A Longitudinal Case Study: Language Abilities of Minority Autistic Young Adults

By

Teresa Girolamo
M.A., Child Language, 2016
M.S.T., Pace University, 2013
M.A., University of Washington, 2011
B.A., Kenyon College, 2008

Submitted to the graduate degree program in Child Language and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Chair: Dr. Mabel L. Rice

____

Dr. Steven F. Warren

____

Dr. Meghan Davidson

____

Dr. Karrie Shogren

____

Dr. Matthew Gillispie

Date Defended: May 4, 2021

The dissertation committee for Teresa Girolamo certifies that this is the approved version of the following dissertation:

A Longitudinal Case Study: Language Abilities of Minority Autistic Young Adults

____

Chair: Dr. Mabel L. Rice

Date Approved: May 6, 2021

Abstract

One aim in clinical research is furthering understanding of linguistic variation across populations. While important for understanding language acquisition in the context of human development, little is known about the language abilities of autistic adolescents and young adults, especially those who are racial/ethnic minorities or whose language is below typical. These knowledge gaps limit our understanding of the experiences of autistic individuals.

To address these gaps, a longitudinal case study investigated the language abilities of eleven autistic minority young adults (1 female, 10 males; 17 – 23 years). Specifically, it explored: (a) whether individuals changed in relative performance levels within the group across three years; (b) whether speech sound disorder interfered with language assessment; (c) individual differences in expressive language and receptive language; (d) individual differences in expressive and receptive vocabulary; (e) individual differences in use of finiteness markers and judgments of finiteness errors. The author administered standardized tests online on overall language, vocabulary, morphosyntax, nonverbal intelligence, and nonword repetition. Analysis included descriptives and nonparametric statistics.

Results revealed that participants did not change in relative performance levels and had articulation such that speech sound disorder likely did not interfere with assessment performance. Most participants performed near floor on standardized assessments of overall language and vocabulary, with limited individual receptive-expressive differences. Participants varied in their use of finiteness markers and judgments of morphosyntax.

This study adds longitudinal knowledge on the language abilities of autistic adolescents and young adults. Findings support longitudinal consistency of language assessment outcomes in this age range, such that participants had persistent language impairment.

Acknowledgements

I would like to thank my advisor, Dr. Mabel L. Rice, for her guidance and support throughout graduate school, which extended well beyond graduate advising to all aspects of life as an academic researcher. I would also like to thank Dr. Steven F. Warren for his mentorship on this project and in career development. I would like to thank Dr. Meghan Davidson, Dr. Karrie Shogren, and Dr. Matt Gillispie for their support as members of my dissertation committee. I would like to acknowledge Dr. Nancy Brady, as well as other faculty affiliated with child language, for their support of my research and career development.

This study would not have been possible without the participation and support of participants and their families. I would like to thank the members of the Language Acquisition Studies Laboratory, who provided insight on language acquisition. I especially want to acknowledge Denise Perpich and Dr. Kathleen Kelsey Earnest for their fellowship and advice.

Last, I would like to thank those who have supported me in other ways. Thank you to Erin Andres, Samantha Ghali, and Brittany Williams, for enriching my graduate school experience. Thank you to Steve Politzer-Ahles for suggesting years ago that I contact Mabel about the Child Language Doctoral Program. Thank you to my geriatric pups Charlie and Mickey for keeping me on a strict schedule. Many thanks to Catherine, Zachary, and Camila for supporting me throughout graduate school and beyond.

Table of Contents

List of Figures

List of Tables

Introduction

One aim in clinical research is furthering understanding of the variation in language phenotypes across groups. An objective in autism research is evaluating the centrality of a language delay. These goals interface within the broader context of understanding language acquisition as a dimension of human development. Although early language is predictive of later outcomes in autistic individuals, little is known about the language abilities of autistic adolescents and young adults (Howlin & Magiati, 2017). Even less is known about language in racial and ethnic minority (hereafter, minority), autistic young adults (Interagency Autism Coordinating Committee, 2017). Most autism research, including that to develop diagnostic assessments, has primarily included high SES, Caucasian children (Durkin et al., 2015). This knowledge gap has implications for understanding the experiences and needs of autistic individuals across the lifespan, as well as the phenotypic variability of language in Autism Spectrum Disorder (ASD). To address this gap, the present study reports findings from a longitudinal case study on the language abilities of minority autistic young adults.

An underlying consideration in this report is that the definition of autism has changed over time, such that a diagnosis no longer requires a language delay and may occur with language impairment (LI; American Psychological Association [APA], 2013). Although one of the core characteristics of autism includes difficulties with the use of language for social communication, competing models disagree whether all autistic individuals have difficulties with structural language or whether only those with LI face such difficulties (Boucher, 2012; Eigsti et al., 2011; Lord & Bishop, 2015). This report endorses the definition of the DSM-5, such that a diagnosis of ASD may occur with LI (APA, 2013), as well as identity-first language, which self-advocates support over person-first language (Bottema-Beutel et al., 2020; Howlin, 2021).

General Language Development and Adulthood Outcomes

General language development theories posit that language may be innate or learned and that language may be modular or part of some global ability (Dennis et al., 2009). These theories have implications for interpreting individual differences in language abilities. If language is learned, then children should acquire language primarily as a function of input, which might be indicated by factors such as maternal level of education (MLE) or years of education. Second, if language is modular, then language abilities may dissociate with cognition; if language is part of some global ability, then language should correlate with cognitive ability (Dennis et al., 2009).

Review of the literature on language in autism reveals knowledge gaps. First, it is unknown what generalizations about language development hold for autistic individuals. Second, it is unclear what to make of language in autism due to confounding diagnostic criteria. Addressing these gaps has implications for understanding the experiences of autistic adolescents and young adults. Autistic adults are likely to remain dependent upon others and have limited social activities outside the home (Duncan & Bishop, 2015; Howlin et al., 2004, 2014; Mason et al., 2020; Taylor & Seltzer, 2011). Females, autistic individuals with ID, and autistic individuals from marginalized backgrounds are more likely to experience adverse outcomes (Howlin, 2021; Lord et al., 2015). Previous work has examined childhood predictors of later outcomes (e.g., Howlin et al., 2004), but less is known about the role of adolescence and adulthood predictors in concurrent and later outcomes. One consideration is that autistic individuals constitute a heterogeneous population with variation in measures, informants, and time between initial assessment and follow-up across studies (Howlin, 2021). Altogether, it is unclear how later language abilities relate to other adulthood outcomes (Duncan & Bishop, 2015; Magiati et al., 2014).

Table 1.

Language of Children with a Diagnosis of Autism

Study	Autistic Participant Characteristics				Language Areas Assessed					Design	ASD Outcomes
	N	Age	race/eth.	dx	Rec	Exp	Sp	Voc	Morph		
Ambridge et al. (2020)	15	7;11 (6-9)	N	3	N	Y	N	N	N	ASD, TD peers matched on FSIQ	ASD = TD on passives but showed more reversal errors
Bartolucci et al. (1980)	10	10.85 (*SD* = 2.28)	Y[a]	1	N	Y	N	N	Y	ID, ASD, TD NVMA peers	ASD more likely to omit morphemes in obligatory contexts
Barton-Hulsey & Sterling (2020	22	13.45 (9.42-16.75)	Y[b]	4	N	N	Y	Y	Y	ASD	Difficulty with past-tense irregular, GJ dropped marker & auxiliary DO
Bennett et al. (2008)	64	5.6 (4-6)	N	1	Y	Y	N	N	N	Asperger syndrome, "HFA"	Some had persistent difficulties with expressive and receptive language
Bennett et al. (2014)	330	3.18 (2-4;11)	N	2	Y	Y	N	N	N	ASD, ALI, ASD+ID	ALI resembled late talkers who caught up linguistically
Brynskov et al. (2016)	21	6.7 (5-6)	N	2	Y	Y	N	Y	Y	ASD, ASD+early language delay, TD NVIQ peers	ASD showed difficulties with syntax, morphology, and vocabulary
Condouris et al. (2003)	44	7;3 (4-14;2)	N	1	Y	Y	N	Y	N	ASD	Some had difficulties with overall language and vocabulary
Eigsti & Bennetto (2009)	21	13.4 (10-16)	N	1	N	N	N	Y	Y	ASD & TD age peers	ASD showed difficulties on 3s, present progressive *-ing* GJ & longer sentences
Eigsti et al. (2007)	16	4.7 (3.2-6.6)	Y[c]	1	N	Y	N	Y	Y	ASD, developmental delay, TD NVIQ peers	ASD showed syntactic difficulties (e.g., verb phrases, sentence structures)
Ellis Weismer et al. (2010)	257	2.6 (2-3)	Y[d]	1	Y	Y	N	N	N	ASD, developmental delay	ASD had receptive and expressive language difficulties
Girolamo et al. (2020)	10	15-21	Y[e]	1	Y	Y	Y	N	Y	ASD	All had overall language difficulties; some had ceiling effects on TEGI
Haebig & Sterling (2017)	50	9.3-16.8	N	4	N	N	N	Y	N	ASD, ASD+fragile X syndrome+ID	36% ASD had lower receptive than expressive vocabulary
Howlin (1984)	16	10.52 (*SD* = 2.99)	N	1	N	Y	N	N	Y	ASD	All showed structural language difficulties (e.g., present progressive, cop/aux BE, 3s)
Huang & Finestack (2020)	15	6;3 (4;4-9;8)	Y[a]	4	Y	Y	N	N	Y	ALI, DLD	ALI showed difficulties on expressive syntactic measures
King & Palikara (2018)	15	12.4 (11-14)	N	4	Y	Y	N	Y	N	ASD, TD age peers	Some had structural language issues but typical NVIQ and vocabulary
Kjelgaard & Tager-Flusberg (2	89	7.4 (4-14)	N	1	Y	Y	Y	Y	N	ASD	Some had difficulties with overall language and receptive vocabulary
Kover et al. (2013)	49	7.61 (4-11)	N	2	N	N	N	Y	N	ASD, TD peers	ASD had difficulties with vocabulary (receptive < expressive)
Lindgren et al. (2009)	52	10.35 (7.75-12.95)	N	4	Y	Y	N	Y	N	ASD, ALI, SLI	ALI affected on expressive and receptive language
Loucas et al. (2008)	97	9-14	N	4	Y	Y	N	Y	N	ASD, ALI, SLI	ALI had equal difficulties with both expressive and receptive language
McGregor et al. (2012)	33	11.0 (*SD* = 2.3)	N	4	Y	Y	N	Y	N	ASD, ALI, SLI, TD age peers, TD syntax peers	ALI had difficulties with expressive and receptive lexical ability
Modyanova et al. (2017)	83	4.35-16.3	N	1	Y	Y	Y	Y	Y	ASD, ALI, TD NVMA peers, TD verbal MA peers	ALI had difficulties with tense-marking and sometimes used wrong tense
Paul et al. (2008)	37	1.8 (1.3-2.1)	N	4	Y	Y	N	N	N	ASD	Some had expressive and receptive language difficulties at two year follow-up
Plesa Skwerer et al. (2016)	19	12.5 (5.75-21.1)	N	4	N	N	N	Y	N	MV ASD	Some at floor on receptive vocabulary but showed variability on other methods
Reindal et al. (2021)	148	12.3 (4-18)	N	4	Y	Y	Y	N	N	ASD, non-ASD peers evaluated for ASD	Both groups showed structural language and pragmatic difficulties
Riches et al. (2010)	17	14;8 (14;0-15;4)	N	4	Y	Y	N	N	N	SLI, ALI, TD age+NVIQ peers	ALI had more difficulties on sentence repetition with more syntactically complex sentences
Riches et al. (2011)	16	14;8 (14.4-16.6)	N	4	Y	Y	N	Y	N	ASD, SLI, TD age peers	ALI had relative difficulties with NWR, expressive vocabulary, expressive fluency
Roberts et al. (2004)	62	5-15	N	1	Y	Y	Y	Y	Y	ASD, ASD + borderline LI, ALI	ASD + borderline LI and ALI had vocabulary & morphosyntax difficulties
Schaeffer (2018)	27	10 (5-14)	N	1	Y	Y	N	N	Y	"HFA," SLI, TD age peers	HFA had pragmatic but not grammatical difficulties
Shriberg et al. (2001)	30	21.2 (*SD* = 10.9)	N	1	N	N	Y	N	N	"HFA," Asperger syndrome, TD speakers	HFA and AS had difficulties with articulation, phrasing, stress, and resonance
Sterling (2018)	18	13.4 (*SD* = 2)	Y[f]	4	N	Y	Y	Y	Y	fragile X syndrome+ASD, ASD	ASD had difficulties with TEGI past-tense, DO, & expressive and receptive vocabulary
Tager-Flusberg (2006)	35	10.3 (*SD* = 1.75)	N	1	N	N	N	Y	Y	ASD, ALI, SLI	ALI had difficulties with expressive vocabulary, expressive fluency, and 3s
Tager-Flusberg (2015)	38	10.32 (*SD* = 1.9)	N	1	Y	Y	N	N	N	ASD, ALI, SLI, TD	ALI had difficulties with NWR but not nonword discrimination tasks
Talli (2020)	33	6-12	N	2	N	Y	N	Y	N	ASD, ASD+ID	Majority had overall language difficulties, NVIQ not tied to expressive vocabulary or syntax
Tek et al. (2014)	17	2.7 (*SD* = .1)	N	4	Y	Y	N	N	Y	ASD, TD expressive language peers	Some had persistent expressive language and morphological difficulties
Thomas et al. (2020)	87	5.0 (*SD* = 1.8)	Y[g]	4	Y	Y	N	N	N	ASD, TD peers ages 2-5	ASD had syntactic difficulties on a semistructured language sampling task
Thurm et al. (2007)	59	2.5 (*SD* = .4)	Y[h]	1	Y	Y	N	N	N	ASD, PDD-NOS, developmental disabilities	ASD showed greatest expressive and receptive language difficulties at age 5
Wittke et al. (2017)	51	5.7 (*SD* = 1.1)	N	4	N	Y	N	Y	Y	ASD, ALI+typical vocabulary, ALI+ID	ALI had lower accuracy of morpheme use in obligatory contexts
Woynaroski et al. (2015)	87	3.3 (2-4)	N	1	Y	Y	N	N	N	MV ASD+ID	All had greater receptive than expressive vocabulary difficulties

Note. Race/eth. = race/ethnicity reported. Autism dx = diagnosis before or after DSM-5. Rec = receptive language. Exp = expressive language. Sp = speech. Voc = vocabulary. Morph = morphosyntax. 1 = before, 2 = after, 3 = unknown, 4 = unreported. Y = yes. N = no. TD = typically developing. FSIQ = full scale IQ. ID = Intellectual disability. NV = nonverbal. MA = mental age. "HFA" = "high-functioning autism." LI + lanugage impairment. NVIQ = nonverbal IQ. PPVT = Peabody Picture Vocabulary Test. SES = socioeconomic status, DLD = developmental language disorder (used synonymously with SLI). MV = minimally verbal. 3s = third person singular.

[a] 100% Caucasian/white; [b] 86% white, 9% multiracial, 5% other, 5% Hispanic/Latinx; [c] 88% white, 6% Black/African American, 6% Hispanic/Latinx; [d] 75% white, 22% Black/African American, 2% Asian/Pacific Islander, 1% biracial; [e] 60% Black/African American, 10% white/Caucasian, 10% multiracial, 10% Puerto Rican, 10% don't know; 50% Hispanic/Latinx; [f] 84% white, 3% Black/African American, 8% multiracial, 3% not reported, 3% other; [g] 65% Caucasian, 13% biracial, 8% Asian, 4% African American, 10% other/unknown; [h] 51% Black/African American, 2% Asian American, 2% Hispanic American, 46% white

An Overview of Language in Autism Research

Diagnosis

As Table 1 shows, most studies on language in autism have used a definition prior to the DSM-5 (APA, 2013). In contrast to the DSM-5, which characterizes autism based on social communication impairment and restricted and repetitive behaviors and interests as two factors, the DSM-4 also required a language delay (APA, 2004, 2013). Of the 38 studies in Table 1, 18 (47%) reported using a definition prior to the DSM-5, four (11%) have reported using the DSM-5 definition, one (3%) reported that exact diagnosis was unknown, and 15 (39%) did not report exact diagnostic criteria. Thus, it is unclear what to make of previous diagnostic category outcomes in this area of research.

Further, it is unclear how to identify LI in autistic individuals, especially in in comparison to intellectual disability (ID). LI may be operationally defined by a cutoff of -1 *SD* on standardized language assessments (Tomblin et al., 1996, 1997), with finiteness marking impairment as a clinical marker (Rice & Wexler, 1996). Under the DSM-5, if an intellectual disability (ID) better explains language difficulties, autistic individuals may not have a diagnosis of LI (APA, 2013). However, the criteria for differential diagnosis are unclear. Previous work has used various strategies to identify ID versus LI in autistic individuals, such as a full-scale IQ (FSIQ) cutoff score of <70 (e.g., Bennett et al., 2014) or a nonverbal IQ (NVIQ) cutoff score of <70 (e.g., Bennett et al., 2008). Given that LI may confound performance on verbal IQ (VIQ) assessments due to the language skills necessary (Kjelgaard & Tager-Flusberg, 2001; Rice, 2016; Talli, 2020), considering NVIQ rather than VIQ may be a best practice (Franklin, 2017).

Design

Instability in diagnostic criteria complicates understanding research design. As Table 1 shows, studies have varied in their groups of participants. Some of these groups, such as autism, Asperger syndrome, "high-functioning autism," and developmental delay, may have varied as a function of available diagnoses and their corresponding diagnostic criteria. For example, individuals could previously receive a diagnosis of Asperger syndrome, whereas currently, they would receive a diagnosis of ASD (APA 2004, 2013). Moreover, studies have varied in their identification criteria for LI in autistic individuals. For example, Roberts and colleagues (2004) identified LI in autistic children as 2 *SD* or more below the mean and borderline LI as 1 to 2 *SD* below the mean on receptive vocabulary. In contrast, Huang & Finestack (2020) identified LI in autistic children as 95 or below on a standardized grammatical language measure. A third dimension along which studies vary is sample size and analysis. As Table 1 shows, studies vary widely in their sample size of autistic individuals, which determines the analyses conducted across studies. Altogether, these differences make it unclear how to interpret findings on LI in autistic individuals across studies.

Demographics

In addition to instability of diagnostic criteria, current language in autism research has primarily focused on younger age ranges. Of the 38 studies in Table 1, 16 (42%) explicitly included at least one adolescent (i.e., age 13) or older, and seven (18%) explicitly had a mean participant age of 13 or older. Thus, few studies focused on language in older age ranges. In parallel, nine studies (24%) reported participant race and ethnicity, of which two were 100% white/Caucasian, five were at least 75% white/Caucasian, one was predominantly minorities, and one was 100% racial or ethnic minorities. These demographics suggest there may be little minority or adolescent and young adult representation in this area of research.

Summary

In all, when reading the contemporary literature, it is necessary to consider when children received a diagnosis and under what criteria. It is similarly important to consider knowledge gaps in terms of ages, race, and ethnicity.

Language in Autism Relative to the DSM-5

Children are more likely to receive a diagnosis of ASD alone without LI under the DSM-5 (APA, 2013). While a central point of discussion is the distinction between receptive and expressive language abilities, only more recent work has probed morphosyntactic knowledge.

Receptive and Expressive Language

As Table 1 shows, most studies included measures of receptive language (24 of 38, or 63%) and expressive language (31 of 38, or 82%). Early work found that autistic children, who, by nature of their diagnosis, had a language delay, were likely to omit grammatical morphemes, (Bartolucci et al., 1980; Howlin, 1984). Subsequent work has since documented that language in autism is heterogeneous (Magiati et al., 2014; Wittke et al. 2017). Some autistic children may show language delays in early childhood that disappear later (Bennett et al., 2014; Boucher, 2012). Those with LI perform lower than their peers with typical language on expressive and receptive standardized and non-standardized measures (Huang & Finestack, 2020; Kjelgaard & Tager-Flusberg, 2001; McGregor et al., 2012; Roberts et al., 2004; Tager-Flusberg, 2006). Both autistic individuals with and without LI may face challenges in higher-order expressive and receptive language, especially when semantics and pragmatics are implicated in language tasks (Ambridge et al., 2020; Brynskov et al., 2017; Eigsti et al., 2007; King & Palikara, 2018; Kjelgaard & Tager-Flusberg, 2001; Loucas et al., 2008; Tek et al., 2014).

What is less clear is the nature of receptive-expressive language differences in autistic children. Some autistic children, both with and without LI, have shown stronger expressive language than receptive language (e.g., Ellis Weismer et al., 2001; Kjelgaard & Tager-Flusberg, 2001; McGregor et al., 2012). Others have equivalent receptive and expressive language abilities (e.g., Condouris et al., 2003; Girolamo et al., 2020; Lindgren et al., 2009 ; Loucas et al., 2008), and minimally verbal children with autism have stronger receptive than expressive language (Woynaroski et al., 2015). In addition to variation in diagnostic criteria, the sensitivity of standardized instruments to variability may help explain discrepancies. Comprehensive assessment of receptive language in minimally verbal autistic children revealed heterogeneity across participants, with standardized receptive vocabulary measures underestimating vocabulary relative to parent report and eye-tracking (Plesa Skwerer et al., 2016). Thus, the floor on standardized assessments may collapse variability for autistic individuals with LI (Girolamo et al., 2020). Further, autistic children with typical expressive language in one area (e.g., sentence recall) have shown delays in others (word structure; Brynskov et al., 2017).

The role of child characteristics to overall language is also unclear. First, if pairing with developmental level, age may influence overall language in autistic children (Kwok et al., 2015; Woynaroski et al., 2015). This might explain cases where there are no age effects on receptive-expressive language differences, in that age may not always correspond to developmental level (Kover et al., 2013), as well as cases where receptive-expressive differences decrease over time (e.g., Davidson & Ellis Weismer, 2014). Alternatively, differences in reporting of results, such as selectively reporting findings by removing outliers from the data, may influence discrepancies across studies. Second, autism characteristics may be predicted by concurrent and later language abilities (Bennett et al., 2008; Paul et al., 2008). Autism characteristics have also correlated with receptive and expressive language in young children (Ellis-Weismer et al., 2010). However, some studies have found autism characteristics to not predict language abilities (Kover et al., 2013; Lindgren et al., 2009) and no differences in levels of autism characteristics between autistic individuals with and without LI (e.g., Loucas et al., 2008). Third, IQ measures vary across studies, such that it is difficult to interpret outcomes due to variation in measurement. For example, NVIQ has dissociated with language abilities in autistic children (Kjelgaard & Tager-Flusberg, 2001) and may not predict receptive-expressive language differences (Kwok et al., 2015; Thurm et al., 2007). Yet some studies have found NVIQ to associate with language abilities (Paul et al., 2008; Talli, 2020), as well as FSIQ to predict word definition task but not sentence production scores (McGregor et al., 2012). In all, there is a need for further work to understand the role of child characteristics in language.

Receptive and Expressive Vocabulary

As Table 1 shows, most studies (20 of 38, or 53%) included vocabulary measures. Previous work has used expressive and receptive vocabulary to differentiate autistic individuals with and without LI (e.g., Roberts et al., 2004). One focal point is the receptive-expressive profile of autistic children with and without LI. Some studies have found autistic children to show similar receptive-expressive levels of vocabulary (Barton-Hulsey & Sterling, 2020; Kjelgaard & Tager-Flusberg, 2001; Loucas et al., 2008; Modyanova et al., 2017; Sterling, 2018; Wittke et al., 2017). Other studies have found subsets of autistic children with stronger expressive vocabulary than receptive vocabulary (Haebig & Sterling, 2017; Kover et al., 2013; Woynaroski et al., 2015) and stronger receptive than expressive vocabulary (Haebig & Sterling, 2017). While the reasons for varying profiles are unclear, one consideration is that sensitivity of the instrument may vary by extent of LI. Minimally verbal autistic children have shown particularly low receptive vocabulary in comparison to expressive vocabulary and other receptive language measures (e.g., parent report; Plesa Skwerer et al., 2016; Wittke et al., 2017).

Previous work has found that the relationship of lexical ability to other language outcomes is unclear. Expressive and receptive vocabulary have associated with and predicted grammaticality judgment (GJ) tasks, language sampling outcomes (e.g., IPSyn), nonword repetition, and tense-marking, in autistic children of varying language ability and NVIQ (Condouris et al., 2003; Eigsti & Bennetto, 2009; Eigsti et al., 2007; Roberts et al., 2004; Tager-Flusberg, 2006). Expressive vocabulary may predict receptive vocabulary (Haebig & Sterling, 2017). In turn, receptive vocabulary may predict expressive vocabulary and develop at a slower rate than expressive vocabulary (Haebig & Sterling, 2017; Kover et al., 2013). Given weaker associations between early receptive vocabulary and subsequent expressive vocabulary size than between early expressive vocabulary and subsequent receptive vocabulary size for minimally verbal autistic preschoolers, it may be that receptive vocabulary development is a function of expressive vocabulary size (Woynaroski et al., 2015).

The role of child characteristics in vocabulary is also unclear. Autism characteristics and age have not predicted receptive-expressive vocabulary difference scores or vocabulary outcomes (Haebig & Sterling, 2017; Kover et al., 2013; Lindgren et al. 2009; Loucas et al., 2008). Some studies have found NVIQ to associate with and predict receptive-expressive difference, receptive vocabulary, and expressive vocabulary scores (Haebig & Sterling, 2017; Kover et al., 2013; McGregor et al., 2012; Plesa Skwerer et al., 2016; Talli, 2020). However, some studies have not found NVIQ to predict expressive vocabulary (McGregor et al., 2012). Thus, while the relationship between vocabulary and child characteristics, such as IQ, may be nonlinear (Kwok et al., 2015), reported outcomes do not replicate across studies.

Speech Ability

As Table 1 shows, eight of 38 (21%) studies included measures of speech ability, despite the potential of speech sound disorder (SSD) to interfere in language assessment performance. Autistic adolescents and adults may have near-ceiling intelligibility and only slight differences in speech and prosody relative to non-autistic peers (Shriberg et al., 2001). Autistic individuals with and without LI have shown typical scores on standardized single-word articulation (Kjelgaard & Tager-Flusberg, 2001) and speech assessments (Reindal et al., 2021). Further, autistic children, adolescents and young adults with LI have shown few to no limitations in producing the sounds for tense-marking in English (Barton-Hulsey & Sterling, 2020; Girolamo et al., 2020; Modyanova et al., 2017; Roberts et al., 2004; Sterling, 2018). Thus, while the current literature reporting SSD status suggests SSD may be unlikely to interfere with language assessment performance in autistic individuals, it is also possible that selection bias in language studies may exclude autistic individuals with SSD.

Limitations in verbal working memory may inform performance on speech and language assessment in autistic individuals with LI (Lindgren et al., 2009; Tager-Flusberg, 2015). For example, Autistic individuals with LI have shown lower performance on nonword repetition and verbal working memory tasks than autistic age peers with typical speech and language and comparable performance to peers with SLI (Lindgren et al., 2009; McGregor et al., 2012; Riches et al., 2011; Tager-Flusberg, 2015). On the Syllable Repetition Task (SRT; Shriberg et al., 2009), autistic children had high intelligibility but below-typical competence and encoding scores, as well as variable memory scores (Shriberg & Lohmeier, 2008; Shriberg & Mabie, 2017). In all, it is important to consider the role of SSD relative to working memory in language assessment of autistic individuals.

Individual Differences in the Use of Finiteness Markers and Judgments of Finiteness Errors

As Table 1 shows, 15 of 38 (39%) studies included morphosyntactic measures, most of which (9 of 15, or 60%) used expressive measures (e.g., language sampling). Other studies (6 of 15, or 40%) have probed morphosyntactic knowledge through elicited production of finiteness marking and judgments of finiteness errors. The latter studies are of interest and reported here.

The use of finiteness marking may identify LI in autistic individuals. Roberts and colleagues (2004) found that autistic children with LI were less accurate than autistic peers with typical language on finiteness marking probes and more likely to omit tense markers for third-person singular present and past tenses, as do children with SLI (Rice & Wexler, 1996). Subsequent research replicated this finding and also documented autistic children with LI were likely to omit auxiliary BE, copula BE, and auxiliary DO (Girolamo et al., 2020; Modyanova et al., 2017). Areas of strength for autistic adolescents with and without LI may be third-person singular, auxiliary BE, and copula BE (Barton-Hulsey & Sterling, 2020; Girolamo et al., 2020; Sterling, 2018; Tager-Flusberg, 2006). Difficulties may be past-tense and auxiliary DO (Barton-Hulsey & Sterling, 2020; Girolamo et al., 2020; Modyanova et al., 2017; Roberts et al., 2004; Sterling, 2018).

Judgments of finiteness marking errors may also identify LI in autistic individuals. Autistic children have performed lower than non-autistic age peers on GJ sentences with tense-marking errors that are hallmarks of LI (e.g., third-person singular present and present progressive *-ing*; Eigsti & Bennetto, 2009). However, reduced sensitivity to these errors only occurred in long sentences, such that memory may have affected performance (Eigsti & Bennetto, 2009). Elsewhere, older autistic children and adolescents of varying NVIQ and language abilities have performed below-ceiling on GJ sentences probing omitted tense-marking, agreement errors, and omitted *-ing* (Barton-Hulsey & Sterling, 2020; Girolamo et al., 2020).

Child characteristics may differently influence use of finiteness marking and judgments of finiteness errors. Age has differentiated GJ task performance, such that older autistic children were more sensitive than younger ones on aspect marking, past-tense, and auxiliaries (Eigsti & Bennetto, 2009). Some autistic older children and adolescents with LI have shown mastery of the use of finiteness markers and judgments of finiteness errors, such that the manifestation of LI may change with age (Barton-Hulsey & Sterling, 2020; Girolamo et al., 2020; Tager-Flusberg, 2006). The role of IQ may differ by whether language is included. VIQ has associated with GJ task performance (Eigsti & Bennetto, 2009; Roberts et al., 2004). In contrast, NVIQ has not explained variance in or associated with the use of finiteness markers or GJ tasks in autistic children of varying IQ (Barton-Hulsey & Sterling, 2020; Roberts et al., 2004). In all, considering possible effects of child characteristics is important.

Summary

Overall, while findings on language in autism have highlighted the heterogeneity of language abilities, less is known about grammatical abilities in older ages. This motivates further study of language and morphosyntactic abilities in autistic adolescents and young adults.

Replication of Assessment Performance

Replication Research Design

Replicated research is scarce in the literature, despite a need for generalizable findings for clinical populations. Aims of replicated research include generalizing findings across studies and laying ground for future explanatory and theoretical research (Lindsay & Ehrenberg, 1993; Mackey, 2012). Replication studies may vary by to what extent they differ from the original study. In close replication, nearly all conditions of the original study are very similar in order to establish whether a new finding holds up across studies (Brandt et al., 2014; Lindsay & Ehrenberg, 1993). In differentiated replication, the study varies in major aspects of the original study, with the aims of extending the generalizability of previous findings (Lindsay & Ehrenberg, 1993). Studies may also vary by whether replication is between or within-person; the latter should yield less error variance (Sullivan, 2008). If there is failure to replicate, potential reasons include: (a) the original or current study may have been flawed; (b) the topic is complex; (c) sample sizes are small; (d) the original study was overconfident in statistical results (Amrhein et al., 2019; Brandt, 2014; Lindsay & Ehrenberg, 1993; Makel & Plucker, 2015).

Additional considerations in replication are the validity of assessments and the assessment environment. First, while nonstandardized measures vary widely, standardized assessments have the advantages of using the same procedures across individuals and being familiar and clinically useful (Selin et al., 2019). Potential disadvantages include the testing format, which may be too structured for autistic individuals, previous testing experiences which may affect willingness to complete assessment, and validity of using measures with populations outside the norming sample (McCallum & McCallum, 2017; Thomas et al., 2020). Second, while online assessment may be valid for overall language (Raman et al., 2019; Sutherland et al., 2019; Waite et al., 2010b), vocabulary (Eriks Brophy et al., 2008), speech (Ciccia et al., 2011; Waite et al., 2006), cognitive ability (Hodge et al., 2019), and autism characteristics (Parmanto et al., 2013; Schutte et al., 2015; Smith et al., 2017), it depends on the testing environment (McCallum & McCallum, 2017). Participants must have reliable internet and equipment, with good sound recording quality (Raman et al., 2019; Waite et al., 2010a), be able to navigate online assessment (Iacono et al., 2016), and perceive online assessment as acceptable (Hodge et al., 2019).

Instantiated Cases of Replicated Autism Research

Instantiated cases of replicated autism research come from Howlin (2004) and Miniscalco and Carlsson (2021). In Howlin (1984), the aim was to evaluate morphological acquisition of autistic children. The original study found that autistic children were more likely than non-autistic peers with typical speech and typical language to omit morphemes, including auxiliaries and copulas, past-tense, third-person present-tense, and present progressives (Bartolucci et al., 1980). Limitations of Bartolucci and colleagues (1980) that motivated replication were: a) the original study used frequency counts of morpheme usage errors; b) children were matched to non-autistic peers with typical speech and language based on mental age rather than utterance length; c) differences in morpheme usage of the groups were large enough that they were likely at different places with regard to language acquisition; d) the original study was cross-sectional rather than within-person. Nonetheless, Howlin (1984) maintained these study features for replication. Participants had the same mental age as those of Bartolucci and colleagues (1980) but were two and a half years younger ($M_{age} = 7.9$), completed different language assessments, and generated a mean of 114 utterances instead of 197. Findings showed that there was a significant correlation between the percentage of different morphemes used and the level of internal consistency with those reported by Bartolucci and colleagues (1980).

Miniscalco and Carlsson (2021) implemented a longitudinal case study to track language development in six autistic children from ages 3 to 8 years. Participants included three boys and three girls, one from each subgroup: autism, autism plus LI, and autism plus LI and ID. Outcomes were receptive language, receptive vocabulary, expressive language, speech sound production, nonword repetition, narrative skills, communicative functioning, NVIQ, autism characteristics, and adaptive behavior, although some of the measures varied across timepoints. Analysis revealed that: a) participants had spiky assessment profiles; b) many had persistent linguistic difficulties, and; c) the subgrouping remained stable over time, which is perhaps indicative of the persistence of language difficulties at young ages. In all, this within-person longitudinal study highlighted the importance of thorough assessment of speech and language abilities, as well as the need for further research on language development in older age ranges.

The Current Study

Taking together what is known about language in autism, the current study investigated language abilities in minority autistic adolescents and young adults. Research questions were:

1. Do participants change in relative levels of performance within the group, as a type of test-retest reliability?
2. Does assessment performance differ between in-person and online assessment?
3. Does speech sound disorder interfere with language assessment performance?
4. Is there individual variation in their overall oral language performance, including individual differences in expressive and overall language?
5. Are there individual differences in expressive and receptive vocabulary?
6. Are there individual differences in use of or judgments of finiteness errors?

Hypotheses

Given what is known about language abilities in autistic individuals and findings from Girolamo et al. (2020), the hypotheses are:

1. Given that participants are well beyond the age range of dynamic change in language acquisition, participants will not change in within-group relative levels of performance.
2. Online assessment will be comparable to in-person assessment performance.
3. Speech sound disorder will not interfere with language assessment performance as long as there is suitable sound recording quality.
4. Overall language performance will be below age expectations, with limited individual

differences in expressive and receptive overall language.

5. Participants will show limited individual differences in their expressive and receptive vocabulary.
6. Participants will show significant individual variation in use of finiteness markers and judgment of finiteness errors, with some performing at or near ceiling.

Method

Changes to the Original Study

This study replicated and extended Girolamo et al. (2020), which examined potentially useful measures for minority autistic adolescents and young adults. Key differences are:

1. The present study included repeated measures data for a total of three timepoints (i.e., T1, T2 and T3). The original study included only T1.
2. Recruitment at T2 took place in-person and recruitment at T3 took place remotely. Recruitment for the original study, or T1, took place in-person.
3. Because the study at T3 focused on comprehensive language assessment, the protocol included assessments on overall language, grammar, vocabulary, NVIQ, autism traits, and adaptive behavior. T1 and T2 were feasibility studies and included assessments on overall language, grammar, NVIQ, and working memory.
4. Data collection at T3 took place online due to the COVID-19 outbreak. Data collection at T1 and at T2 took place in person in community settings.

Participants

Inclusion and Exclusion Criteria

To be in this study, participants had to: (a) be a racial or ethnic minority, (b) have a diagnosis of ASD, (c) be of ages 14 years to 25 years, (d) be native speakers of mainstream American English, and (e) have received their education in a 100% self-contained, specialized setting at the time of recruitment.

Participants who did not: (a) have normal hearing, (b) use verbal language, (c) have normal or corrected-to-normal vision, or (d) have the ability to complete testing activities, as determined by consultation with participants and their caregivers prior to enrolling in the study, were excluded from the study. In the case that potential participants who were not minorities were interested in the study, the plan was to assess them for the purpose of including them in the broader research program but exclude them from the present analysis.

Participant Characteristics

All participants were a racial and/or an ethnic minority. At T1, 70% were Black/African American, 10% multiracial, and 50% Hispanic/Latinx. At T2, 57% were Black/African American, 14% multiracial, and 71% Hispanic/Latinx. At T3, 73% were Black/African American, 9% multiracial, and 45% Hispanic/Latinx. All participants were male at T1 and T2, and one of 11 (9.1%) participants was female at T3. As Table 2 shows, maternal level of education was variable, with about half having a high school degree as the highest education level completed. At T3, all participants had considerable experience in the public education system. Four received their education in specialized classes, one had moved into a GED class since T1, and six had exited public education. One of the six was enrolled in community college, and five did not receive educational programming. Given their ages and time of diagnosis, participants received a pre-DSM-5 diagnosis. On average, participants had a below-typical NVIQ (*M* = 100, *SD* = 15) and moderate autism characteristics (i.e., *t*-score = 66-75).

Table 2.

Participant Demographics

ID	MLE	Ed Yrs.	NVIQ SS	SRS-2 total	CA		
					T1	T2	T3
1	N/A				21.1		
2	HS	15.92	93	61	16.6	18.45	18.9
3	HS	18.79	104	45	20.1	21.91	22.5
4	MA	19.5	100	73	19.4	21.29	21.9
5	BA	12.64	75	64	15.3	17.08	17.6
6	HS	14.28	73	84	15.9	17.73	18.3
7	AS	14.49	52	62	17.2	18.93	19.5
8	BA	16.07	93	63	18.9	20.69	21.3
9	MA	17.82	75	62	21.3		23.6
10	N/A				17.3		
11	HS	11.41	72	77			17.4

12	HS	14.25	75	66			20.3
13	HS	19.38	63	67			21.6
M		15.87	79.55	65.82	18.31	19.44	20.3
SD		2.62	15.29	9.56	2.04	1.72	1.98
max.		19.5	104	84	21.3	21.91	23.6

Note. MLE = mother's level of education. Ed. Yrs. = years of education. N/A = not available. HS = high school. MA = master's. BA = bachelor's. AS = associate's degree or some college. Black/AA = Black/African American. Multi = multiracial. Ed. Yrs. = years of education in the public education system. NVIQ SS = Raven's 2 standard score (*Raven's Progressive Matrices–Second Edition: Manual,* 2018). SRS-2 total = Social Responsivity Scale-Second Edition total *t*-score (Constantino & Gruber, 2012). CA = chronological age. T1 = time 1. T2 = time 2. T3 = time 3. Blank space = participant not seen at timepoint.

Procedures

Ethics

This study received institutional review board approval and followed all ethical guidelines.

Sampling and Participant Selection

T1. As Figure 1 shows, the author recruited participants in-person using a participant-centered approach (Girolamo et al., 2020). Central to this approach was making the completion of study activities accessible to participants and their families by shifting the burden of completing study activities from participants and their families to the first author. A brief summary of procedures follows.

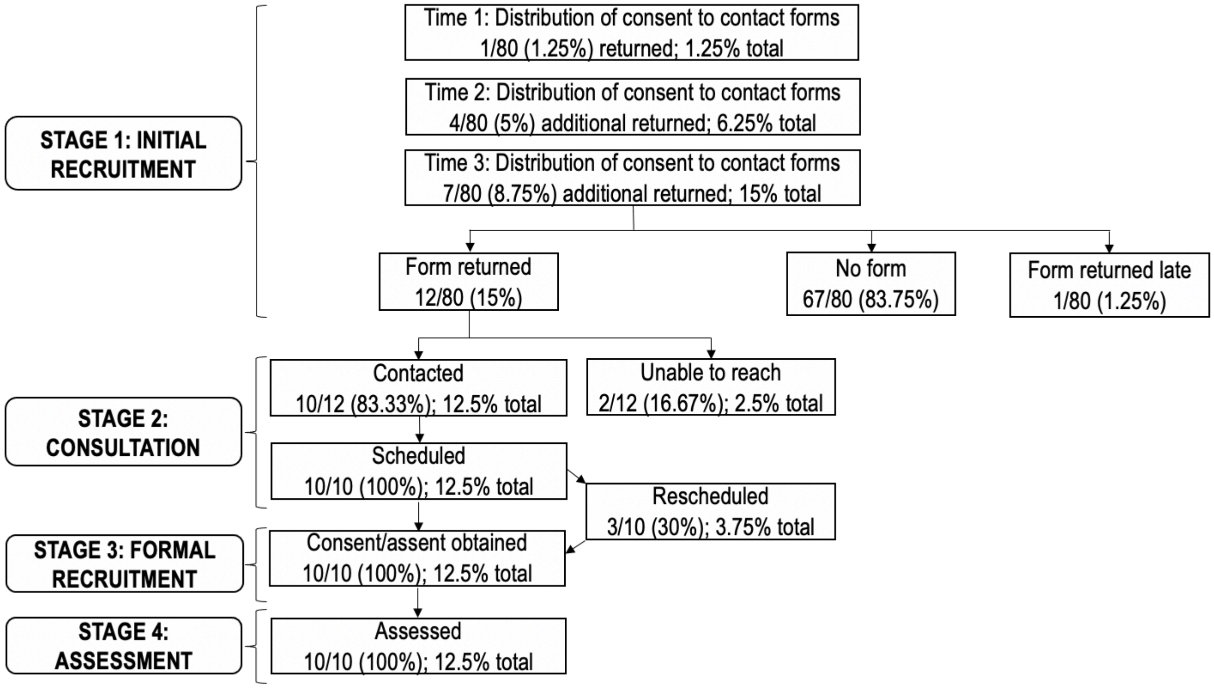

Figure 1. Participant-centered approach implemented at T1

Initial recruitment took place in partnership with a community organization in a major urban area in the Northeast which provided specialized support to adolescents and young adults with a diagnosis of ASD. Most were racial/ethnic minorities who qualified for Title I funding, spoke English as a first language, and were exempt from state standardized testing. They received their education in a special education setting, with 6, 8, or 12 students per class. To develop this partnership, the first author, who was a former teacher, traveled from Kansas to meet with organizational leaders and staff. Over the course of six meetings from 2015 to 2018, the parties discussed broader study aims and the logistics of initial recruitment, which consisted of the author distributing and collecting consent-to-contact forms to and from potential participants. Visits took place early in the morning or late in the afternoon to minimize disruptions to programming. All recruitment materials were written in English at a fourth-grade literacy level. The sample was self-selecting, as the expectation was potential participants would bring forms home to their caregivers, discuss their interest in participating, and bring the form back to the organization for pick-up or contact the author.

In consultation, the author contacted each participant by phone or email to provide information about the study. The author did not assume families had prior research experience or knowledge about research studies and provided a jargon-free overview of the study and a step-by-step explanation of what participation entailed. In addition, the author encouraged families to ask questions and voice concerns. Only after families expressed comfort with proceeding did they schedule a time and place for informed consent and assent and assessment. Families selected a time and place convenient to them outside of school, which generally meant providing informed consent and assent, as well as completing assessment, on nights and weekends in a community location, such as a public library.

T2 and T3. At T2 and T3, the sample was self-selecting. At T2, the author recruited participants by contacting previous participants. At T3, the author recruited by re-recruiting participants, by contacting an individual who had returned a consent-to-contact form but who had not formally enrolled in the study, and by receiving referrals from previous participants. Participants provided assent, and caregivers provided informed consent. Participants received a $20 gift card and their caregiver $40 upon completing the protocol.

Sample Size, Power, and Precision

T1. In the original study (Girolamo et al., 2020), the primary consideration was determining the efficacy of individualized assessment for minority autistic adolescents and young adults. The determination of sample size did not include power analysis. The target and achieved sample size were each 10 participants.

T2. The primary considerations of T2 were to carry out a replicated case study and to continue building trust and rapport with participants. The determination of sample size did not include power analysis. The target sample size was 10 participants, and specifically the same 10 participants from T1, for a replication of T1 (Girolamo et al., 2020) and to help build a sustainable relationship with the participant community. As Table 2 shows, the achieved sample size was seven participants.

T3. At T3, the primary considerations were the feasibility of recruiting participants online during the COVID-19 outbreak and the collection of data for a longitudinal case study, with a focus on comprehensive language assessment. The determination of sample size did not include power analysis. The target sample size was a range of 10 participants, which would allow for a replication of Girolamo et al. (2020) and T2, to 20 participants, which would allow for more rigorous statistical analysis. The plan was to stop after collecting data from 20 participants or after recruitment halted (i.e., no new recruitment after contacting recruitment sources three times over two months). As Table 2 shows, the achieved sample size was 11 participants.

Data Collection

T1. The author administered direct behavioral assessments in-person to participants in community settings. Caregivers sat next to or in the vicinity of participants. Assessment required about 60 minutes. Conducting assessment in familiar settings maximized the likelihood of valid assessment outcomes, in that research activities took place in a socially appropriate and accessible manner (Girolamo et al., 2020).

T2 and T3. At T2, data collection took place in-person and used the same procedures as in T1. At T3, data collection took place online on a Zoom platform. Participants and their caregivers used a home tablet or computer to complete research activities, and the author used a university-issued desktop paired with a webcam and external audio recorder in their office to administer research activities. Participants completed direct behavioral assessments, and their caregivers completed rating scales. Prior to data collection, caregivers helped their child join Zoom, and the author coached some parents through how to download and use Zoom. Once on Zoom, the author informed participants and their families of when recording would begin and end, what to do if an uninvited person joined the meeting, and checked for privacy on each end of the call. Next, the author shared their screen with participants on Zoom. Immediately following assessment, the author administered rating scales for autism characteristics to caregivers. Participation compensation followed the procedures of T1 and T2. On average, the protocol required two to two and a half hours. All participants completed the protocol in a single session without any demonstrable stress or need for breaks other than to use the bathroom.

Quality of Measurements

The author was the only data collector and used the same procedures to enhance the quality of measurements for all timepoints. The author completed a rigorous training protocol as a Graduate Research Assistant in the LAS Laboratory (PI: Mabel L. Rice), an NIH-funded laboratory which has consistently received funding for over 25 years and conducted a longitudinal pedigree study on specific language impairment. This training included how to successfully implement a participant-centered approach to carry out assessment in community settings and how to administer a behavioral assessment protocol that required two and a half to four hours. Training required passing reliability checks as a data collector for each of the following steps and demonstrating the ability to administer assessments in a standardized manner following manual instructions to a variety of participants.

Specific training procedures are as follows. First, the author read each assessment manual and had to demonstrate knowledge of how to administer assessments using standardized instructions from the manual. Second, the author practiced administering each assessment to their lab mate. Third, the author practiced administering each assessment to seasoned examiners in the lab, some of whom had worked as full-time examiners for years. Fourth, the author administered assessments to a practice child, or a child who was not a participant and completed assessments explicitly to help train new examiners, in an observation room of the Schiefelbusch Speech, Language, and Hearing Clinic at the University of Kansas. During this time, expert examiners and the PI observed the author administer assessments through a mirrored wall and provided feedback. Additionally, all practice sessions were audio- and video-taped. Fifth, the author accompanied a seasoned examiner to observe them administer the assessment protocol to an actual participant. Sixth, the author administered the assessment protocol to an actual participant while a seasoned examiner observed the data collection session. Furthermore, as all data collection sessions in the LAS Lab were, this trial run was audio- and video-taped. Only after passing all of these steps was the author allowed to independently collect data; in addition, the author had to pass ongoing checks for reliability. Thus, while there were not multiple observers, the training process was extensive and continuous.

Measures

Instrumentation

Table 3 displays an overview of the materials used at the three time points. At T1, the primary considerations were administering tests the author had trained on in an order that kept participants engaged and avoided possible negative order effects that could create frustration or fatigue (Girolamo et al., 2020). The order was: (a) CELF-3 (Semel et al., 1995); (b) TEGI (Rice & Wexler, 2001); (c) Columbia Mental Maturity Scale (Burgemeister et al., 1972); (d) Wechsler Intelligence Scale for Children-Third Edition Digit Span (WISC-III D-Span; Wechsler, 1991).

Table 3.

Outcome Measures and Materials Used at Times 1,2, and 3

Outcome measures	Test materials	T1	T2	T3
Expressive language	Clinical Evaluation of Language Fundamentals—Third Edition	X	X	
	Clinical Evaluation of Language Fundamentals—Fifth Edition			X
Receptive language	Clinical Evaluation of Language Fundamentals—Third Edition	X	X	
	Clinical Evaluation of Language Fundamentals—Fifth Edition			X
Expressive vocabulary	Expressive Vocabulary Test—Third Edition			X
Receptive vocabulary	Peabody Picture Vocabulary Test—Fifth Edition			X
Sentence repetition	Clinical Evaluation of Language Fundamentals—Third Edition	X	X	
	Clinical Evaluation of Language Fundamentals—Fifth Edition			X
Non-word repetition	Syllable Repetition Task			X
Speech sound ability	Test of Early Grammatical Impairment	X	X	X
Nonverbal intelligence	Columbia Mental Maturity Scale	X	X	
	Raven's Progressive Matrices—Second Edition			X
Autism characteristics	Social Responsiveness Scale—Second Edition			X
Working memory	Wechsler Intelligence Scales for Children Digit Span—Third Edition	X	X	

T2 and T3. The order of assessments at T2 was the same order as in T1. At T3, in addition to the need to consider assessment order, it was also important to minimize switching between presentation formats, or between digitized assessments on a screen and live-action items. Thus, the order of assessments at T3 was: (a) Syllable Repetition Task (SRT; Shriberg et al., 2009); (b) Clinical Evaluation of Language Fundamentals-Fifth Edition (CELF-5; Wiig et al., 2013); (c) Peabody Picture Vocabulary Test-Fifth Edition (PPVT-5; Dunn, 2019); (d) Expressive Vocabulary Test-Third Edition (EVT-3; Williams, 2019); (e) Raven's Progressive Matrices-Second Edition (Raven's 2; *Raven's Progressive Matrices-Second Edition: Manual*, 2018); (f) Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001); and (g) Social Responsiveness Scale-Second Edition (SRS-2; Constantino & Gruber, 2012).

Outcomes

Primary outcomes were: (a) SRT percentage of consonants correct (PCC), (b) CELF core language standard score, (c) CELF expressive language standard score; (d) CELF receptive language standard score; (e) PPVT-5 standard score, (d) EVT-3 standard score, (e) TEGI third-person singular (3s) probe score, (f) TEGI third-person singular past-tense (past) probe score, (g)

TEGI Be/Do probe score, (h) TEGI A' score for Dropped marker, (i) TEGI A' score for Agreement, and (j) TEGI A' score for Dropped *-ing*. Secondary measures were the (a) Raven's 2 standard score, (b) the SRS-2 total *t* score (i.e., the standard score provided in the manual).

Psychometrics

All measures are standardized with precedent in the literature as informative and reliable for autistic individuals. All are untimed and provide an individual outcome relative to age expectations.

SRT. The SRT (Shriberg et al., 2009) is a nonword repetition task where speakers listen to and repeat a nonword of 2, 3, or 4 syllables. Each nonword is comprised of a combination of 4 consonants (i.e. /b/, /d/, /m/, /n/) and one vowel (i.e., /a/), which are acquired early. In this way, the SRT eliminates confounds that may arise if a speaker does not have the target speech sounds in their phonetic inventory or has misarticulation issues. The SRT has acceptable to good internal reliability coefficients for children with typical speech (.71-.88), good to excellent coefficients for children with speech disorders (.83-.92), and high interrater reliability for transcription (.88; Shriberg & Lohmeier, 2008). Six-year-old autistic children with a FSIQ of at least 70 had a lower competence (i.e., percentage consonants correct [PCC]; $M = 92.0$, $SD = 5.9$), encoding score ($M = 64.5$, $SD = 30.1$), memory ($M = 89.8$, $SD = 15.4$), and transcoding scores ($M = 93.4$, $SD = 8.2$; Shriberg & Mabie, 2017) than non-autistic peers with typical speech and language.

CELF. The CELF (Semel et al., 1995; Wiig et al., 2013) is an overall oral language measure for ages 5 years to 21;11. The CELF-3 is composed of six untimed subtests, and the CELF-5 is composed of seven untimed subtests, respectively. Each subtest raw score can be transformed into a standard score ($M = 10$, $SD = 3$). Some stimuli may be repeated. Subtest standard scores can be transformed into composite scores for expressive language, receptive language, and a total language score (*M* = 100, *SD* = 15). The CELF-3 has a floor of 50, and the CELF-5 has a floor of 40. Examinees do tasks such as looking at pictures and talking about them, responding to questions, and repeating words and sentences. Both versions are available in a physical format, and the CELF-5 is also available in a digital format.

Because this study used the most recent version of the test, this section reports the norming sample for the CELF-5. The CELF-5 norming sample included over 3,000 English speakers representative of the U.S. population ages 5 years to 21 years in terms of age, sex, race/ethnicity, geographic region, and parent/caregiver education level. Performance varied little across adolescence and young adulthood, with good to excellent reliability for the 17;00-21;22 group (*r* = .82-.95). Test-retest reliability for the oldest age band in the norming sample (i.e., 12;00-16;11) was variable: (a) excellent for Word Classes (.90) and the Expressive Language Index (.91), (b) good for the Core Language Score (.88), Recalling Sentences (.87), Receptive Language Index (.86), and Sentence Assembly (.84), and (c) acceptable for Understanding Spoken Paragraphs (.79), Semantic Relationships (.72) and Formulated Sentences (.69). The CELF-5 included a sample of autistic individuals ($n = 69$, $M_{age} = 10.4$, $SD_{age} = 3.5$), ranging in age from 5;0-21;11 and with an IQ over 60. The autistic sample performed lower on all subtests and index scores than non-autistic peers matched on age, race/ethnicity, sex, and parent education level. Subtest reliability coefficients were excellent and ranged from .91 to .97.

Previous work has employed the CELF to assess language abilities in autistic children with and without LI (Bennett et al., 2014; Kjelgaard & Tager-Flusberg, 2001; Riches et al., 2010), in comparison to autistic children with ID (Bennett et al., 2014) and children with SLI (Schaeffer, 2018), and in minority autistic adolescents and young adults (Girolamo et al., 2020). CELF scores have correlated with measures of spontaneous speech in autistic children and may measure the same underlying linguistic constructs (Condouris et al., 2003).

PPVT-5. The PPVT-5 (Dunn, 2019) is a receptive vocabulary assessment. It is designed for ages 2;6 years (years; months) to over 90 years. Raw scores can be translated into a standard score (*M* = 100, *SD* = 15). Examinees look at a set of pictures and point at or verbally indicate which picture corresponds to a one-word verbal stimulus. The PPVT-5 is available in a physical format or a digital format that can be presented on a screen.

The norming sample included 2,720 English speakers representative of the U.S. population ages 2;6 to over 90 in terms of age, sex, race/ethnicity, geographic region, and parent/caregiver education level. Reliability was excellent for ages 14;0 to 24;11 (r = .95-.97). Test stability reliability for ages 12;0-24;11 was good (r = .86). Autistic individuals with language disorder (n = 118) comprised 0.7% of the norming sample (M_{age} = 11.0, *SD* = 4.0), 11.9% of whom were African American, 8.5% Hispanic, 10.2% Hispanic, and 11.9% female. None had ID or neurological conditions. Performance on the PPVT-5 was significantly lower than non-autistic peers matched for age, race/ethnicity, parent education level, and sex. PPVT scores have differentiated autistic children with LI and autistic children without LI (Condouris et al., 2003; Kjelgaard & Tager-Flusberg, 2001), as well as correlated with CELF total, EVT, and nonword repetition scores (Condouris et al., 2003).

EVT-3. The EVT-3 (Williams, 2019) is an expressive vocabulary assessment for individuals ages 2;6 years to over 90 years. Raw scores can be translated into a standard score (*M* = 100, *SD* = 15). Examinees look at a picture and provide a one-word label or synonym in response to a one-word verbal stimulus. The EVT-3 is available in a physical and a digital format via Q-Global that can be presented on a screen.

The demographics of the norming sample are the same as the demographics of the PPVT-

5. Reliability was excellent for the total norming sample (.97) and age ranges 14;0 to 24;11 (range = .95-.97). For ages 12;0-24;11, test-retest reliability was excellent (.93). Autistic individuals performed lower than their non-autistic peers matched for age, race/ethnicity, parent education level, and sex. EVT scores have differentiated autistic children with and without LI (Condouris et al., 2003; Kjelgaard & Tager-Flusberg, 2001), varied together with FSIQ in autistic children with and without LI (Kjelgaard & Tager-Flusberg, 2001), and correlated with CELF total and PPVT-III standard scores (Condouris et al., 2003).

Raven's 2. The Raven's 2 (*Raven's Progressive Matrices–Second Edition: Manual*, 2018) is a nonverbal test of general cognitive ability for ages 4 years to 90 years. It uses minimal verbal instruction and is untimed, making it more widely accessible (Franklin, 2017; Nader et al., 2016). In contrast to measures of NVIQ, using verbal IQ tests or IQ tests requiring more verbal language abilities may confound performance in autistic children, especially those with below-typical language (Grondhuis et al., 2018). The Raven's 2 is available in both a physical and a digital format with options for a short and long form. In this assessment, participants look at a set of images and select an image to complete the set. Because the digital version uses item response theory, the digital Raven's 2 produces only ability scores and standard scores (M = 100, SD = 15).

The norming sample included 2,275 individuals representative of the U.S. population ages 4;0 to 90;11 in terms of age, education, race/ethnicity, geographic region, and gender. Autistic individuals comprised 0.2% of the total sample. Reliability for the digital long form was good (r =.89) and the same as that of the paper form for ages 17;0-19;11 and 20;0-24;11. Test-retest reliability of the digital long form for ages 17-54 was good and the same as that of the paper form (.89).

TEGI. The TEGI (Rice & Wexler, 2001) provides information on grammatical abilities. It is designed for children in an age range of dynamic change in accuracy of morphosyntax beginning at 3 years and likely to show mastery by 8;11. It includes a phonological probe to demonstrate that examinees are able to produce the speech sounds marking tense in English, probes on third-person singular present tense, third-person singular past tense, and Be/Do third-person forms, which can be summarized into an elicited grammar composite score and a screener score. The TEGI also includes a GJ subtest, which examines dropped marker, agreement errors, and dropped progressive *-ing*. Examinees name pictures, look at pictures and talk about them, respond to questions or stimuli, and listen to sentences. On the BE/DO probe, examinees respond to stimuli by talking to an inanimate object (i.e., a puppet). Adultlike performance is about 95%.

The norming sample included 837 speakers of mainstream American English with adequate hearing and vision, the ability to complete a standardized test, and with the ability to produce sounds that mark finiteness in English. Children with typical language (n = 393, ages 3;0 to 6;11) were representative of the U.S. population in terms of age, gender, race/ethnicity, geographic region, and parent education level. Children with a language disorder (n = 444, ages 3;0 to 8;11) were stratified by age only. Test-retest reliability was good to excellent for all probes (range = .82-.92), excellent for the elicited grammar composite (.95), and variable for the GJ task. Reliability was lower for 4;00 to 4;05 age group (range = .37-.43) than the 4;06 to 4;11 age group (range = .65-.82), and both age groups were younger than the expected age of mastery.

Previous work has employed the TEGI to investigate interindividual variation in autistic children with and without LI (Barton-Hulsey & Sterling, 2020; Modyanova et al., 2017; Roberts et al., 2004; Sterling, 2018; Sterling et al., 2012). The TEGI has been sensitive to interindividual variation in autistic children, adolescents and young adults of varying NVIQ from near floor to the typical range (Barton-Hulsey & Sterling, 2020; Girolamo et al., 2020). Given that the TEGI compares morphosyntactic performance relative to the adult grammar and that tense marking is a clinical marker of LI (Rice & Wexler, 1996), it is an informative measure.

SRS-2. The SRS-2 (Constantino & Gruber, 2012) is a measure of social impairment relevant to ASD for ages 2;6 years through adulthood. The respondent rates 65 items on behaviors to indicate severity level. Raw scores are transformed into a total *t* score (i.e., a standard score), as well as subtest *t* scores for social awareness, social cognition, social communication, social motivation, and RRBI. Each *t* score corresponds to a level: within normal limits, mild, moderate, and severe. The SRS-2 is available in both a physical and a digital format.

The norming sample included 2,025 school-age reports on 1,014 children ages 4 to 18, as well as 2,210 adult reports on 702 adults ages 18 to 89. Both groups were representative of the U.S. population in terms of gender, race/ethnicity, parent educational level, and geographic region. Both reports had very strong internal consistency ($M = .95$). The clinical sample included 4,891 individuals with a diagnosis of ASD or of a related condition of ages 4 to 18, 2.7% of whom were African American, 4.5% Hispanic or Latino, and 4.7% of multiple racial/ethnic backgrounds. Internal consistency was very strong (.95), with higher scores for the clinical sample than non-autistic peers.

Previous work has found the SRS-2 to have high diagnostic agreement with the CARS-2 (Schopler et al., 2010) yet to measure different constructs as the CARS-2 (Chen et al., 2018). In addition, the SRS may be valid for adults, in that SRS factors strongly associated with autism characteristics and behavioral measures and in that SRS factors differently related to dimensions of social or behavioral domains (i.e., domains of autism characteristics; Chan et al., 2017).

Analysis

Data Diagnostics

All scorable responses were included in the analysis at T1, T2, and T3. For example, if a participant completed a subtest of the TEGI but had unscorable responses, their responses would not be included in analyses other than descriptive analysis to demonstrate the proportion of scorable, unscorable, and other types of responses. Because the author administered assessments to participants, there were no criteria for deciding when to infer missing data.

Data Transformations

Data transformations allowed for the comparison outcomes across assessments. Any missing data was due to participants providing unscorable responses, or on the TEGI in rare instances, no response. Such instances were left as missing data. All scoring followed the standardized instructions and used the standard scores provided in the assessment manuals.

SRT. The author used raw scores to calculate a competency score, an encoding score, a memory score, and a transcoding score. A competency score was the total percentage of consonants correct (PCC). An encoding score was the number of within-class substitutions, where class referred to the number of syllables (i.e., classes of 2, 3, or 4 syllables), divided by the number of substitution errors; this total was multiplied by 100. To calculate the memory score, the author completed the following calculation: 100*(1+(LN(3 syllable PCC/2 syllable PCC))). Last, the transcoding score was the difference between 100 and the percentage of total items with additions. Analysis compared scores from participants with those of 17-year-olds with typical speech and typical language in the norming sample (PCC: M = 96.4, SD = 3.9, encoding: M = 62.5, SD = 30.6; memory 3/2: M = 97.1, SD =3.0; transcoding: M = 97.8, SD = 2.9; Lohmeier & Shriberg, 2011).

CELF-5. Subtest raw scores were transformed into standard scores, including composite standard scores for the Expressive Language Index (ELI), Receptive Language Index (RLI), and Core Language Score (CLS). Individual difference scores were also calculated by subtracting the RLI from the ELI following instructions in the assessment manual.

PPVT-5 and EVT-3. Raw scores were transformed into standard scores. To calculate individual difference scores, each PPVT-5 standard score was subtracted from EVT-3 standard scores (Williams, 2019). Positive values marked a stronger expressive than receptive vocabulary, and negative values marked a stronger receptive than expressive vocabulary.

Raven's 2. Because the online version of this assessment uses item-response theory, each participant saw a different set of items. Thus, only standard scores were available.

TEGI. The percentages correct for the third-person singular (3s), past tense probe (past), BE, and DO probe were calculated following tables from the assessment manual. Raw scores from the GJ task were transformed into three A' scores for Dropped Marker, Agreement, and Dropped *-ing* following tables from the assessment manual.

SRS-2. Raw scores for subtest scales (i.e., social awareness, social cognition, social communication, social motivation, restricted interests and repetitive behavior) were transformed into *t* scores and a composite, total *t* score. As the standard scores provided in the assessment manual, *t* scores corresponded to severity level: within typical limits, mild, moderate, and severe.

Analytic Strategies

Analyses included descriptives, including Shapiro-Wilk tests (Shapiro & Wilk, 1965) and nonparametric measures, including Freidman tests (Friedman, 1937) and Spearman's rank-order correlations. Interpretation of effect sizes was as follows: very strong (.8-.99), strong (.6-.79), and moderate (.4-.59). In addition, analyses used an a priori significance level of $p < .05$. A description of analyses per hypothesis follows.

RQ1 & RQ2. To test whether participants changed in relative levels of performance within the group across timepoints and assessment methods, analysis included descriptives and inspection of outcomes across timepoints and methods. To compare CELF and TEGI scores across T1, T2, and T3, analysis included descriptives and a Friedman test (1937).

RQ3. To test whether speech sound disorder interfered with language assessment performance, analysis included inspection of outcomes on the TEGI phonological probe and the SRT. If the participant failed the TEGI phonological probe, the TEGI is invalid.

RQ4. To test whether participants had below typical overall language performance with limited individual differences in expressive and receptive language, analysis included descriptives of group and individual outcomes for the CELF-5 total language, expressive language, receptive language, and receptive-expressive language difference scores, as well as correlations with demographic and cognitive variables.

RQ5. To test whether participants had below typical vocabulary with limited individual differences in expressive and receptive vocabulary, analysis included descriptives of group and individual EVT-3 standard score, PPVT-5 standard score, and receptive-expressive vocabulary difference score, as well as correlations with demographic and cognitive variables.

RQ6. Assuming participants passed the TEGI phonological probe, to test whether participants showed individual variation in use of finiteness markers and judgment of finiteness errors, analysis included descriptives of group and individual outcomes on the TEGI 3s, past, BE, and DO, probes and on GJ A' scores for Dropped Marker, Agreement, and Dropped *-ing*. Analysis also included correlations of TEGI outcomes with demographic and cognitive variables.

Results

Participant Flow

At T2, re-recruiting participants from T1 resulted in seven participants who completed assessment; two were unavailable during the time of data collection. At T3, recruitment resulted in 11 participants, eight of whom were previous participants. Of the eight previous participants at T3, seven had completed assessment at all three timepoints, and eight had completed assessment at two timepoints (i.e., T1 and T3). Of the new participants at T3, two were participant referrals and one had previously turned in a consent-to-contact form but not participated in the study. Attrition occurred for unknown reasons for four participants from T1.

Recruitment

Recruitment and data collection for T1, T2, and T3 took place from 2018 to 2020. There was approximately one year between timepoints.

Missing Data

There were no missing data for the timepoints of this study (i.e., T2 or T3). However, at T2, one participant was unavailable for assessment, and thus, only completed assessment at T1 and T3.

An Overview of Assessment Profiles

As Figure 2 shows, on average, at T3, participants performed below age expectations on language and NVIQ assessments. All participants had expressive and receptive language scores below age expectations. Participants 3 and 6 were relatively high performers in the group. Yet while participant 3 had typical expressive and receptive vocabulary, as well as typical NVIQ, participant 6 had below-typical vocabulary and NVIQ scores. Participants 2 and 4 had ceiling effects on most TEGI subtests and typical NVIQ but scored below age expectations on other language tasks. Participant 13 had a ceiling effect only on the TEGI GJ tasks (i.e., TEGI DM, TEGI -*ing*, and TEGI AGR) and an NVIQ < 70. Participants 5, 7, 8, 9, 11, and 12 performed at below adultlike levels on the TEGI, but participant 5 had typical receptive vocabulary, and participant 8 had a typical NVIQ. Participants 7, 8, 9, 11, 12, and 13 had assessment scores below age expectations, with participant 5 having an NVIQ < 70. Thus, while language abilities were typically lower than age expectations, there was heterogeneity in outcomes.

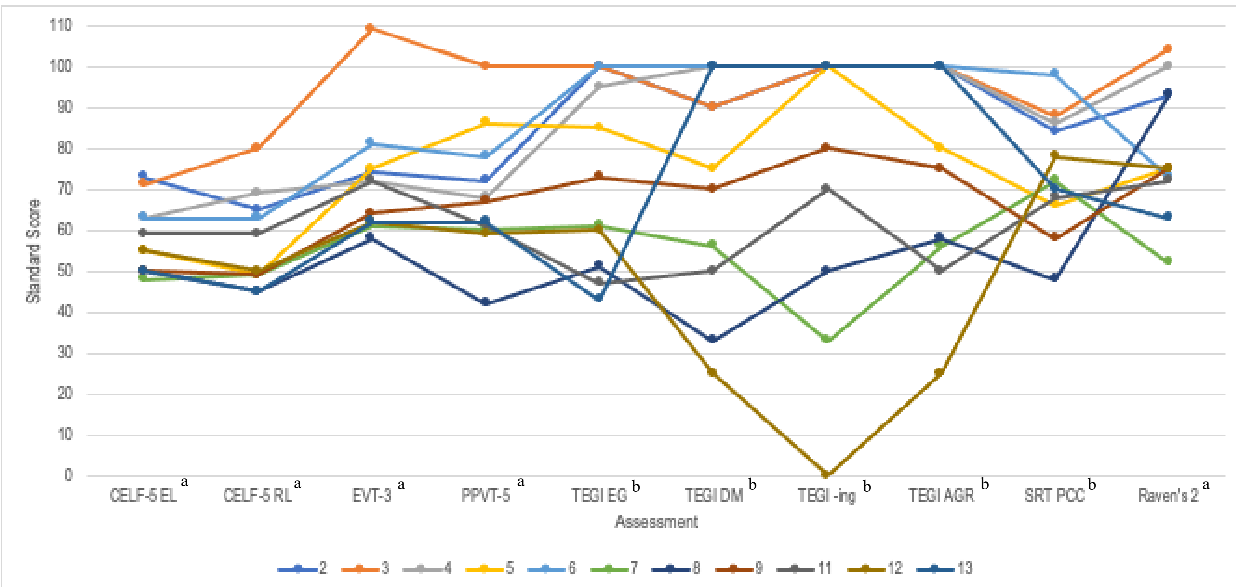

Figure 2. Individual Assessment Profiles as Standard Scores at T3.
[a] CELF, EVT, PPVT, and Raven's M = 100, SD =15. [b] TEGI and SRT max. = 100%. EG = elicited grammar. DM = GJ dropped marker. AGR= GJ agreement. PCC = percentage consonants correct.

RQ1 and RQ2: Do Participants Change in Relative Performance Levels Within the Group?

Overall, Expressive, and Receptive Language

On average, participants did not change in relative performance levels within the group. A Friedman test revealed that CELF total scores did not significantly differ from T1 (*Mdn* = 50.00), to T2 (*Mdn* = 50.00), to T3 (*Mdn* = 53.00), $\chi^2(2) = 1.182$, $p = .554$ (Friedman, 1937). CELF receptive language scores did not significantly differ from T1 (*Mdn* = 50.00), to T2 (*Mdn* = 50.00), to T3 (*Mdn* = 59.00), $\chi^2(2) = 1.200$, $p = .549$. CELF expressive language scores did not significantly differ from T1 (*Mdn* = 50.00), to T2 (*Mdn* = 50.00), to T3 (*Mdn* = 56.50), $\chi^2(2)$ = 1.368, *p* = .504, either. Therefore, there was no change over time greater than chance.

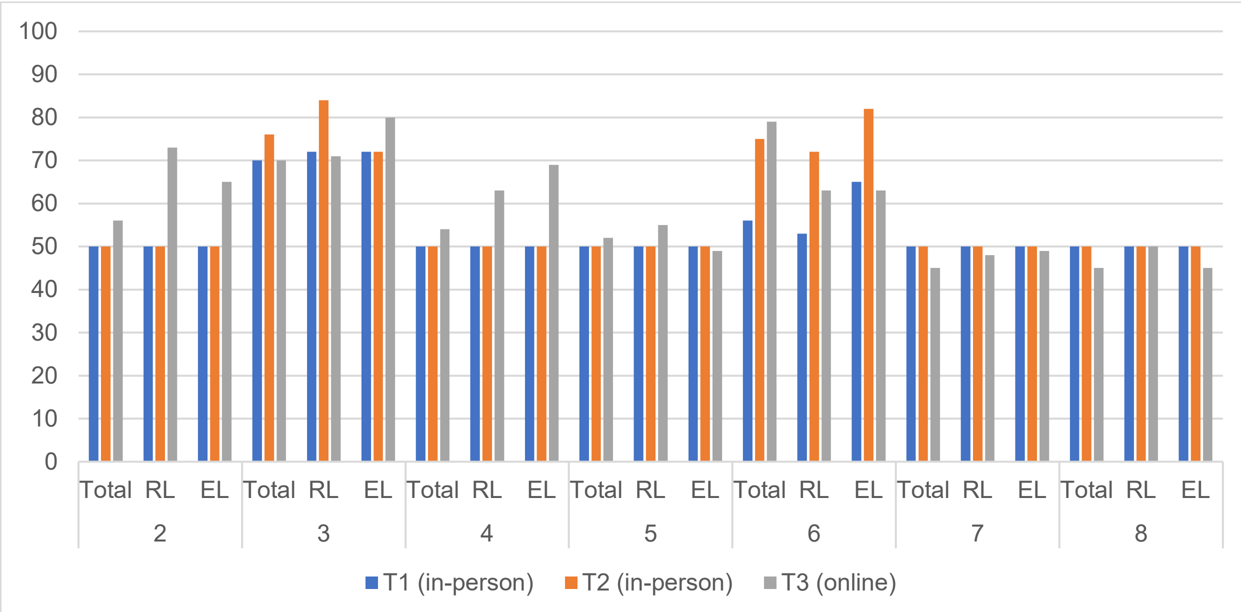

Figure 3. CELF total, receptive language (RL), and expressive language (EL) standard scores by participant across T1, T2, and T3

As Figure 3 shows, some participants had stable CELF scores and some had scores which varied across timepoints and assessment methods. One consideration is that the CELF-3, which was used at T1 and T2, has a floor of 50, whereas the CELF-5, which was used at T3, has a floor of 40. Therefore, the floor level varies from one test to the other. One way of interpreting individual differences is by floor effects within the same version of the CELF. Participants 2, 4, 5, 7, and 8 showed floor effects at T1 and T2. Participants 5, 7, and 8 had stable scores at or near floor across all timepoints. In contrast, participants 2 and 4 showed increases in CELF scores between T2 and T3, which coincided with differences in assessment method. Participants 3 and 6 had scores which varied over time. Participant 3 had higher scores for total language and receptive language at T2 compared to T1 and T3, but their expressive score at T3 was higher compared to T1 and T2. Participant 6 had increases in their total language score from T1 to T3, and higher performance at T2 for receptive and expressive language compared to T1 and T3.

Grammar

On average, participants did not change in relative performance levels within the group. The TEGI screener, or the average of the third-person singular present and past-tense probes, is a short form of the elicited grammar (EG) score, which is the average of the third-person singular present, past-tense, BE, and DO probes. A Friedman test showed that screener scores did not significantly differ from T1 (*Mdn* = 98.50), to T2 (*Mdn* = 97.00), to T3 (*Mdn* = 93.00), $\chi^2(2) = .353$, $p = .838$ (Friedman, 1937). TEGI EG scores did not significantly differ from T1 (*Mdn* = 98.00), to T2 (*Mdn* = 98.00), to T3 (*Mdn* = 90.00), $\chi^2(2) = .240$, $p = .887$.

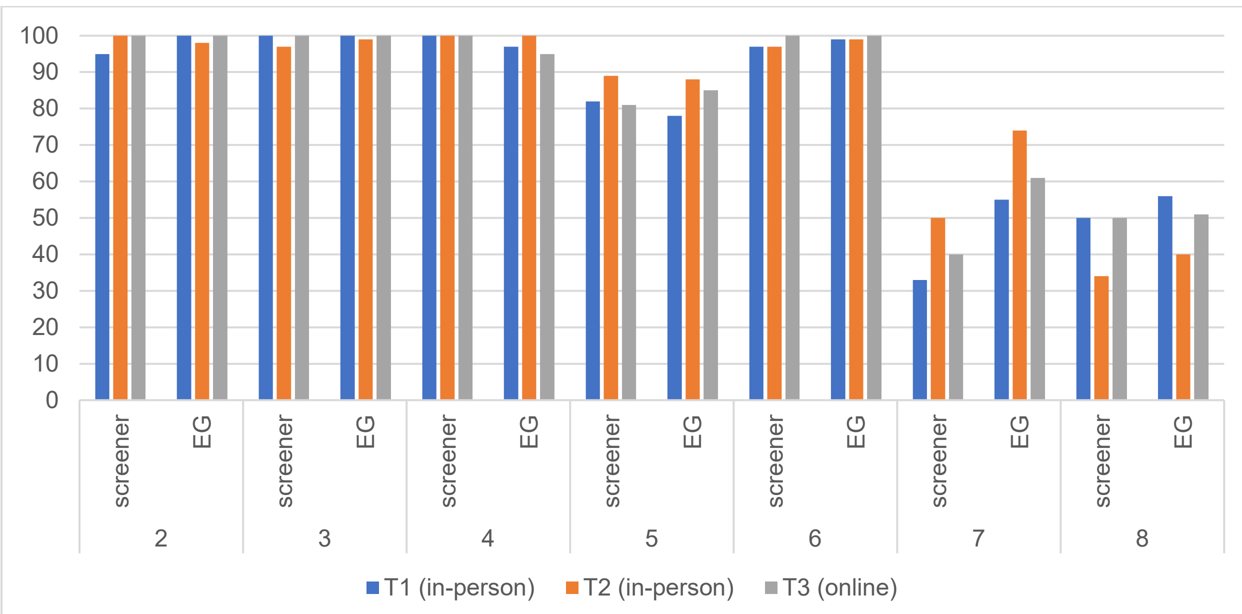

Figure 4. TEGI outcomes by participant across T1, T2, and T3

As Figure 4 shows, some participants had stable TEGI scores and some had scores which varied across timepoints and assessment methods, although their relative ranking was stable. One way of interpreting individual differences is by ceiling effects. Participants 2, 3, 4, and 6 had stable, near-ceiling scores across T1, T2, and T3. In contrast, participants 5, 7, and 8, had variable outcomes, although the differences may not be clinically meaningful. Participant 5 scores changed slightly from T1 to T2 to T3. Participants 7 and 8 had significantly lower scores.

Participant 7 scores increased from T1 to T2 before decreasing from T2 to T3. Participant 8 scores decreased from T1 to T2 before increasing from T2 to T3. Overall, without significant differences across timepoints and methods, subsequent sections report out outcomes from T3.

RQ3: Does Speech Sound Disorder Interfere with Language Assessment Performance?

As Table 4 shows, all participants had 100% accuracy on the TEGI phonological probe and were able to produce the sounds for finiteness-marking in English. However, most SRT scores were mostly below ceiling. Since most participants showed few transcoding difficulties, individual differences in verbal working memory (VWM) rather than neuromotor difficulties may have influenced performance (Shriberg et al., 2012).

Table 4.

Score Summary for TEGI Phonological Probe and SRT

ID	CA	TEGI Phono	SRT			
			PCC	ENCOD	MEM	TRANS
M	20.26	100.00	74.18	68.42	80.48	97.98
SD	1.98	0.00	13.76	15.49	23.87	3.57
max	23.63	100	98	100	100	100

Note. CA = chronological age. TEGI = Test of Early Grammatical Impairment (Rice & Wexler, 2001). TEGI Phono = phonological probe. SRT = Syllable Repetition Task (Shriberg et al., 2009). SRT administered at T3 only. PCC = percent consonants correct (competence score). ENCOD = encoding score. MEM = memory score. TRANS = transcoding score.

Analysis explored relationships between SRT scores and language assessment outcomes. Because there was a very strong, positive relationship between competence scores and encoding scores, $r_s(11) = .909$, $p < .001$, subsequent analysis used only competence scores (i.e., PCC). As Table 5 shows, there was a very strong, positive relationship between competence scores and CELF-5 expressive language and TEGI 3s scores. There was a strong, positive relationship between competence scores and CELF-5 receptive language and TEGI dropped marker scores. SRT memory and transcoding scores had fewer effects. There was a strong, positive relationship between transcoding scores and TEGI DO scores. In all, scores indicative of overall mastery and correct auditory-perceptual representations, rather than VWM or speech motor planning, tended to associate with language assessment scores.

Table 5.

Correlations Between Language Assessment Outcomes

	EL	PPVT-5	EVT-3	PCC	Mem	Trans	3s	Past	Be	Do	DM	AGR	-ING
RL	.73*	.65*	.81**	.69*	.21	.44	.81**	.57	.34	.38	.64*	.67*	.75**
EL		.55	.66*	.94**	.58	.51	.82**	.47	.46	.32	.65*	.58	.53
PPVT-5			.93**	.59	.15	.39	.73*	.23	.48	.36	.76**	.73*	.87**
EVT-3				.51	.13	.31	.69*	.14	.63*	.37	.85**	.84**	.90**
PCC					.49	.25	.82**	.48	.48	.10	.64*	.56	.43
MEM						.25	.45	-.05	-.01	-.10	.20	.06	.06
TRANS							.17	-.05	.14	.80**	.37	.37	.50
3s								.63	.61*	.16	.76**	.74**	.68*
PAST									.21	.38	.32	.49	.31
BE										.27	.73*	.76**	.57
DO											.39	.54	.56
DM												.95**	.90**
AGR													.92**

Note. * = $p < .05$. ** = $p < .001$ RL = CELF-5 Receptive Language. EL = CELF-5 Expressive Language. PCC = SRT percentage consonants correct (i.e., competence score). MEM = SRT memory score. TRANS = SRT transcoding score. 3s = TEGI third-person present-tense probe. PAST = TEGI third-person past-tense probe. TEGI BE = Be probe. DO = TEGI Do probe. DM = TEGI GJ dropped marker. AGR = TEGI GJ agreement. -ING = TEGI GJ dropped -ing.

RQ4: Are There Individual Differences in Expressive and Receptive Language Ability?

Table 6.

Summary of CELF-5, PPVT-5, and EVT-3 Standard Scores

	CELF-5				PPVT-5	EVT-3	Vocab Diff.
	Total	RL	EL	Diff			
M	52.27	57.91	56.64	-1.27	68.64	71.82	3.18
SD	12.22	8.69	15.97	5.26	15.38	14.30	7.18
Max	79	73	91	9	100	109	16

Note. RL = CELF-5 receptive language index score. EL = CELF-5 expressive language index score. CELF-5 difference score = expressive standard score - receptive language score. Difference score = EVT-3 standard score - PPVT-5 standard score.

Receptive

As Table 6 shows, receptive language scores were typically near floor. As Table 5 shows, receptive language associated with other language outcomes. There was a very strong positive relationship with expressive vocabulary, which was consistent with test norming data (CELF-5 RLI and EVT-2 r = .72; Wiig et al., 2013), as well as with TEGI 3s scores. There was also a strong positive relationship with: (a) expressive language, which was consistent with test norming data (CELF-5 ELI and RLI r = .76; Wiig et al., 2013); (b) receptive vocabulary, which was consistent with test norming data (CELF-5 RLI and PPVT-5 r = .72; Dunn, 2019); (c) SRT competence; (d) TEGI GJ dropped marker; (e) TEGI GJ agreement, and; (f) TEGI GJ dropped -*ing* scores. As Table 7 shows, there were no significant effects of child variables on receptive language scores.

Table 7.

Correlations between Language Outcomes and Child Variables

Language	Age	Ed. Yrs.	MLE	SRS-T	Raven's 2
CELF-5 RL	-.15	.02	-.39	.01	.60
CELF-5 EL	-.12	.01	-.33	.05	.33
CELF-5 Diff	.34	.31	.09	.08	.08
PPVT-5	.00	.13	-.09	-.15	.37
EVT-3	-.18	-.10	-.28	.01	.37
Vocab Diff	.40	.03	-.21	.16	.39
TEGI 3s	.17	.22	-.17	-.08	.54
TEGI Past	.38	.48	-.18	-.02	.73*
TEGI BE	.36	.43	.03	-.44	.23
TEGI DO	-.19	-.02	.13	-.25	.52
TEGI DM	.11	.37	.10	.03	.38
TEGI AGR	.21	.46	.09	-.16	.56
TEGI -ING	-.04	.17	.10	-.04	.53

Note. * = p < .05. RL = receptive language. EL = expressive language. CELF-5 Diff = CELF-5 difference score. Vocab Diff = Vocabulary difference score. EG = elicited grammar composite. -ing = Dropped -ing. DM = dropped marker. AGR = agreement. Ed. Yrs. = years of education. MLE = maternal level of education. SRS-T = SRS total *t*-score.

Expressive

As Table 6 shows, expressive language scores were generally near floor. As Table 7 shows, there were no significant effects of child variables on expressive language scores. However, as Table 5 shows, expressive language associated with other language outcomes. There was a very strong, positive relationship with SRT competence scores, as well as with TEGI 3s scores. In addition, there was a strong, positive relationship with expressive vocabulary, which was consistent with test norming data (CELF-5 and EVT-2 $r = .70$; Wiig et al., 2013), as well as with TEGI GJ dropped marker scores.

Individual Receptive-Expressive Language Differences

As Table 6 shows, on average, there were no significant receptive-expressive language differences, although individual outcomes varied. Participant 3 had significantly stronger expressive than receptive language at the $p < .05$ level (difference score = 9; Wiig et al., 2013) As Table 7 shows, there were no significant effects of child variables, and there were no significant relationships with other language scores.

RQ5: Are There Individual Differences in Expressive and Receptive Vocabulary?

Receptive

As Table 6 shows, receptive vocabulary scores were generally below typical, with interindividual variation. As Table 5 shows, receptive vocabulary associated with other language assessment outcomes. There was a very strong, positive relationship with expressive vocabulary, which was consistent with test norming data (PPVT-5 and EVT-3 $r = .77$; Williams, 2019), as well as with TEGI GJ dropped *-ing* scores. There was a strong, positive relationship with: (a) TEGI 3s; (b) TEGI GJ dropped marker; (c) TEGI GJ agreement, and; (d) CELF-5 receptive language scores. As Table 7 shows, there were no significant effects of child variables on receptive language scores.

Expressive

As Table 6 shows, EVT-3 scores were typically below age expectations, with individual variability. As Table 5 shows, expressive vocabulary associated with other language outcomes. There was a very strong, positive relationship with (a) CELF-5 receptive language; (b) receptive vocabulary; (c) TEGI GJ dropped marker; (d) TEGI GJ agreement, and; (e) TEGI dropped *-ing* scores. In addition, there was a strong, positive relationship with: (a) TEGI 3s; (b) TEGI BE, and; (c) CELF-5 expressive language scores. As Table 7 shows, there were no effects of child variables.

Individual Receptive-Expressive Vocabulary Differences

As Table 6 shows, on average, there were no significant receptive-expressive vocabulary differences, although individual outcomes varied. Participants 3 (difference score = 9), 8 (difference score = 16), and 11 (difference score = 11) had statistically significant receptive-expressive vocabulary differences at the $p < .05$ level, such that their expressive vocabulary was stronger than their receptive vocabulary (William, 2019). Conversely, participant 5 had a statistically significant receptive-expressive vocabulary difference at the $p < .05$ level (difference score = -11), such that their receptive vocabulary was greater than their expressive vocabulary (Williams, 2019). There were no significant relationships with other language assessment outcomes. As Table 7 shows, there were no significant effects of child variables on receptive-expressive vocabulary differences.

RQ6: Are There Individual Differences in Use of Finiteness Markers or Judgments of Finiteness Markers?

Use of Finiteness Markers

As Table 8 shows, finiteness-marking probe scores for 3s, third-person past tense, BE, and DO, as well as the elicited grammar composite (i.e., the average of the four probe scores) score, were typically below ceiling. To compare, young children ages 6;6-6;11 typically show ceiling effects on these probes (Rice & Wexler, 2001). However, there was variability. As Table 5 shows, there was a strong, positive relationship between TEGI 3s scores and: (a) TEGI BE; (b) GJ dropped marker; (c) GJ agreement scores, and; (d) dropped *-ing* scores, consistent with test norming data (Rice & Wexler, 2001). There was also a very strong, positive relationship with CELF-5 receptive language, CELF-5 expressive language, and; SRT competence scores, as well as strong, positive relationships with expressive vocabulary and receptive vocabulary scores. Further, there was a strong, positive relationship between TEGI BE scores and TEGI dropped marker, TEGI dropped *-ing*, TEGI 3s, and expressive vocabulary scores. There was only a strong, positive relationship between TEGI DO scores and SRT transcoding scores.

As Table 7 shows, there were nearly no significant effects of child variables. There was only a strong, positive relationship between TEGI past tense probe scores and NVIQ, which was unexpected relative to test norming data, $r_s(11) = .728$, $p < .05$ (Rice & Wexler, 2001), and twice the effect size reported in Roberts and colleagues (2004; past and NVIQ $r = .359$, $p < .01$).

Table 8.

Score Summary of TEGI Outcomes

Measure	*M*	*SD*	Max
Grammatical production			
Third-person singular	72.5	30.4	100
Past-tense overall	83.9	26.5	100

Auxiliary & copula BE	81.1	19.2	100
Auxiliary DO	63.7	40.5	100
Elicited grammar composite	74.1	21.7	100
GJ dropped marker	68.5	24.5	100
GJ dropped -ing	73.5	29.4	100
GJ agreement	70.8	29.3	100

Note. Elicited grammar composite = average of four grammatical production probes, or third-person singular, past-tense, BE and DO.

Judgments of Finiteness Errors

As Table 8 shows, although GJ scores for finiteness errors were, on average, below ceiling, there was variability. As Table 7 shows, there were many relationships between each of the GJ scores and other TEGI scores. There was a very strong, positive relationship between GJ dropped marker scores and GJ agreement scores, which was higher than the correlation reported in Barton-Hulsey & Sterling (2020) of $r = .69$, $p < .01$, as well as GJ dropped *-ing* scores. There was also a very strong, positive relationship between GJ agreement and GJ dropped *-ing*, which is higher than the correlation reported in Barton-Hulsey & Sterling (2020) of $r = .73$, $p < .01$. In addition, there was a strong, positive relationship between each of the GJ scores and the 3s probe scores. There was a strong, positive relationship between the GJ dropped marker and GJ agreement scores, respectively, and BE scores.

In addition to within-TEGI relationships, there were also relationships between the GJ scores and other language assessments. TEGI GJ dropped marker scores had a very strong, positive relationship with expressive vocabulary scores, as well as a strong, positive relationship with: (a) CELF-5 receptive language; (b) CELF-5 expressive language; (c) receptive vocabulary, and; (d) SRT competence scores. TEGI GJ agreement scores had a very strong, positive relationship with expressive vocabulary scores, as well as a strong relationship with CELF-5 receptive language and receptive vocabulary scores. TEGI GJ dropped *-ing* scores had a very strong, positive relationship with receptive vocabulary scores and expressive vocabulary scores, as well as a strong relationship with receptive language scores. As Table 7 shows, there were no significant effects of child variables. Overall, these findings support the validity of TEGI as a diagnostic assessment for LI in autistic individuals.

Summary

In all, participants did not change in their relative performance within the group across timepoints or assessment methods. Overall language assessment was below age expectations, with limited differences between expressive and receptive language. While vocabulary outcomes were typically below age expectations, over one-third of participants had significant differences between their receptive and expressive vocabulary, such that 27% had stronger expressive than receptive vocabulary and 9% had stronger receptive than expressive vocabulary. On average, the use of finiteness markers and judgments of morphosyntax were below ceiling, although there was interindividual variation. These findings add to the empirical evidence base supporting the TEGI as an informative measure for this population. Finally, there were nearly no effects of child variables (i.e., age, years of education, MLE, autism characteristics, and NVIQ) on language assessment outcomes, and participants were capable of producing the sounds for finiteness-marking in English.

Discussion

This report investigated the language abilities of autistic minority young adults. Administering an extensive assessment protocol was informative, replicated findings from Girolamo et al. (2020), and extended what is known about longitudinal stability across multiple dimensions of language. These findings have implications for understanding the validity and reliability of assessment for autistic participants such as the ones in this study, as well as for understanding the nature of language impairment in adolescence and young adulthood.

Longitudinal Consistency of Assessment

Participants had consistent overall language, vocabulary, and morphosyntactic assessment outcomes in terms of relative group rank across three timepoints over three years. Performance on the CELF was consistent and near floor, with scores not changing significantly across timepoints. While TEGI outcomes were more variable across participants, performance did not change significantly across timepoints. Consistent with previous work (Girolamo et al., 2020), speech production likely did not interfere with language assessment performance. On one hand, participants showed mastery of the speech sounds used for tense-marking in English. In addition, while overall competence scores on a nonword repetition task were below ceiling, participants showed near-ceiling transcoding scores, which suggest they did not have neuromotor difficulties as related to speech sound planning (Shriberg et al., 2012). Overall, comprehensive language assessment resulted in longitudinal consistency in a sample of autistic adolescents and young adults.

This study highlights the importance of repeated measures data, despite few longitudinal studies of comprehensive language assessment rather than broader assessment of academic skills (Kim et al., 2018). Findings were consistent with Minicalsco and Carlsson (2021), who showed that autistic children had persistent group membership from ages 3 to 8 years in terms of autism plus typical language and autism plus LI. In contrast, findings ran counter to studies showing that some autistic children show structural language difficulties in early childhood and lose them in later age ranges (e.g., Bennett et al., 2014; Reindal et al., 2021). Participants here experienced persistent difficulties with language.

LI in Autistic Adolescents and Young Adults

In this study, nearly all participants had LI, as evidenced by overall language, lexical, and morphosyntactic outcomes. The mean overall language and vocabulary scores were near floor, with limited individual receptive-expressive differences. On average, performance on the use of finiteness markers and judgments of morphosyntactic well-formedness was below ceiling, consistent with previous work on autistic adolescents with LI (e.g., Barton-Hulsey & Sterling, 2020; Brynskov et al., 2016; Wittke et al., 2017). However, some participants performed at ceiling on the TEGI. Given that children with typical speech and language ages 6;6 to 6;11 master these skills (Rice & Wexler, 2001), the nature of LI in some participants may not include early acquired morphosyntactic knowledge – or finiteness-marking as an obligatory property of the grammar that does not express semantic meaning. Rather, their persistent language difficulties may lie in more semantic or complex tasks. On the other hand, other participants persisted in well below-ceiling performance on the TEGI. In all, these findings support the TEGI as a valid assessment for identifying LI in autistic adolescents and young adults.

The findings of this study support language independent of NVIQ (Dennis et al., 2009) and of demographic factors, such as MLE or years of education. This is consistent with previous findings that NVIQ did not differentiate the performance of autistic children on morphosyntactic measures (Barton-Hulsey & Sterling, 2020; Brynskov et al., 2016; Eigsti et al., 2007, Thomas et al., 2020), predict receptive-expressive vocabulary difference scores (Haebig & Sterling, 2017), or predict vocabulary outcomes (Plesa Skwerer et al., 2016). However, this finding runs counter to previous findings of NVIQ as a predictor of expressive vocabulary (Haebig & Sterling, 2017; McGregor et al., 2012; Talli, 2020) and receptive-expressive difference scores (Kover et al., 2013). The cause of these discrepancies is unknown. Further, the lack of other child characteristic effects is consistent with previous work finding language outcomes to be independent of age and autism characteristics (Haebig & Sterling, 2017; Kover et al, 2013; Lindgren et al., 2009; Loucas et al., 2008).

Limitations

This study encountered limitations. First, longitudinal data were available for only some participants. Second, this study had sample bias. Four of the 13 (23.5%) participants who enrolled dropped from the study. Although the exact reasons for attrition are unknown, there are still barriers to making research engaging and accessible to a wider range of autistic minority young adults. Third, some standardized instruments, namely the CELF-5, may have collapsed variation due to floor effects. Further, some tasks, such as GJs, may have presented format challenges to autistic individuals. Following Barton-Hulsey & Sterling (2020), integrating background information in language comprehension may present difficulty to some autistic individuals due to cognitive flexibility demands (i.e., individuals must separate the semantic and grammatical nature of GJ tasks).

Future Directions

This report lays ground for future research. First, future work should replicate this research in longitudinal designs. Additional timepoints or participant groups may inform our understanding of the experiences of autistic individuals across the life span. Such information is necessary to understand whether findings of this study hold up in other studies (Brandt et al., 2014). Second, future research should focus on collecting additional measures to more holistically understand the abilities and characteristics of autistic young adults. Subsequent research might integrate measures of linguistic processing, as well as measures on attention and socio-pragmatics, into assessment. Last, working with autistic individuals to develop research and broaden representation to include older age ranges and minorities is an important future direction for research (Bottema-Beutel et al., 2020). The participant-centered approach of this study suggests that, with intentional planning, such initiatives are acceptable or welcome. In all, these are a few of many avenues for furthering understanding of the experiences of autistic individuals across the lifespan.

Conclusion

In implementing a longitudinal case study, this study added unprecedented information to the literature on the language and morphosyntactic abilities of autistic adolescents and young adults. Here, participants had consistent language assessment outcomes over time and persistent LI. Questions for future research include how linguistic mechanisms and the full array of individual characteristics inform individual language outcomes.

Bibliography

Ambridge, B., Bidgood, A., & Thomas, K. (2021). Disentangling syntactic, semantic and pragmatic impairments in ASD: Elicited production of passives. *Journal of Child Language*, *48*(1), 184-201. https://doi.org/10.1017/S0305000920000215

American Psychiatric Association. (2004). *Diagnostic and statistical manual of mental disorders* (4th ed.). Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, *73*(S1), 262-270. https://doi.org/10.1080/00031305.2018.1543137

Bartolucci, G., Pierce, S. J., & Streiner, D. (1980). Cross-sectional studies of grammatical morphemes in autistic and mentally retarded children. *Journal of Autism and Developmental Disorders, 10*(1), 39-50. https://doi.org/10.1007/BF02408431

Barton-Hulsey, A., & Sterling, A. (2020). Grammatical judgment and production in male participants with idiopathic autism spectrum disorder. *Clinical Linguistics & Phonetics, 34*(12), 1088-1111. https://doi.org/10.1080/02699206.2020.1719208

Bennett, T., Szatmari, P., Bryson, S., Volden, J., Zwaigenbaum, L., Vaccarella, L., et al. (2008). Differentiating autism and Asperger syndrome on the basis of language delay or impairment. *Journal of Autism and Developmental Disorders, 38*(4), 616–625. https://doi.org/10.1007/s10803-007-0428-7

Bennett, T. A., Szatmari, P., Georgiades, K., Hanna, S., Janus, M., Georgiades, S., Duku, E., Bryson, S., Fombonne, E., Smith, I. M., Mirenda, P., Volden, J., Waddell, C., Roberts, W., Vaillancourt, T., Zwaigenbaum, L., Elsabbagh, M., Thompson, A., & Pathways in ASD Study Team. (2014). Language impairment and early social competence in preschoolers with autism spectrum disorders: a comparison of DSM-5 profiles. *Journal of Autism and Developmental Disorders*, *44*(11), 2797-2808. https://doi.org/10.1007/s10803-014-2138-2

Boucher, J. (2012). Research review: structural language in autistic spectrum disorder– characteristics and causes. *Journal of Child Psychology and Psychiatry*, *53*(3), 219-233. https://doi.org/10.1111/j.1469-7610.2011.02508.x

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217-224. https://doi.org/10.1016/j.jesp.2013.10.005

Brynskov, C., Eigsti, I.-M., Jørgensen, M., Lemcke, S., Bohn, O. S., & Krøjgaard, P. (2017). Syntax and morphology in Danish-speaking children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *47*(2), 373-383. https://doi.org/10.1007/s10803-016-2962-7

Burgemeister, B. B., Blum, L. H., & Lorge, I. (1972). *Columbia Mental Maturity Scale*. Harcourt Brace Jovanovich.

Chan, W., Smith, L. E., Hong, J., Greenberg, J. S., & Mailick, M. R. (2017). Validating the Social Responsiveness Scale for adults with autism. *Autism Research*, *10*(10), 1663-1671. https://doi.org/10.1002/aur.1813

Chen, K. L., Lin, C. H., Yu, T. Y., Huang, C. Y., & Chen, Y. D. (2018). Differences between the Childhood Autism Rating Scale and the Social Responsiveness Scale in assessing symptoms of children with Autistic Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(9), 3191-3198. https://doi.org/10.1007/s10803-018-3585-y

Ciccia, A.H., Whitford, B., Krumm, M., & McNeal, K. (2011). Improving the access of young urban children to speech, language and hearing screening via telehealth. *Journal of Telemedicine and Telecare*, *17*(5), 240-244. https://doi.org/10.1258%2Fjtt.2011.100810

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology, 12*(3), 349-358. https://doi.org/10.1044/1058-0360(2003/080)

Constantino, J. N. & Gruber, C. P. (2012). *Social Responsiveness Scale, Second Edition: Manual*. Western Psychological Services.

Davidson, M. M., & Ellis Weismer, S. (2017). A discrepancy in comprehension and production in early language development in ASD: Is it clinically relevant? *Journal of Autism and Developmental Disorders, 47*(7), 2163-2175. https://doi.org/10.1007%2Fs10803-017-3135-z

Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society*, *15*(3), 331-343. https://doi.org/10.1017/S1355617709090481

Duncan, A. W., & Bishop, S. L. (2015). Understanding the gap between cognitive abilities and daily living skills in adolescents with autism spectrum disorders with average intelligence. *Autism*, *19*(1), 64-72. https://doi.org/10.1177/1362361313510068

Dunn, D. M. (2019). *Peabody Picture Vocabulary Test, Fifth Edition: Manual*. Pearson.

Durkin, M. S., Elsabbagh, M., Barbaro, J., Gladstone, M., Happe, F., Hoekstra, R. A., Lee, L. C., Rattazzi, A., Stapel-Wax, J., Stone, W. L., Tager-Flusberg, H., Thurm, A., Tomlinson, M., & Shih, A. (2015). Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide. *Autism Research, 8*(5), 473–476. https://doi.org/10.1002/aur.1575

Edwards, M., Stredler-Brown, A., & Houston, K.T. (2012). Expanding use of telepractice in speech-language pathology and audiology. *Volta Review*, *112*(3), 227-242. https://doi.org/10.17955/tvr.112.3.m.704

Eigsti, I.-M., & Bennetto, L. (2009). Grammaticality judgments in autism: Deviance or delay. *Journal of Child Language*, *36*(5), 999-1021. https://doi.org/10.1017/s0305000909009362

Eigsti, I.-M., Bennetto, L., & Dadlani, M. B. (2007). Beyond pragmatics: Morphosyntactic development in autism. *Journal of Autism and Developmental Disorders*, *37*(6), 1007-1023. https://doi.org/10.1007/s10803-006-0239-2

Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders, 5*(2), 681–691. https://doi.org/10.1016/j.rasd.2010.09.001

Ellis Weismer, S., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers with developmental delay. *Journal of Autism and Developmental Disorders, 40*(10), 1259–1273. https://doi.org/10.1007/s10803-010-0983-1

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *43*(4), 865-878. https://doi.org/10.1044/jslhr.4304.865

Eriks-Brophy, A., Quittenbaum, J., Anderson, D., & Nelson, T. (2008). Part of the problem or part of the solution? Communication assessments of Aboriginal children residing in remote communities using videoconferencing. *Clinical Linguistics & Phonetics*, *22*(8), 589-609. https://doi.org/10.1080/02699200802221737

Franklin, T. (2017). Best practices in in multicultural assessment of cognition. In R. S. McCallum (Ed)., *Handbook of Nonverbal Assessment* (2nd ed., pp. 39-46). Springer.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*(200), 675-701. https://doi.org/10.1080%2F01621459.1937.10503522

Girolamo, T. M., Rice, M. L., & Warren, S. F. (2020). Assessment of language abilities in minority adolescents and young adults With Autism Spectrum Disorder and extensive special education needs: A pilot study. *American Journal of Speech-Language Pathology*, *29*, 804-818. https://doi.org/10.1044/2020_AJSLP-19-00036

Grzadzinski, R., Huerta, M., & Lord, C. (2013). DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes. *Molecular Autism, 4*(1), 12. https://doi.org/10.1186/2040-2392-4-12

Grondhuis, S. N., Lecavalier, L., Arnold, L. E., Handen, B. L., Scahill, L., McDougle, C. J., & Aman, M. G. (2018). Differences in verbal and nonverbal IQ test scores in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *49*, 47-55. https://doi.org/10.1016/j.rasd.2018.02.001

Haebig, E., & Sterling, A. (2017). Investigating the receptive-expressive vocabulary profile in children with idiopathic ASD and comorbid ASD and fragile X syndrome. *Journal of Autism and Developmental Disorders, 47*, 260–274. https://doi.org/10.1007/s10803-016-2921-3

Hodge, M.A., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., Detheridge, J., Drevensek, S., Edwards, L., Everett, M., Ganesalingam, K., Geier, P., Kass, C., Mathieson, S., McCabe, M., Micallef, K., Molomby, K., Ong, N., Pfeiffer, S., Pope, S., Tait, F., Williams, M., Young-Dwarte, L., & Silove, N. (2019). Agreement between telehealth and face-to-face assessment of intellectual ability in children with specific learning disorder. *Journal of Telemedicine and Telecare*, *25*(7), 431-437. https://doi.org/10.1177%2F1357633X18776095

Howlin, P. (1984). The acquisition of grammatical morphemes in autistic children: A critique and replication of the findings of Bartolucci, Pierce, and Streiner, 1980. *Journal of Autism and Developmental Disorders*, *14*, 127-136. https://doi.org/10.1007/BF02409656

Howlin, P. (2021). Adults with autism: Changes in understanding since DSM-111. *Journal of Autism and Developmental Disorders*, 1-18. https://doi.org/10.1007/s10803-020-04847-z

Howlin, P., & Magiati, I. (2017). Autism spectrum disorder: Outcomes in adulthood. *Current Opinion in Psychiatry*, *30*(2), 69-76. https://doi.org/10.1097/yco.0000000000000308

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *The Journal of Child Psychology and Psychiatry*, *45*(2), 212–229. https://doi.org/10.1111/ j.1469-7610.2004.00215.x

Howlin, P., Savage, S., Moss, P., Tempier, A., & Rutter, M. (2014). Cognitive and language skills in adults with autism: a 40-year follow-up. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(1), 49–58. https://doi.org/10.1111/jcpp.12115

Huang, T., & Finestack, L. (2020). Comparing morphosyntactic profiles of children with developmental language disorder or language disorder associated with Autism Spectrum Disorder. *American Journal of Speech-Language Pathology*, *29*(2), 714-731. https://doi.org/10.1044/2019_AJSLP-19-00207

Iacono, T., Dissanayake, C., Trembath, D., Hudry, K., Erickson, S., & Spong, J. (2016). Family and practitioner perspectives on telehealth for services to young children with autism. In A.J. Maeder, K. Ho, & A. Marcelo (Eds.), *The promise of new technologies in an age of new health challenges* (pp. 63-73). https://doi.org/ 10.3233/978-1-61499-712-2-63

Interagency Autism Coordinating Committee. (2017). 2016–2017 *Interagency Autism Coordinating Committee Strategic Plan for Autism Spectrum Disorder*. https://iacc.hhs.gov/publications/strategic-plan/2017

Kim, S. H., Bal, V. H., & Lord, C. (2018). Longitudinal follow-up of academic achievement in children with autism from age 2 to 18. *Journal of Child Psychology and Psychiatry*, *59*(3), 258-267. https://doi.org/10.1111/jcpp.12808

King, D., & Palikara, O. (2018). Assessing language skills in adolescents with autism spectrum disorder. *Child Language Teaching and Therapy*, *34*(2), 101-113. https://doi.org/10.1177/0265659018780968

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, *16*(2-3), 287-308. https://doi.org/10.1080/01690960042000058

Kover, S. T., McDuffie, A. S., Hagerman, R. J., & Abbeduto, L. (2013). Receptive vocabulary in boys with autism spectrum disorder: Cross-sectional developmental trajectories. *Journal of Autism and Developmental Disorders*, *43*(11), 2696-2709.

https://doi.org/10.1007/s10803-013-1823-x

Kwok, E. Y., Brown, H. M., Smyth, R. E., & Cardy, J. O. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *9*, 202-222. https://doi.org/10.1016/j.rasd.2014.10.008

Lindsay, R. M., & Ehrenberg, A. S. (1993). The design of replicated studies. *The American Statistician*, *47*(3), 217-228. https://doi.org/10.1080/00031305.1993.10475983

Lohmeier, H. L., & Shriberg, L. D. (2011). *Reference data for the Syllable Repetition Task (SRT)* (Report No. 17). Phonology Project, Waisman Center, University of Wisconsin-Madison. https://www2.waisman.wisc.edu/phonology/techreports/TREP17.pdf

Lord, C., & Bishop, S. L. (2015). Recent advances in autism research as reflected in DSM-5 criteria for autism spectrum disorder. *Annual Review of Clinical Psychology, 11*, 53-70. https://doi.org/10.1146/annurev-clinpsy-032814-112745

Lord, C., & Rutter, M. (2012). *Autism Diagnostic Observation Schedule, Second Edition*. Western Psychological Services.

Loucas, T., Charman, T., Pickles, A., Simonoff, E., Chandler, S., Meldrum, D., et al. (2008). Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry, 49*(11), 1184–1192. https://doi.org/10.1111/j.1469-7610.2008.01951.x.

Mackey, A. (2012). Why (or why not), when and how to replicate research. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 34-69). Cambridge University Press.

Magaña, S., Lopez, K., Aguinaga, A., & Morton, H. (2013). Access to diagnosis and treatment services among Latino children with autism spectrum disorders. *Intellectual and Developmental Disabilities, 51*(3), 141–153. https://doi.org/10.1352/1934-9556- 51.3.141

Magiati, I., Tay, X. W., & Howlin, P. (2014). Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. *Clinical Psychology Review*, *34*(1), 73-86. https://doi.org/10.1016/j.cpr.2013.11.002

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304-316. https://doi.org/10.3102/0013189X14545513

McCallum, R. E., & McCallum, R. S. (2017). Psychological and physiological influences on multicultural and nonverbal assessment. In R. S. McCallum (Ed)., *Handbook of Nonverbal Assessment* (2nd ed., pp. 77-104). Springer.

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without ASD and language impairment. *Journal of Autism and Developmental Disorders*, *42*(1), 35-47. https://doi.org/10.1007/s10803-011-1210-4

Miniscalco, C., & Carlsson, E. (2021). A longitudinal case study of six children with autism and specified language and non-verbal profiles. *Clinical Linguistics & Phonetics*, 1-19. https://doi.org/10.1080/02699206.2021.1874536

Modyanova, N., Perovic, A., & Wexler, K. (2017). Grammar is differentially impaired in subgroups of autism spectrum disorders: evidence from an investigation of tense marking and morphosyntax. *Frontiers in Psychology*, *8*, 320. https://doi.org/10.3389/fpsyg.2017.00320

Nader, A. M., Courchesne, V., Dawson, M., & Soulières, I. (2016). Does WISC-IV underestimate the intelligence of autistic children? *Journal of Autism and Developmental*

Disorders, *46*(5), 1582-1589. https://doi.org/ 10.1007/s10803-014-2270-z

Paul, R., Chawarska, K., Cicchetti, D., & Volkmar, F. (2008). Language outcomes of toddlers with autism spectrum disorders: A two year follow-up. *Autism Research*, *1*(2), 97-107. https://doi.org/10.1002/aur.12

Parmanto, B., Pulantara, I. W., Schutte, J. L., Saptono, A., & McCue, M. P. (2013). An integrated telehealth system for remote administration of an adult autism assessment. *Telemedicine and e-Health*, *19*(2), 88-94. https://doi.org/10.1089/tmj.2012.0104

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, *20*(5), 591-604. https://doi.org/10.1177/1362361315600146

Raman, N., Nagarajan, R., Venkatesh, L., Monica, D.S., Ramkumar, V., & Krumm, M. (2019). School-based language screening among primary school children using telepractice: A feasibility study from India. *International Journal of Speech-Language Pathology*, *21*(4), 425-434. https://doi.org/10.1080/17549507.2018.1493142

Raven's Progressive Matrices–Second Edition: Manual. (2018). Pearson.

Reindal, L., Nærland, T., Weidle, B., Lydersen, S., Andreassen, O. A., & Sund, A. M. (2021). Structural and pragmatic language impairments in children evaluated for Autism Spectrum Disorder (ASD). *Journal of Autism and Developmental Disorders*, 1-19. https://doi.org/10.1007/s10803-020-04853-1

Reynolds, A.L., Vick, J.L., & Haak, N.J. (2009). Telehealth applications in speech-language pathology: A modified narrative review. *Journal of Telemedicine and Telecare*, *15*(6), 310-316. https://doi.org/10.1258/jtt.2009.081215

Rice, M. L. (2016). Specific language impairment, nonverbal IQ, attention-deficit/hyperactivity disorder, autism spectrum disorder, cochlear implants, bilingualism, and dialectal variants: Defining the boundaries, clarifying clinical conditions, and sorting out causes. *Journal of Speech, Language, and Hearing Research*, *59*(1), 122-132. https://doi.org/10.1044/2015_JSLHR-L-15-0255

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1239-1257. https://doi.org/10.1044/jshr.3906.1239

Rice, M. L., & Wexler, K. (2001). *Test of Early Grammatical Impairment: Manual*. The Psychological Corporation.

Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: An investigation of complex syntax. *International Journal of Language & Communication Disorders*, *45*(1), 47-60. https://doi.org/10.3109/13682820802647676

Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2011). Non-word repetition in adolescents with specific language impairment and autism plus language impairments: A qualitative analysis. *Journal of Communication Disorders*, *44*(1), 23-36. https://doi.org/10.1016/j.jcomdis.2010.06.003

Roberts, J. A., Rice, M. L., & Tager-Flusberg, H. (2004). Tense marking in children with autism. *Applied Psycholinguistics, 25*(3), 429-448. https://doi.org/10.1017.S0142716404001201

Schaeffer, J. (2018). Linguistic and cognitive abilities in children with specific language impairment as compared to children with high-functioning autism. *Language Acquisition*, *25*(1), 5-23. https://doi.org/10.1080/10489223.2016.1188928

Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *Childhood Autism Rating Scale, Second Edition: Manual*. Western Psychological Services.

Schutte, J. L., McCue, M. P., Parmanto, B., McGonigle, J., Handen, B., Lewis, A., Pulantara, W., & Saptono, A. (2015). Usability and reliability of a remotely administered adult autism assessment, the autism diagnostic observation schedule (ADOS) module 4. *Telemedicine and e-Health*, *21*(3), 176-184. https://doi.org/10.1089/tmj.2014.0011

Semel, E., Wiig, E. H., & Secord, W. A. (1995). *The Clinical Evaluation of Language Fundamentals-Third Edition (CELF-3): Technical manual*. The Psychological Corporation.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3-4), 591-611. https://doi.org/10.1093/biomet/52.3-4.591

Shattuck, P. T., Lau, L., Anderson, K. A., & Kuo, A. (2018). A national research agenda for the transition of youth with autism. *Pediatrics, 141*(Suppl. 4), S355–S361. https://doi.org/10.1542/ peds.2016-4300M

Shriberg, L. D., & Lohmeier, H. L. (2008). *The Syllable Repetition Task* (Report No. 14). Phonology Project, Waisman Center, University of Wisconsin-Madison. http://www.waisman.wisc.edu/phonology/techreports/TREP14.pdf

Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., Green, J. R., & Moore, C. (2009). A nonword repetition task for speakers with misarticulations: The Syllable Repetition Task (SRT). *Journal of Speech, Language, and Hearing Research, 52*(5), 1189-1212. https://doi.org/10.1044/1092-4388(2009/08-0047)

Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in childhood apraxia of speech. *Clinical Linguistics &*

Phonetics, *26*(5), 445-482. https://doi.org/10.3109%2F02699206.2012.655841

Shriberg, L. D. & Mabie, H. L. (2017). *Speech and motor speech assessment findings in eight complex neurodevelopmental disorders* (Report No. 24). Phonology Project, Waisman Center, University of Wisconsin-Madison. www.waisman.wisc.edu/phonology/techreports/TREP24.PDF

Smith, C. J., Rozga, A., Matthews, N., Oberleitner, R., Nazneen, N., & Abowd, G. (2017). Investigating the accuracy of a novel telehealth diagnostic approach for Autism Spectrum Disorder. *Psychological Assessment*, *29*(3), 245-252. https://doi.org/10.1037/pas0000317

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C.A. (2016). *Vineland Adaptive Behavior Scales-Third Edition: Manual.* Pearson.

Sterling, A. (2018). Grammar in boys with idiopathic autism spectrum disorder and boys with fragile X syndrome plus autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, *61*(4), 857-869. https://doi.org/10.1044/2017_JSLHR-L-17-0248

Sterling, A. M., Rice, M. L., & Warren, S. F. (2012). Finiteness marking in boys with fragile X syndrome. *Journal of Speech, Language, and Hearing Research, 55*(6), 1704-1715. https://doi.org/10.1044/1092-4388(2012/10-0106)

Sullivan, L. M. (2008). Repeated measures. *Circulation, 117*(9), 1238-1243. https://doi.org/10.1161/CIRCULATIONAHA.107.654350

Sutherland, R., Trembath, D., Hodge, M.A., Rose, V., & Roberts, J. (2019). Telehealth and autism: Are telehealth language assessments reliable and feasible for children with autism? *International Journal of Language & Communication Disorders*, *54*(2), 281-291. https://doi.org/10.1111/1460-6984.12440

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience*

Research, *6*(3-4), 219-224. https://doi.org/10.1016/j.cnr.2006.06.007

Tager-Flusberg H. (2015). Defining language impairments in a subgroup of children with autism spectrum disorder. *Science China Life Sciences*, *58*(10), 1044-1052. https://doi.org/10.1007/s11427-012-4297-8

Talli, I. (2020). Verbal short-term memory as language predictor in children with Autism Spectrum Disorder. *Journal of Behavioral and Brain Science*, *10*(5), 200. https://doi.org/10.4236/jbbs.2020.10513

Taylor, J. L., & Seltzer, M. M. (2011). Employment and post-secondary educational activities for young adults with autism spectrum disorders during the transition to adulthood. *Journal of Autism and Developmental Disorders, 41*(5), 566–574. https://doi.org/10.1007/s10803-010-1070-3

Taylor, O.D., Armfield, N.R., Dodrill, P., & Smith, A.C. (2014). A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment. *Journal of Telemedicine and Telecare*, *20*(7), 405-412. https://doi.org/10.1177%2F1357633X14552388

Tek, S., Mesite, L., Fein, D., & Naigles, L. (2014). Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(1), 75-89. https://doi.org/10.1007/s10803-013-1853-4

Thomas, H. R., Rooney, T., Cohen, M., Bishop, S. L., Lord, C., & Kim, S. H. (2020). Spontaneous expressive language profiles in a clinically ascertained sample of children with Autism Spectrum Disorder. *Autism Research*. https://doi.org/10.1002/aur.2408

Thurm, A., Lord, C., Lee, L. C., & Newschaffer, C. (2007). Predictors of language acquisition in preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(9), 1721-1734. https://doi.org/10.1007/s10803-006-0300-1

Tomblin, J.B., Records, N.L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245-1260. https://doi.org/10.1044/jslhr.4006.1245

Tomblin, J.B., Records, N.L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 39*(6), 1284-1294. https://doi.org/10.1044/jshr.3906.1284

Valicenti-McDermott, M., Hottinger, K., Seijo, R., & Shulman, L. (2012). Age at diagnosis of autism spectrum disorders. *The Journal of Pediatrics, 161*(3), 554–556. https://doi.org/10.1016/ j.jpeds.2012.05.012

Waite, M.C., Cahill, L.M., Theodoras, D.G., Busuttin, S., & Russell, T.G. (2006). A pilot study of online assessment of childhood speech disorders. *Journal of Telemedicine and Telecare*, *12*(3_suppl), 92-94. https://doi.org/10.1258/135763306779380048

Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010a). Assessment of children's literacy via an internet-based telehealth system. *Telemedicine and e-Health*, *16*(5), 564-575. https://doi.org/10.1089/tmj.2009.0161

Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010b). Internet-based telehealth assessment of language using the CELF–4. *Language, Speech, and Hearing Services in Schools*. https://doi.org/10.1044/0161-1461(2009/08-0131)

Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2012). Assessing children's speech intelligibility and oral structures, and functions via an Internet-based telehealth system. *Journal of Telemedicine and Telecare*, *18*(4), 198-203. https://doi.org/10.1258/jtt.2012.111116

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. The Psychological Corporation.

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals-Fifth Edition (CELF-5): Technical manual.* Pearson.

Williams, K. T. (2019). *Expressive Vocabulary Test, Third Edition: Manual*. Pearson.

Williams, D., Botting, N. and Boucher, J. (2008). Language in autism and specific language impairment: Where are the links? *Psychological Bulletin*, *134*(6), 944-963. https://doi.org/10.1037/0013743

Winter, A. S., Fountain, C., Cheslack-Postava, K., & Bearman, P. S. (2020). The social patterning of autism diagnoses reversed in California between 1992 and 2018. *Proceedings of the National Academy of Sciences*, *117*(48), 30295-30302. https://doi.org/10.1073/pnas.2015762117

Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., & Naigles, L. R. (2017). Grammatical language impairment in autism spectrum disorder: Exploring language phenotypes beyond standardized testing. *Frontiers in Psychology*, *8*, 532. https://doi.org/10.3389/fpsyg.2017.00532

Woynaroski, T., Yoder, P., & Watson, L. R. (2015). Atypical cross- modal profiles and longitudinal associations between vocabulary scores in initially minimally verbal children with ASD. *Autism Research, 9*, 301–310. https://doi.org/10.1002/aur.1516