time 2	52.93 (7.33) 43.63-82.59	53.91 (8.1) 43.67-72.85	53.30 (7.57) 43.64-82.59	$t(57) = .547$, $p = .642$
	Frequency (%) DD	Frequency (%) DLD	Total	
Participant Gender				
Male	31(83.7%)	13 (62%)	44	
Female	6(16.3%)	8 (38%)	14	
Total	37	21	58	$x^2 (1, N = 58) = 3.5$, $p = .0614$
Ethnicity				
White	27(77%)	18(85.7%)	45	
Black/African American	5(14.3%)	2(9.5%)	7	
Latino/Hispanic	2(5.7%)	0	2	
Asian/Biracial/Other	1(3%)	1(4.8%)	2	
Total	35	21	56	$x^2 (3, N = 56) = .980$, $p = .914$
Maternal Education				
No degree or diploma	2 (6.5%)	0	2	
High school diploma or GED	8 (25.8%)	7 (33.3%)	15	

Vocational, technical, or associates degree	2 (6.5%)	1(5.3%)	3	
College degree	11 (35.5%)	6 (32%)	17	
Masters Level degree	7(22.6%)	3(15.8%)	10	
Ph.D., MD, JD level degree	1(3.2%)	2(10.5%)	3	
Total	31	19	50	x^2 (5, N = 50) = .696, p = .755
Yearly Income				
<$10,000	0	1 (6%)	1	
$10,000-$20,000	1(3.7%)	0	1	
$20,000-$30,000	6 (22.2%)	2 (11%)	8	
$30,000-$40,000	1 (3.7%)	2 (11%)	3	
$40,000-$50,000	2 (7.4%)	1 (6%)	3	
$50,000-$60,000	1 (3.7%)	2 (11%)	3	
$60,000-$70,000	6 (22.2%)	1 (6%)	7	
$70,000-$80,000	1(3.7%)	1(6%)	2	
$80,000-$90,000	0	1(6%)	1	
$90,000-$100,000	2 (7.4%)	3 (17%)	5	
>$100,000	7 (26%)	4 (22%)	11	
Total	27	18	45	x^2 (10, N = 45) = 9.156, p = .517

Table 2. Diagnostic Stability, DD vs. DLD – All Diagnostic Outcomes

Time 1 Diagnosis	Time 2 Diagnosis						x^2	*p*	Φ (Phi)
	Retain Diagnosis	Reverse Diagnosis	No Diagnosis	Other Diagnosis	ASD	Total			
DD	19 (51.4%)	1 (2.7%)	9 (24.3%)	3 (8.1%)	5 (13.5%)	37	12.334	.015	.461
DLD	2 (9.5%)	2 (9.5%)	12 (57.1%)	3 (14.3%)	2 (9.5%)	21			

Table 3. Diagnostic Stability, DD vs. DLD – Retain Diagnosis vs. Any Diagnosis

Time 2 Diagnosis	Time 1 Diagnosis				
	DD	DLD	x^2	*p*	Φ (Phi)
Retain	19 (51.4%)	2 (9.5%)	10.15	.0014	.4183
Any Other Diagnosis	18 (48.6%)	19 (89.5%)			

Table 4. Diagnostic Stability, DD vs. DLD – No Diagnosis vs. Any Diagnosis

Time 2 Diagnosis	Time 1 Diagnosis				
	DD	DLD	x^2	*p*	Φ (Phi)
No Diagnosis	9 (24.3%)	12 (57.1%)	6.25	.0124	.3283
Any Other Diagnosis	28 (75.7%)	9 (42.9%)			

Table 5. Diagnostic Stability, DD vs. DLD – Retain Diagnosis vs. No Diagnosis

Time 2 Diagnosis	Time 1 Diagnosis				
	DD	DLD	x^2	*p*	Φ (Phi)
Retain	19 (67.9%)	2 (14.3%)	10.714	.01	.505
No Diagnosis	9 (32.1%)	12 (85.79%)			

Table 6. Developmental Progress between Evaluations – DD vs. DLD

	DD	DLD		
	Mental Growth Mean (SD)	Mental Growth Mean (SD)	*t*	*p*
Visual Reception	.82 (.51)	1.06 (.25)	1.171	.248
Language (Expressive + Receptive)	.92 (.56)	1.09 (.29)	2.223	.032**

Table 7. Time 2 Diagnosis by Initial Diagnosis and Participant Gender

	DD		DLD	
	Males	Females	Males	Females
Retain	17	2	0	2
Reverse	1	0	2	0
No Diagnosis	5	4	9	3
ASD	3	0	1	2
Other Diagnosis	5	0	1	1
Total	31	6	13	8
	37		21	

Table 8. Males and Diagnostic Stability between DD and DLD groups

Time 2 Diagnosis	Time 1 Diagnosis				
	DD	DLD	x^2	*p*	Φ (Phi)
Retain Diagnosis	17 (54.8%)	0 (0%)	11.618	.001	.514
Any Other Diagnostic Outcome	14 (45.2%)	13 (100%)			

Table 9. DD and DLD Diagnostic Stability between Genders

Time 2 Diagnosis	Gender				
	Males	Females	x^2	*p*	Φ (Phi)
Retain Diagnosis	17 (38.6%)	4 (28.6%)	.466	.49	.09
Any Other Diagnostic Outcome	27 (61.4%)	10 (71.4%)			

Appendix A

DSM-IV TR

Expressive Language Disorder

A. The scores obtained from standardized individually administered measures of expressive language development are substantially below those obtained from standardized measures of both nonverbal intellectual capacity and receptive language development. The disturbance may be manifest clinically by symptoms that include having a markedly limited vocabulary, making errors in tense, or having difficulty recalling words or producing sentences with developmentally appropriate length or complexity.

B. The difficulties with expressive language interfere with academic or occupational achievement or with social communication.

C. Criteria are not met for Mixed Receptive-Expressive Language Disorder or a Pervasive Developmental Disorder.

D. If Mental Retardation, a speech-motor or sensory deficit, or environmental deprivation is present, the language difficulties are in excess of those usually associated with these problems.

Mixed Expressive-Receptive Language Disorder

A. The scores obtained from a battery of standardized individually administered measures of both receptive and expressive language development are substantially below those obtained from standardized measures of nonverbal intellectual capacity. Symptoms include those for Expressive Language Disorder as well as difficulty understanding words, sentences, or specific types of words, such as spatial terms.

B. The difficulties with receptive and expressive language significantly interfere with academic or occupational achievement or with social communication.

C. Criteria are not met for a Pervasive Developmental Disorder.

D. If Mental Retardation, a speech-motor or sensory deficit, or environmental deprivation is present, the language difficulties are in excess of those usually associated with these problems.

Appendix B

DSM V

Language Disorder Diagnostic Criteria 315.39 (F80.9)

1. Persistent difficulties in the acquisition and use of language across modalities (i.e., spoken, written, sign language, or other) due to deficits in comprehension or production that include the following:
    - Reduced vocabulary (word knowledge and use).
    - Limited sentence structure (ability to put words and word endings together to form sentences based on the rules of grammar and morphology).
    - Impairments in discourse (ability to use vocabulary and connect sentences to explain or describe a topic or series of events or have a conversation).
2. Language abilities are substantially and quantifiably below those expected for age, resulting in functional limitations in effective communication, social participation, academic achievement, or occupational performance, individually or in any combination.
3. Onset of symptoms is in the early developmental period.
4. The difficulties are not attributable to hearing or other sensory impairment, motor dysfunction, or another medical or neurological condition and are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay.

Intellectual Disability Diagnostic Criteria 319

Intellectual disability (intellectual developmental disorder) is a disorder with onset during the developmental period that includes both intellectual and adaptive functioning deficits in conceptual, social, and practical domains. The following three criteria must be met:

1. Deficits in intellectual functions, such as reasoning, problem solving, planning, abstract thinking, judgment, academic learning, and learning from experience, confirmed by both clinical assessment and individualized, standardized intelligence testing.
2. Deficits in adaptive functioning that result in failure to meet developmental and sociocultural standards for personal independence and social responsibility. Without ongoing support, the adaptive deficits limit functioning in one or more activities of daily life, such as communication, social participation, and independent living, across multiple environments, such as home, school, work, and community.
3. Onset of intellectual and adaptive deficits during the developmental period.

Note: The diagnostic term *intellectual disability* is the equivalent term for the ICD-11 diagnosis of *intellectual developmental disorders.*

Appendix C

Diagnostic Criteria for DD, DLD

Developmental Delay

_____ Delay of at least 1.5 standard deviations on AT LEAST ONE of the following ("Non-language"):

_____ **Mullen** Visual Reception (T-score=35 or less)
_____ **Mullen** Fine Motor (T-score=35 or less)
_____ **Vineland** Motor Skills (SS=77 or less)

AND

_____ Delay of at least 1.5 standard deviations on AT LEAST ONE of the following ("Language"):

_____ **Mullen** Expressive Language (T-score=35 or less)
_____ **Mullen** Receptive Language (T-score=35 or less)
_____ **Vineland** Communication (SS=77 or less)

AND

_____ At least one from the 2 categories above must be a delay on the **Mullen**

Developmental Language Disorder (DLD)

_____ Delays of at least 1.5 standard deviations on AT LEAST TWO of the following:

_____ Mullen Expressive Language (T-score=35 or less)
_____ Mullen Receptive Language (T-score=35 or less)
_____ Vineland Communication (SS=77 or less)

OR

_____ Delay of at least two standard deviations on ONLY ONE of the following:

_____ Mullen Expressive Language (T-score=30 or less)
_____ Mullen Receptive Language (T-score=30 or less)
_____ Vineland Communication (SS=70 or less)

AND

_____ **No delays** of greater than 1.5 standard deviations on **any other** subscales or domains:

_____ Mullen Visual Reception (T-score=36 or higher)
_____ Mullen Fine Motor (T-score=36 or higher)
_____ Vineland Motor Skills (SS=76 or higher)