Language impairment in autistic adolescents and young adults: Variability by definition

Teresa Girolamo[1], Lindsay Butler[2], Julia Parish-Morris[3,4]

[1]San Diego State University, [2]University of Connecticut, [3]University of Pennsylvania School of Medicine, [4]Children's Hospital of Philadelphia

Key words: language impairment, autism, adolescents, adults, language assessment

Author Note

CRediT statement: TG: conceptualization, methodology, formal analysis, investigation, resources, data curation, writing – original draft, project administration, funding acquisition. LB: formal analysis, data curation, writing – review & editing, visualization. JPM: interpretation, writing – review and editing

Author Information: Teresa Girolamo, School of Speech, Language, and Hearing Sciences, San Diego, CA, San Diego State University. Lindsay Butler, Department of Speech, Language, and Hearing Sciences, Storrs, CT, University of Connecticut.

Funding: TG was supported by an ASHFoundation New Investigators Research Grant (PI: Girolamo). Research reported in this publication was supported by the National Institute on Deafness and other Communication Disorders of the National Institutes of Health under award number R21DC021769-01A. This award supported authors Girolamo and Butler to conduct this research. Research reported in this publication was supported by the National Institute on Deafness and other Communication Disorders of the National Institutes of Health under award number R01DC018289 (PI: Parish-Morris). This award supported co-author Parish-Morris to conduct this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Corresponding Author: Teresa Girolamo, 5201 Campanile Mall, San Diego State University, San Diego, CA, 07103. Phone: 619-594-4024. Email: tgirolamo@sdsu.edu.

Conflict of Interest: The authors have no financial or nonfinancial disclosures.

Data Availability Statement: Data are unavailable due to ethical restrictions. Information on data structure and analysis can be shared upon reasonable request to the corresponding author.

Abstract

Purpose: Co-occurring language impairment (LI) in autism is common and predicts long-term academic, occupational, and social outcomes. Yet little is known about LI in autism beyond childhood. One challenge to closing this gap is the lack of consensus regarding how LI should be operationally defined in adolescents and adults. This study examines how different epidemiological definitions of LI influence clinical classification and observed language profiles in speaking autistic adolescents and young adults.

Method: Participants (*N* = 75; ages 13-30) varying in levels of autism traits completed norm-referenced measures of overall expressive language, overall receptive language, receptive vocabulary, expressive vocabulary, nonword repetition, and nonverbal intelligence. Scores were compared to epidemiological definitions for LI varying in criteria and cutoffs from -1 *SD* to -1.5 *SD*. Data were analyzed using descriptives and clustering.

Results: More stringent definitions yielded a greater proportion of participants meeting LI criteria, and more stringent cutoffs for LI yielded greater overall consistency in clinical classification across individual language measures, but there was no one-to-one ratio between cutoff and clinical classification. Clustering indicated three profiles differentiated on the basis of language and nonverbal cognitive skills, but each cluster was heterogeneous. Individual performance also varied across language measures.

Discussion: Findings support multi-domain approaches to characterizing language skills in autistic adolescents and adults, including those with LI. Future work is needed to understand language skills in autism beyond childhood and how to develop effective assessment practices.

Language impairment in autistic adolescents and young adults: Variability by definition

There is growing evidence of heterogeneity in language abilities across the autism spectrum (Bal et al., 2016; Butler et al., 2023; Pizzano et al., 2024), but little is known about language beyond childhood (Howlin & Taylor, 2015). Language skills are associated with long-term academic, occupational, and social outcomes (Brignell et al., 2018; Magiati et al., 2014), underscoring the importance of accurately characterizing language skills across development. In the United States, the transition to adulthood is a key developmental stage, as autistic youth age out of child- and school-based services ("Individuals with Disabilities Education Improvement Act [IDEIA] of 2004," 2018). Many lack access to services and supports (Eilenberg et al., 2019; Roux et al., 2024; Weir et al., 2022). Despite these challenges, language in autistic adolescents and young adults remains poorly understood.

More than 50% of autistic individuals have structural language impairment (Boucher, 2012), and over 60% show broader language delays (Baird et al., 2006; Levy et al., 2010). However, operational definitions of language impairment (LI) vary substantially (Girolamo et al., 2023), limiting comparisons across studies and developmental stages. In this report, LI refers to difficulties in one or more domains of structural language (e.g., morphosyntax, semantics, phonology) (Schaeffer et al., 2023), and is treated as a construct versus a diagnostic category. This distinction is important, as existing classification systems provide limited operational guidance for identifying LI in older individuals. The *Diagnostic and Statistical Manual of Mental Disorders, 5th Ed.* (DSM-5) includes LI as a specifier but does not define measurement criteria or clarify how LI should be operationalized (American Psychiatric Association [APA], 2013). Similarly, the *International Classification of Diseases, 11th Revision* (ICD-11) includes severity qualifiers for functional language impairment without specifying assessment procedures

(World Health Organization [WHO], 2022).

As a result, the literature uses varying assessment batteries and cutoff thresholds (Girolamo et al., 2023; Koegel et al., 2020; Kwok et al., 2015; Magiati et al., 2014), often relying on too little information, such as using verbal IQ as an indicator of overall cognitive ability (McCauley et al., 2020), which confounds language with cognition (Grondhuis et al., 2018), or using a label like “LI” as an indicator of language skills across domains (Schaeffer et al., 2023). These practices complicate the interpretation and comparison of findings across studies and influence how language abilities are documented. A secondary implication of lack of operational guidance is that after publication of the DSM-5 (APA, 2013), students with a primary disability label of autism may not receive speech-language evaluation or access to services (Musgrove, 2015), even though U.S. legislation mandates access to services in all areas associated with a disability and their unique needs (“IDEIA of 2004,” 2018).

The present study examines how different epidemiological definitions of LI influence classification outcomes and observed language profiles in speaking autistic adolescents and young adults. By applying multiple empirical definitions to the same sample, a focus is on how measurement decisions influence classification and interpretation. Consistent with recent calls for multidimensional approaches to assessment (Kover & Abbeduto, 2023; Schaeffer et al., 2023), this study evaluates language performance across multiple language domains and examines how classification differs across definitions. To better understand heterogeneity within this population, this study also explores patterns of language and nonverbal intelligence.

Structural LI in Autistic Adolescents and Adults

Characterizing LI in autism beyond childhood presents both conceptual and methodological challenges. While there are validated clinical language assessments for youth

(Girolamo et al., 2022; Nitido & Plante, 2020), far fewer measures are normed for adults over age 21 (Manenti et al., 2023). Understanding performance on norm-referenced assessments is important, as they are often used to document language skills in service systems (Burke et al., 2024; Selin et al., 2019). In practice, studies often use assessments designed for other clinical populations. These practices introduce uncertainty regarding interpretation of receptive and expressive language profiles, as operational definitions of LI use specific cutoffs.

Limitations of available adult language measures further complicate interpretation. Some assessments were normed decades ago on non-representative samples, such as adults with a college education (Hammill et al., 2007), limiting the generalizability of norms. Receptive and expressive vocabulary tests normed for younger populations have yielded age equivalent scores outside basal and ceiling ranges when administered to autistic adults ranging widely in age (ages 21-64 years), leaving it unclear how to interpret assessment results (Howlin et al., 2004, 2014; Mawhood et al., 2000). At the same time, reliance on a single language domain, such as vocabulary, may underestimate broader structural language difficulties, independent of autism (Calder et al., 2023; Schaeffer et al., 2023). Elsewhere, omnibus (i.e., overall receptive-expressive) language batteries developed for adults with acquired language disorders (e.g., aphasia) have been applied to autistic adults without brain injury (Lewis et al., 2008), raising questions about construct validity. These examples illustrate the limitations of available tools for measuring LI as a multidimensional construct.

A second challenge involves variability in how LI is operationalized in adolescent and adults. Studies have used differing combinations of measures and cutoffs, often extending norms to those over age 21 years. Definitions have ranged from -2 *SD* on omnibus language composite (Botting, 2020), to -1.5 *SD* on omnibus language composite (Clegg et al., 2021; Norbury et al.,

2016), to -1 *SD* on receptive vocabulary or omnibus language composite scores, or -2 *SD* on ≥ 2 subtests of an omnibus language measure (Johnson et al., 1999b; Poll et al., 2010). In autistic samples, LI has been also defined using -1.25 *SD* on two or more measures spanning overall receptive language, overall expressive language, vocabulary, and phonological domains (Girolamo & Rice, 2022), consistent with epidemiological approaches developed in nonautistic youth (Tomblin et al., 1996). Across studies, there is no consensus on which measures should be included, which cutoffs are most appropriate, or how norms should be applied in adulthood. Characteristics of prior adult samples are summarized in Supplementary Table 1.

Independent validation studies also demonstrate that there is no single operational definition for LI in adulthood. In the Ottawa longitudinal study, differences between test norming samples and the study cohort motivated the development of local norms to define LI, inclusive of autism (Johnson et al., 1999b). A -1 *SD* cutoff on receptive vocabulary or omnibus language yielded higher LI estimates (12.6%) than expert clinical judgment (8.3%) (Johnson et al., 1999a), emphasizing how classification thresholds can meaningfully alter prevalence estimates. In a separate study, multi-domain combinations of phonological, semantic, and syntactic measures yielded the highest classification accuracy for developmental LI in nonautistic adults (Fidler et al., 2011, 2013). In contrast, single measures of narration, nonword repetition, sentence production, and grammatical judgment were inconsistent in identifying LI (Fidler et al., 2011). While narration can confound structural language skills with autism traits (Baixauli et al., 2016), prior work suggests that LI and autism traits are separable constructs (Loucas et al., 2008; Manenti et al., 2024; Silleresi et al., 2020). Together, findings support use of multi-domain, multi-method assessment.

A related methodological issue concerns the use of intelligence quotient (IQ) cutoffs in

studies of LI in autism. Many studies exclude individuals below a specified IQ threshold (Girolamo et al., 2023), despite current diagnostic frameworks allowing co-occurring autism and intellectual disability (APA, 2013; WHO, 2022). The use of IQ exclusion criteria may constrain observed patterns of LI and limit representation of the full autism spectrum (Russell et al., 2019; Shaw et al., 2025). To this end, recent work in autistic adults demonstrated no one-to-one correspondence between structural language and nonverbal IQ (NVIQ), with heterogeneous profiles indicating individuals with LI and average of high NVIQ (≥110), as well as individuals with lower NVIQ (<80) and without LI (Manenti et al., 2024). Importantly, LI in that study was defined using cutoff thresholds (i.e., moderate at -1.25 *SD* and severe at -1.65 *SD*) on non-normed nonword repetition and sentence repetition measures relative to a comparison sample of non-autistic adults (Manenti et al., 2024), illustrating how operational decisions influence observed patterns. Overall, variability in LI cutoffs, inclusion criteria, and measurement choices motivates a dimension approach to characterizing language in autistic adolescents and adults.

The Current Study

In the absence of consensus of how to operationalize LI in adolescence and adulthood, examining how classification varies by definition can clarify the influence of measurement decisions. This exploratory, observational study applies established epidemiological definitions of LI to the same sample of speaking autistic adolescents and young adults. By varying only the operational criteria in terms of measures included and cutoff thresholds, this approach serves to isolate how different definitions shape classification outcomes and interpretation. Consistent with prior epidemiological work (Johnson et al., 1999a; Norbury et al., 2016; Tomblin, 1996), this study evaluates LI using normed measures across multiple language domains, including omnibus expressive and receptive language, vocabulary, and phonology. Following current

diagnostic frameworks that recognize variability across domains (APA, 2013; WHO, 2022), this study did not use NVIQ exclusion criteria. Research questions were:

1) To what extent does clinical classification of LI differ by definition in speaking autistic adolescents and young adults: (a) -1 *SD* on omnibus language or receptive vocabulary measures, or -2 *SD* on ≥2 subtests of an omnibus language measure (Johnson et al., 1999a); (b) -1.25 *SD* on ≥2 measures of overall receptive language, overall expressive language, receptive vocabulary, expressive vocabulary, and nonword repetition, or on an omnibus language measure (Tomblin et al., 1996); and c) -1.5 *SD* on ≥2 measures from (b) (Norbury et al., 2016), and by individual language measure?
2) How does performance on language measures and NVIQ relate within this population?

Based on prior validation studies indicating that stricter cutoffs yield greater concordance with clinical judgment (Johnson et al., 1999a), it was expected that more stringent definitions would yield lower LI estimates. Given evidence that vocabulary may be a relative linguistic strength in autism and may not fully capture structural language difficulties (Arunachalam & Luyster, 2016; Schaeffer et al., 2023), it was expected that fewer participants would meet LI cutoffs on receptive and expressive vocabulary relative to broader language measures. Finally, consistent with recent work showing dissociation between language and NVIQ in autistic adults (Manenti et al., 2024), it was expected that participants would show heterogeneous profiles.

Methods

Participants

Participants were speaking autistic adolescents and young adults aged 13 to 30 years. This age range was selected to capture the transition to adulthood period, including adolescents approaching transition planning and young adults beyond eligibility for school-based services

("Every Student Succeeds Act," 2015; "IDEIA of 2004," 2018). Inclusion criteria were: (a) meeting DSM-5 criteria for autism spectrum disorder (APA, 2013), confirmed through a clinical best estimate (CBE) approach (Bishop & Lord, 2023); (b) age between 13 and 30 years; (c) monolingual English speaker, as study activities were in English; (d) adequate hearing and vision for completing audiovisual assessment tasks; and (e) primary use of spoken language to communicate, as activities required verbal responses. Autism diagnosis was confirmed using a CBE procedure that triangulated documentation of a prior professional autism diagnosis, Social Responsiveness Scale-2 (SRS-2) (Constantino & Grubler, 2012) overall *t*-scores, and expert clinical judgment by a trained clinician. SRS-2 caregiver and self-report forms were combined, because mean overall T-scores did not significantly differ across form types ($p = .061$): caregiver student forms ($n = 37$), caregiver adult forms ($n = 15$), and adult self-report forms ($n = 21$).

Procedures

This study received institutional review board approval from San Diego State University (HS-2023-0225-SMT) and recruited participants using a multi-step process. The team shared study flyers with organizations providing services to autistic individuals, provided personalized consultation about the study to potential participants and caregivers, and obtained informed consent and assent prior to data collection. Participants provided informed consent if they were their own legal guardian; otherwise, caregivers provided consent and participants provided assent. Recruitment and data collection took place from 2022 to 2024 remotely on HIPAA-compliant Zoom. Participants completed assessments using the digital version of assessments, in accordance with test publisher telepractice guidance (Pearson, 2025). Examiners followed manualized administration procedures and completed training prior to data collection. Language and NVIQ measures were administered in a standardized order within one or more sessions,

depending on participant fatigue and scheduling needs. A trained research assistant independently verified scoring accuracy for all measures, except those automatically scored through publisher platforms (e.g., SRS-2, Raven's 2).

Measures

Demographics

Participants or caregivers reported participant race and ethnicity using U.S. Census categories (Office of Management and Budget, 1997). Respondents could select multiple categories or provide write-in responses. Respondents also reported participant sex assigned at birth and gender, with optional write-in responses.

Autism Traits

Autism traits were assessed using SRS-2 (Constantino & Grubler, 2012) caregiver or self-report forms as appropriate for age. The SRS-2 consists of 65 items rated on a four-point Likert scale and yields standardized T-scores (M = 50, SD = 10), reflecting social communication and interaction and restricted interests and repetitive behaviors. Internal consistency for forms ranges from α = .95 to .97. T-scores are categorized as subclinical (59), mild (60-65), moderate (66-75), and severe (≥76). Per study procedures, SRS-2 scores were used as part of the CBE process for confirming autism status and were also analyzed dimensionally.

Language Skills

Language was assessed across multiple domains using norm-referenced, untimed, direct behavioral assessments. Omnibus language (a composite measure derived from multiple subtests) was assessed using six subtests of the Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5) (Wiig et al., 2013). The CELF-5 provides scaled scores (M = 10, SD = 3) and composite indices (M = 100, SD = 15) for expressive, receptive, and overall expressive-receptive

language (composite reliability *r* = .95-.96). Expressive subtests included: Formulated Sentences (sentence production; *r* = .86), Recalling Sentences (sentence repetition; *r* = .94), and Semantic Relationships (sentence-level semantic processing; *r* = .89). Receptive subtests included: Word Classes (semantic associations; *r*: .90), Understanding Spoken Paragraphs (discourse-level comprehension; *r* = .85), and Sentence Assembly (sentence-level integration; *r* = .93). For participants over age 21 (*n* = 24), age 21 norms were used, consistent with prior work (Botting, 2020; Clegg et al., 2021; Fidler et al., 2011).

Receptive vocabulary was assessed using the Peabody Picture Vocabulary Test–Fifth Edition (Dunn, 2019), and expressive vocabulary using the Expressive Vocabulary Test–Third Edition (EVT-3) (Williams, 2019). These co-normed measures provide standard sores (*M* = 100, *SD* = 15; reliability *r* = .97) and are normed from early childhood through >90 years. The PPVT-5 requires participants to select one of four images corresponding to a spoken word. The EVT-3 requires participants to generate a single-word response to an auditory question and visual stimulus.

Phonological processing was assessed using the Syllable Repetition Task (SRT) (Shriberg et al., 2009a). Participants repeat 18 nonwords of two to four syllables, which use early-acquired phonemes to minimize articulation and speech motor confounds (/b/, /d/, /m/, /n/, and /a/) (Shriberg et al., 2009a). Percent accuracy was used as the primary outcome variable. Reliability for 16-year-olds, the oldest norming group, ranges from *r* = .69 to .89 (*M* = 93.2, *SD* = 3.0) (Shriberg et al., 2009b; Shriberg & Mabie, 2017).

NVIQ

NVIQ was assessed using the Raven's Progressive Matrices–Second Edition (Raven et al., 2018), long form (*M* = 100, *SD* = 15; reliability *r* = .88-.91). The Raven's is untimed and

minimizes verbal demands by requiring participants to identify the missing element in a visual matrix without any spoken instructions. This measure reduces confounding of language ability with cognitive assessment (Grondhuis et al., 2018).

Data Processing and Analysis

All analyses were conducted using SPSS 31.0 (IBM Corp., 2025) using an α level of .05. Because several variables violated assumptions of normality, nonparametric measures were used.

Data Preparation

A trained research assistant independently checked scoring for all measures except those automatically scored by publisher platforms (e.g., SRS-2, Raven's 2). Next, data were examined for missingness. As missing data were minimal (<5%), autism trait scores were replaced using single imputation with predictive mean matching (Little & Rubin, 2019). One participant missing both language and NVIQ scores was excluded from analyses. Prior to analysis, variables were examined for outliers using standardized *z*-scores ($|z| > 3.29$), boxplots, and Mahalanobis distance for multivariate analyses. No outliers were identified. All observations were retained to maintain representation of variability within the sample. For analyses examining relationships across measures, language, autism trait, and NVIQ scores were standardized (*z*-scored) to allow comparability across scales (Jolliffe & Cadima, 2016).

Analyses

To examine whether LI classification differed by definition, participants were classified (yes/no) under each of three epidemiological definitions: (a) -1 *SD* on CELF-5 core language or PPVT-5, or < -2 *SD* on ≥2 CELF-5 core language subtests (Johnson et al., 1999a); (b) -1.25 *SD* on ≥2 measures: CELF-5 Receptive Language Index, CELF-5 Expressive Language Index, PPVT-5, EVT-3, and SRT accuracy (Tomblin et al., 1996); and (c) -1.5 *SD* on ≥2 measures from

(b) (Norbury et al., 2016). Cochran's Q tests (Cochran, 1950) evaluated differences in classification rates. To test if LI classification differed by individual language measure, participants were also classified at -1 *SD*, -1.25 *SD*, or -1.5 *SD* thresholds on each language measure: PPVT-5, EVT-3, SRT accuracy, and CELF-5 subtests (Formulated Sentences, Recalling Sentences, Semantic Relationships, Word Classes, Understanding Spoken Paragraphs, Sentence Assembly). Pairwise comparisons were conducted using Dunn's (1964) procedure, with a Bonferroni (1936) correction for multiple comparisons.

To explore patterns among language, autism trait, language, and NVIQ, analysis used principal component analysis (PCA) followed by clustering. Given sample size and to avoid selective reporting (Dalmaijer et al., 2022), analyses were exploratory and descriptive (Husson et al., 2010). Prior to PCA, assumptions of linearity, sampling adequacy, multicollinearity, and multivariate outliers were evaluated (variance inflation factor ≥ 10). No cases were removed. PCA was conducted on *z*-scored SRS-2 domains, CELF-5 subtests, PPVT-5, EVT-3, and SRT scores (Abdi & Williams, 2010). Component retention was guided by eigenvalues > 1, scree plot inspection (Cattell, 1966), and interpretability. Varimax rotation was used to aid interpretation.

As autism traits loaded weakly on the primary component, clustering analyses were conducted on language and NVIQ variables only. Agglomerative hierarchical clustering using Ward's method (1963) and squared Euclidean distance was conducted first to examine cluster structure (Bouguettaya et al., 2015; Day & Edelsbrunner, 1984). Inspection of the dendrogram and agglomerative schedule informed the number of solutions and initial centers entered into *k*-means clustering (Hartigan & Wong, 1979; Lu et al., 2008). *K*-means clustering was conducted with 999 iterations and no running means to ensure stability of cluster assignment. Internal cluster validity was evaluated using the KO_MACROS (Orlov, 2024). External validity was

evaluated via discriminant function analysis and Cohen's κ (1960) comparing *k*-means and hierarchical solutions (Dudoit & Fridlyand, 2002). Patterns were interpreted descriptively based on relative mean *z*-scores across language and NVIQ measures. Effects were interpreted as: small at .25, medium at .55, and large at .95 (Gaeta & Brydges, 2020). Because clustering was exploratory and sample dependent, findings were interpreted as illustrative of heterogeneity versus evidence of discrete profiles. Participants were classified into NVIQ bands based on test developer guidance and clinical criteria indicating that an IQ of 70 to 75 represents clinically significant differences (American Association on Intellectual and Developmental Disabilities, 2024; Raven et al., 2018): low (≤75), borderline (76-89), average (90-109), and high (≥110).

Results

Sample Characteristics

The final sample included 75 participants aged 13 to 30 years; see Tables 1 and 2. One participant missing both language and NVIQ scores was excluded from analysis. SRS-2 scores were missing for two participants. Given minimal missingness (Little & Rubin, 2019), scores were imputed using predictive means matching. Participants represented diverse racial and ethnic backgrounds. The sample had a female-to-male ratio for sex at birth and gender of one to 2-2.1 in the overall sample, one to 2.2-2.3 among participants 21 years or younger, and of one to 1.6 among participants over age 21. The mean SRS-2 overall T-score fell within the "moderate" range, indicating clinically elevated autism traits.

LI Classification by Definitions and Measure

Clinical classification of LI differed significantly by definition, $\chi^2(2)=19.54$, $p < .0001$. Classification rates were: (a) 70.27% using the -1 *SD* on CELF-5 core language score or PPVT-5 scores, or -2 *SD* on ≥2 CELF-5 core language subtests (Johnson et al., 1999a); (b) 62.16% using

the -1.25 *SD* on ≥2 measures (CELF-5 Receptive Language Index, CELF-5 Expressive Language Index, PPVT-5, EVT-3, and SRT overall accuracy) (Tomblin et al., 1996); and (c) 52.70% using the -1.5 *SD* on ≥2 measures from definition (b) (Norbury et al., 2016). Yet only the decrease from (a) to (c) was significant (17.57%), $p < .0001$. Thus, greater cutoff stringency was associated with lower LI classification, but differences were not uniformly proportional across cutoff thresholds.

More stringent cutoffs generally yielded greater consistency in LI classification across language measures, as indicated by fewer significant pairwise comparisons: -1 *SD* ($n = 18$), -1.25 *SD* ($n = 13$), and -1.5 *SD* ($n = 9$); see Supplementary Table 2 and Supplementary Figure 1. However, classification varied by measure; see Figure 1. CELF-5 Understanding Spoken Paragraphs yielded higher LI classification rates at less stringent thresholds (-1 *SD* and -1.25 *SD*) relative to several other measures, such as EVT-3 (Δ: 33.78-40.54%), PPVT-5 (Δ: 25.68-35.13%), CELF-5 Sentence Assembly (Δ: 18.92-29.73%), and CELF-5 Word Classes (Δ: 21.62-32.43%). In contrast, EVT-3 yielded comparatively lower LI classification rates across cutoffs, including CELF-5 Sentence Repetition (Δ: 20.27-21.62%). Across cutoffs, no single measure consistently classified participants as meeting or not meeting LI definitional criteria. Findings indicate that LI classification is sensitive to cutoff and to specific language domains assessed.

Language Performance in NVIQ Patterns

PCA assumptions were met. All variables had at least one $r > 0.3$, sampling was adequate (KMO = 0.89) (Kaiser, 1974), and Bartlett's (1951) test was significant ($p < .0001$). Two components with eigenvalues >1 accounted for 73.28% of the total variance. Following Varimax rotation, language measures and NVIQ loaded strongly on Component 1 (50.32%), while autism traits loaded strongly on Component 2 (24.99%); see Supplementary Table 3 and Supplementary

Figure 2. Because autism traits weakly associated with the primary language and NVIQ component ($r < .16$), clustering analyses were run on only language and NVIQ variables.

Hierarchical clustering suggested a three-cluster solution, which was confirmed using *k*-means clustering; see Supplementary Figure 3 and Supplementary Table 4. Internal validity metrics and near-perfect agreement between clustering methods, $\kappa = .873$, 95% CI [.775, .971], $p < .0001$, supported selection of the three-cluster solution; see Supplementary Table 5. Discriminant function analysis classified 93.24% of cases currently under cross-validation. Of 14 participants expected to be in Cluster 1, one was in Cluster 3; of 31 expected to be in Cluster 2, one was in Cluster 3, and; of 29 expected to be in Cluster 3, three were in Cluster 2.

Clusters differed descriptively in mean language, NVIQ, and sex at birth distribution; see Table 3. Cluster 1 ($n = 14$) demonstrated the lowest mean language scores (< -2 *SD*) and NVIQ scores (< -1.5 *SD*). Cluster 2 ($n = 29$) had language scores (< -1 *SD*) and NVIQ slightly below average (~-1 *SD*). Cluster 3 ($n = 31$) had language scores at or above the mean (≥ -1 *SD*) and average NVIQ. Despite differences in cluster means, there was substantial within-cluster variability; see Figure 2 and Supplementary Figure 4. When examining clinical categories descriptively, there was overlap between LI and NVIQ levels across clusters. Participants meeting more stringent LI criteria (-1.5 *SD* on ≥2 measures) were distributed across NVIQ bands, including low NVIQ (Cluster 1: $n = 6$, Cluster 2: $n = 5$), borderline NVIQ (Cluster 1: $n = 5$, Cluster 2: $n = 4$, Cluster 3: $n = 1$), and average NVIQ (Cluster 1: $n = 3$, Cluster 2: $n = 11$, Cluster 3: $n = 3$). A small number of participants with high NVIQ also met LI criteria at the -1.25 *SD* on ≥2 measures threshold (Cluster 2: $n = 3$, Cluster 3: $n = 1$). These distributions indicate that classification did not align completely with NVIQ category.

There was also observed within-cluster variability. In Cluster 1, most participants scored

< -1.5 *SD* on all language measures, with 64.29% (*n* = 9) scoring < -1.5 *SD* across all language measures and 21.43% (*n* = 3) scoring < -1.5 *SD* on all but two of nine measures. However, some participants showed more selective profiles, with relative strengths on vocabulary (EVT-3, PPVT-5) or sentence-level tasks (e.g., CELF-5 Formulated Sentences). Similarly, although most Cluster 2 participants met LI criteria at the < -1.5 *SD* on ≥2 measures (*n* = 26, or 89.66%), some met criteria only at the -1.25 *SD* threshold or showed mixed profiles across measures. In Cluster 3, over two-thirds of participants scored within the typical range on all language measures (*n* = 11, or 35.48%) or all but one measure (*n* = 12, or 38.71%). A subset showed relatively lower scores at -1.25 *SD* or -1.5 *SD* on two to three measures. Across clusters, individual performance varied across language measures and domains, indicating heterogeneity even within groups defined by similar mean language and NVIQ levels.

Discussion

This study examined how different epidemiological definitions of LI influence classification outcomes and observed language profiles in a sample of speaking autistic adolescents and young adults. Findings indicate that LI classification varied depending upon cutoff, measures, and specific language domains assessed. Exploratory multivariate analyses further supported heterogeneity in profiles, including dissociation between structural language performance and NVIQ in a subset of participants.

LI Classification and Assessment Performance

LI definitions with more stringent criteria yielded lower LI estimates. However, decreases were not proportional across thresholds, and no single language measures consistently identified participants as meeting LI criteria. These findings align with prior validation studies demonstrating that operational thresholds and measure selection influence LI classification in

adulthood (Fidler et al., 2011; Johnson et al., 1999a). LI classification rates in the present sample (52.70%-70.27%) were comparable to epidemiological estimates of structural LI (>50%) and overall LI (63%) in autistic youth (ages 8-10 years) (Baird et al., 2006; Levy et al., 2010). However, they were higher than those in a recent adult sample (ages 18-56 years; 26.32%) (Manenti et al., 2024). Differences may reflect several factors: the younger age range in this sample (13-30 years), use of norm-referenced measures rather than control-based *z*-scores, broader inclusion of language domains beyond sentence repetition and nonword repetition tasks, and use of an NVIQ measure that did not require use of language. Notably, the 70.27% rate under the -1 *SD* definition (Johnson et al., 1999a) aligns with prior work showing that local norms near this threshold may yield higher LI estimates than expert clinical judgment, particularly when impairment is not present across multiple domains (Johnson et al., 1999b). Similarly, validation studies in non-autistic adults demonstrate that individual measures vary in discriminatory ability and that multi-domain combinations outperform single measures (Fidler et al., 2011), with adult samples not replicating adolescent-based definitions (Tomblin, 2008). Together, these findings suggest that variation in LI estimates across studies reflects operational decisions rather than just developmental change.

Although more stringent cutoffs reduced LI classification rates in this study, decreases were not uniform across thresholds or measure. Vocabulary measures yielded lower LI classification rates relative to omnibus expressive-receptive language indices and sentence-level tasks, consistent with concerns that single-domain measures may underestimate broader structural language difficulties (Schaeffer et al., 2023). However, domain-level differences were not consistently significant, and their clinical relevance remains unclear. Determining clinical significance requires a “gold standard” definition for LI in adults, which remains elusive (Fidler

et al., 2011; Howlin & Taylor, 2015), and corresponding measurement approaches.

Patterns of language and NVIQ further illustrate heterogeneity. Consistent with prior work (Manenti et al., 2024; Silleresi et al., 2020), language performance and NVIQ did not align uniformly. Descriptively, participants showed different patterns from Manenti et al. (2024), where no autistic adults showed high NVIQ and LI. A small number of participants with high NVIQ met LI criteria, whereas others demonstrated strengths across several language domains. Here, two participants with high NVIQ met LI criteria at < -1.25 *SD* and < -1.5 *SD* on 2 or more measures: CELF-5 Formulated Sentences, Recalling Sentences, Semantic Relationships, and Understanding Spoken Paragraphs. While the first three measures are sensitive to LI in nonautistic youth (Calder et al., 2023), their clinical utility could differ for older autistic individuals, as tasks like sentence production require pragmatic skills (Wiig et al., 2013). Overall, these patterns support a dimensional perspective on language in autism and caution against assuming a singular operational definition of LI beyond childhood within the field.

Implications for Evidence-Based Practice

Findings have implications for practitioners and researchers who interpret language assessment results in autistic adolescents and adults, with implications for building the evidence base (Durkin et al., 2015). While relevant for service eligibility (Selin et al., 2019), performance on norm-referenced measures indicated that classification outcomes were sensitive to operational definitions for LI. No single measure consistently identified LI across definitions. Contrary to expectations, receptive vocabulary and nonword repetition each had two discrepant classification outcomes from other language measures at -1 *SD* and -1.25 *SD*. Sentence repetition also had more discrepant classification outcomes – and the same number of discrepant classification outcomes as expressive vocabulary – at -1 *SD* than at -1.25 *SD* or -1.5 *SD*.

Results caution against relying on a single measure to make inferences about language skills, consistent with best-practice recommendations emphasizing converging evidence (Sackett et al., 1996). Structural language performance may vary independently of autism traits and NVIQ, as observed in this sample and prior work (Manenti et al., 2024; Silleresi et al., 2020). Though such heterogeneity is not new information, it is important to document that it exists. In contexts where assessment informs documentation and eligibility decisions (Musgrove, 2015), awareness of how definitions and measurement choices influence classification may contribute to more precise interpretation of results, and ultimately, access to services. At the same time, standardized assessments capture only one component of language skills and must consider personal goals and desired supports for language and communication (Burke et al., 2024; Cummins et al., 2020). Interpretation of language performance should consider domains assessed, cutoff, and the broader assessment context, especially given the limited availability of adult-normed language measures.

Limitations

This study encountered several limitations. First, the sample size was modest, and clustering analyses were exploratory. Hence, findings are highly sample dependent and may not replicate (Cox & Sosine, 2023). Although internal and external validity indices supported the three-cluster solution, replication in larger, multi-site samples is needed before making population-level inferences through age-controlled groups or replication cohorts. Second, the CELF-5 was normed only through age 21. While use of age 21 norms in this study aligned with previous work (Botting, 2020; Clegg et al., 2021; Fidler et al., 2011; Poll et al., 2010), the lack of a consensus definition of LI and of adult-normed language measures limit precision in interpreting language performance beyond early adulthood and calculating psychometrics (e.g.,

sensitivity, specificity). Third, the study included speaking participants with a professional autism diagnosis, which hinders generalizability. Autistic individuals who use minimal spoken language have clinically relevant variability in their language skills (Bal et al., 2016; Butler et al., 2023), but language assessments requiring verbal responses are not designed for all communication profiles. Finally, this study used direct standardized measures, which do not fully capture functional communication across contexts (Barokova et al., 2020; Butler et al., 2022). Beyond standardized measures, integrating functional and person-centered, naturalistic measures, as well as medical and intervention history, is needed for a more comprehensive understanding of language abilities in adolescence and adulthood (Sackett et al., 1996).

Future Directions

Future research should prioritize development and validation of language measures normed for autistic adults, including those without a professional diagnosis. Limited availability of adult-appropriate assessments constrains research comparisons and interpretation. Studies comparing direct assessment with clinical judgment, functional communication, and person-centered measures may help clarify LI beyond childhood and how to assess it (Johnson et al., 1999a; Schaeffer et al., 2023). Given evidence that language, NVIQ, and autism traits may vary independently (Manenti et al., 2024), larger and more diverse samples are needed to examine patterns across adulthood. Such sampling would enable robust analytic approaches, including hold-out samples or replication cohorts (Lombardo et al., 2019), would strengthen confidence in classification frameworks and facilitate building consensus toward a definition for LI beyond childhood. Finally, integrating direct clinical assessment with interaction-based measures may clarify how language skills relate to communication experiences and outcomes (Crompton et al., 2020; Nikolaus & Fourtassi, 2023). Such would support precision-driven approaches to

assessment and developing supports that better reflect the priorities of autistic individuals across the lifespan.

Conclusion

In evaluating language assessment across linguistic domains in autistic adolescents and young adults, this study showed that clinical classification of LI differed by operational definition, measure, and cutoff severity. Levels of language and NVIQ dissociated in approximately one-third of the sample. Exploratory analysis indicated heterogeneity in sample performance, even within participants grouped on the basis of language and NVIQ levels. Overall, findings support use of multi-domain, multi-measure language assessment and a need for adult clinical language measures.

Data Availability Statement

Data are unavailable due to ethical restrictions. Information on data structure and analysis can be shared upon reasonable request to the corresponding author and IRB approval.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

American Association on Intellectual and Developmental Disabilities. (2024). *Defining criteria for intellectual disability*. Retrieved June 2, 2025, from https://www.aaidd.org/intellectual-disability/definition

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

Arunachalam, S., & Luyster, R. J. (2016). The integrity of lexical acquisition mechanisms in autism spectrum disorders: A research review. *Autism Research*, *9*(8), 810–828. https://doi.org/https://doi.org/10.1002/aur.1590

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *The Lancet*, *368*(9531), 210–215. https://doi.org/10.1097/DBP.0b013e3181d5d03b

Baixauli, I., Colomer, C., Roselló, B., & Miranda, A. (2016). Narratives of children with high-functioning autism spectrum disorder: A meta-analysis. *Research in Developmental Disabilities*, *59*, 234–254. https://doi.org/10.1016/j.ridd.2016.09.007

Bal, V. H., Katz, T., Bishop, S. L., & Krasileva, K. (2016). Understanding definitions of

minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *57*(12), 1424–1433. https://doi.org/10.1111/jcpp.12609

Barokova, M. D., Hassan, S., Lee, C., Xu, M., & Tager-Flusberg, H. (2020). A comparison of natural language samples collected from minimally and low-verbal children and adolescents with autism by parents and examiners. *Journal of Speech, Language, and Hearing Research*, *63*(12), 4018–4028. https://doi.org/10.1044/2020_jslhr-20-00343

Bartlett, M. S. (1951). The effect of standardization on a Chi-square approximation in factor analysis. *Biometrika*, *38*, 337–344. https://doi.org/10.1093/biomet/38.3-4.337

Bishop, S. L., & Lord, C. (2023). Commentary: Best practices and processes for assessment of autism spectrum disorder–the intended role of standardized diagnostic instruments. *Journal of Child Psychology and Psychiatry*, *64*(5), 834–838. https://doi.org/10.1111/jcpp.13802

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, *8*, 3–62.

Botting, N. (2020). Language, literacy and cognitive skills of young adults with developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, *55*(2), 255–265. https://doi.org/10.1111/1460-6984.12518

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

Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015). Efficient agglomerative hierarchical clustering. *Expert Systems with Applications*, *42*(5), 2785–2797. https://doi.org/10.1016/j.eswa.2014.09.054

Brignell, A., Morgan, A. T., Woolfenden, S., Klopper, F., May, T., Sarkozy, V., & Williams, K. (2018). A systematic review and meta-analysis of the prognosis of language outcomes for individuals with autism spectrum disorder. *Autism & Developmental Language Impairments*, *3*, 2396941518767610. https://doi.org/10.1177/2396941518767610

Burke, M. M., Cheung, W. C., Best, M., DaWalt, L. S., & Taylor, J. L. (2024). Measuring what matters: Considerations for the measurement of services for individuals with autism. *Journal of Developmental and Physical Disabilities*, *36*, 423–439. https://doi.org/10.1007/s10882-023-09916-6

Butler, L. K., LaValle, C., Shen, L. S., Palana, J., Schwartz, S., Liu, C., Peterman, N., & Tager-Flusberg, H. (2022). Remote natural language sampling of parents and children with autism spectrum disorder: Role of activity and language level. *Frontiers in Communication*,, *7*, 820564. https://doi.org/10.3389/fcomm.2022.820564

Butler, L. K., Shen, L., Chenausky, K. V., La Valle, C., Schwartz, S., & Tager-Flusberg, H. (2023). Lexical and morphosyntactic profiles of autistic youth with minimal or low spoken language skills. *American Journal of Speech-Language Pathology*, *32*(2), 733–747. https://doi.org/10.1044/2022_AJSLP-22-00098

Calder, S. D., Brennan-Jones, C. G., Robinson, M., Whitehouse, A., & Hill, E. (2023). How we measure language skills of children at scale: A call to move beyond domain-specific tests as a proxy for language. *International Journal of Speech-Language Pathology*, *32*(2), 440–448. https://doi.org/10.1080/17549507.2023.2171488

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Clegg, J., Crawford, E., Spencer, S., & Matthews, D. (2021). Developmental Language Disorder (DLD) in young people leaving care in England: A study profiling the language, literacy and communication abilities of young people transitioning from care to independence. *International Journal of Environmental Research and Public Health*, *18*(8), 4107. https://doi.org/10.3390/ijerph18084107

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*(3/4), 256–266. https://doi.org/10.1093/biomet/37.3-4.256

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Constantino, J. N., & Grubler, C. P. (2012). *Social Responsiveness Scale-Second Ed*. Western Psychological Services.

Cox, D., & Sosine, J. (2023). Influence of sample size, feature set, and algorithm on cluster analyses for patients with autism spectrum disorders. *PsyArXiv*. https://doi.org/10.21203/rs.3.rs-3351792/v1

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., & Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. *Autism*, *24*(7), 1704–1712. https://doi.org/10.1177/1362361320919286

Cummins, C., Pellicano, E., & Crane, L. (2020). Autistic adults' views of their communication skills and needs. *International Journal of Language & Communication Disorders*, *55*(5), 678–689. https://doi.org/10.1111/1460-6984.12552

Dalmaijer, E. S., Nord C. L., & Astle D. E. (2022). Statistical power for cluster analysis. *BMC Bioinformatics*, *23*(1), 205. https://doi.org/10.1186/s12859-022-04675-1

Day, W. H., & Edelsbrunner, H. (1984). Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of classification*, *1*(1), 7–24. https://doi.org/10.1007/BF01890115

Dudoit, S., & Fridlyand, J. (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology*, *3*(7), research0036.0031. https://doi.org/10.1186/gb-2002-3-7-research0036

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

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*(3), 241–252. https://doi.org/10.1080/00401706.1964.10490181

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. (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

Eilenberg, J. S., Paff, M., Harrison, A. J., & Long, K. A. (2019). Disparities based on race, ethnicity, and socioeconomic status over the transition to adulthood among adolescents and young adults on the autism spectrum: A systematic review. *Current Psychiatry Reports*, *21*(5), 32. https://doi.org/10.1007/s11920-019-1016-1

Every Student Succeeds Act, Pub. L. No. 114-95, § 1001-1605 (2015).

Fidler, L. J., Plante, E., & Vance, R. (2011). Identification of adults with developmental language impairments. *American Journal of Speech-Language Pathology*, *20*(1), 2–13. https://doi.org/10.1044/1058-0360(2010/09-0096)

Fidler, L. J., Plante, E., & Vance, R. (2013). Erratum. *American Journal of Speech-Language Pathology*, *22*(3), 577–577. https://doi.org/10.1044/1058-0360(2013/13-0018)

Gaeta, L., & Brydges, C. R. (2020). An examination of effect sizes and statistical power in

speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1572–1580. https://doi.org/10.1044/2020_JSLHR-19-00299

Girolamo, T., Ghali, S., Campos, I., & Ford, A. (2022). Interpretation and use of language assessments for diverse school-age individuals. *Perspectives of the ASHA special interest groups*. https://doi.org/10.1044/2022_PERSP-21-00322

Girolamo, T., & Rice, M. L. (2022). Language impairment in autistic young adults. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2022_JSLHR-21-00517

Girolamo, T., Shen, L., Monroe Gulick, A., Rice, M. L., & Eigsti, I. M. (2023). Studies assessing domains pertaining to structural language in autism vary in reporting practices and approaches to assessment: A systematic review. *Autism*, 13623613231216155. https://doi.org/10.1177/13623613231216155

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

Hammill, D. D., Brown, V. L., Larsen, S. C., & Wiederholt, J. L. (2007). *Test of Adolescent and Adult Language-Fourth Edition (TOAL-4)*. Pro-Ed.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *28*(1), 100–108. https://doi.org/10.2307.2346830

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *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*, *55*(1), 49–58. https://doi.org/10.1111/jcpp.12115

Howlin, P., & Taylor, J. L. (2015). Addressing the need for high quality research on autism in adulthood. *Autism*, *19*(7), 771–773. https://doi.org/10.1177/1362361315595582

Husson, F., Josse, J., & Pages, J. (2010). *Principal component methods-hierarchical clustering-partitional clustering: Why would we need to choose for visualizing data* (Technical Report of the Applied Mathematics Department, Issue. Agrocampus Ouest. http://www.agrocampus-oeust.fr/math/

IBM Corp. (2025). *IBM SPSS Statistics for Macintosh*. In (Version 31.0.0.0) IBM Corp.

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 et seq. (2018).

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., Brownlie, E. B., Douglas, L., Taback, N., & Lam, I. (1999a). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research*, *42*(3), 744–760. https://doi.org/10.1044/jslhr.4203.744

Johnson, C. J., Taback, N., Escobar, M., Wilson, B., & Beitchman, J. (1999b). Local norming of the Test of Adolescent/Adult Language-3 in the Ottawa Speech and Language Study. *Journal of Speech, Language, and Hearing Research*, *42*(3), 761–766. https://doi.org/10.1044/jslhr.4203.761

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent

developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Koegel, L. K., Bryan, K. M., Su, P. L., Vaidya, M., & Camarata, S. (2020). Definitions of nonverbal and minimally verbal in research for autism: A systematic review of the literature. *Journal of Autism and Developmental Disorders*, *50*, 2957–2972. https://doi.org/10.1007/s10803-020-04402-w

Kover, S. T., & Abbeduto, L. (2023). Toward equity in research on intellectual and developmental disabilities. *American Journal on Intellectual and Developmental Disabilities*, *128*(5), 350–370. https://doi.org/10.1352/1944-7558-128.5.350

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

Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., Nicholas, J., Reaven, J., & Rice, C. E. (2010). Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States. *Journal of Developmental & Behavioral Pediatrics*, *31*(4), 267–275. https://doi.org/10.1097/DBP.0b013e3181d5d03b

Lewis, F. M., Woodyatt, G. C., & Murdoch, B. E. (2008). Linguistic and pragmatic language skills in adults with autism spectrum disorder: A pilot study. *Research in Autism Spectrum Disorders*, *2*(1), 176–187. https://doi.org/10.1016/j.rasd.2007.05.002

Little, R. J., & Rubin, D. B. (2019). Single imputation methods. In *Statistical analysis with missing data* (Third edition ed., Vol. 793, pp. 67–81). John Wiley & Sons. https://doi.org/10.1002/9781119482260

Lombardo, M. V., Lai, M.-C., & Baron-Cohen, S. (2019). Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular Psychiatry*, *24*(10), 1435–1450. https://doi.org/10.1038/s41380-018-0321-0

Loucas, T., Charman, T., Pickles, A., Simonoff, E., Chandler, S., Meldrum, D., & Baird, G. (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

Lu, J.-F., Tang, J., Tang, Z.-M., & Yang, J.-Y. (2008). Hierarchical initialization approach for K-Means clustering. *Pattern Recognition Letters*, *29*(6), 787–795. https://doi.org/10.1016/j.patrec.2007.12.009

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

Manenti, M., Ferré, S., Tuller, L., Houy-Durand, E., Bonnet-Brilhault, F., & Prévost, P. (2024). Profiles of structural language and nonverbal intellectual abilities in verbal autistic adults. *Research in Autism Spectrum Disorders*, *114*, 102361. https://doi.org/10.1016/j.lingua.2023.103598

Manenti, M., Tuller, L., Houy-Durand, E., Bonnet-Brilhault, F., & Prévost, P. (2023). Assessing structural language skills of autistic adults: Focus on sentence repetition. *Lingua*, *294*,

103598. https://doi.org/10.1016/j.rasd.2024.102361

Mawhood, L., Howlin, P., & Rutter, M. (2000). Autism and developmental receptive language disorder—A comparative follow-up in early adult life. I: Cognitive and language outcomes. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 547–559. https://doi.org/10.1111/1469-7610.00642

McCauley, J. B., Pickles, A., Huerta, M., & Lord, C. (2020). Defining positive outcomes in more and less cognitively able autistic adults. *Autism Research*, *13*(9), 1548–1560. https://doi.org/10.1002/aur.2359

Musgrove, M. (2015). *Dear Colleague Letter: Speech and language services for students with autism spectrum disorder*. U.S. Department of Education. https://sites.ed.gov/idea/idea-files/osep-dear-colleague-letter-on-speech-and-language-services-for-students-with-autism-spectrum-disorder/

Nikolaus, M., & Fourtassi, A. (2023). Communicative feedback in language acquisition. *New Ideas in Psychology*, *68*, 100985. https://doi.org/10.1016/j.newideapsych.2022.100985

Nitido, H., & Plante, E. (2020). Diagnosis of developmental language disorder in research studies. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2777–2778. https://doi.org/10.1044/2020_JSLHR-20-00091

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Office of Management and Budget. (1997). *Revisions to the standards for the classification of federal data on race and ethnicity* (Federal Register Notice, Issue. https://obamawhitehouse.archives.gov/omb/fedreg_1997standards

Orlov, K. (2024). *KO Macros*. In *Kirill's SPSS Macros*, Raynald's SPSS Tools,. https://www.spsstools.net/en/macros/KO-spssmacros/

Pearson. (2025). *Staying connected through telepractice*. Retrieved January 16 from https://www.pearsonassessments.com/professional-assessments/digital-solutions/telepractice/about.html

Pizzano, M., Shire, S., Shih, W., Levato, L., Landa, R., Lord, C., Smith, T., & Kasari, C. (2024). Profiles of minimally verbal autistic children: Illuminating the neglected end of the spectrum. *Autism Research*, *17*(6), 1218–1229. https://doi.org/10.1002/aur.3151

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *Journal of Speech, Language & Hearing Research*, *53*, 414–429. https://doi.org/10.1044/1092-4388(2009/08-0016)

Raven, J., Rust, J., Chan, F., & Zhou, X. (2018). *Raven's 2 Progressive Matrices, Clinical Edition*. Pearson.

Roux, A. M., Chvasta, K., McLean, K. J., Carey, M., Perez Liz, G., Tomczuk, L., Lopez, K., Assing-Murray, E., Shattuck, P. T., & Shea, L. L. (2024). Challenges and opportunities in transitioning autistic individuals into adulthood. *Pediatrics*, *154*(5), e2024067195. https://doi.org/10.1542/peds.2024-067195

Russell, G., Mandy, W., Elliott, D., White, R., Pittwood, T., & Ford, T. (2019). Selection bias on intellectual ability in autism research: A cross-sectional review and meta-analysis. *Molecular Autism*, *10*(1), 1–10. https://doi.org/10.1186/s13229-019-0260-x

Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ*, *312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

Schaeffer, J., Abd El-Raziq, M., Castroviejo, E., Durrleman, S., Ferré, S., Grama, I., Hendriks, P., Kissine, M., Manenti, M., Marinis, T., Meir, N., Novogrodsky, R., Perovic, A., Panzeri, F., Silleresi, S., Sukenik, N., Vicente, A., Zebib, R., Prévost, P., & Tuller, L. (2023). Language in autism: Domains, profiles and co-occurring conditions. *Journal of Neural Transmission*, *130*(3), 433–457. https://doi.org/10.1007/s00702-023-02592-y

Selin, C. M., Rice, M. L., Girolamo, T., & Wang, C. J. (2019). Speech-language pathologists' clinical decision making for children with specific language impairment. *Language, Speech, and Hearing Services in Schools*, *50*(2), 283–307. https://doi.org/10.1044/2018_LSHSS-18-0017

Shaw, K. A., Williams, S., Patrick, M. E., Valencia-Prado, M., Durkin, M. S., Howerton, E. M., Ladd-Acosta, C. M., Pas, E. T., Bakian, A. V., & Bartholomew, P. (2025). Prevalence and early identification of autism spectrum disorder among children aged 4 and 8 years-Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022. *MMWR Surveillance Summaries*, *74*(2), 1–22. https://doi.org/10.15585/mmwr.ss7402a1

Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., Green, J. R., & Moore, C. A. (2009a). A nonword repetition task for speakers with misarticulations: The Syllable Repetition Task (SRT). https://doi.org/10.1044/1092-4388(2009/08-0047)

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

Shriberg, L. D., & Mabie, H. L. (2017). *Speech and motor speech assessment findings in eight complex neurodevelopmental disorders* (24). www.waisman.wisc.edu/phonology/techreports/TREP24.PDF

Silleresi, S., Prévost, P., Zebib, R., Bonnet□Brilhault, F., Conte, D., & Tuller, L. (2020). Identifying language and cognitive profiles in children with ASD via a cluster analysis exploration: Implications for the new ICD□11. *Autism Research*, *13*(7), 1155–1167. https://doi.org/10.1002/aur.2268

Tomblin, J. B. (1996). The big picture of SLI: Epidemiology of SLI. Symposium on Research in Child Language Disorders, Madison, WI,

Tomblin, J. B. (2008). Validating diagnostic standards for specific language impairment using adolescent outcomes. In C. F. Norbury, J. B. Tomblin, & D. V. M. Bishop (Eds.), *Understanding developmental language disorders: From theory to practice* (pp. 93–116). Psychology Press.

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

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Weir, E., Allison, C., & Baron-Cohen, S. (2022). Autistic adults have poorer quality healthcare and worse health based on self-report data. *Molecular Autism*, *13*(1), 23.

https://doi.org/10.1186/s13229-022-00501-w
Wiig, E. H., Semel, E., & Secord, W. (2013). *Clinical Evaluation of Language Fundamentals-5th Ed.* Pearson.
Williams, K. T. (2019). *Expressive Vocabulary Test-Third Edition: Manual*. Pearson.
World Health Organization. (2022). *International Classification of Diseases (ICD)-11*. https://icd.who.int/en

Figure Legend

Figure 1. Percent participants meeting language impairment cutoffs across language measures. LI = language impairment. Receptive vocabulary was assessed by the Peabody Picture Vocabulary Test-5 (Dunn, 2019). Expressive vocabulary was assessed by the Expressive Vocabulary Test-3 (Williams, 2019). Formulated Sentences, Recalling Sentences, Sentence Assembly, Word Classes, Understanding Spoken Paragraphs, and Semantic Relationships were assessed by the Clinical Evaluation of Language Fundamentals-Fifth Edition (Wiig et al., 2013). Nonword repetition was assessed by the Syllable Repetition Task overall percent accuracy (Shriberg et al., 2009).

Figure 2. *Z*-scores of individual participants across language and NVIQ measures by cluster. PPVT-5 = Peabody Picture Vocabulary Test-5 (Dunn, 2019). EVT-3 = Expressive Vocabulary Test-3 (Williams, 2019). FS = Formulated Sentences, RS = Recalling Sentences, SA = Sentence Assembly, WC = Word Classes, USP = Understanding Spoken Paragraphs, and SR = Semantic Relationships from the Clinical Evaluation of Language Fundamentals-5 (Wiig et al., 2013). SRT = Syllable Repetition Task (Shriberg et al., 2009). NVIQ = Raven's 2 nonverbal intelligence quotient (Raven et al., 2018).

Table 1

Sample Characteristics (N = 75)

Variable	Full Sample (*N* = 75)				≤ 21 years (*n* = 52)				> 21 years (*n* = 23)			
	n	%	*M* (*SD*)	Range	*n*	%	*M* (*SD*)	Range	*n*	%	*M* (*SD*)	Range
Age	75	-	19.70 (4.07)	13.27-30.43	52	-	17.11 (2.61)	13.27-21.75	23	-	23.55 (2.41)	22.28-30.43
Raven's NVIQ standard score	47	-	90.26 (16.46)	47-135	35	-	87.90 (16.55)	47-135	12	-	95.48 (15.35)	66-126
SRS-2 social communication impairment *t*-score	37	-	71.58 (11.19)	50-90	31	-	70.80 (11.94)	50-90	6	-	73.36 (9.23)	-
SRS-2 restricted & repetitive behaviors *t*-score	37	-	72.07 (12.85)	43-90	26	-	69.86 (13.29)	43-90	13	-	77.18 (10.35)	55-90
CELF-5 Expressive Language Index	47	-	76.35 (19.29)	45-116	31	-	75.75 (18.87)	45-116	16	-	79.35 (20.31)	45-110
CELF-5 Receptive Language Index	37	-	80.88	45-	25	-	77.76	45-	12	-	87.78	48-118
Race												
American Indian, Native American, or Alaska Native	12	16.00	-	-	6	11.54	-	-	6	26.07	-	-
Asian	19	25.33	-	-	16	30.77	-	-	3	13.04	-	-
Black or African American	39	52.00	-	-	26	50.00	-	-	13	56.52	-	-
Multiracial	26	34.67	-	-	18	34.62	-	-	8	34.78	-	-
Pacific Islander	4	5.33	-	-	3	5.77	-	-	1	4.35	-	-
White or Caucasian	21	28.00	-	-	14	26.92	-	-	7	30.43	-	-
Other	9	12.00	-	-	7	13.46	-	-	2	8.70	-	-
Hispanic/Latine: Yes	15	20.00	-	-	11	21.15	-	-	4	17.39	-	-
Sex at Birth												
Female	24	32	-	-	15	28.85	-	-	9	39.13	-	-
Male	51	68	-	-	37	71.15	-	-	14	60.87	-	-
Gender												
Female	25	33.33	-	-	16	30.77	-	-	9	39.13	-	-
Male	50	66.67	-	-	36	69.23	-	-	14	60.87	-	-

Note. Age presented as *M* (*SD*), and range. All other variables are presented as percentages. Multiracial is non-exclusive with other categories. Other: Latine (*n* = 7), Middle Eastern (*n* = 1), and Puerto Rican (*n* = 1).

Table 2

Sample Clinical Assessment Scores (N = 75)

	4	(21.03)	118	1	(19.66)	114		(22.73)	
Peabody Picture Vocabulary Test-5	7	90.59	40-	5	89.92	40-	23	92.09	56-123
standard score	4	(21.41)	129	1	(21.77)	129		(20.97)	
Expressive Vocabulary Test-3	7	92.54	40-	5	91.67	49-	23	94.48	57-128
standard score	4	(18.77)	128	1	(18.45)	122		(19.73)	
CELF-5 Recalling Sentences scaled	7	6.47	1-14	5	6.27	1-14	23	6.91	1-13
score	4	(3.74)		1	(3.85)			(3.53)	
Syllable Repetition Task percent	7	89.54	48-	5	89.22	48-	23	90.26	66-100
accuracy	4	(10.54)	100	1	(10.55)	100		(10.74)	

Note. Peabody Picture Vocabulary Test-5th Ed. (Dunn, 2018). Expressive Vocabulary Test-3rd Ed. (Williams, 2018). CELF-5 = Clinical Evaluation of Language Fundamentals-5th Ed. (Wiig et al., 2013). Syllable Repetition Task (Shriberg et al., 2009). Two participants had missing SRS-2 scores. One participant had missing language and NVIQ scores.

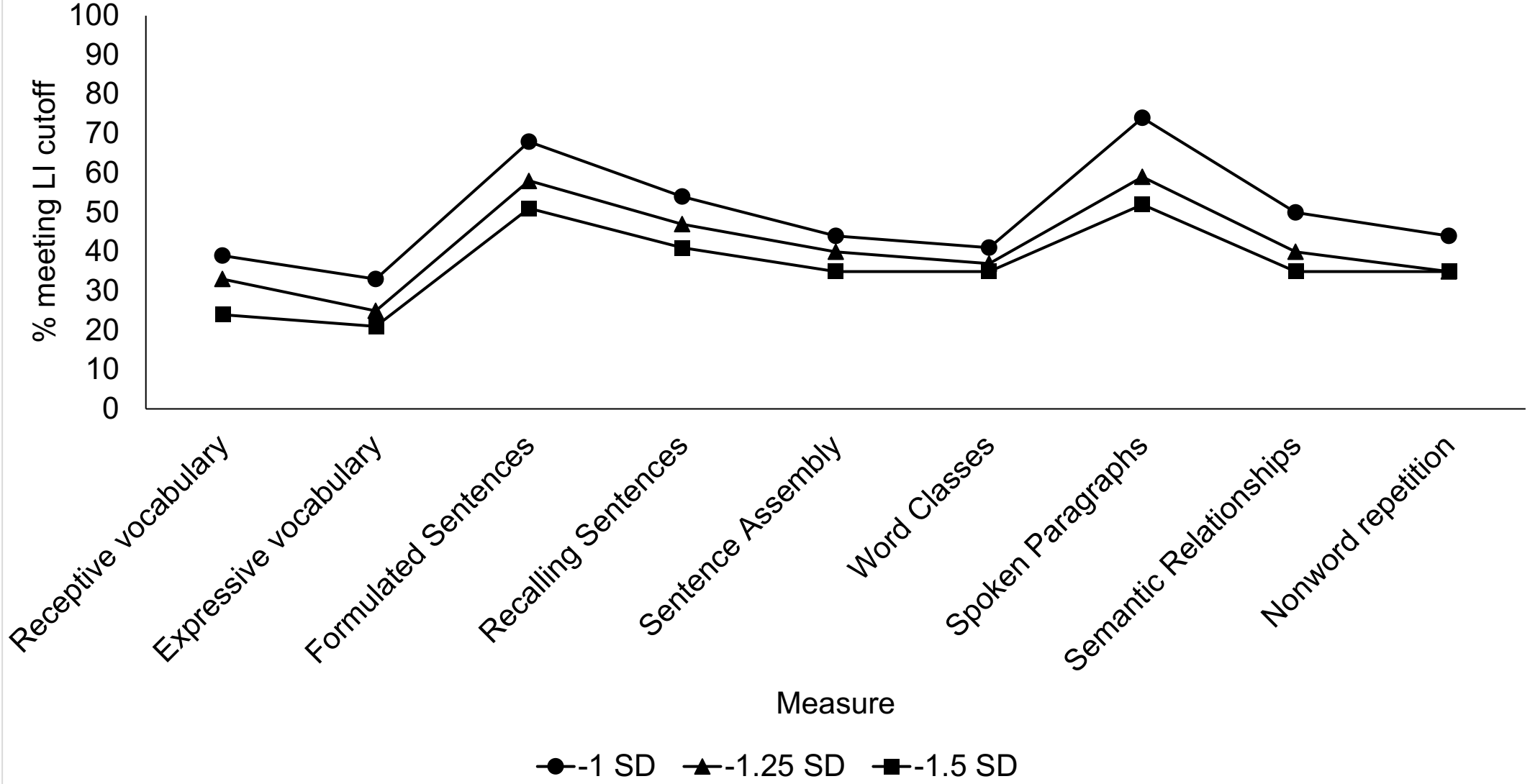
Table 3

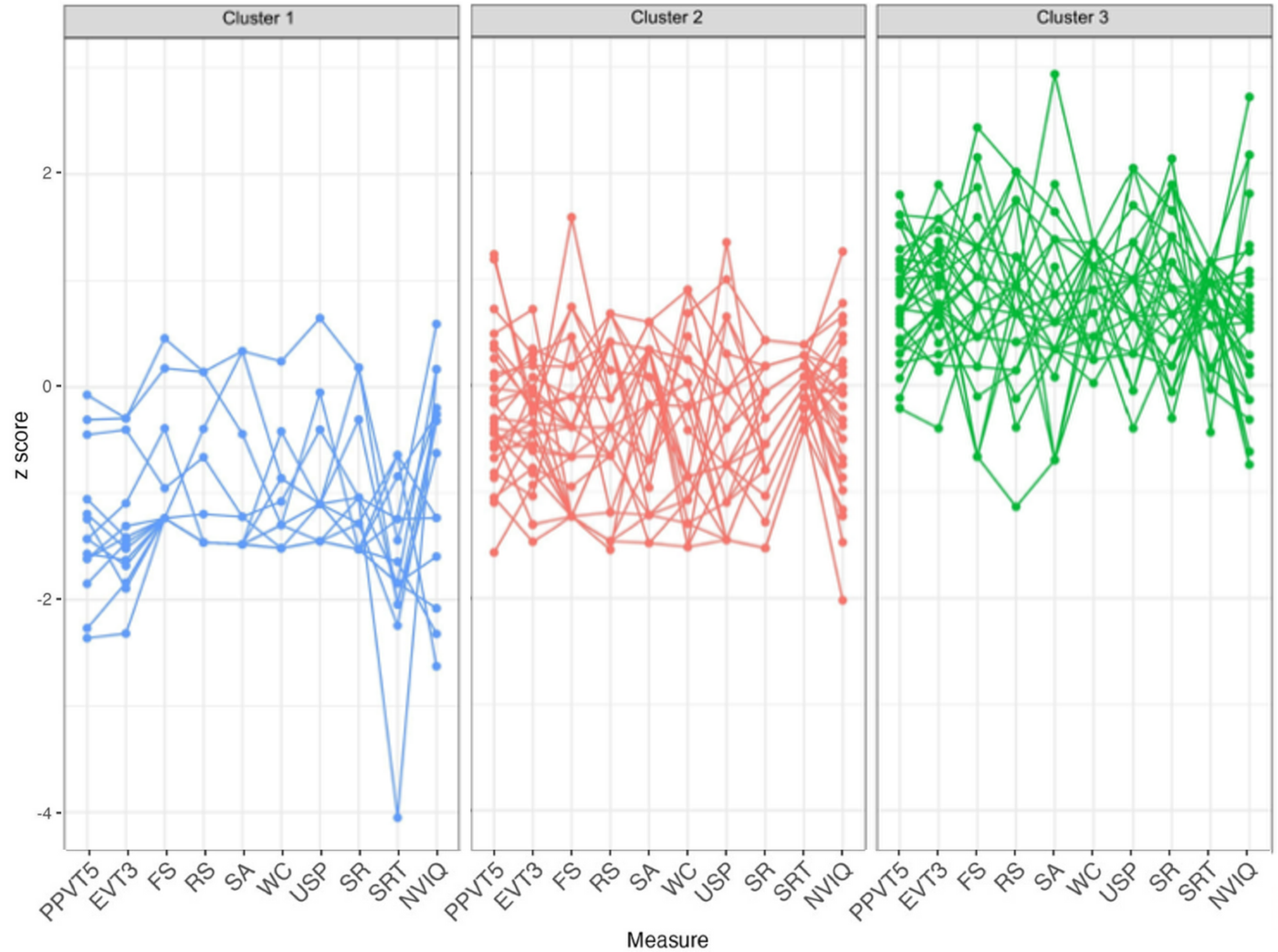
K-Means Three-Cluster Solution Cluster Means, Minimum, and Maximum Scores

Measure	Cluster 1 (*n* = 14)	Cluster 2 (*n* = 29)	Cluster 3 (*n* = 31)
Sex at birth ratio (female:male)	3:11	8:21	12:19
Age	20.03 (13.48-29.67)	19.09 (13.27-29.03)	20.26 (13.35-30.43)
NVIQ standard score	76.36 (47-100)	85.97 (57-111)	100.55 (78-135)
Receptive vocabulary standard score	62.36 (40-89)	85.48 (57-117)	108.13 (86-129)
Expressive vocabulary standard score	67.64 (49-87)	86.24 (65-106)	109.68 (85-128)
Nonword repetition percent accuracy	75.23 (48-96)	90.36 (68-100)	94.52 (80-100)
Word Classes scaled score	2.79 (1-9)	6.03 (1-12)	11.97 (8-14)
Formulated Sentences scaled score	2.07 (1-7)	3.93 (1-11)	8.23 (3-14)
Recalling Sentences scaled score	2.43 (1-7)	5.16 (1-9)	9.52 (2-14)
Understanding Spoken Paragraphs scaled score	2.36 (1-7)	3.86 (1-9)	7.61 (4-11)
Sentence Assembly scaled score	2.43 (1-8)	5.72 (1-9)	9.55 (4-18)
Semantic Relationships scaled score	2.93 (1-8)	5.48 (1-9)	10.87 (6-16)

Note. Sex at birth presented as counts. All other variables presented as means, minimum, and maximum. NVIQ, receptive vocabulary, and expressive vocabulary standard score *M* = 100, *SD* = 15. Nonword repetition percent accuracy: *M* = 93.2, *SD* = 3.0. Word Classes, Formulated Sentences, Recalling Sentences, Understanding Spoken Paragraphs, Sentence Assembly, and

Semantic Relationships scaled score: *M* = 10, *SD* = 3. Participants over 21: *n* = 5 in Cluster 1, *n* = 13 in Cluster 2, and *n* = 5 in Cluster 3.


100
90
80
70
60
50
40
30
20
10
0
% meeting LI cutoff
Receptive vocabulary
Expressive vocabulary
Formulated Sentences
Recalling Sentences
Sentence Assembly
Word Classes
Spoken Paragraphs
Semantic Relationships
Nonword repetition
Measure
-1 SD
-1.25 SD
-1.5 SD


Cluster 1
Cluster 2
Cluster 3
z score
2
0
-2
-4
PPVT5
EVT3
FS
RS
SA
WC
USP
SR
SRT
NVIQ
Measure