HHS Public Access

Author manuscript

Autism. Author manuscript; available in PMC 2024 July 01.

Published in final edited form as:

Autism. 2024 July ; 28(7): 1602–1621. doi:10.1177/13623613231216155.

Studies assessing domains pertaining to structural language in autism vary in reporting practices and approaches to assessment: A systematic review

Teresa Girolamo[1,2], **Lue Shen**[3], **Amalia Monroe Gulick**[4], **Mabel L Rice**[4], **Inge-Marie Eigsti**[2]

[1]San Diego State University, USA

[2]University of Connecticut, USA

[3]Boston University, USA

[4]The University of Kansas, USA

Abstract

Language in autism is heterogeneous, with a significant proportion of individuals having structural language difficulties and inclusion of language impairment as a specifier under *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.) criteria for autism. This systematic review asked: What are the reporting patterns of variables pertaining to structural language in autism prior to and after publication of the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.)? What norm-referenced assessments does research use to characterize the language abilities of autistic individuals with respect to language impairment? This preregistered review (PROSPERO: CRD42021260394) followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. Searches took place in September 2022 and included Linguistics and Language Behavior Abstracts, PsycINFO, PubMed, and the Directory of Open Access Journals. Search terms included three essential concepts: autism, language, and age. Two coders independently screened and evaluated articles. Searches yielded 57 qualifying studies, with mostly consistent reporting practices prior to and after the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Studies varied in how they defined language groups and in what norm-referenced measures they used. Interpreting research on structural language in autism requires attention to diagnostic and grouping criteria. Although inconsistency in reporting in original studies limited this review, better understanding the available information on structural language in autistic individuals aged 3–21 years may support identification of language needs.

Corresponding author: Teresa Girolamo, School of Speech, Language, and Hearing Sciences, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA. teresa.girolamo@uconn.edu.

Author contributions

Contributor Roles Taxonomy (CRediT) Statement: T.G.: conceptualization, methodology, formal analysis, data curation, writing—original draft, visualization, project administration, supervision; L.S.: methodology, formal analysis, data curation, writing—review & editing; A.M.G.: methodology, investigation, writing—review & editing; M.L.R.: conceptualization, writing—review & editing, funding acquisition; I.-M.E.: writing—review & editing, funding acquisition.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

Lay abstract

Under the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.), language impairment can co-occur with autism. It is not yet clear how research defines, reports, and characterizes structural language abilities of autistic individuals eligible for school-based special education services (aged 3–21 years) in the United States. In the United States, students typically must be formally diagnosed to be eligible for services and supports. However, the quality of diagnosis is only as good as the research evidence on which diagnosis depends. To evaluate evidence quality, we examined how studies of school-aged autistic individuals report assessments of language ability. This systematic review included 57 studies using English language age-referenced assessments used to measure structural language. Findings showed many differences across studies in how language abilities were measured and reported. Also, none of the studies fully reported the variables relevant to characterizing language impairment. Outcomes were similar across versions of the *Diagnostic and Statistical Manual of Mental Disorders*. Findings indicate that researchers and clinicians should pay attention to reporting diagnostic and grouping criteria. Carefully interpreting research evidence is critical for ensuring that diagnostic criteria and supports are representative of and accessible to autistic individuals and relevant parties.

Keywords

adolescents; autism spectrum disorders; behavioral measurement; communication and language; intellectual measurement; school-aged children

Despite the fact that language in autism spectrum disorder (ASD) is heterogeneous (Magiati et al., 2014) and language impairment (LI) is common in autism (Boucher, 2012; Kwok et al., 2015), limitations in knowledge about the structural language abilities of autistic individuals across the spectrum limit the ability to provide supports (Interagency Autism Coordinating Committee, 2020). LI refers to difficulties with structural language, such as morphology, syntax, and grammar (Schaeffer et al., 2023). LI in autistic and nonautistic individuals is tied to negative educational, health, occupational, and social outcomes (Johnson et al., 2010; Magiati et al., 2014). Thus, addressing structural language in autism is important to improving the quality of the evidence base informing assessment and service delivery (Russell et al., 2019; Tager-Flusberg & Kasari, 2013).

First, the inclusion of LI in the diagnostic criteria for autism has changed over time (American Psychiatric Association (APA), 1980, 1994, 2013). Second, LI can manifest differently across domains in assessment, such as receptive and expressive overall language, vocabulary, and grammar (Calder et al., 2023; Norbury et al., 2016; Tomblin et al., 1997). We note that while LI can influence the use of language for social communication, the underlying difficulties in LI involve structural language and not pragmatics (Andreou et al., 2022). This merits attention to how studies use norm-referenced assessments, which provide an outcome relative to a nationally representative sample of age peers and are commonly used to determine service eligibility in at least the United States for children aged 3–21 years who are eligible for special education services (hereafter, school-age; Individuals with Disabilities Education Improvement Act [IDEIA], 2004; Selin et al.,

2022). Understanding heterogeneity of structural language requires transparent reporting of approaches to measurement (Koegel et al., 2020).

Changes in the diagnostic criteria of autism

Although the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) provides criteria to facilitate reliable diagnoses of autism (Surís et al., 2016), autism as a concept has changed over time (Rosen et al., 2021). Autism first appeared under schizophrenia in the *Diagnostic and Statistical Manual of Mental Disorders* (1st ed.; DSM-I; APA, 1952) and *Diagnostic and Statistical Manual of Mental Disorders* (2nd ed.; DSM-II; APA, 1968). In the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.; DSM-III), autism became an independent diagnosis, amid a pivot from etiological definitions to concrete criteria (Surís et al., 2016). DSM-III criteria specified early onset of a lack of interest in people, gross deficits in language development, peculiar speech patterns, bizarre responses to the environment, and absence of delusions as in schizophrenia (APA, 1980). Early findings documented evidence of structural LI in autism (Bartolucci et al., 1980; Howlin, 1984). The *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev; DSM-IV-TR) included an early spoken language delay as a criterion in the communication domain for autistic disorder (APA, 1994, 2000). In contrast, the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., DSM-5) has no communication domain and does not include early language delay as a criterion in the social communication and social interaction domain for ASD; rather, autism includes LI as a specifier (APA, 2013). While the role of a language delay in autism is an ongoing topic of debate (Cirnigliaro et al., 2023), these changes underline a need for precision in reporting structural language benchmarked against diagnostic criteria.

Characterizing structural language

A second consideration in characterizing structural language in autism involves what domains to assess. To our knowledge, there are no population studies of LI in autism. However, there are population studies of LI in nonautistic youth, which have a more significant “weight of the evidence,” as they draw from samples representative of the population (vs a convenience sample which may not be representative). Findings from both the autism literature and these population studies support assessment of non-verbal intelligence (NVIQ), speech sound production, and of multiple language domains.

NVIQ

Findings from autism support no one-to-one correspondence between LI and NVIQ, including those who are minimally speaking (Munson et al., 2008; Slušná et al., 2021) and those with LI (Girolamo & Rice, 2022). Some samples have shown a “radical dissociation” between language and NVIQ in nonspeaking and minimally speaking autistic individuals (Munson et al., 2008; Slušná et al., 2021). A meta-analysis of 54 studies tracking language outcomes in autistic individuals aged 17 months to 38 years found IQ did not moderate language outcomes or language growth (Brignell et al., 2018). Some nonspeaking and minimally speaking autistic individuals have age-appropriate range NVIQ, indicating that language profiles in autism cannot be fully explained by intellectual disability (Slušná et al., 2021).

Population studies of LI in nonautistic children also do not support a universal co-occurrence between structural language skills and NVIQ. Tomblin and colleagues (1996) used a cutoff of NVIQ ⩾ 87 to ascertain kindergartners with specific LI, with the cutoff indicative of the specificity of difficulties to structural language. Yet, 12% of the sample had an NVIQ below this cutoff and typical language (Rice, 2017; Tomblin et al., 1997). Norbury and colleagues (2016) examined language outcomes in children (aged 4–5) with LI, finding no difference in children whether NVIQ was within −1 *SD* or −1 to −2 *SD*. The only difference was that children with NVIQ < 70 performed lower on overall production but not the four other composite scores (Norbury et al., 2016). Thus, LI in autistic and nonautistic youth can dissociate with NVIQ.

Speech sound production

Structural language difficulties and speech sound disorders can co-occur (Shriberg et al., 1999), and each can cause expressive disruptions in fluency (Zhang & Tomblin, 2000). Yet, the ability to produce speech sounds required for language assessment may be due to limitations in articulation, which comprise one aspect of speech sound development, or phonology; this ability is distinct from structural language skills (Dodd et al., 2018; Fey, 1992).

In 42 autistic youth aged 4–7 years without intellectual disability, the mean percentage of consonants correct on a measure of syllable repetition was nearly 92%; however, 17% showed a speech delay (Shriberg et al., 2011; Shriberg & Mabie, 2017). In population studies of LI in nonautistic youth, 5%–8% of those with specific LI over age 9 have shown a speech delay (Shriberg et al., 1999; Tomblin et al., 1997). A separate study of nonautistic youth found children with LI had lower accuracy than age peers without LI, though all had >95% consonants correct (Norbury et al., 2016). Therefore, speech sound disorder can co-occur with autism and LI.

Assessment of structural language across domains

In autism, language skills can vary by language domain and across the spectrum. Indeed, autistic individuals (aged 6–21) who produce fewer than 20 or 200 spoken words show lexical and morphosyntactic variability (Butler et al., 2023). Variability in assessment outcomes may arise from the nature of assessment. Relying on tasks that draw on social communication from a neurotypical perspective may not be useful for assessing structural language in autism. For instance, autistic children perform lower on narration tasks than nonautistic peers (Baixauli et al., 2016; Geelhand et al., 2020).

In addition, assessing NVIQ independent of verbal IQ (VIQ) or full-scale IQ (FSIQ; which includes both NVIQ and VIQ) is important. LI can conflate difficulties accessing structural language used in cognitive assessment with cognitive abilities. Autistic children (aged 4–14) score lower on VIQ versus NVIQ, with a discrepancy of nearly −1 *SD* in LI (Grondhuis et al., 2018). In turn, autism studies vary in whether they use VIQ, NVIQ, or FSIQ (Russell et al., 2019).

Given variability in structural language and IQ in autism, one question might be how and what to assess. Studies of LI in nonautistic youth have found ascertaining LI using

two or more of five composite scores to be clinically useful: expressive and receptive vocabulary, expressive and receptive grammar, expressive and receptive narration, overall comprehension, and overall expression (Norbury et al., 2016; Tomblin et al., 1996). Yet, in a third population study, Calder and colleagues (2023) found that individual measures were inconsistent in their ability to identify LI. Children with LI had scores within age expectations on a norm-referenced measure of receptive vocabulary, and only some Clinical Evaluation of Language Fundamentals, 3rd ed. (CELF-3; Semel et al., 1995) subtests predicted LI: semantic relationships, recalling sentences, and sentence assembly (Calder et al., 2023). Altogether, findings support the use of multi-domain assessment versus relying on a single domain to profile language.

Summary

Prior studies of LI support assessment of NVIQ, speech sound production, and multi-domain language assessment in characterizing structural language (Calder et al., 2023; Norbury et al., 2016; Tomblin et al., 1997). Together with findings in autism, this approach aligns to recent calls to appreciate linguistic heterogeneity across the autism spectrum rather than forming coarse groupings of typical language, LI, or minimally verbal individuals (Schaeffer et al., 2023).

Impact of assessment on access to supports

A third consideration in characterizing structural language in autism pertains to understanding the assessments studies use; this process has significant real-world implications. In the United States, assessment often serves as the point of access to supports for school-aged children (Adlof & Hogan, 2019; IDEIA, 2004). Within this system, US-based speech-language pathologists report that norm-referenced assessments are frequently part of the eligibility criteria for services and that they use a cutoff on one or two assessments to determine eligibility (Selin et al., 2022). Commonly used measures include assessments of overall receptive-expressive language or vocabulary, namely, the CELF (Semel et al., 2003), Preschool Language Scale, 4th ed. (PLS-4; Zimmerman et al., 2002), Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007), and Expressive One-Word Picture Vocabulary Test (EOWPVT; Betz et al., 2013; Brownell, 2000). Yet, studies of LI in nonautistic individuals do not consistently use validated assessments, instead relying on assessments without evidence-based cut scores that maximize diagnostic accuracy of structural language difficulties (Nitido & Plante, 2020). Understanding how research assesses structural language in autism is important for informing evidence-based practice, including assessment, that provides access to services and supports.

Post-DSM-5, the Office of Special Education and Rehabilitative Services of the United States Department of Education issued a letter identifying concerns that a growing number of autistic children are failing to receive appropriate services to address their communication needs (Musgrove, 2015). Specifically, the concern was that special education programs were not longer providing speech-language pathology services to autistic children or including speech-language pathologists in assessment or eligibility determinations (Musgrove, 2015). This letter contrasts with pre-DSM-5 data. Analysis of the National Longitudinal Transition

Survey-2, a 10-year longitudinal study of a nationally representative sample of students with disabilities from 2000 to 2009, found 75% of autistic youth in their last year of high school received speech-language therapy or communication services (Newman et al., 2011). Since a language delay including LI is no longer an eligibility criterion for diagnosis of autism, autistic children may face reduced access to assessment and services to meet their communication needs.

The current study

Amid diagnostic changes in autism, it is critical to understand how studies report norm-referenced outcomes when considering the linguistic dimensions of structural language pertaining to structural language and LI. To address this gap, this systematic review aimed to characterize studies in school-aged autistic individuals that used norm-referenced measures for linguistic domains pertaining to LI. We asked:

1. What are the patterns of reporting of variables, namely, clinical diagnosis, language groups, speech sound production, overall receptive-expressive language, vocabulary, grammar, and NVIQ, relevant to LI in ASD prior to and after publication of the DSM-5?
2. What norm-referenced assessments does the research literature use to characterize the language abilities of autistic individuals with respect to LI?

Method

This systematic review was preregistered with PROSPERO (CRD42021260394). We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocol (Page et al., 2021).

Search procedures

The third author conducted database searches on 20 September 2022. Prior to the searches, the third author searched for existing relevant reviews and protocols to avoid redundancy: *Cochrane Database of Systematic Reviews*, *Campbell Systematic Reviews*, *JBI Evidence Synthesis*, and ProQuest PsycINFO. Next, we reviewed test searches and used key articles to determine efficacy before finalizing the search strategy. The search included the following databases: PubMed, PsycINFO, ProQuest Linguistics, and Language Behavior Abstracts. To reduce publication bias and identify gray literature, the search also included the Directory of Open Access Journals. The final search included terms related to three essential concepts without limits on language or publication year: autism, language, and age; see Appendix. Search results were uploaded into Covidence (Veritas Health Innovation, 2021), which automatically removed duplicate records. The first and second authors independently screened titles and abstracts, followed by full texts, discussing agreements until they reached consensus.

Selection criteria

This review included empirical studies published between 1980 (i.e. when autism became an independent diagnosis) and 2021 in English. We also included studies primarily involving

evaluation of dimensions of language relevant to LI (overall expressive-receptive language, vocabulary, speech sound production, and morphosyntax) and using at least one norm-referenced assessment. Finally, this review included studies with at least one individual aged 3–21 with a diagnosis of autism, with no exclusion of co-occurring diagnoses (e.g. fragile X syndrome). This age-range coincides with eligibility for school-based special education services, where children become eligible in the year they turn 3 (IDEIA, 2004). Furthermore, as studies often only included group-level data, excluding individual participants out of this age range was impossible. We also excluded studies that focused on areas other than assessment (e.g. intervention or neuroimaging) or on dimensions of language that are not structural language (e.g. pragmatics). Finally, this review excluded studies using only experimental measures, or assessments in other languages or specific variants of English (e.g. British English), as language communities differ in their use of language and norms.

Quality review

This review appraised bias in studies reporting using one or more norm-referenced measures to evaluate structural language and related areas (e.g. speech sound production, cognitive ability). Here we considered systematic error from the truth in the presentation of results. Because Cochrane (2022) has no risk of bias tool ready for implementation in systematic reviews of observational, non-intervention studies, we considered the quality of reporting for internal validity and completeness of reporting (Viswanathan et al., 2012). Internal validity included whether studies assessed language abilities across domains relevant to LI in autism. Completeness of reporting entailed (a) whether studies provided sufficient information about participants, including clinical diagnosis, NVIQ, and any relevant definitions for grouping, to interpret the findings, and (b) whether studies selectively reported outcome measures rather than reporting all outcomes. Studies with the following characteristics received a high-quality rating:

1. Sample size: studies with autistic participants $n \geqslant 20$, following estimates for power analysis for speech-language pathology (Gaeta & Brydges, 2020);
2. Selective outcome reporting: studies providing information on all participant outcomes or stating that participants were excluded for reasons unrelated to performance (e.g. attrition). An example of selective outcome reporting is a study *post hoc* excluding participants who impact results in a way the authors did not intend as outliers. Such exclusion without explanation of how results and inferences differed would bias results;
3. Clinical diagnosis: studies providing a specific diagnostic label, such as “DSM-5 autism,” or a reference to the version of DSM used;
4. Grouping: studies providing an operational definition for grouping of autistic participants (e.g. minimally verbal), when applicable. While some support moving away from such grouping (Schaeffer et al., 2023), operational definitions for groups provide precision in understanding participant characteristics (Koegel et al., 2020);

5. NVIQ: studies providing NVIQ, and not just VIQ or FSIQ, which can yield scores nearly 1 *SD* lower than NVIQ in autistic individuals with LI (Grondhuis et al., 2018);
6. Speech sound production: studies providing information about articulation and phonological abilities that confirm the ability to produce speech sounds for language assessment (Zhang & Tomblin, 2000);
7. Overall language ability: studies providing information about overall expressive–receptive language ability. Here we considered composite scores and did not appraise studies by whether they provided domain or subtest scores;
8. Grammar: studies providing information about grammar abilities (expressive, receptive, or both);
9. Vocabulary ability: studies providing information on vocabulary abilities (expressive, receptive, or both).

Data extraction and synthesis

The authors analyzed studies in terms of participants, assessments, and findings. To describe the language variables reported in studies, the first and second authors independently screened and extracted data in Covidence (Veritas Health Innovation, 2021), discussing disagreements until they reached a consensus. The authors extracted and synthesized information on age, population, sample size, cutoff for LI, provision of information on IQ, domains of language assessed, and results. The authors also compiled information on measures of speech sound production, vocabulary, grammar, overall language, and IQ. Comparisons in reporting patterns across pre-DSM-5 (DSM-III, DSM-IV-TR) and post-DSM-5 studies used descriptive analyses (e.g. frequencies, chi-square tests of homogeneity when sample size was sufficient and Fisher's exact test when $n < 5$) with an *a priori* significance value of $p < 0.05$ (Blalock, 1972; Marascuilo & McSweeney, 1977).

Community involvement

This systematic review included external reviewers varying in their relationship to autism research and practice. The research team included autistic individuals who supported the publication and dissemination of this review.

Results

Study selection

Searches yielded 7913 results from databases; see Figure 1. After removing 2051 duplicates, screening 5862 studies' titles and abstracts led to the exclusion of 5735 studies. The remaining 127 studies were assessed for full-text eligibility, with exclusion for various reasons: published in a language other than English ($n = 24$), used language measures with norms other than American English (e.g. Australian or British English; $n = 14$), no use of norm-referenced language assessments (i.e. experimental measures only; $n = 14$), primary outcome other than language (e.g. neuroimaging; $n = 13$), insufficient information for inclusion (e.g. no inclusion of diagnostic labels; $n = 2$), non-observational study design

(e.g. intervention; $n = 1$), and wrong age range (e.g. <3 years; $n = 2$). The 57 studies that qualified for the review included 7915 autistic individuals, with sample sizes ranging from 1 to 2047 participants. However, participants were not all unique. For instance, some came from larger samples or longitudinal studies; Ellis Weismer et al., 2010, report that about one-third of participants overlapped with Luyster et al., 2007). Of all 57 studies, 36 (63%) used pre-DSM-5 criteria or were published prior to the DSM-5, with one DSM-III study, 34 DSM-IV-TR studies, and 22 DSM-5 studies.

Quality analysis: reporting practices prior to and after the DSM-5

To address our first research question, we rated reporting of variables relevant to LI in autism pre- and post-DSM-5. For a summary of the quality analysis, see Table 1. Because ratings for pre- and post-DSM-5 studies did not differ on any criteria except clinical diagnosis, we report frequencies for the entire sample; exact frequencies are in Table 2.

Sample size and selective outcome reporting.—A total of 46 of 57 (81%) of studies received high-quality ratings for sample size. Of the 11 studies that received low-quality ratings, 9 (82%) had samples of $N < 20$, and 2 (18%) were case studies. Nearly all studies (55 of 57, or 97%) received high-quality ratings for selective outcome reporting, though some studies may not have reported exclusion. One DSM-IV study mentioned exclusion of 14 participants who had borderline LI and did not meet selection criteria for autism with or without LI (Modyanova et al., 2017). The one DSM-III-R study excluded data from 20 of 82 children who scored at floor on language measures (Rapin et al., 2009).

Clinical diagnosis.—Most studies (37 of 57, or 65%) received high-quality ratings for diagnostic information, but the expected proportion of studies receiving high-quality ratings differed by DSM version. Pre-DSM-5 studies (27 of 35, or 77%) were more likely to receive a high-quality rating than post-DSM-5 studies (10 of 22, or 46%), $p = 0.015$. Studies also differed in the DSM versions they used to characterize participants: DSM-III-R ($n = 1$), DSM-III-R or DSM-IV ($n = 1$), DSM-IV ($n = 23$; 14 DSM-IV and 9 DSM-IV-TR), and DSM-5 ($n = 7$); see Supplementary Table 1.

ASD.: A majority of studies reported ASD as a diagnostic label ($n = 36$ of 57, or 63%). However, DSM-5 studies (21 of 22, or 96%) were more likely to report ASD as a diagnostic label than DSM-III or DSM-IV studies (15 of 35, or 43%), $p < 0.001$. Note that while DSM-5 studies used ASD as an umbrella diagnosis, DSM-IV studies varied in definitions of ASD. For instance, Paul and colleagues (2008) defined ASD as inclusive of autism and PDD-NOS; in contrast, Anderson and colleagues (2007) treated ASD and PDD-NOS as separate groups. Furthermore, while all DSM-5 studies used DSM-5 criteria to confirm diagnosis, six (27%) reported DSM-IV diagnoses now included under ASD: Asperger syndrome ($n = 2$), autistic disorder ($n = 1$), and PDD-NOS ($n = 4$). These differences in reporting impact the consistency of reporting participant characteristics across studies.

ASD plus co-occurring genetic conditions.: Few studies reported ASD plus co-occurring genetic conditions regardless of pre- or post-DSM-5 status ($n = 4$, or 7%). No pre-DSM-5 study reported such diagnoses versus four DSM-5 studies (18%) that reported ASD plus

either fragile X syndrome ($n = 3$) or "chromosomal abnormalities" ($n = 1$), $p = 0.019$. Although further discussion is beyond the scope of this report, the recency of genetic testing and creation of a fragile X participant database may contribute to this difference (Sherman et al., 2017).

Autistic disorder.: Few studies reported autistic disorder: 8 of 35 (23%) pre-DSM-5 studies reported autistic disorder compared with 1 of 22 (5%) post-DSM-5 studies; this difference was not significant. While autistic disorder is not a DSM-5 diagnosis, some DSM-5 studies analyzed data from databases with data collected prior to the DSM-5. In turn, in studies using DSM-III-R criteria, autistic disorder would be the only possible diagnosis (e.g. Minshew et al., 1995; Rapin et al., 2009). Appreciating these differences in how samples were originally ascertained is relevant for understanding who comprises the evidence base.

Asperger syndrome.: Like autistic disorder, few studies reported Asperger syndrome, a diagnosis that only existed in the DSM-IV-TR. Four of 35 pre-DSM-5 studies (11%; all DSM-IV studies) included Asperger syndrome compared with two DSM-5 studies (9%). This difference was not significant. Studies varied in how they operationalized this term. While acknowledging that DSM-IV criteria stated a child with Asperger syndrome who met criteria for autism would receive a diagnosis of autism, Bennett and colleagues (2008) differentiated Asperger syndrome and "high-functioning autism" on the basis of having an early, significant language delay. In general, pre-DSM-5 studies likely assumed that individuals with Asperger syndrome would have age-appropriate and unimpaired language abilities.

Autism.: Pre-DSM-5 studies were more likely than DSM-5 studies to report autism as a diagnostic label, with the caveat that studies differed in their use of the term. For instance, autism may be an umbrella term or refer to a specific diagnosis, such as autistic disorder (autism) in the DSM-IV. While no DSM-III-R study referred to autism, 14 DSM-IV studies (40%) did compare with 2 of 22 (9%) DSM-5 study; this difference was statistically significant, $p = 0.011$. Many studies used research definitions and instruments that refer to "autism" (see Supplementary Table 1) to qualify participants. Yet, they were not uniform in clearly stating whether autism was tied to a specific instrument, clinical cutoff, or detailed evaluation.

PDD-NOS.: Studies did not differ in rates of reported inclusion of PDD-NOS ($n = 13$, or 23%): 9 of 35 DSM-IV studies (26%) reported PDD-NOS compared with four of 22 (18%) DSM-5 studies; this difference was not significant.

Summary.: Differences in diagnostic labels paralleled changes. in the DSM, from autistic disorder in the DSM-III-R study to autistic disorder and related in DSM-IV studies to ASD in DSM-5 studies. Transparency in reporting diagnostic labels and criteria is key for enhancing the accessibility of research findings.

Grouping.—We analyzed grouping criteria relevant to language. Of the 32 studies that used grouping criteria, 28 (88%) received a high-quality rating. Studies focused on four groups with no differences in the expected proportion of studies per group: autism without

further specifiers (*n* = 28, or 49%), autism plus LI (*n* = 14, or 25%), "high functioning autism" (*n* = 6, or 11%), and minimally verbal (*n* = 5, or 9%); see Table 3 for frequencies and Supplementary Table 2 for details.

Autism.: Most studies focused on language in autism broadly (*n* = 28, or 49%). These studies included 18 pre-DSM-5 studies (1 DSM-III-R, 17 DSM-IV; 51%) and 10 DSM-5 studies (46%); this difference was not significant. Although these studies did not use specifiers, some used grouping criteria to profile language in individuals aged 3–18, including 1 DSM-III-R study, 4 DSM-IV studies, and 1 DSM-5 study.

Pre-DSM-5 studies typically grouped participants to characterize developmental trajectories. The one qualifying DSM-III-R study assessed multiple domains of language and used these assessments to group participants by language profiles; however, the study did not operationally define each of the language profiles: mixed receptive-expressive language disorders, higher-order language processing disorders, and expressive phonology with or without grammar disorders (Rapin et al., 2009). Two DSM-IV studies grouped participants on the basis of early language and communicative regression (Gagnon et al., 2021; Prescott & Ellis Weismer, 2022), and an additional two grouped participants by language level: (a) low language, or being administered ADOS Module 1 (Lord et al., 1999), and an overall receptive–expressive language standard score of ⩽50 (Ellis Weismer & Kover, 2015); and (b) spoken language benchmarks in children using age-equivalent scores on an overall receptive–expressive language assessment of <15 months for prelinguistic, 15–23 months for first words, 24–35 months for word combinations, and >35 months for sentences (Ellawadi & Ellis Weismer, 2015). The one DSM-5 study that grouped participants defined "high verbal" as within −1 *SD* or higher and "mid-verbal" as −1 *SD* or lower on an expressive language measure (Jyotishi et al., 2017). Overall, only some of these definitions focus on structural language.

Autism plus LI.: One-quarter of studies explicitly examined the co-occurrence of LI and autism, with no significant differences by DSM version: 10 (29%) DSM-IV studies and 4 (18%) DSM-5 studies. Studies used 10 different definitions and cutoffs that typically spanned −1 to −2 *SD* on one or more language measures in individuals aged 4–21.

DSM-IV studies used seven different definitions of LI, with two using a single measure or subtests from one domain: receptive vocabulary at −2 *SD* in individuals (Roberts et al., 2004) or −1 *SD* on sentence production and sentence repetition subtests (McGregor et al., 2012). Other definitions used a cutoff of −2 *SD* on a receptive vocabulary measure or an overall receptive–expressive language measure (Kjelgaard & Tager-Flusberg, 2001). Still other definitions referenced multiple specific language domains: (a) <10th percentile on at least two measures of receptive grammar, receptive vocabulary, and expressive vocabulary using a subtest for VIQ from a brief intelligence test (Modyanova et al., 2017; Perovic et al., 2013); (b) <10th percentile on at least two measures of receptive grammar, narration, subtests for sight word and phonemic decoding, overall communication, nonword repetition, and sentence repetition (Whitehouse et al., 2008); (c) expressive vocabulary <10th percentile and receptive grammar <50th per-centile (Botting & Conti-Ramsden, 2003); and (d) history of a language delay and −1 *SD* on an overall receptive–expressive language or nonword

repetition plus VIQ >50 (Lindgren et al., 2009). Amid these varying definitions, some studies used a minimum IQ for group comparisons (Botting & Conti-Ramsden, 2003; McGregor et al., 2012; Whitehouse et al., 2008) or treated IQ < 70 as mutually exclusive with LI (Bennett et al., 2008).

DSM-5 studies also tended to define LI using multiple assessments. One study used a cutoff of ⩽95 on a norm-referenced expressive grammar test shown to have good sensitivity to LI and NVIQ of at least 70 (Huang & Finestack, 2020). All other DSM-5 definitions included multiple domains: (a) NVIQ of at least 70 plus a standard score at −1.5 *SD* on an overall expressive–receptive language test (Bennett et al., 2014), and (b) −1.25 *SD* on at least two measures of overall language, receptive vocabulary, expressive vocabulary, grammar, and nonword repetition (Girolamo & Rice, 2022). In all, these 10 definitions of LI aligned to studies of LI in non-autistic individuals, with about half restricting LI to NVIQ of >70.

“High functioning” autism.: Six studies examined language in “high functioning” autism (“HFA”); the quotes refer to use of this term in the original reports. Five DSM-IV (14%) studies and one DSM-5 study (5%) reported “HFA,” which did not significantly differ. These studies used five different definitions involving IQ in individuals aged 7 to >18.

Two DSM-IV studies defined “HFA” that included language domains relevant to LI: (a) NVIQ of at least 68 or 70 plus an average within −1.5 *SD* on expressive and receptive grammar subtests from an overall receptive–expressive language assessment (Bennett et al., 2008); and (b) FSIQ, VIQ, and a receptive vocabulary standard score within −1 *SD* (Eigsti & Bennetto, 2009). Other definitions used FSIQ but not language measures: (a) FSIQ >80 (Landa & Goldberg, 2005), and (b) an early language delay plus VIQ and FSIQ > 70 plus at least a second-grade reading, spelling, and arithmetic level (Minshew et al., 1995). Thus, these pre-DSM-5 definitions of “HFA” mostly considered verbal and nonverbal abilities together.

The one DSM-5 study used a similar approach as in DSM-IV studies to define “HFA”: FSIQ of at least 85 and an age equivalent of at least 48 months on a parent report measure of expressive language (Burton et al., 2020). In sum, pre- and post-DSM-5 definitions considered language and cognition together. Using broad measures of language and cognition as proxies for overall abilities does not reflect the full heterogeneity of abilities and unmet needs of individuals across the autism spectrum (Grondhuis et al., 2018; Waizbard-Bartov et al., 2023).

Minimally speaking.: Five (9%) studies focused on minimally speaking autistic individuals and used nine different definitions in individuals aged 2–21; one study accounted for five of these definitions (Bal et al., 2016). There were no statistically significant differences in the expected proportion of pre-DSM-5 ($n = 2$) and ($n = 3$) DSM-5 studies focusing on this population. This discrepancy in definitions is consistent with prior work (Koegel et al., 2020). Here, we focus on the aspects of language relevant to LI, particularly with regard to vocabulary.

DSM-IV studies used different operational definitions of minimally speaking. Woynaroski and colleagues (2016) used a continuous variable of ⩽20 words on the MacArthur-Bates Communicative Developmental Inventories (CDI) Words and Gestures checklist (Fenson et al., 2007) plus no more than five different word roots produced during a 15-min language sample in children who qualified for a diagnosis of ASD aged 2–4. Another definition used a categorical approach paired with direct observation, or a rating corresponding to no speech, single words, or occasional phrases on Autism Diagnostic Observation Schedule (ADOS) item A1, or "Overall level of non-echoed spoken language" (Lord et al., 2000; Thurm et al., 2015).

DSM-5 studies also defined minimally speaking using parent report and categorical criteria. These definitions included no use of phrase speech on a daily basis or parent report of <30 spoken words or phrases (Plesa Skwerer et al., 2016), as well as a comparison of minimally speaking status across five instruments: (a) being administered ADOS Module 1 (Bal et al., 2016, 2020; Lord et al., 2000); (b) parent estimate of ⩽25 spoken words; (c) expressive language age equivalent of <18 months per parent report; (d) no functional three-word phrases daily per item 30 of the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003), and (e) no use of phrases or sentences per Item 1 of the Social Communication Questionnaire (Bal et al., 2016; Rutter, Bailey, & Lord, 2003). Only Biller and Johnson (2020) referenced mental age, defining minimally speaking as parent estimate of ⩽25 spoken words and a nonverbal mental age of at least 12 months. In all, it is unclear to what extent such categorical criteria might collapse linguistic heterogeneity in autistic individuals who are minimally speaking (Butler et al., 2023).

Summary.: Overall, pre-DSM-5 and DSM-5 studies were consistent in the broad domains used to define groups: oral language for minimally speaking, structural language for LI, and FSIQ with or without language measures for "HFA." Yet, because the exact definitions differed, the ability to make direct comparisons across studies is limited.

Inclusion of measures of nonverbal cognitive ability.—Most studies received high-quality ratings for using nonverbal cognitive ability (91%) to characterize participants, with no differences by pre-DSM-5 or DSM-5 status; see Tables 2 and 3, and Supplementary Table 3. Some studies reported more than one type of cognitive measure and did not differ by DSM status in reporting of nonverbal (n = 49, or 86%) or verbal cognitive ability (n = 13, or 23%). There were differences by DSM status, however, in the expected proportion of studies reporting full-scale measures of cognitive ability. Ten (29%) DSM-IV studies versus one (5%) DSM-5 study reported full-scale measures of cognitive ability, p = 0.037. In all, findings indicate a tendency of studies to align to best practices for use of IQ in autism research (Grondhuis et al., 2018).

Use of measures of speech sound production, overall language, grammar, and vocabulary.—A majority of studies received high-quality ratings for overall language (63%) and vocabulary (60%), but they were unlikely to report measures of speech sound production (12 of 57, 21%) or specific information on grammar (32%); see Table 3 and Supplementary Table 4. In addition, few studies reported measures of expressive (n = 7, or 20%) or receptive grammar (n = 6, or 17%), or expressive vocabulary (n = 27, or 47%).

In contrast, studies were likely to report overall receptive (n = 35, or 61%) and expressive language (n = 35, or 61%), as well as receptive vocabulary (n = 29, or 51%). While many studies used norm-referenced measures of overall receptive–expressive language, some studies used it as an outcome and did not report actual values, and only some reported specific subtests with information on grammar (e.g. Burton et al., 2020; Worth & Reynolds, 2008).

Summary.—There were nearly no differences in the reporting practices of studies by pre- or post-DSM-5 status. While this analysis does not evaluate the overall quality of study design, inconsistency in reporting of clinical diagnosis and definitions of language groups prevent more fully understanding participant characteristics and who comprises the evidence base.

Use of norm-referenced assessments in pre-DSM-5 and DSM-5 studies

In our second research question, we examined assessments in studies by DSM-5 status. Given that assessments differ in their sensitivity to structural language (Calder et al., 2023) and the heterogeneity of LI across IQ (Norbury et al., 2016), we examined cognitive assessments and language assessments by domain; see Supplementary Table 4 and Figures 1 and 2.

Cognitive abilities.—Studies varied in the cognitive measures they used. The most common measure was Mullen Scales of Early Learning (Mullen, 1995; n = 19 of 57, or 33%; 11 pre- and 8 post-DSM-5 studies), followed by a version of the Wechsler Intelligence Scales (n = 14, or 25%): Wechsler Intelligence Scales for Children (Wechsler, 1974, 1991, 2003, 2014; n = 11; 9 pre- and 2 post-DSM-5 studies) or Wechsler Adult Intelligence Scales (Wechsler, 1981, 1997; n = 3; all pre-DSM-5). These assessments benchmark performance against time and require fine motor skills, which precludes accessibility for all autistic students (e.g. Kasari et al., 2013). The third most common measure was the Differential Abilities Scales (Elliott, 1990, 2007; n = 9, or 16%; 10 pre- and 2 post-DSM-5 studies). Eleven studies (19%) reported age equivalent scores or proxies for mental age (n = 6 pre- and 5 post-DSM-5 studies), which is common when standard scores are unavailable; eight of these studies examined language in early childhood (i.e. upper M age = 5 years).

Speech sound production.—In studies reporting measures of articulation and phonology, some limited assessment to the speech errors that confound pronunciation of finiteness-marking in English with finiteness-marking and used the Test of Early Grammatical Impairment phonological probe (Rice & Wexler, 2001; n = 1 pre- and 3 post-DSM-5 studies). In this case, studies did not report outcomes other than pass/fail rates, which is the outcome of the probe. Norm-referenced measures that assess articulation and phonology more comprehensively, such as the Goldman Fristoe Test of Articulation (Goldman & Fristoe, 1986; n = 2), or speech motor issues, such as the NEPSY oromotor test (Korkman et al., 1998; n = 1) and Voice Motor Production Assessment for Children (Hayden & Square, 1999; n = 1), were less common.

Overall language.—Common measures of receptive-overall language included direct behavioral assessments (vs parent report), such as the Preschool Language Scales (Zimmerman et al., 2002, 2011; *n* = 9, or 16%; 5 pre- and 4 post-DSM-5 studies) and the CELF (Semel et al., 1995, 2003; Wiig et al., 1992, 2013; Zimmerman et al., 2002, 2011; *n* = 14, or 25%). Because these assessments are not designed for all ages or profiles, other common measures, such as for minimally verbal individuals, were more general measures of development. These included the Vineland Adaptive Behavior Scales (Sparrow et al., 2005, 2016; *n* = 12, or 21%; 4 pre- and 8 post-DSM-5 studies), which is a parent report, and the Mullen Scales of Early Learning (Mullen, 1995; *n* = 11, or 19%; 6 pre- and 5 post-DSM-5 studies).

Receptive and expressive vocabulary.—Direct behavioral assessments of vocabulary frequently included a version of the Peabody Picture Vocabulary Test (Dunn, 2019; Dunn & Dunn, 1997, 1981, 2007) for receptive vocabulary (*n* = 20, or 35%; 14 pre- and 6 post-DSM-5 studies) and the Expressive Vocabulary Test (Williams, 1997, 2007, 2019; *n* = 9, or 16%; 4 pre- and 5 post-DSM-5 studies). Some studies used an indirect measure of vocabulary using words understood (*n* = 5 or 9%; 3 pre- and 2 post-DSM-5 studies) or words produced (*n* = 11, or 19%; 5 pre- and 6 post-DSM-5 studies) per parent report on the CDI (Fenson et al., 2007).

Grammar.—While most studies used a measure of overall language, which includes grammar, few studies used a norm-referenced grammar measure or reported grammar-specific information. Here, measures included the Test of Receptive Grammar (Bishop, 1982, 2003b, 2005; *n* = 5, or 9%; all pre-DSM-5 studies) and the Test of Early Grammatical Impairment (Rice & Wexler, 2001; *n* = 5, or 9%; 2 pre- and 3 post-DSM-5 studies). Less common was indirect assessment via parent report: the Children's Communication Checklist (Bishop, 1998, 2003a; *n* = 2) and grammatical complexity items from the CDI (Fenson et al., 2007; *n* = 1).

Summary.: Pre-DSM-5 and DSM-5 studies mostly did not vary in their reporting practices of criteria relevant to structural language in autism. Studies varied more in how they defined groups of participants in terms of clinical diagnosis and language profiles. Studies also differed in reporting of norm-referenced assessment outcomes across language domains and cognitive abilities.

Discussion

This systematic review identified differences in reporting of information relevant to characterizing structural language in studies of language in autism using one or more norm-referenced assessments in school-aged individuals. While previous reviews focused on empirical findings on structural language in autism (Andreou et al., 2022), this review underlines the importance of appreciating consistency in reporting of approaches to assessment.

Variation in reporting practices

In this review, no study received high-quality ratings across all criteria. While not representative of the quality of study design, these ratings indicate that the approach to characterizing structural language has been piecemeal—both heterogeneous and incomplete. Effectively, this limits the ability to understand sample characteristics. Reporting in terms of exclusion and grouping, as well as diagnostic labels, also has implications for understanding linguistic heterogeneity in autism.

First, few studies reported exclusion of participants, and when they did, it was on the basis of participants showing heterogeneous language abilities versus ones that aligned to grouping criteria. In excluding about one-quarter of 82 participants who scored at floor on norm-referenced assessments, Rapin and colleagues (2009) did not report the reason for their performance: inaccessibility, noncompliance, or some other reason. Modyanova and colleagues (2017) similarly excluded 14 of 97 autistic participants who had borderline LI, did not meet criteria for the LI or non-LI group, and were too few in number compared with other groups to create a third group. While appropriate for these individual group designs, such exclusion is consistent with broader trends in autism research. At a systemic level, there is a tendency to focus on groups differentiated by levels of spoken language and structural LI, which perpetuates the masking of linguistic heterogeneity in autism (Schaeffer et al., 2023).

Furthermore, 65% of studies received high-quality ratings for reporting an exact diagnostic label yet differently operationalized the same label. For example, DSM-IV studies varied in whether they included PDD-NOS under ASD (e.g. Anderson et al., 2007; Paul et al., 2008). While many studies independently confirm diagnosis (e.g. with the ADOS, Lord et al., 2000; see Supplementary Table 1), being clear about the exact source of diagnostic labels is important. In this review, 27% of DSM-5 studies that analyzed data from databases, longitudinal studies, and participants who had originally received a DSM-IV diagnosis both independently confirmed diagnosis using DSM-5 criteria and listed the original diagnoses of participants. Providing precise diagnostic information is necessary to understand whether participants were ascertained on the basis of having a language delay, and in turn, how phenotypic variability relates to neurological differences (Cirnigliaro et al., 2023).

Variation in assessments

As for assessments, findings, which primarily came from US-based studies, mimicked clinical practice approaches in the United States (Betz et al., 2013). Studies were the most likely to use norm-referenced assessments for overall receptive–expressive language ability (63%) and vocabulary (60%), coinciding with clinician report of most commonly using measures of overall language and vocabulary in practice (Betz et al., 2013). The most commonly used language measures were the (a) CELF (25% of studies here vs 67% of clinicians using the CELF-4 in practice at least sometimes in Betz et al., 2013; Semel et al., 1995, 2003; Wiig et al., 1992, 2013; Zimmerman et al., 2002, 2011); and (b) Preschool Language Scales (16% of studies here vs 50% of clinicians using the PLS-4 in practice at least sometimes in Betz et al., 2013; Zimmerman et al., 2002, 2011).

In contrast, few studies in this review reported subtests providing information on specific linguistic domains like grammar (which 32% of studies reported). Recall that Calder and colleagues (2023) found that a norm-referenced receptive vocabulary measure overestimated the abilities of nonautistic children with LI and that only some subtests of the Clinical Evaluation of Language Fundamentals-3 (CELF-3; Semel et al., 1995) predicted LI status. The authors used a cut point of −1.5 *SD* on the CELF-3 (Semel et al., 1995) derived from their population mean, as this test was not normed on Australian children and was published nearly 30 years ago. While every study cannot determine its own population mean, these methods underline the importance of precision in reporting assessments, including interpretation and use. Here, examining the clinical validity of measures was beyond the scope of this review (Nitido & Plante, 2020). Yet, documenting how these measures function across various samples of autistic individuals is crucial to enhancing the quality of the evidence base informing best practices in assessment.

Recommendations for reporting

Given the variability in reporting and assessments of studies in this review, one question is how to report information on structural language in autism research. Here, we are guided by real-world implications for autistic individuals. Through our experiences with autistic individuals in research and on our research team, having access to resources to reach their goals is a priority. Access depends on autistic individuals and relevant parties (previously called “stakeholders”) having information about norm-referenced language assessments, including how and whether they are meaningful to them. In the real world, autistic youth face disparities in accessing and receiving speech-language services (Taylor & Henninger, 2015). Research may contribute to this disparity, as insufficient transparency in reporting hinders understanding who is and is not included in the evidence base informing development of evidence-based practices, supports, and our understanding of autism. Per Adlof and Hogan (2019), without assessment of all relevant areas of language, it is impossible to understand all potential areas of unmet access needs.

We call for replicable reporting that allows for full evaluation of linguistic heterogeneity. In a review on definitions of nonspeaking and minimally speaking autistic children, Koegel and colleagues (2020) identified several key needs for quality reporting in future intervention studies, including: (a) clearly identifying participant language profiles using systematic assessment, (b) use of norm-referenced receptive–expressive vocabulary and language tests when possible (as some assessments do not have adaptations for individuals who do not use spoken language) along with careful interpretation of the validity of the findings, and (c) inclusion of both verbal and nonverbal cognitive ability. We realize that including heterogeneity in studies of structural language in autism when many standardized assessments are not developed for nonspeaking or minimally speaking individuals presents unique considerations. Workable solutions might include precisely stating what assessments were empirically shown to be accessible or not to an individual (vs assuming accessibility), using standardized assessments in accessible formats (e.g. touch screen for receptive vocabulary; Plesa Skwerer et al., 2016), and developing more broadly accessible standardized measures, with the idea that measures can be replicable and harmonized across studies versus assuming there is one idealized norm for spoken language

development. In addition, while every study clearly did not aim to comprehensively assess the domains of structural language pertaining to LI, studies might consider the following as a start.

Following the can of Koegel and colleagues (2020), studies should precisely characterize participants. We take this to entail reporting clear diagnostic information, language subtest outcomes (vs only summary scores), information on when data were collected, and information on to what extent the test norming sample was similar to participants; much of this information aligns to current best practices for reporting. Overall, precision in reporting original work comprises one part of transparency in research reporting. Providing this information is crucial for facilitating understanding of research findings and has real-world relevance for consumers of research within and beyond the ivory tower.

Limitations

This systematic review had several limitations. First, the search was limited to records in English and outcomes using mostly verbal norm-referenced assessments in English, with no specific variants of English. This exclusion prevents a broader understanding of LI in autistic individuals. Second, our search may have overlooked autistic participants, who, in earlier studies, may have had other diagnoses (e.g. “mental retardation”; Croen et al., 2002). Searches did not include education-focused databases such as ERIC. However, with the preliminary test searches used to develop our strategy, benchmark articles that were not found in PubMed or PsycInfo were found in Linguistic and Language Behavior Abstracts (LLBA). This indicated a potential for expanding the search results, because of the likelihood of more overlap with an education-specific database. Third, focusing on studies that primarily looked at language as an outcome using norm-referenced assessment, versus experimental measures or interventions, may have resulted in overlooking studies that used norm-referenced assessments in the target populations. Yet, evaluating assessment of structural language in these studies would not have been feasible, as they each have different requirements for quality analysis that extend beyond this review. Finally, while the search allowed for flagging reports with mentions of “language,” it is impossible to estimate the number of papers on autism post-DSM-5 without any information about language.

Future research

Our findings highlight directions for further work. Although studies mostly did not differ by DSM version, including reporting practices and frequency of assessments, there is a need to better understand how these approaches to assessing structural language in research align to actual clinical practice. Specific areas in need of clarification are to what extent these approaches are inclusive of the diverse autistic population in terms of test norming (Nitido & Plante, 2020). Furthermore, understanding to what extent interpretation and use of these assessments is relevant to autistic individuals is important. Although beyond the scope of this review, future work should also examine whether autism research post-DSM-5 is as likely to report information on participant language abilities, with the goal of providing advocacy for individuals who may want language supports. Finally, it is a question of how studies cite structural language in autism studies that use norm-referenced assessments, as that shapes our understanding of autism. For example, Russell and colleagues (2019)

found 91% of 187 publications citing original autism studies treated original study findings as being broadly applicable to all autistic individuals, even though 94% of these original studies were estimated to exclude autistic individuals with intellectual disability (who comprise a significant proportion of the autistic population, with estimates ranging from 38% to 50%; Charman et al., 2011; Loomes et al., 2017; Maenner et al., 2023). There are many more future directions, but these next steps will help strengthen the evidence base and its relevance to autistic individuals.

Conclusion

In documenting reporting practices prior to and post-DSM-5, this review advocates for greater detail and clarity in reporting of diagnostic labels and language assessment outcomes. Together with exclusion of racially and ethnically minoritized autistic individuals in research (Girolamo et al., 2023), there is a need to critically evaluate findings on language in autistic individuals across the spectrum, lifespan, and walks of life.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: T.G. was supported by an American Speech-Language-Hearing Foundation Grant and University of Kansas Research Excellence Initiative Grant (Girolamo). T.G. and M.L.R. were supported by NIH-NIDCD T32DC000052 (Rice). T.G. and I.-M.E. were supported by NIH-NIDCD T32DC017703 (Eigsti & Myers). L.S. was supported by NIH-NIDCD P50DC018006 (Kasari/Tager-Flusberg). M.L.R. was supported by NIH-NIDCD R01DC001803 (Rice). I.-M.E. was supported by NIH-NIDCD R01MH112678 (Eigsti).

References

*Study included in review.

Adlof SM, & Hogan TP (2019). If we don’t look, we won’t see: Measuring language development to inform literacy instruction. Policy Insights from the Behavioral and Brain Sciences, 6(2), 210–217. 10.1177/2372732219839075 [PubMed: 32864426]

American Psychiatric Association. (1952). Diagnostic and statistical manual: Mental disorders (1st ed.).

American Psychiatric Association. (1968). Diagnostic and statistical manual of mental disorders (2nd ed.).

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.).

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed., Rev.).

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.).

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text rev.).

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

*Anderson DK, Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Welch K, & Pickles A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. Journal

of Consulting and Clinical Psychology, 75(4), 594–604. 10.1037/0022-006X.75.4.594 [PubMed: 17663613]

Andreou G, Lymperopoulou V, & Aslanoglou V. (2022). Developmental Language Disorder (DLD) and Autism Spectrum Disorder (ASD): Similarities in pragmatic language abilities. A systematic review. International Journal of Developmental Disabilities, 1–15. 10.1080/20473869.2022.2132669 [PubMed: 35173959]

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. 10.1016/j.ridd.2016.09.007 [PubMed: 27643571]

*Bal VH, Fok M, Lord C, Smith IM, Mirenda P, Szatmari P, Vaillancourt T, Volden J, Waddell C, & & Zwaigenbaum L. (2020). Predictors of longer-term development of expressive language in two independent longitudinal cohorts of language-delayed preschoolers with autism spectrum disorder. Journal of Child Psychology and Psychiatry, 61(7), 826–835. 10.1111/jcpp.13117 [PubMed: 31429087]

*Bal VH, Katz T, Bishop SL, & 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. 10.1111/jcpp.12609 [PubMed: 27473432]

Bartolucci G, Pierce SJ, & 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. 10.1007/BF02408431 [PubMed: 6927677]

*Bennett T, Szatmari P, Bryson S, Volden J, Zwaigenbaum L, Vaccarella L, Duku E, & Boyle M. (2008). Differentiating autism and Asperger syndrome on the basis of language delay or impairment. Journal of Autism and Developmental Disorders, 38(4), 616–625. 10.1007/s10803-007-0428-7 [PubMed: 17721697]

*Bennett T, Szatmari P, Georgiades K, Hanna S, Janus M, Georgiades S, Duku E, Bryson S, Fombonne E, & Smith I. (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. 10.1007/s10803-014-2138-2 [PubMed: 24865586]

Betz SK, Eickhoff JR, & Sullivan SF (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. Language, Speech, and Hearing Services in Schools, 44(2), 133–146. 10.1044/0161-1461(2012/12-0093) [PubMed: 23633640]

*Biller MF, & Johnson CJ (2020). Examining useful spoken language in a minimally verbal child with autism spectrum disorder: A descriptive clinical single-case study. American Journal of Speech-Language Pathology, 29(3), 1361–1375. 10.1044/2020_AJSLP-19-00085 [PubMed: 32479743]

Bishop DVM (1982). Test for reception of grammar. The University of Manchester.

Bishop DVM (1998). Development of the Children's Communication Checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 39(6), 879–891. 10.1017/S0021963098002832 [PubMed: 9758196]

Bishop DVM (2003a). The Children's Communication Checklist (2nd ed.). Harcourt Assessment.

Bishop DVM (2003b). Test for reception of grammar (2nd ed.). Harcourt Assessment.

Bishop DVM (2005). Test for reception of grammar: Electronic. Harcourt Assessment.

Blalock H. (1972). Social statistics. McGraw-Hill.

*Botting N, & Conti-Ramsden G. (2003). Autism, primary pragmatic difficulties, and specific language impairment: Can we distinguish them using psycholinguistic markers? Developmental Medicine & Child Neurology, 45(8), 515–524. 10.1111/j.1469-8749.2003.tb00951.x [PubMed: 12882530]

Boucher J. (2012). Research review: Structural language in autistic spectrum disorder–characteristics and causes. Journal of Child Psychology and Psychiatry, 53(3), 219–233. 10.1111/j.1469-7610.2011.02508.x [PubMed: 22188468]

Brignell A, Morgan AT, 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. 10.1177/2396941518767610

*Broome K, McCabe P, Docking K, Doble M, & Carrigg B. (2021). Speech abilities in a heterogeneous group of children with autism. Journal of Speech, Language, and Hearing Research, 64(12), 4599–4613. 10.1044/2021_JSLHR-20-00651

*Broome K, McCabe P, Docking K, Doble M, & Carrigg B. (2023). Speech development across subgroups of autistic children: A longitudinal study. Journal of Autism and Developmental Disorders, 53, 2570–2586. 10.1007/s10803-022-05561-8 [PubMed: 35438437]

Brownell R. (2000). Expressive One-Word Picture Vocabulary Test (3rd ed.). Academic Therapy Publications.

*Burton JM, Creaghead NA, Silbert N, Breit-Smith A, Duncan AW, & Grether SM (2020). Social communication and structural language of girls with high-functioning autism spectrum disorder. Language, Speech, and Hearing Services in Schools, 51(4), 1139–1155. 10.1044/2020_LSHSS-20-00004 [PubMed: 32916076]

Butler LK, Shen L, Chenausky KV, 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. 10.1044/2022_AJSLP-22-00098 [PubMed: 36706456]

Calder SD, Brennan-Jones CG, 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, 25, 440–448. 10.1080/17549507.2023.2171488 [PubMed: 36786688]

*Charman T, Drew A, Baird C, & Baird G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). Journal of Child Language, 30(1), 213–236. 10.1017/S0305000902005482 [PubMed: 12718299]

Charman T, Pickles A, Simonoff E, Chandler S, Loucas T, & Baird G. (2011). IQ in children with autism spectrum disorders: Data from the Special Needs and Autism Project (SNAP). Psychological Medicine, 41(3), 619–627. 10.1017/S0033291710000991 [PubMed: 21272389]

Cirnigliaro M, Chang TS, Arteaga SA, Pérez-Cano L, Ruzzo EK, Gordon A, Bicks LK, Jung JY, Lowe JK, Wall DP, & Geschwind DH (2023). The contributions of rare inherited and polygenic risk to ASD in multiplex families. Proceedings of the National Academy of Sciences of the United States of America, 120(31), e2215632120. 10.1073/pnas.2215632120

Cochrane. (2022). A tool for assessing Risk of Bias due to Missing Evidence in a synthesis (ROB-ME). https://methods.cochrane.org/bias/resources/rob-me

*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, 349–358. 10.1044/1058-0360(2003/080) [PubMed: 12971823]

Croen LA, Grether JK, Hoogstrate J, & Selvin S. (2002). The changing prevalence of autism in California. Journal of Autism and Developmental Disorders, 32(3), 207–215. [PubMed: 12108622]

Dodd B, Reilly S, Ttofari Eecen K, & Morgan AT (2018). Articulation or phonology? Evidence from longitudinal error data. Clinical Linguistics & Phonetics, 32(11), 1027–1041. 10.1080/02699206.2018.1488994 [PubMed: 29969299]

Dunn DM (2019). Peabody Picture Vocabulary Test, Fifth Edition: Manual. Pearson.

Dunn LM, & Dunn DM (1997). Peabody Picture Vocabulary Test (3rd ed.). American Guidance Service.

Dunn LM, & Dunn DM (2007). Peabody Picture Vocabulary Test (4th ed.). Pearson.

Dunn LM, & Dunn LM (1981). Peabody Picture Vocabulary Test—Revised. American Guidance Service.

*Eigsti IM, & Bennetto L. (2009). Grammaticality judgments in autism: Deviance or delay. Journal of Child Language, 36(5), 999–1021. [PubMed: 19224652]

*Eigsti IM, Bennetto L, & Dadlani MB (2007). Beyond pragmatics: Morphosyntactic development in autism. Journal of Autism and Developmental Disorders, 37(6), 1007–1023. 10.1007/s10803-006-0239-2 [PubMed: 17089196]

*Ellawadi AB, & Ellis Weismer SE (2015). Using spoken language benchmarks to characterize the expressive language skills of young children with autism spectrum disorders. American Journal of Speech-Language Pathology, 24(4), 696–707. 10.1044/2015_AJSLP-14-0190 [PubMed: 26254475]

Elliott CD (1990). Differential Ability Scales. The Psychological Corporation.

Elliott CD (2007). Differential Ability Scales (2nd ed.). Pearson.

*Ellis Weismer S, Gernsbacher MA, Stronach S, Karasinski C, Eernisse ER, Venker CE, & Sindberg H. (2011). Lexical and grammatical skills in toddlers on the autism spectrum compared to late talking toddlers. Journal of Autism and Developmental Disorders, 41(8), 1065–1075. 10.1007/s10803-010-1134-4 [PubMed: 21061053]

*Ellis Weismer S, & Kover ST (2015). Preschool language variation, growth, and predictors in children on the autism spectrum. Journal of Child Psychology and Psychiatry, 56(12), 1327–1337. 10.1111/jcpp.12406 [PubMed: 25753577]

*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. 10.1007/s10803-010-0983-1 [PubMed: 20195735]

Fenson L, Dale P, Reznick J, Thal D, Bates E, Hartung J, Pethick S, & Reilly J. (2007). The MacArthur Communicative Development Inventories: User's guide and technical manual. Paul H. Brookes.

Fey ME (1992). Articulation and phonology: Inextricable constructs in speech pathology. Language, Speech, and Hearing Services in Schools, 23(3), 225–232. 10.1044/0161-1461.2303.225

Gaeta L, & Brydges CR (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. 10.1044/2020_JSLHR-19-00299

*Gagnon D, Zeribi A, Douard É, Courchesne V, Rodríguez-Herreros B, Huguet G, Jacquemont S, Loum MA, & Mottron L. (2021). Bayonet-shaped language development in autism with regression: A retrospective study. Molecular Autism, 12(1), Article 35. 10.1186/s13229-021-00444-8

Geelhand P, Papastamou F, Deliens G, & Kissine M. (2020). Narrative production in autistic adults: A systematic analysis of the microstructure, macrostructure and internal state language. Journal of Pragmatics, 164, 57–81. 10.1016/j.pragma.2020.04.014

Girolamo TM, Shen L, Gulick A, Rice ML, & Eigsti IM (2023). Studies pertaining to language impairment in autism underreport participant sociodemographics: A systematic review. Autism, 27, 2218–2240. 10.1177/13623613231166749 [PubMed: 37157821]

*Girolamo TM, & Rice ML (2022). Language impairment in autistic adolescents and young adults. Journal of Speech, Language, and Hearing Research, 65(9), 3518–3530. 10.1044/2022_JSLHR-21-00517

*Girolamo TM, Rice ML, & Warren SF (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(2), 804–818. 10.1044/2020_AJSLP-19-00036 [PubMed: 32324427]

Goldman R, & Fristoe M. (1986). Goldman-Fristoe Test of Articulation. American Guidance Service.

Grondhuis SN, Lecavalier L, Arnold LE, Handen BL, Scahill L, McDougle CJ, & Aman MG (2018). Differences in verbal and nonverbal IQ test scores in children with autism spectrum disorder. Research in Autism Spectrum Disorders, 49, 47–55. 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(2), 260–274. 10.1007/s10803-016-2921-3 [PubMed: 27796729]

*Hart CM, & Curtin S. (2023). Trajectories of vocabulary development in children with autism spectrum disorder across multiple measures. Journal of Autism and Developmental Disorders, 53, 1347–1362. 10.1007/s10803-021-05379-w [PubMed: 34817769]

*Hartley SL, Buckendorf GR, Haines K, Hall TA, & Sikora DM (2008). The Oral and Written Language Scales: Is it useful for older children with autism spectrum disorder? Research in Autism Spectrum Disorders, 2(1), 137–146. 10.1016/j.rasd.2007.04.003

Hayden D, & Square P. (1999). Verbal Motor Assessment of Children. Pearson.

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(2), 127–136. 10.1007/BF02409656 [PubMed: 6746504]

*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. 10.1044/2019_AJSLP-19-00207 [PubMed: 32182436]

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

Interagency Autism Coordinating Committee. (2020). IACC strategic plan for Autism Spectrum Disorder (ASD) 2018–2019 update. National Institute for Occupational Safety & Health, U.S. Department of Health and Human Services. http://iacc.hhs.gov/strategic-plan/2019/

*Jiménez E, Haebig E, & Hills TT (2021). Identifying areas of overlap and distinction in early lexical profiles of children with autism spectrum disorder, late talkers, and typical talkers. Journal of Autism and Developmental Disorders, 51(9), 3109–3125. 10.1007/s10803-020-04772-1 [PubMed: 33156473]

Johnson CJ, Beitchman JH, & Brownlie EB (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. American Journal of Speech-Language Pathology, 19, 51–65. 10.1044/1058-0360(2009/08-0083) [PubMed: 19644128]

*Jokel A, Armstrong E, Gabis L, & Segal O. (2021). Associations and dissociations among phonological processing skills, language skills and nonverbal cognition in individuals with autism spectrum disorder. Folia Phoniatrica et Logopaedica, 73(3), 222–232. 10.1159/000505744 [PubMed: 32829327]

*Jyotishi M, Fein D, & Naigles L. (2017). Investigating the grammatical and pragmatic origins of wh-questions in children with autism spectrum disorders. Frontiers in Psychology, 8, Article 319. 10.3389/fpsyg.2017.00319

Kasari C, Brady N, Lord C, & Tager-Flusberg H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. Autism Research, 6(6), 479–493. 10.1002/aur.1334 [PubMed: 24353165]

*Kjelgaard MM, & Tager-Flusberg H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. Language and Cognitive Processes, 16(2–3), 287–308. 10.1080/01690960042000058 [PubMed: 16703115]

*Klusek J, Martin GE, & Losh M. (2014). A comparison of pragmatic language in boys with autism and fragile X syndrome. Journal of Speech, Language, and Hearing Research, 57(5), 1692–1707. 10.1044/2014_JSLHR-L-13-0064

Koegel LK, Bryan KM, Su PL, 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. 10.1007/s10803-020-04402-w [PubMed: 32056115]

Korkman M, Kirk U, & Kemp SI (1998). NEPSY: A developmental neuropsychological assessment. The Psychological Corporation.

*Kover ST, & Ellis Weismer S. (2014). Lexical characteristics of expressive vocabulary in toddlers with autism spectrum disorder. Journal of Speech, Language, and Hearing Research, 57(4), 1428–1441. 10.1044/2014_JSLHR-L-13-0006

*Kover ST, Haebig E, Oakes A, McDuffie A, Hagerman RJ, & Abbeduto L. (2014). Sentence comprehension in boys with autism spectrum disorder. American Journal of Speech-Language Pathology, 23(3), 385–394. 10.1044/2014_AJSLP-13-0073 [PubMed: 24687049]

*Kover ST, McDuffie AS, Hagerman RJ, & 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. 10.1007/s10803-013-1823-x [PubMed: 23588510]

Kwok EY, Brown HM, Smyth RE, & Cardy JO (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. Research in Autism Spectrum Disorders, 9, 202–222. 10.1016/j.rasd.2014.10.008

*Landa RJ, & Goldberg MC (2005). Language, social, and executive functions in high functioning autism: A continuum of performance. Journal of Autism and Developmental Disorders, 35(5), 557–573. 10.1007/s10803-005-0001-1 [PubMed: 16211332]

*Lindgren KA, Folstein SE, Tomblin JB, & Tager-Flusberg H. (2009). Language and reading abilities of children with autism spectrum disorders and specific language impairment and their first-degree relatives. Autism Research, 2(1), 22–38. 10.1002/aur.63 [PubMed: 19358305]

Loomes R, Hull L, & Mandy WPL (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. Journal of the American Academy of Child & Adolescent Psychiatry, 56(6), 466–474. 10.1016/j.jaac.2017.03.013

Lord C, Rutter M, DiLavore PC, & Risi S. (1999). Autism Diagnostic Observation Schedule. Western Psychological Services.

*Luyster RJ, Lopez K, & Lord C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur-Bates Communicative Development Inventory (CDI). Journal of Child Language, 34(3), 623–654. 10.1017/S0305000907008094 [PubMed: 17822142]

Maenner MJ, Warren Z, Williams AR, Amoakohene E, Bakian AV, Bilder DA, … & Shaw KA (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years —Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. MMWR Surveillance Summaries, 72(2), 1–14. 10.15585/mmwr.ss7202a1

Magiati I, Tay XW, & 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. 10.1016/j.cpr.2013.11.002 [PubMed: 24424351]

Marascuilo LA, & McSweeney M. (1977). Nonparametric and distribution-free methods for the social sciences. Brooks/Cole.

*McGregor KK, Berns AJ, Owen AJ, Michels SA, Duff D, Bahnsen AJ, & 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. [PubMed: 21365354]

*Minshew NJ, Goldstein G, & Siegel DJ (1995). Speech and language in high-functioning autistic individuals. Neuropsychology, 9(2), 255–261. 10.1037/0894-4105.9.2.255

*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, Article 320. 10.3389/fpsyg.2017.00320

Mullen EM (1995). Mullen Scales of Early Learning. Pearson.

Munson J, Dawson G, Sterling L, Beauchaine T, Zhou A, Koehler E, Lord C, Rogers S, Sigman M, Estes A, & Abbott R. (2008). Evidence for latent classes of IQ in young children with autism spectrum disorder. American Journal on Mental Retardation, 113(6), 439–452. 10.1352/2008.113:439-452 [PubMed: 19127655]

Musgrove M. (2015, July 6). Dear colleague letter on speech and language services for students with autism spectrum disorder. Office for Civil Rights, 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/

*Nadig A, & Mulligan A. (2017). Intact non-word repetition and similar error patterns in language-matched children with autism spectrum disorders: A pilot study. Journal of Communication Disorders, 66, 13–21. 10.1016/j.jcomdis.2017.03.003 [PubMed: 28349875]

*Nevill R, Hedley D, Uljarević M, Sahin E, Zadek J, Butter E, & Mulick JA (2019). Language profiles in young children with autism spectrum disorder: A community sample using multiple assessment instruments. Autism, 23(1), 141–153. 10.1177/1362361317726245 [PubMed: 29126347]

Newman L, Wagner M, Huang T, Shaver D, Knokey AM, Yu J, Contreras E, Ferguson K, Greene S, Nagle K, & Cameto R. (2011). Secondary school programs and performance of students with disabilities: A special topic report of findings from the National Longitudinal Transition

Study-2 (NLTS2) (NCSER 2012–3000). National Center for Special Education Research. https://files.eric.ed.gov/full-text/ED526242.pdf

Nitido H, & Plante E. (2020). Diagnosis of developmental language disorder in research studies. Journal of Speech, Language, and Hearing Research, 63(8), 2777–2788. 10.1044/2020_JSLHR-20-00091

Norbury CF, 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. [PubMed: 27184709]

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, . . .Moher D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLOS Medicine, 18(3), Article e1003583. 10.1371/journal.pmed.1003583

*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. 10.1002/aur.12 [PubMed: 19360656]

*Perovic A, Modyanova N, & Wexler K. (2013). Comparison of grammar in neurodevelopmental disorders: The case of binding in Williams syndrome and autism with and without language impairment. Language Acquisition, 20(2), 133–154. 10.1080/10489223.2013.766742 [PubMed: 25170241]

*Plesa Skwerer D, Jordan SE, Brukilacchio BH, 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. 10.1177/1362361315600014 [PubMed: 26408635]

*Prescott KE, & Ellis Weismer S. (2022). Children with ASD and communication regression: Examining pre-loss skills and later language outcomes through the preschool years. Journal of Autism and Developmental Disorders, 52(5), 1956–1970. 10.1007/s10803-021-05098-2 [PubMed: 34061309]

*Rapin I, Dunn MA, Allen DA, Stevens MC, & Fein D. (2009). Subtypes of language disorders in school-age children with autism. Developmental Neuropsychology, 34(1), 66–84. 10.1080/87565640802564648 [PubMed: 19142767]

*Reinhartsen D, Tapia A, Watson L, Crais E, Bradley C, Fairchild J, Herring A, & Daniels J. (2019). Expressive dominant versus receptive dominant language patterns in young children: Findings from the study to explore early development. Journal of Autism and Developmental Disorders, 49(6), 2447–2460. 10.1007/s10803-019-03999-x [PubMed: 30937735]

Rice ML (2017, November 1). What causes specific language impairment? Open Access Government. https://www.openaccessgovernment.org/causes-specific-language-impairment/39203/

Rice ML, & Wexler K. (2001). Rice Wexler Test of Early Grammatical Impairment. Hove.

*Riley E, Paynter J, & Gilmore L. (2019). Comparing the Mullen Scales of Early Learning and the Preschool Language Scale—Fifth Edition for young children with autism spectrum disorder. Advances in Neurodevelopmental Disorders, 3(1), 29–37. 10.1007/s41252-018-0084-2

*Roberts JA, Rice ML, & Tager-Flusberg H. (2004). Tense marking in children with autism. Applied Psycholinguistics, 25(3), 429–448. 10.1017/S0142716404001201

Rosen NE, Lord C, & Volkmar FR (2021). The diagnosis of autism: From Kanner to DSM-III to DSM-5 and beyond. Journal of Autism and Developmental Disorders, 51, 4253–4270. 10.1007/s10803-021-04904-1 [PubMed: 33624215]

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), Article 9. 10.1186/s13229-019-0260-x

Rutter M, Bailey A, & Lord C. (2003). Social communication questionnaire. Western Psychological Services.

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

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. 10.1007/s00702-023-02592-y [PubMed: 36922431]

Selin CM, Rice ML, Girolamo TM, & Wang CJ (2022). Work setting effects on speech-language pathology practice: Implications for identification of children with specific language impairment. American Journal of Speech-Language Pathology, 31(2), 854–880. 10.1044/2021_AJSLP-21-00024 [PubMed: 35120298]

Semel EM, Wiig EH, & Secord W. (1995). Clinical Evaluation of Language Fundamentals (3rd ed.). The Psychological Corporation, Harcourt Brace.

Semel EM, Wiig EH, & Secord W. (2003). Clinical Evaluation of Language Fundamentals (4th ed.). Pearson.

Sherman SL, Kidd SA, Riley C, Berry-Kravis E, Andrews HF, Miller RM, . . .Brown WT. (2017). FORWARD: A registry and longitudinal clinical database to study fragile X syndrome. Pediatrics, 139(Suppl. 3), S183–S193. 10.1542/peds.2016-1159E [PubMed: 28814539]

Shriberg LD, & Mabie HL (2017, December). Speech and motor speech assessment findings in eight complex neurodevelopmental disorders (Technical Report No. 24). https://waismanphonology.wiscweb.wisc.edu/wp-content/uploads/sites/532/2018/05/TREP24.pdf

Shriberg LD, Paul R, Black LM, & van Santen JP (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. Journal of Autism and Developmental Disorders, 41, 405–426. 10.1007/s10803-010-1117-5 [PubMed: 20972615]

Shriberg LD, Tomblin JB, & McSweeny JL (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. Journal of Speech, Language, and Hearing Research, 42(6), 1461–1481. 10.1044/jslhr.4206.1461

Slušná D, Rodríguez A, Salvadó B, Vicente A, & Hinzen W. (2021). Relations between language, non-verbal cognition, and conceptualization in non-or minimally verbal individuals with ASD across the lifespan. Autism & Developmental Language Impairments, 6, 23969415211053264. 10.1177/23969415211053264

Sparrow SS, Cicchetti DV, & Balla DA (2005). Vineland Adaptive Behavior Scales (2nd ed.). APA PsycTests.

Sparrow SS, Cicchetti DV, & Saulnier CA (2016). Vineland Adaptive Behavior Scales (3rd ed.). 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. 10.1044/2017_JSLHR-L-17-0248

Surís A, Holliday R, & North CS (2016). The evolution of the classification of psychiatric disorders. Behavioral Sciences, 6(1), 5. 10.3390/bs6010005 [PubMed: 26797641]

Tager-Flusberg H, & Kasari C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. Autism Research, 6(6), 468–478. 10.1002/aur.1329 [PubMed: 24124067]

Taylor JL, & Henninger NA (2015). Frequency and correlates of service access among youth with autism transitioning to adulthood. Journal of Autism and Developmental Disorders, 45, 179–191. 10.1007/s10803-014-2203-x [PubMed: 25081594]

*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. 10.1007/s10803-006-0300-1 [PubMed: 17180717]

*Thurm A, Manwaring SS, Swineford L, & Farmer C. (2015). Longitudinal study of symptom severity and language in minimally verbal children with autism. Journal of Child Psychology and Psychiatry, 56(1), 97–104. 10.1111/jcpp.12285 [PubMed: 24961159]

*Thurman AJ, & Hoyos Alvarez C. (2020). Language performance in preschool-aged boys with nonsyndromic autism Spectrum disorder or fragile X syndrome. Journal of Autism and Developmental Disorders, 50(5), 1621–1638. 10.1007/s10803-019-03919-z [PubMed: 30783899]

Tomblin JB, Records NL, 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.

Tomblin JB, Records NL, & 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. 10.1044/jshr.3906.1284

Veritas Health Innovation. (2021). Covidence. https://www.covidence.org

Viswanathan M, Ansari M, Berkman N, Chang S, Hartling L, McPheeters M, & Treadwell J. (2012). Assessing the risk of bias of individual studies in systematic reviews of health care interventions: Methods guide for comparative effectiveness reviews. Agency for Healthcare Research and Quality. http://www.effectivehealthcare.ahrq.gov/

*Volden J, Smith IM, Szatmari P, Bryson S, Fombonne E, Mirenda P, Roberts W, Vaillancourt T, Waddell C, & & Zwaigenbaum L. (2011). Using the Preschool Language Scale, Fourth Edition to characterize language in preschoolers with autism spectrum disorders. American Journal of Speech-Language Pathology, 20(3), 200–208. 10.1044/1058-0360(2011/10-0035) [PubMed: 21478278]

Waizbard-Bartov E, Fein D, Lord C, & Amaral DG (2023). Autism severity and its relationship to disability. Autism Research, 16(4), 685–696. 10.1002/aur.2898 [PubMed: 36786314]

Wechsler D. (1974). Wechsler Intelligence Scale for Children—Revised. The Psychological Corporation.

Wechsler D. (1981). Wechsler Adult Intelligence Scale—Revised. The Psychological Corporation.

Wechsler D. (1991). The Wechsler Intelligence Scale for Children (3rd ed.). The Psychological Corporation.

Wechsler D. (1997). Wechsler Adult Intelligence Scale (3rd ed.). The Psychological Corporation.

Wechsler D. (2003). Wechsler Intelligence Scale for Children (4th ed.). The Psychological Corporation.

Wechsler D. (2014). Wechsler Intelligence Scale for Children (5th ed.). Pearson.

Whitehouse AJ, Barry JG, & Bishop DV (2008). Further defining the language impairment of autism: Is there a specific language impairment subtype? Journal of Communication Disorders, 41(4), 319–336. 10.1016/j.jcomdis.2008.01.002 [PubMed: 18295779]

Wiig EH, Secord W, & Semel E. (1992). Clinical evaluation of language fundamentals-preschool. The Psychological Corporation.

Wiig EH, Semel E, & Secord W. (2013). Clinical evaluation of language fundamentals (5th ed.). NCS Pearson.

Williams KT (1997). Expressive Vocabulary Test. American Guidance Service.

Williams KT (2007). Expressive Vocabulary Test (2nd ed.). Pearson.

Williams KT (2019). Expressive Vocabulary Test: Manual (3rd ed.). Pearson.

*Worth S, & Reynolds S. (2008). The assessment and identification of language impairment in Asperger's syndrome: A case study. Child Language Teaching and Therapy, 24(1), 55–71. 10.1177/0265659007084568

*Woynaroski T, Yoder P, & Watson LR (2016). Atypical cross-modal profiles and longitudinal associations between vocabulary scores in initially minimally verbal children with ASD. Autism Research, 9(2), 301–310. 10.1002/aur.1516 [PubMed: 26180010]

Zhang X, & Tomblin JB (2000). The association of intervention receipt with speech-language profiles and social-demographic variables. American Journal of Speech-Language Pathology, 9(4), 345–357.

Zimmerman IL, Steiner VG, & Pond RE (2002). Preschool Language Scale (4th ed.). The Psychological Corporation.

Zimmerman IL, Steiner VG, & Pond RE (2011). Preschool Language Scales (5th ed.). Pearson.

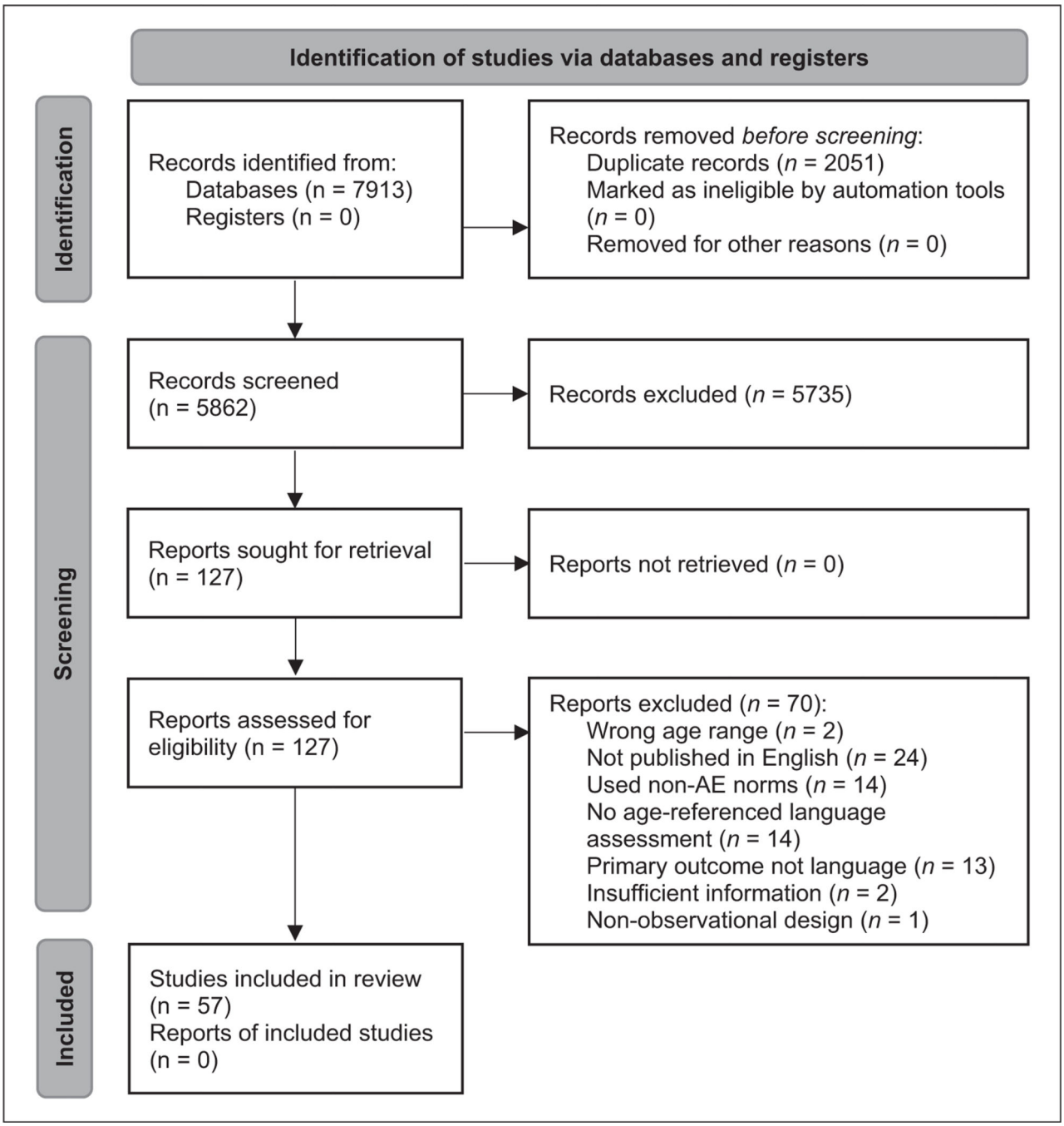

Figure 1.
PRISMA flow chart of searches of databases and other sources (Page et al., 2021).

Table 1.

Quality analysis of studies reporting use of age-referenced measures to assess structural language in autism.

Reference	*N*	Reporting	Dx	Group	NVIQ	Artic/Sp	Overall	Grammar	Vocab
DSM-III/DSM-III-R									
Rapin et al. (2009)	118	Low	High	High	High	High	High	Low	High
DSM-IV/DSM-IV-TR									
Anderson et al. (2007)	98	High	High	N/A	High	Low	High	Low	Low
Bal et al. (2020)	267	High	Low	High	High	Low	High	Low	Low
Bennett et al. (2008)	64	High	High	High	High	Low	Low	High	Low
Bennett et al. (2014)	330	High	High	High	Low	Low	High	Low	Low
Botting and Conti-Ramsden (2003)	13	High	High	High	High	Low	High	High	High
Charman et al. (2003)	134	High	High	N/A	High	Low	Low	Low	High
Condouris et al. (2003)	44	High	High	High	High	Low	High	High	High
Eigsti et al. (2007)	16	High	High	High	High	Low	Low	Low	High
Eigsti and Bennetto (2009)	21	High	High	High	High	Low	Low	Low	High
Ellawadi and Ellis Weismer (2015)	105	High	High	High	High	Low	High	Low	Low
Ellis Weismer and Kover (2015)	129	High	High	High	High	Low	High	Low	High
Ellis Weismer et al. (2010)	257	High	High	N/A	High	Low	High	Low	Low
Ellis Weismer et al. (2011)	40	High	High	N/A	High	Low	Low	High	High
Gagnon et al. (2021)	2047	High	High	High	High	Low	High	Low	High
Hartley et al. (2008)	53	High	High	N/A	High	Low	High	Low	Low
Jyotishi et al. (2017)	20	High	Low	High	High	Low	High	Low	High
Kjelgaard and Tager-Flusberg (2001)	89	High	High	High	High	High	High	Low	High
Kover and Ellis Weismer (2014)	57	High	High	N/A	High	Low	Low	Low	High
Kover et al. (2013)	49	High	Low	N/A	High	Low	Low	Low	High
Landa and Goldberg (2005)	19	High	Low	High	High	Low	Low	High	Low
Lindgren et al. (2009)	52	High	Low	High	High	Low	High	Low	High
Luyster et al. (2007)	93	High	Low	N/A	High	Low	Low	Low	High
McGregor et al. (2012)	33	High	Low	High	High	Low	High	High	High
Minshew et al. (1995)	62	High	High	High	High	Low	Low	Low	Low
Modyanova et al. (2017)	83	Low	High	High	High	High	Low	High	High

Reference	*N*	Reporting	Dx	Group	NVIQ	Artic/Sp	Overall	Grammar	Vocab
Paul et al. (2008)	37	High	High	N/A	High	Low	High	Low	High
Perovic et al. (2013)	48	High	High	High	High	Low	Low	High	High
Prescott and Ellis Weismer (2022)	126	High	High	High	High	Low	High	Low	Low
Riley et al. (2019)	24	High	High	N/A	High	Low	High	Low	Low
Roberts et al. (2004)	62	High	High	High	High	High	Low	High	High
Thurm et al. (2007)	59	High	Low	N/A	High	Low	High	Low	High
Thurm et al. (2015)	70	High	High	High	High	Low	High	Low	Low
Volden et al. (2011)	294	High	High	N/A	High	Low	High	Low	Low
Whitehouse et al. (2008)	34	High	High	High	High	High	High	High	Low
Worth and Reynolds (2008)	1	High	Low	Low	Low	Low	High	High	Low
Woynaroski et al. (2016)	87	High	High	High	Low	Low	High	Low	High
DSM-5									
Bal et al. (2016)	1470	High	Low	High	High	Low	High	Low	Low
Biller and Johnson (2020)	1	High	High	High	High	High	High	Low	High
Broome et al. (2021)	22	High	High	N/A	High	High	High	Low	High
Broome et al. (2023)	23	High	High	N/A	High	High	High	Low	High
Burton et al. (2020)	16	High	High	High	High	Low	High	High	Low
Girolamo et al. (2020)	10	High	Low	Low	High	High	High	High	Low
Girolamo and Rice (2022)	13	High	High	High	High	High	High	High	High
Haebig and Sterling (2017)	50	High	Low	N/A	High	Low	Low	Low	High
Hart and Curtin (2023)	20	Low	High	N/A	Low	Low	Low	Low	High
Huang and Finestack (2020)	15	High	Low	High	High	Low	High	High	Low
Jiménez et al. (2021)	118	High	Low	High	High	Low	Low	Low	High
Jokel et al. (2021)	21	High	High	Low	High	Low	High	Low	Low
Klusek et al. (2014)	67	High	Low	N/A	High	Low	Low	Low	High
Kover et al. (2014)	45	High	Low	N/A	High	Low	Low	High	High
Nadig and Mulligan (2017)	9	High	Low	N/A	High	Low	High	Low	Low
Nevill et al. (2019)	104	High	High	N/A	High	Low	High	Low	Low
Plesa Skwerer et al. (2016)	19	High	Low	High	High	Low	High	Low	High
Reinhartsen et al. (2019)	695	High	Low	N/A	High	Low	High	Low	Low
Sterling (2018)	37	High	Low	N/A	High	High	Low	High	High

Reference	*N*	Reporting	Dx	Group	NVIQ	Artic/Sp	Overall	Grammar	Vocab
Thurman and Hoyos Alvarez (2020)	25	High	Low	High	High	Low	High	Low	High

Note. Broome et al. (2021, 2023) and Jokel et al. (2021) included participants with a DSM-IV or DSM-5 diagnosis of ASD. Minshew et al. (1995) included participants with a DSM-III-R/DSM-IV diagnosis. Reporting = complete reporting of outcomes; Dx = reported clinical diagnostic label; Group = provided an operational definition for grouping when applicable; NVIQ = used age-referenced nonverbal intelligence measure; Artic/Sp = used age-referenced articulation/speech measure; Overall = used age-referenced overall language measure; Grammar = used age-referenced grammar measure; Vocab = used age-referenced vocabulary measure; High = autistic $N \geqslant 20$; complete outcome reporting; provided a specific diagnostic label; provided an operational definition for grouping autistic participants; provided NVIQ; provided information about articulation and speech abilities; provided information about overall expressive-receptive language ability; provided information about grammar abilities; provided information about vocabulary abilities; Low = autistic $N < 20$; selective outcome reporting (e.g. excluding participants who impacted results in a way the authors did not intend as outliers); did not provide a specific diagnostic label; did not provide an operational definition for grouping autistic participants; did not provide NVIQ; did not provide information about articulation and speech abilities; did not provide information about overall expressive-receptive language ability; did not provide information about grammar abilities; did not provide information about vocabulary abilities; N/A = not applicable; DSM-III/III-R/IV/IV-TR/5 = *Diagnostic and statistical manual of mental disorders*-3rd/3rd-revised/4th/4th-text revision-5th edition (American Psychiatric Association, 1980, 1987, 1994, 2000, 2013); ASD = autism spectrum disorder.

Table 2.

Proportions of studies receiving high-quality ratings per criterion pre- and post-DSM-5.

Criterion	Pre-DSM-5 (*n* = 35)		Post-DSM-5 (*n* = 22)		Total (*n* = 57)		*p*
	n	%	*n*	%	*n*	%	
Autism sample size	31	88.6	15	68.2	46	80.7	0.086
Selective outcome reporting	33	94.3	22	100.0	55	96.5	0.518
Clinical diagnosis	**27**	**77.1**	**10**	**45.5**	**37**	**64.9**	**0.015**
Grouping *	19	86.4	9	90.0	28	87.5	1.000
NVIQ	31	88.6	21	95.5	52	91.2	0.639
Articulation/speech	5	14.3	6	27.3	11	19.3	0.305
Overall language ability	21	60.0	15	68.2	36	63.2	0.533
Grammar	12	34.3	6	27.3	18	31.6	0.579
Vocabulary	21	60.0	13	59.1	34	59.6	0.946

Note. Significant differences in bolded text. DSM-III study not reported separately, as there was one DSM-III study and 35 DSM-IV studies. Sample size, selective outcome reporting, NVIQ, articulation and speech, and grammar used Fisher's exact test due to small sample size. Clinical diagnosis, overall language ability and vocabulary used chi-square tests of homogeneity. DSM-5 = Diagnostic and statistical manual of mental disorders-5th edition (American Psychiatric Association [APA], 2013); NVIQ = used age-referenced nonverbal intelligence measure; DSM-III/IV/5 = Diagnostic and statistical manual of mental disorders-3rd/4th (APA, 1980, 1994).

* Total *n* for grouping = 32. Pre-DSM-5 studies that used grouping criteria *n* = 22, and post-DSM-5 studies that used grouping criteria *n* = 10.

Table 3.

Frequencies of diagnostic labels, groups, cognitive abilities, and language domains assessed.

	Pre-DSM-5 (*n* = 35)		**Post-DSM-5 (*n* = 22)**		**total (*n* = 57)**		***p***
	n	**%**	***n***	**%**	***n***	**%**	
Diagnostic labels							
ASD	**15**	**42.9**	**21**	**95.5**	**36**	**63.2**	**<0.001**
ASD + genetic diagnoses	**0**	**0.0**	**4**	**18.2**	**4**	**7.0**	**0.019**
Asperger syndrome	4	11.4	2	9.1	6	10.5	1.000
Autism	**14**	**40.0**	**2**	**9.1**	**16**	**28.1**	**0.011**
Autistic disorder	8	22.9	1	4.5	9	15.8	0.132
PDD-NOS	9	25.7	4	18.2	13	22.8	0.509
Grouping							
Autism—no specifiers	18	51.4	10	45.5	28	49.1	0.661
Autism + language impairment	10	28.6	4	18.2	14	24.6	0.375
“High functioning” autism	5	14.3	1	4.5	6	10.5	0.389
Minimally speaking	2	5.7	3	13.6	5	8.8	0.364
Cognitive ability							
Nonverbal	30	85.7	19	86.4	49	86.0	1.000
Verbal	11	31.4	2	9.1	13	22.8	0.050
Full scale/overall	**10**	**28.6**	**1**	**4.5**	**11**	**19.3**	**0.037**
Language domains							
Articulation/speech	5	14.3	7	31.8	12	21.1	0.181
Expressive grammar	7	20.0	5	22.7	12	21.1	1.000
Receptive grammar	6	17.1	4	18.2	10	17.5	1.000
Overall receptive	19	54.3	16	72.7	35	61.4	0.164
Overall expressive	21	60.0	14	63.6	35	61.4	0.784
Receptive vocabulary	18	51.4	11	50.0	29	50.9	0.916
Expressive vocabulary	15	42.9	12	54.5	27	47.4	0.390

Note. Significant differences in bolded text. PDD-NOS = Pervasive Developmental Disorder–Not Otherwise Specified. Full scale/overall = verbal plus nonverbal together. ASD + genetic diagnoses, Asperger syndrome, autistic disorder, “high functioning” autism, minimally verbal, nonverbal cognitive ability, full scale/overall cognitive ability, articulation/speech, expressive grammar, and receptive grammar used Fisher’s exact test due to small sample size. ASD, Autism, PDD-NOS, autism—no specifiers, autism + language impairment, verbal cognitive ability, overall receptive, overall expressive,

receptive vocabulary, and expressive vocabulary used chi-square tests of homogeneity. DSM-5 = *Diagnostic and statistical manual of mental disorders*-5th edition (American Psychiatric Association, 2013); ASD = autism spectrum disorder.