

Article

Not peer-reviewed version

Assessing Language Skills in Children Aged 4 to 6 Years with Autism Spectrum Disorder: A Prospective Study

[Jade Mériaux](#)^{*}, Sandrine Foin, [Abdessadek El Ahmadi](#), Christine Assaiante, [Pascale Colé](#)^{*}

Posted Date: 30 September 2025

doi: 10.20944/preprints202509.2489.v1

Keywords: Autism Spectrum Disorder; preschool; children; lexicon; comprehension; phonology; articulation; standardized test



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Assessing Language Skills in Children Aged 4 to 6 Years with Autism Spectrum Disorder: A Prospective Study

Jade Mériaux ^{1,2,*}, Sandrine Foin ², Abdessadek El Ahmadi ², Christine Assaïante ² and Pascale Colé ^{2,*}

¹ Saint-Louis Early Medico-Social Action Center and Rabelais-Littoral Day Hospital, Edouard Toulouse Hospital Center, 118 chemin de Mimet 13015 Marseille, France

² Psychology and Neuroscience Research Center, mixed research Unit 7077, Aix-Marseille University-National Center for Scientific Research, Saint-Charles Center, Pôle 3C, Case C, 3 Place Victor Hugo 13331 Marseille Cedex 03, France

* Correspondence: jade.meriaux@ch-edouard-toulouse.fr (J.M.); pascale.cole@univ-amu.fr (P.C.)

Abstract

Background/Objectives: Language impairments are highly prevalent in children with Autism Spectrum Disorder (ASD). In preschoolers (3-6 years), language development predicts future social outcomes. Despite the availability of standardized tests for typically developing children, few studies have specifically examined language impairments in preschool-aged children with ASD using these tools. This study aimed to comprehensively assess receptive and expressive lexicon, receptive comprehension, phonology and articulation using standardized tools, and to evaluate their feasibility. A secondary goal was to compare the results obtained with standardized tests to those from developmental batteries and hetero-assessments (caregiver reports). **Methods:** Forty-seven children with ASD aged 4 to 6 years participated. Assessments included standardized language tests, developmental batteries and hetero-assessments. The dichotomous Rasch model evaluated feasibility and item performance of standardized tests. Concordance across methods was analyzed via Pearson correlations and stepwise linear regression. **Results:** Standardized assessments were feasible for most participants despite wide variability in language abilities. Partial but non-equivalent concordance was found among assessment methods, each providing complementary insights into language profiles. **Conclusions:** Combining multiple assessment methods is crucial to capture the complexity of language development in children with ASD. Standardized tests can be adapted and provide more precise profiles than developmental batteries or hetero-assessments alone. A multimodal approach is essential to accurately identify language strengths and therapeutic targets in preschool-aged children with ASD.

Keywords: Autism Spectrum Disorder; preschool; children; lexicon; comprehension; phonology; articulation; standardized test

Introduction

The Diagnostic and Statistical Manual of Mental Disorders, fifth Edition (DSM-5) [1] identifies language impairment as a potential indicator of Autism Spectrum Disorder (ASD), noting that receptive language is often more affected than expressive language. For language assessment in ASD, it recommends a multi-informant approach that includes clinical observations, developmental history and three main categories of assessment: standardized tests (normed on typically developing individuals), developmental batteries (which assess broad development domains and estimate developmental age) and informant-based measures, such as self-reports (when applicable) and third-party reports from caregivers or teachers (hetero-assessments). These same categories are frequently

used in research contexts to assess language skills in children with ASD. However, the DSM-5 recommendations bring to light two major challenges. Firstly, the issue of consistency across different types of assessments must be addressed. Secondly, clinical practice faces a significant limitation due to the lack of standardized tools specifically adapted for assessing early language in children with ASD [2–4] – especially in the 4–6-year age group targeted by this study.

Language development in ASD is highly heterogeneous, ranging from a complete absence of verbal output to the use of complex sentences [5]. While qualitative communication deficits – in both verbal and nonverbal modalities (e.g., gestures, posture, eye gaze) – are common, quantitative speech delays (e.g., no single words by age 2, no two-word phrases by age 3; [6]) are not universal. Nevertheless, language difficulties are typically observable in around 80% of children with ASD by 18–24 months [7,8] and are often among the earliest concerns reported by caregivers [9,10]. Early language development in the preschool years (ages 3–6) is a robust predictor of later social functioning [8]. Delays during this period are associated with more pronounced autistic symptoms [11], whereas acquiring functional language by age 5 – which encompasses phonological, lexical, morphosyntactic and discourse-level abilities in both expressive and receptive modalities – is associated with more favorable long-term outcomes [12]. This period is also critical as it coincides with entry into formal schooling and the onset of extensive social interactions, when language, literacy and numeracy skills develop rapidly [13]. Despite the availability of standardized tests for typically developing children, assessing language skills in preschool-aged children with ASD remains challenging. This is partly due to delayed diagnoses, which often occur between ages 3 and 6 [14,15], despite recommendations for identification by age 2 [16], and to the fact that core ASD symptoms peak in expression between ages 4 and 5 [17].

This study pursued several interrelated objectives. First, we aimed to assess the feasibility and reliability of administering standardized language assessments to children with ASD aged 4 to 6 years. Second, we examined the level of agreement between standardized tests and more commonly used tools in clinical and research contexts – namely, developmental batteries and hetero-assessments. We also explored the degree of concordance between developmental batteries and hetero-assessments. Direct assessments (standardized tests and developmental batteries) offer clear advantages over hetero-assessments: they provide objective, real-time observation of language behavior in controlled settings, thereby reducing biases related to environmental influences or caregiver interpretation. In particular, standardized language tests yield more precise and consistent data than developmental batteries [18]. Nevertheless, their use in ASD remains limited – rarely administered before age 3, but more commonly used only after age 7.

Systematic reviews and meta-analyses (e.g., [19–22]) report that studies of high-risk infants (i.e., siblings of children with ASD) tend to focus on early developmental windows (5–24 months), with ASD diagnoses typically confirmed between 24 and 36 months. These studies mostly rely on developmental batteries and hetero-assessments [23–26], while standardized language tests [27] are rarely employed. Language delays in both expressive and receptive modalities have been reported as early as 8 [28], 12 [29], 18 [30] or 36 months [31–33], and persist through ASD diagnosis at 36 months. However, cross-study comparisons are difficult due to significant heterogeneity in assessment tools, which often results in inconsistent findings [18]. These discrepancies may arise from floor effects [33] or the limited variability of early expressive language measures [31].

A major limitation of this body of research is the reliance on broad developmental tools that do not isolate specific language components – such as vocabulary or syntax – needed to distinguish between receptive and expressive language profiles [18]. While standardized tests offer age-normed scoring and greater specificity, developmental batteries often lack transparency in item selection, limiting the granularity of analysis. To our knowledge, only Yirmiya and colleagues [27] combined a developmental battery (Reynell Developmental Language Scales – RDLS –; [34]) with a standardized test (Clinical Evaluation of Language Fundamentals – CELF –Preschool; [35]). They found lower expressive and receptive language in high-risk toddlers by 24 months, persisting to ASD diagnosis at 36 months. However, their reporting of only composite scores – without analysis of individual

subtests – yielded a global profile comparable to that of a developmental battery alone, thereby precluding more detailed interpretation.

In older children and young adults with ASD (up to 19 years), standardized language assessments are more common [20], but this skews samples toward children with greater verbal abilities, potentially excluding the 25-30% with limited language [36–38]. Moreover, matching participants with ASD and those with typical development on nonverbal cognition can complicate interpretation. Findings across studies are mixed, depending on the measures and age groups studied: while some report preserved receptive and expressive abilities [39], others describe deficits in vocabulary, pronoun use or syntax [40,41]. Importantly, syntactic impairments appear to affect only a subgroup of children with ASD, some of whom exhibit profiles similar to developmental language disorder [42,43].

The present study sought to evaluate the suitability and reliability of standardized language assessments in children with ASD aged 4 to 6, a population that remains underrepresented in language assessment research [20]. To our knowledge, only two studies – by Kover and colleagues [18] and Thurm and colleagues [44] – have used standardized language tools in preschoolers with ASD. Both studies found greater impairments in receptive than expressive language, a pattern that diverges from typical development.

Kover and colleagues [18] specifically examined lexical development in 49 verbal children with ASD (ages 4-11), using the Peabody Picture Vocabulary Test, fourth edition (PPVT-4; [45]) for receptive and the Expressive Vocabulary Test, second edition (EVT-2; [46]) for expressive vocabulary. They found that both receptive vocabulary and nonverbal cognition were delayed relative to chronological age, and that receptive vocabulary had a slower developmental trajectory than expressive vocabulary – although this difference was reduced after controlling for nonverbal cognitive ability –, suggesting that nonverbal cognition plays a key role in receptive language development.

Thurm and colleagues [44] assessed receptive and expressive language in 59 children with ASD using hetero-assessments at age 3, including the Vineland Adaptive Behaviour Scales (VABS; [47]) and the Sequenced Inventory of Communication Development (SICD)-parent report [48]. At age 5, language assessment was conducted using standardized tests, including the Verbal comprehension and Naming vocabulary subtests of the Differential Abilities Scales (DAS; [49]), or alternatively, depending on the child's developmental level, a battery developmental, either the Infant MSEL [50], or the MSEL [51]. They found that early communication skills and nonverbal Intellectual Quotient (IQ) at age 3 predicted both expressive and receptive language outcomes at age 5, with ASD diagnosis predicting poorer receptive language, aligning with observed lower scores.

Importantly, our study also examined the concordance between standardized language tests, developmental batteries and hetero-assessments. To date, only Thurm and colleagues [44] combined these assessment approaches in a sample of preschool-aged children with ASD. They investigated the predictive value of Communication and Socialization subscale scores from the VABS at age 3 on verbal and language outcomes measured by the DAS and the MSEL at age 5. Results showed that early communication skills significantly predicted both expressive and receptive language development at age 5, whereas socialization skills predicted only receptive language development. However, while informative, Thurm and colleagues' study [44] did not examine the concordance of receptive and expressive language scores across different assessment methods. Nor did it explore the consistency of specific linguistic domains – such as receptive and expressive vocabulary, verbal comprehension, phonology and articulation – across standardized tests, developmental batteries and hetero-assessments.

Importantly, both studies converged in identifying language difficulties in ASD using standardized tests. However, these conclusions are drawn from only two studies – one of which included participants older than preschool age. Furthermore, comparison between the two studies is complicated by the use of different assessment tools.

Thus, our study addresses two primary research questions: (1) Can standardized tests normed on typically developing children to evaluate various language skills (receptive and expressive lexicon, receptive comprehension, phonology, articulation) reliably assess children ASD aged 4 to 6 with severe language impairments, a population with limited standardized language assessment research? This age group is particularly understudied, in part because 50-74% of preschoolers with ASD have minimal or no functional verbal language [52]. Nonetheless, despite the challenges in assessing language skills in this population, language difficulties are widely regarded as a core and reliable marker of ASD [44]. (2) To what extent are results consistent across standardized language tests, developmental batteries and hetero-assessments? In light of its limited sample size and exploratory design, this investigation should be regarded as a prospective, hypothesis-generating study, intended to inform and refine future research directions and clinical assessment practices.

Materials and Methods

Participants

The study involved 47 French preschool-aged children with ASD (36 boys, 11 girls, $M = 5$ years 9 months \pm 8 months, age range = 4 years 0 months to 6 years 9 months). They were recruited locally through various medical and school institutions located in the Provence-Alpes-Côte d'Azur and Auvergne-Rhône-Alpes regions¹. Our study did not include a control group because we used standardized tests with established norms that provide comparative metrics. Moreover, our research questions focused on evaluating the appropriateness and reliability of standardized language assessments, as well as the concordance between language assessment tools within the ASD population, rather than on comparisons between groups.

Approval for the study was granted by the South Mediterranean I Personal Protection Committee under the reference 2020-A02980-39 on 02 December 2020. Before taking part in the study, at least one parent of each participant reviewed the written information sheet together with the examiner and signed a statement of non-opposition.

Children were assessed in their regular care environment and were included if they had received a diagnosis of ASD from the referring physician based on the criteria outlined in the DSM-5 [1].

In addition to the diagnosis, children were required to have undergone a comprehensive assessment that included a hetero-assessment using the VABS, second edition (VABS-II; [53]) and a direct developmental assessment – such as the Psychoeducational Profile, third edition (PEP-3; [54]) – or a cognitive evaluation adapted to each child's profile.

Children under the age of 4 were not included in the study, since an official ASD diagnosis is rarely made before this age and because the inclusion criteria required both a formal diagnosis and a comprehensive evaluation.

¹ Saint-Louis Early Medico-Social Action Center, Saint-Jérôme Day Hospital and Pythéas Medico-Psychological Center; Edouard Toulouse Hospital Center (Marseille); Salvator Polyvalent Early Medico-Social Action Center: Salvator Hospital (Marseille); North Early Medico-Social Action Center: North Hospital (Marseille); La Bricarde Autism Nursery School Unit, Early Medico-Social Action Center and Jean Maurel Autism Nursery School Unit: Regional Association for the Integration of people with disabilities or in difficulty (Marseille, Orange and Puyricard); Florian Departemental Medico-Psycho-Pedagogical Center: General Council of Bouches-du-Rhône (Marseille), Les Tamaris Autism Nursery School Unit: Association of Parents of Disabled Children (Bollène), Aiguebelle Autism Nursery School Unit: Association of Pupils of Public Education Sud-Rhône-Alpes (Donzère), Early Medico-Social Action Centers: Union for the Management of Health Insurance Fund Establishments (Hyères and Toulon), Louis Pécout Autism Nursery School Unit: Regional Association for the Education and Placement of Handicapped Young People (La Ciotat) and Early Medico-Social Action Center: George Sand Hospital Center (La Seyne-sur-Mer).

The VABS-II is a semi-structured interview conducted with a parent or caregiver, and it evaluates adaptive functioning in four areas such as communication, daily living skills, socialization and motor skills. The PEP-3 evaluates developmental level and delays in children with developmental disorders, aged from 2 to 7 and a half years. It covers both developmental and maladaptive behavioral areas, including verbal and nonverbal cognition, language, motor skills, oculomotor imitation, affective expression, social reciprocity, and characteristic motor and verbal behaviors.

Descriptive characteristics of the sample are reported in Supplementary File 1 (Table S1).

Four children in the sample had been diagnosed with asthma and were undergoing treatment. One child had been diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). Specifically, 39 children were receiving speech therapy. 35 children had individualized human support at school and 11 children were enrolled in a specialized Autism Nursery School Unit.

Measures

Hetero-Assessments with the Parent

Parents completed the VABS-II and the French inventories of communicative development (or Inventaires Français du Développement Communicatif, IFDC; [55]) The IFDC, adapted from the MacArthur Communicative Development Inventories [56], were completed retrospectively to assess early communicative development, specifically communicative gestures and vocabulary skills at 12, 18 and 24 months. Part 1 (for 12 months) evaluates babbling, the production of 25 gestures and the comprehension or production of 81 words. Part 2 (for 18 months) focuses on the comprehension or production of 97 words and on word combinations. Part 3 (for 24 months) assesses the production of 100 words and the average sentence length, but this last measure was excluded because no child exceeded 2.66 words. Raw scores were used for analysis instead of percentile ranks.

Developmental Battery

Children were assessed with the PEP-3, which provided a direct measure of developmental functioning.

Standardized Tests

All standardized tests used in this study align with French clinical recommendations [16,57,58].

Receptive Vocabulary

Receptive lexicon was assessed using the PPVT (or Echelle de Vocabulaire en Images Peabody, EVIP; [59]) and the Designation from a cue subtest from the oral language assessment 2-6 years old (or Evaluation du Langage Oral 2-6 ans, EVALO 2-6; [60]).

The EVIP assesses lexical breadth in children aged 2 years and 6 months to 18 years. The child points to one of four images that matches a spoken word, with the three remaining images serving as distractors: a phonological distractor (sharing sounds with the target word), a semantic distractor (related in meaning to the target word) and an unrelated distractor (having no relationship to the target word). Correct answers score 1; scoring stops after 6 errors in 8 consecutive items. The raw score corresponds to the total number of correct answers, with a possible maximum score of 170. In our sample, the highest item reached was 65, item 65 being the last to obtain at least one correct response. This raw score is then converted into a standardized score based on the child's age, with a mean of 100 and a Standard Deviation (SD) of 15. 46 children completed this test.

We also administered the Designation from a cue subtest from the EVALO 2-6 language assessment battery. This battery includes two versions: a «young child» version (for ages 2 years 3 months to 4 years 3 months) and an «old child» version (for ages 4 years 3 months to 6 years 3 months). In our sample, 5 children were assessed using the «young child» version, 37 children using

the «old child» version and 5 older children were not assessed with this tool. The Designation from a cue subtest requires children to identify and point to images based on a semantic or taxonomic category provided by the experimenter (e.g., «Show me the animal») and offers an assessment of lexical breadth as well as insight into children's lexical organization. Correct answers score 1; no formal stopping rule but, to prevent the child from experiencing repeated failure, testing was stopped after 5 consecutive errors. The raw score (maximum 22) was converted to z-scores ($M = 0$, $SD = 1$).

Expressive Vocabulary

Expressive lexicon was assessed through the Denomination subtest shortlist from the EVALO 2-6 battery, which requires the child to name pictures. This subtest provides, for each word presented to the children, a procedure that simultaneously evaluates both word knowledge (Denomination-Lexicon scoring) and phonological representation quality (Denomination-Phonology scoring) through two distinct scoring systems. The lexical scoring system (Denomination-Lexicon) provides immediate and precise evaluation of word retrieval accuracy and speed. For instance, if the target word is «chemise» ([shirt]), the child might produce the correct phonological form or a phonologically incorrect variant, such as «temise» (nonword). If the child's response is incorrect – whether it is a semantically related word like «veste» ([jacket]), an unrelated word like «table» or no response at all – the examiner provides a phonemic cue by giving the initial phoneme of the target word and prompts the child to try again. This procedure helps distinguish between difficulties in retrieving words from memory and a true lack of word knowledge. Responses are scored as 2 for a correct initial answer, 1 for a correct answer after oral assistance and 0 for an incorrect response. Raw scores reflect the total number of correct answers, with a maximum of 80 for initially correct responses and 40 for correct responses following assistance.

The phonological scoring system (Denomination-Phonology) assesses the phonological accuracy of the produced word. This procedure distinguishes between naming errors resulting from deficits in word knowledge and those stemming from phonological articulation difficulties. When the child makes an error during picture naming – either producing a word with incorrect phonology, failing to produce the expected target or producing no word –, the examiner offers the first phoneme as a cue and asks the child to name the image; if phonological distortions occur, the examiner models the correct word and prompts the child to repeat it. An initially correct response or one given with oral assistance scores 2, a correct response after repetition scores 1 and an incorrect response scores 0. Raw scores correspond to the total correct answers, with maximum possible scores of 80 (initial or assisted response) and 40 (response after repetition). Although there is no formal stopping rule, the test was stopped after 5 consecutive errors to avoid repeated failure. Raw scores were converted into age-based z-scores ($M = 0$, $SD = 1$).

Oral Comprehension

Oral language comprehension was evaluated using the Understanding of topological terms subtest from the EVALO 2-6 and the syntaxico-semantic comprehension test (or Epreuve de COMpréhension Syntaxico-SEmantique, E.CO.S.SE; [61]).

In the Understanding of topological terms subtest, the child is asked to place a dog figurine relative to a bench based on a spatial instruction from the examiner, such as «put the dog on the bench». Each correct response is scored 1. Although there is no formal stopping rule, the test was ended after 5 consecutive errors to avoid repeated failure. The maximum raw score varies by version: 6 for the «young child» and 9 for the «old child» version. Raw scores are then converted into age-based z-scores with a mean of 0 and a SD of 1.

The E.CO.S.SE evaluates oral sentence comprehension in children aged 4 to 12 years. The child is presented with an oral word or sentence (e.g., «le garçon court» [the boy runs] or «la balle est rouge» [the ball is red]) and must select the matching image from four options. In our sample, 46 children were assessed using this tool. Each error scores 1. Although the author suggests stopping the test after 6-7 failed blocks (24-28 items), we ended it after 5 consecutive errors to avoid repeated

failure. The raw score is the number of errors (maximum 92) and is converted into a z-score adjusted for age, gender and parental socioeconomic status ($M = 0$, $SD = 1$). In our study, we used the number of correct responses as the raw score.

Articulation Skills

Articulation skills were assessed with the Orofacial and lingual praxis subtest from the EVALO 2-6. The child had to imitate bucco-facial movements (e.g., tongue out, lips pursed, cheeks puffed, blowing) and sound-based actions (e.g., phonemes, prolonged kissing sound, tongue click) demonstrated by the examiner. Each correct response scored 1, incorrect responses scored 0. Although there is no formal stopping rule, the test was ended after 5 consecutive errors to avoid repeated failure. The raw score (maximum 18) was converted into a z-score adjusted for age ($M = 0$, $SD = 1$).

Procedure

An initial two-hour session was held with the parent to explain the study, review the written information sheet, obtain signed non-opposition consent and administer the IFDC. It was followed by an average of three weekly sessions with the child, each lasting approximately 1.5 hours, during which the comprehensive language assessment protocol was administered. The number and duration of sessions were adapted to the child's availability. All sessions were conducted either by the principal investigator, a psychologist specialized in neuropsychology, or by trained graduate students in neuropsychology or neuroscience under the investigator's supervision. Each parent-child dyad was typically assessed by the same examiner throughout the study.

Overview of Statistical Analysis

The feasibility of using standardized language tests – normed on typically developing populations – with children aged 4 to 6 years with ASD was first assessed with an item-level analysis using the dichotomous Rasch model [62,63]. This model, based on binary scoring (1 for correct, 0 for incorrect), estimates item difficulty to characterize the relationship between individual ability and the probability of a correct response. According to the Rasch model, the likelihood of a correct answer is determined by the difference between item difficulty and individual skill. Thus, individuals with higher language abilities are more likely to succeed across all items, while easier items are generally answered correctly by most participants. The Wright map provides a visual representation of participant ability and item difficulty along a common scale: participants are ranked by ability on the left and items are ordered by difficulty on the right. Higher positions on the continuum reflect stronger language skills (for participants) or greater difficulty (for items), whereas lower positions indicate lower ability or easier items. The symmetry of the item distribution offers insight into the overall difficulty of the test. This analytical approach helps evaluate the appropriateness of test items (i.e., whether they are neither too easy nor too difficult), and allows detection of potential ceiling or floor effects as well as the overall psychometric quality of the tests. Goodness-of-fit statistics – with acceptable mean square values between 0.7 and 1.3 for our sample [64] – were used to evaluate how well each item fits the underlying construct. Higher fit statistics suggest increased response variability [65]. Person-separation reliability (with values approaching 1 indicating better discrimination) reflects the test's capacity to distinguish among participants with different levels of ability and supports the construct validity of the assessments [65–67].

Then, to examine how the different language components assessed in this study relate to one another and to gain a more integrated understanding of children's overall language competence, a series of Pearson correlation analyses was conducted on Rasch scores across the seven language variables. Particular attention was given to the relationship between receptive and expressive lexical abilities, in light of previous findings indicating greater challenges in receptive language among

children. To further investigate this discrepancy, paired Student's t-tests were used to compare performance in receptive versus expressive lexical domains based on Rasch scores.

Finally, to assess the concordance between standardized tests, developmental battery and hetero-assessment measures, we first conducted Pearson correlation analyses. Subsequently, stepwise linear regression analyses were conducted to assess the level of agreement between the assessment methods by identifying the combination of predictors that best accounted for variance in each language component, thereby quantifying the degree of overlap among the tools.

Results

All data analyses were conducted using the statistical softwares Jamovi, version 2.3.28, JASP, version 0.19.1.0 and R, version 4.4.2.

Q1: Can standardized tests normed on typically developing children reliably assess the language skills of children aged 4 to 6 with ASD?

Rasch Analysis

To enhance clarity and conciseness in the presentation of results, we reported Rasch analyses only for the EVIP in the main text. Complete Rasch analyses for the remaining language assessments are available in Supplementary File 2 (Figures S1 to S8). The person reliability value was high (.910, $p < .001$), indicating that the test reliably measured the intended construct. The majority of item infits fell within the acceptable range [0.7; 1.3], with 9 items < 0.7 and 5 items > 1.3 , suggesting adequate model fit. As shown in Figure 1, there was limited overlap between participant ability and item difficulty distributions: 12 out of 46 participants scored 0 on the test and 12 out of 65 items were not answered correctly by any participant (and were thus excluded from the Rasch analysis).

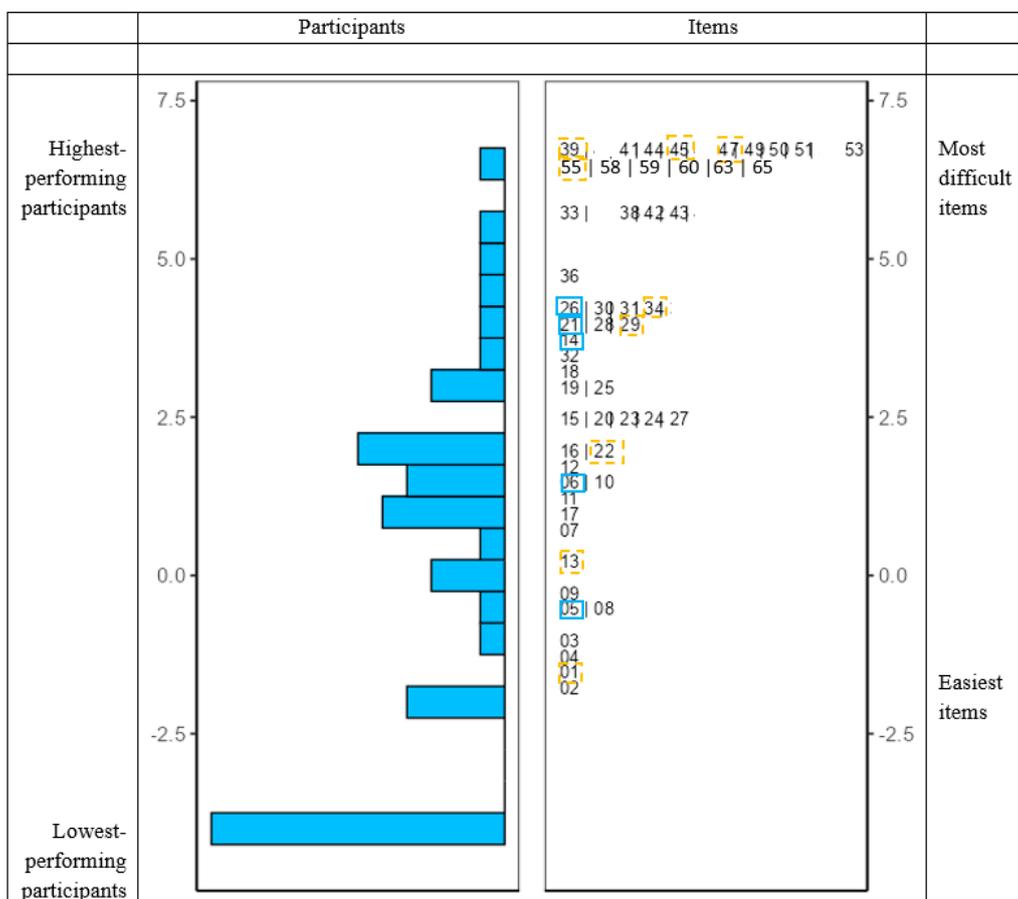


Figure 1. Wright map of the EVIP. Orange (dashed line): item infits < 0.7 ; blue (solid line): item infits > 1.3 .

Excluding the 12 participants who scored 0, the distribution function of receptive vocabulary was approximately normal, with a slight positive skew, suggesting that most children were able to respond variably to the items. The Rasch analysis indicated that the EVIP was appropriately calibrated for 74% of participants, who demonstrated differentiated responses. However, for the remaining 26%, the test proved to be too challenging (lowest negative score on the Rasch scale).

Moreover, the analysis confirmed that items clustered at similar levels on the Wright map were of comparable difficulty within our sample, suggesting that the test could be shortened by retaining a few representative items per level without significant loss of psychometric information. Finally, the item difficulty, as determined by chronological age norms, did not always correspond to observed performance in our atypical sample – highlighting that developmental trajectories in children with ASD may diverge from typical patterns.

A summary of key Rasch outcomes across all tests is provided in Table 1.

Table 1. Summarize of key findings from Rasch analyses.

		Person reliability	Item infits < 0.7	Item infits > 1.3	Items failed by all participants	Participants scored 0	Score distribution	
Lexicon	Receptive lexicon	EVIP	.910***	9	5	12	12 (26%)	Approximately normal with a slight positive skew
		Designation from a cue	.796***	2	1	0	15 (37%)	Uniform
	Expressive lexicon	Denomination-Lex 1	.700***	4	3	3	22 (52%)	Globally uniform
Receptive comprehension		Understanding of topological terms	.619***	5	0	0	14 (34%)	Positive skew
		E.CO.S.SE	.915***	27	6	19	12 (26%)	Approximately normal with a slight positive skew
Phonology		Denomination-Phono 1	.733***	4	1	1	21 (50%)	Globally uniform
Articulation		Orofacial and lingual praxis	.816***	2	0	0	10 (24%)	Approximately normal with a slight negative skew

*** $p \leq .001$; ** $p \leq .01$; * $p \leq .05$.

Q2: To what extent are the results of standardized language tests consistent with one another, as well as with scores from developmental batteries and hetero-assessments?

Pearson Correlation Analyses

Inter-Correlation Matrix of Rasch Scores on Language Tests

Pearson correlation analyses were conducted using Rasch scores rather than standardized scores because they provide objective estimates that are independent of item difficulty (for individuals) and individual ability (for items). This makes them a more robust metric for examining relationships among constructs. The full inter-correlation matrix is presented in Supplementary File 3 (Table S1).

Rasch scores from language assessments were significantly and positively correlated with one another, with coefficients ranging from $r = .450^{**}$ to $r = .911^{***}$. One exception was the Orofacial and Lingual Praxis subtest, which showed only moderate correlations with the Denomination-Lex 1 and Understanding of topological terms subtests (respectively, $r = .450^{**}$ and $r = .477^{**}$). These more moderate correlations suggest that orofacial praxis may reflect partially distinct abilities from those

assessed by language-based measures. Notably, the very strong correlation between Denomination-Lex 1 and Denomination-Phono 1 ($r = .911^{***}$) subtests (which represented alternative scoring methods applied to the same subtest) indicated that they likely captured overlapping underlying constructs. This redundancy suggests that retaining both may be unnecessary and that either score could be selected depending on the specific focus of the assessment.

Focus on Receptive and Expressive Lexicon

Paired-sample t-tests were conducted to compare receptive and expressive lexicon performance. As shown in Table 2, expressive lexicon performance – measured by Denomination-Lex 1 – was significantly lower than receptive lexicon performance, as assessed by both the EVIP ($t(40) = -3.11$, $p < .01$) and the Designation from a cue subtest ($t(40) = -2.21$, $p < .05$). Children demonstrated poorer performance on the expressive lexicon test. Thus, the expressive lexicon test can be considered more challenging than the receptive one.

Table 2. Paired-sample t-tests comparing receptive and expressive lexicon assessments using Rasch scores.

Lexicon	Test	Mean	SD	$t(df = 40)$	P value	Mean differences	
						Mean	SD
Receptive lexicon	EVIP	-.006	2.93	-3.11**	.003	-1.018	.327
	Designation from a cue	.455	2.27	-2.21*	.033	-.557	.252
Expressive lexicon	Denomination-Lex 1	1.012	2.12				

Df (degrees of freedom). *** $p \leq .001$; ** $p \leq .01$; * $p \leq .05$. A negative t-value indicates lower scores on the expressive test relative to the receptive one.

Inter-Correlation Matrix Between Hetero-Assessment and Developmental Battery Scores

For the purpose of Pearson correlation analyses between Rasch scores on language tests and scores from hetero-assessments and developmental batteries, only the subdomains specifically targeting language skills were retained. These included the v-scale scores (standardized scores with a mean of 15 and a SD of 3) for the Receptive and Expressive language subdomains of the VABS-II. From the IFDC, we selected raw scores corresponding to the number of words understood and/or produced at 12 and 18 months, and the number of words produced at 24 months. From the PEP-3, we used the developmental age scores for the Receptive and Expressive language subdomains.

The complete inter-correlation matrix is presented in Supplementary File 3 (Table S2).

Results showed that most scores derived from the IFDC and the VABS-II, both hetero-assessment tools, were not significantly correlated. An exception was observed for the Receptive language subdomain of the VABS-II, which showed a significant moderate positive correlation with the number of words understood on the IFDC at 18 months ($r = .397^{**}$). In contrast, the Receptive and Expressive language subdomains of the PEP-3 were positively correlated with nearly all hetero-assessment measures (r values ranging from $r = .357^*$ to $r = .684^{***}$), with the exception of: (a) the number of words understood at 12 months on the IFDC, and (b) the correlation between the Receptive language subdomain of the PEP-3 and the number of words understood at 18 months on the IFDC.

Inter-Correlation Matrix Between Rasch Scores on Language Tests and Hetero-Assessment Scores

The inter-correlation matrix between Rasch scores on language tests and hetero-assessment scores is presented in Supplementary File 3 (Table S3).

No significant correlations were observed between any of the IFDC scales and the standardized language tests. In contrast, all language tests – except for the Understanding of topological terms subtest – were significantly and moderately positively correlated with the Receptive language subdomain of the VABS-II (r values ranging from $r = .344^*$ to $r = .464^{**}$). Notably, the correlation

between the Orofacial and Lingual Praxis subtest and the Receptive language subdomain of the VABS-II was strong ($r = .542^{***}$). Furthermore, all language tests demonstrated significant and strong positive correlations with the Expressive language subdomain of the VABS-II (r values ranging from $r = .541^{***}$ to $r = .753^{***}$).

Inter-Correlation Matrix Between Rasch Scores on Language Tests and Developmental Battery Scores

The inter-correlation matrix between Rasch scores on language tests and the PEP-3 developmental battery scores is presented in Supplementary File 3 (Table S4).

All language tests showed significant positive correlations with the Receptive language subdomain of the PEP-3 (r values ranging from $r = .504^{**}$ to $r = .751^{***}$), with the exception of the Designation from a cue and Understanding of topological terms subtests, which showed moderate correlations (respectively, $r = .414^*$ and $r = .457^*$). In terms of expressive language, the EVIP, the Denomination-Lex 1, as well as the E.CO.S.SE were strongly and significantly correlated with the Expressive language subdomain of the PEP-3 (r values ranging from $r = .561^{***}$ to $r = .773^{***}$). However, the Designation from a cue and Orofacial and lingual praxis subtests showed only moderate correlations with the Expressive language subdomain of the PEP-3 (respectively, $r = .386^*$ and $r = .466^*$). No significant correlation was observed between the Understanding of topological terms subtest and the Expressive language subdomain of the PEP-3.

Stepwise Linear Regression Analyses

Stepwise linear regression analyses were conducted to identify concurrent predictors of children's language skills across lexicon, receptive comprehension, phonology and articulation components. This approach was used to assess the concordance between standardized language tests, developmental battery and hetero-assessments. The degrees of freedom varied across analyses due to differences in sample size stemming from age restrictions or missing data.

From two complementary perspectives, this analysis addresses the following research questions: (a) To what extent do scores from hetero-assessments and developmental battery predict performance on standardized language tests? In other words, can hetero-assessments alone suffice or is direct assessment still necessary? (b) Conversely, do Rasch scores from standardized language tests predict results obtained from hetero-assessments and developmental battery? That is, are standardized tests sufficiently reliable to predict subjective and informant-based measures? To explore these questions, we performed two sets of stepwise regressions. First, we treated standardized test Rasch scores as dependent variables with hetero-assessment and developmental battery scores as predictors (addressing question a). Next, the roles were reversed to address question (b).

The results presented below focus on question (a). Here, Rasch scores from standardized language assessments were included as dependent variables, while hetero-assessment and developmental battery scores served as predictors. This analysis tests whether data from standardized tests can be effectively approximated or replaced by information from hetero-assessments and developmental battery, thereby evaluating the necessity of direct assessment.

Stepwise Linear Regression Predicting Rasch Scores on Language Tests from Hetero-Assessment and Developmental Battery Scores

The resulting regression models are summarized in Table 3. Detailed statistical outcomes are available in Supplementary File 4.

As shown in Table 3, regression analyses revealed that a single-step model significantly predicted performance on lexical breadth (EVIP), with PEP-3 Receptive language accounting for 43% of the variance. A one-step model also significantly predicted performance on word retrieval accuracy and speed (Denomination-Lex 1 scoring), with PEP-3 Expressive language accounting for 52% of the variance. Lexical breadth and organization (Designation from a cue subtest) were

significantly predicted by a one-step model with PEP-3 Expressive language accounting for 19% of the variance. Oral comprehension of topological terms (Understanding of topological terms subtest) was significantly predicted by a one-step model with PEP-3 Receptive language accounting for 20% of the variance. Oral sentence comprehension (E.CO.S.SE) was significantly predicted by a two-step model including PEP-3 Expressive language and VABS-II Receptive language, which together explained 65% of the variance. Phonological accuracy of the produced word (Denomination-Phono 1 scoring) was significantly predicted by a two-step model including PEP-3 Receptive language and VABS-II Expressive language, jointly accounting for 62% of the variance. Finally, articulation skills (Orofacial and lingual praxis subtest) were significantly predicted by a one-step model with VABS-II Receptive language accounting for 31% of the variance.

Table 3. Summary of linear regression models predicting Rasch scores on language tests from hetero-assessment and developmental battery scores (R^2 values).

Dependent variables	Lexicon						
	Receptive lexicon		Expressive lexicon		Receptive comprehension	Phonology	Articulation
	EVIP	Designation from a cue	Denomination-Lex 1	Understanding of topological terms	E.CO.S.SE	Denomination-Phono 1	Orofacial and lingual praxis
IFDC-12 months	Comprehension						
	Production						
IFDC-18 months	Comprehension						
	Production						
IFDC-24 months	Production						
VABS-II	Receptive language				22%** (simple effect)		31%**
	Expressive language					55%*** (simple effect)	
PEP-3	Receptive language	43%***			20%*		55%*** (simple effect)
	Expressive language		19%*	52%***		60%*** (simple effect)	
Total effect		43%***	19%*	52%***	20%*	65%*** (simultaneous effect)	62%*** (simultaneous effect)

*** $p \leq .001$; ** $p \leq .01$; * $p \leq .05$. R^2 values represent the variance explained by the predictors; simple effects refer to individual predictors, while simultaneous effects refer to combined predictors in the model. Gray cells indicate the absence of regression model.

The following analysis addresses question (b). In this model, scores from the hetero-assessments and developmental batteries were entered as dependent variables, while Rasch scores from the standardized language tests served as predictor variables.

Stepwise Linear Regression to Explain Receptive Language Scores from Hetero-Assessments and Developmental Battery Using Rasch Scores from Receptive Language Standardized Tests

The linear regression models are summarized in Table 4, with detailed results available in Supplementary File 4.

As shown in Table 4, a one-step regression model significantly predicted VABS-II Receptive language, with a single predictor – E.CO.S.SE (oral sentence comprehension) – accounting for 19% of the variance. Similarly, a one-step regression model significantly predicted PEP-3 Receptive language, with E.CO.S.SE again emerging as the sole predictor, explaining 53% of the variance.

Table 4. Summary of linear regression models predicting receptive language scores from hetero-assessment and developmental battery scores using Rasch scores on receptive language tests (R² values).

Predictor variables		Dependent variables	IFDC-12 months	IFDC-18 months	VABS-II	PEP-3
			Comprehension	Comprehension	Receptive language	Receptive language
Lexicon	Receptive lexicon	EVIP				
		Designation from a cue				
	Receptive comprehension	Understanding of topological terms				
		E.CO.S.SE			19%** (simple effect)	53%*** (simple effect)
		Total effect			19%**	53%***

*** $p \leq .001$; ** $p \leq .01$; * $p \leq .05$. R² values represent the variance explained by the predictors; simple effects refer to individual predictors, while simultaneous effects refer to combined predictors in the model. Gray cells indicate the absence of a regression model.

Stepwise Linear Regression to Explain Expressive Language Scores from Hetero-Assessments and Developmental Battery Using Rasch Scores from Expressive Language Standardized Tests

The linear regression models are presented in Table 5, with detailed results available in Supplementary File 4.

As shown in Table 5, a two-step regression model significantly predicted VABS-II Expressive language, with two predictors – Denomination-Lex 1 (word retrieval accuracy and speed) and Orofacial and lingual praxis (articulation skills) subtests – accounting for 66% of the variance. Similarly, a two-step regression model significantly predicted PEP-3 Expressive language, with the same two predictors accounting for 60% of the variance in performance.

Table 5. Summary of linear regression models predicting expressive language scores from hetero-assessment and developmental battery scores using Rasch scores on expressive language tests (R² values).

Predictor variables		Dependent variables	IFDC-12 months	IFDC-18 months	IFDC-24 months	VABS-II	PEP-3
			Production	Production	Production	Expressive language	Expressive language
Lexicon	Expressive lexicon	Denomination-Lex 1				57%*** (simple effect)	52%*** (simple effect)
	Phonology	Denomination-Phono 1					
	Articulation	Orofacial and lingual praxis				39%*** (simple effect)	24%*** (simple effect)
		Total effect				66%*** (simultaneous effect)	60%*** (simultaneous effect)

*** $p \leq .001$; ** $p \leq .01$; * $p \leq .05$. R² values represent the variance explained by the predictors; simple effects refer to individual predictors, while simultaneous effects refer to combined predictors in the model. Gray cells indicate the absence of a regression model.

Discussion

The primary aim of this prospective study was to evaluate the feasibility of a comprehensive, multidimensional language assessment – encompassing receptive and expressive vocabulary, receptive comprehension, phonology and articulation – in children aged 4 to 6 years with ASD, a subgroup underrepresented in the literature [68]. Notably, our sample was intentionally inclusive of children with highly heterogeneous language profiles, the majority of whom were minimally or nonverbal. A secondary objective was to examine the degree of concordance between standardized language assessments and more commonly used tools in this population, namely developmental batteries and hetero-assessments.

Feasibility of Standardized Language Testing in Children with ASD Aged 4 to 6 Years

We employed standardized assessments aligned with French clinical guidelines for preschoolers [16,57,58], ensuring ecological validity and relevance to real-world clinical practice [2]. Rasch modeling confirmed the appropriateness of these tools for most participants, while highlighting varying levels of accessibility depending on the specific test. However, these tools – normed on typically developing children – do not always capture the developmental specificity of ASD profiles. Indeed, the difficulty levels inherent in these assessments, which are based on chronological age norms, did not consistently reflect the actual abilities of our sample. Age-equivalent scoring, as used in some prior studies [18], would likely have reduced variability due to floor effects [69], and such methods may compromise the psychometric robustness of the instruments [70,71].

We found moderate to strong correlations among the various components of language, suggesting a coherent pattern of functionally interdependent language skills. However, strong correlation between, for example, word retrieval accuracy and speed, and phonological accuracy of the produced word suggested potential redundancy, possibly reflecting overlapping language processes (e.g., the speed of phonological representation activation and the use of articulatory codes). Although our initial aim was to conduct a broad language assessment across multiple domains – extending beyond the typical expressive/receptive dichotomy [18,44] –, our findings supported a more focused approach. Test selection should be guided by empirical patterns of performance in this population, given their atypical developmental trajectories [18].

Interestingly, children in our sample performed more poorly on the expressive lexicon test than on the receptive lexicon test – contrary to findings in earlier studies, which typically report more marked deficits in receptive language [18,44] –. Several factors may explain this discrepancy.

Firstly, our study included children across a wide spectrum of verbal abilities and cognitive profiles. Based on parental reports, nearly half of our sample (45%) would be considered nonverbal according to the definition proposed by Rutter and colleagues [17]. In contrast, estimates in the literature range from 50-74% for preschool-aged children [52]. Furthermore, in our sample, 31 children scored 0 on at least one language test and 4 children scored 0 on all tests. This pattern may be indicative of a selection bias associated with recruitment from specialized clinical services, which commonly serve children with more pronounced expressive vocabulary impairments rather than receptive vocabulary deficits. It is important to note that the term «minimally verbal» to describe children who either fail to develop functional verbal language or exhibit only minimal verbal expression is inconsistently defined across studies [22]. Some use a vocabulary threshold [72], others apply the Autism Diagnostic Observation Schedule, second edition (ADOS-2; [73]) Module 1 criteria for children over 30 months who remain at the single-word stage [74] and still others use hetero-assessments like the MacArthur-Bates Communicative Development Inventories (CDI; [75]) to identify children whose expressive skills do not exceed those of an 18-month-old [76]. Consequently, prevalence estimates vary widely – from 13% using CDI criteria to 28% using ADOS-2 Module 1 [22] –. Few studies have specifically targeted this subgroup, underscoring the need for adapted methodologies and dedicated research [22].

Secondly, an alternative explanation could be attributed to the nature of the tests used to assess word knowledge. More generally, our receptive and expressive assessments can both be conceptualized as recognition memory tasks employing different cues [77]. In the EVIP (receptive test), participants can benefit from both the phonological cue (word name) provided by the examiner and the visual referent among four simultaneously presented pictures. Conversely, the Denomination-Lex 1 subtest (expressive test) presents participants with only a visual referent, requiring the explicit retrieval of both phonological and articulatory representations associated with the picture. While both assessment formats provide memory cues, their cognitive demands differ substantially. This methodological distinction may provide an explanation for the observed discrepancy in performance between receptive and expressive vocabulary measures, with children with ASD aged 4 to 6 demonstrating superior performance in receptive vocabulary compared to expressive vocabulary, as evidenced in our study. The observed discrepancy between our findings and those reported in other studies [18,44] is likely attributable to methodological variation.

Exploring the Concordance of Standardized Language Tests, Developmental Battery and Hetero-Assessments

Our findings indicate that while concordance between standardized language tests, developmental battery and hetero-assessments was occasionally observed, it was not systematic. Specifically, we found positive correlations between scores on the Receptive and Expressive language subdomains of the PEP-3 developmental battery and the Receptive and Expressive language subdomains from two hetero-assessments – the IFDC (retrospective) and VABS-II (current). However, correlations between retrospective (IFDC) and current (VABS-II) hetero-assessments were largely absent, raising concerns about the reliability and consistency of parent-reported retrospective data. Notably, both the VABS-II and PEP-3 subdomains for Receptive and Expressive language significantly correlated with several standardized test scores (covering receptive and expressive vocabulary, comprehension, phonology and articulation). These results indicate that hetero-assessments, developmental batteries and standardized tests for evaluating language performance were complementary but not interchangeable. Previous studies show that correlations between direct assessments and parent-reported measures tend to be lower for receptive language than for expressive language in children with ASD [78,79]. This finding highlights the complementary nature of these assessment approaches, emphasizing the importance of combining direct assessments and parental reports to obtain a more comprehensive profile of language abilities in children with ASD. However, in our study, the lack of correlation between retrospective IFDC scores, which depends on parental recall to characterize their child's early development, and standardized tests revealed the limitations of relying on parental recall, which is subject to memory biases, omissions and inaccuracies. Our results reinforced the complementary – but non-interchangeable – nature of hetero-assessments and direct evaluations, emphasizing the necessity of combining both methods to generate a nuanced understanding of language abilities in children with ASD.

Concerning the predictive value of developmental battery and hetero-assessments, linear regression models revealed several distinct patterns: (1) Receptive language assessed through the PEP-3 was the strongest predictor of lexical breadth (EVIP) and oral comprehension of topological terms (Understanding of topological terms subtest) in standardized tests; (2) Expressive language in the PEP-3 best predicted word retrieval accuracy and speed (Denomination-Lex 1 scoring), as well as lexical breadth and organization (Designation from a cue subtest); (3) Receptive language from the VABS-II emerged as the strongest predictor of articulatory skills (Orofacial and lingual praxis subtest), while oral sentence comprehension (E.CO.S.SE) was best predicted by a combination of PEP-3 Expressive language and VABS-II Receptive language; (4) For phonological accuracy of the produced word (Denomination-Phono 1 scoring), both Expressive language from VABS-II and Receptive language from PEP-3 were significant concurrent predictors. Importantly, none of the subdomains of the IFDC predicted performance on any standardized test, confirming its limited validity as a retrospective tool due to its reliance on parental recall to characterize child's early

development. These results support the conclusion that current parent-reported observations (e.g., VABS-II) and developmental batteries (e.g., PEP-3) were predictive of specific language abilities, but their contributions varied across domains. They did not consistently map onto receptive or expressive functions as measured by standardized tests, likely due to differences in test structure, cognitive demand and assessment context.

Our findings highlight that standardized direct assessments cannot be reduced to or replaced by global developmental evaluations or hetero-assessments. While developmental batteries like the PEP-3 offer a broader overview that may combine distinct language processes (e.g., word comprehension, sentence comprehension, instruction following), standardized tests allow for finer-grained, function-specific analysis. This distinction is critical when assessing children with ASD, who often present with atypical developmental trajectories and variable language profiles. As emphasized by Kover and colleagues [18], direct assessments provide real-time, objective observations under controlled conditions, reducing biases associated with parental reports, which may be influenced by environmental factors, subjective experiences or expectations. Likewise, Posar and Visconti [22] advocate for the systematic use of standardized tools to avoid interpretative ambiguities across studies and clinical contexts. While hetero-assessments and developmental batteries offer complementary perspectives, they must be interpreted alongside standardized measures to ensure diagnostic accuracy and appropriate intervention planning.

Additional analyses revealed that specific subtests were particularly predictive of performance on hetero-assessments and developmental battery: 1) Oral sentence comprehension (E.CO.S.SE) subtest was the best predictor of Receptive language in both the VABS-II and PEP-3; and 2) the combination of articulation skills (Orofacial and lingual praxis subtest) and word retrieval accuracy and speed (Denomination-Lex 1 scoring) best predicted Expressive language in these same tools. However, despite these predictive links, global language subdomains in the VABS-II and PEP-3 could not be fully explained by any one group of standardized tests. This finding reflects the broader issue of construct overlap in global assessments, which often conflate multiple linguistic and cognitive processes under a single score. In contrast, standardized tests target specific competencies, allowing for greater diagnostic precision.

Although preliminary, this concordance analysis suggests that relying on a single type of assessment – whether standardized, developmental or parental – risks overlooking important aspects of language function in preschoolers with ASD. A multimodal framework that integrates a standardized testing (for fine-grained, objective measures), developmental batteries (for broader functional insight) and current hetero-assessments (for real-world contextual observations) offers a more comprehensive picture of language capacity [80,81]. However, such an approach is time- and resource-intensive. Consequently, its relevance must be carefully evaluated in light of clinical objectives and research priorities. Future studies with larger samples are needed to refine this integrative model and determine how best to balance thoroughness with feasibility.

Limitations and Future Directions

Our findings should be interpreted in light of several limitations. First, despite the large number of measures collected, our sample size remained small and findings should be interpreted as preliminary. Nevertheless, this study was ambitious in its design, as it investigated a population that is rarely studied using specialized tools that are infrequently employed in research. To achieve adequate statistical power, a sample size of approximately 115 participants was estimated a priori using G*Power (version 3.1.9.7). A larger cohort would enable more sophisticated analyses such as network modeling or evaluation of hierarchical reticular models. Second, our study lacked a directly assessed comparison group, which limits interpretation of relative delays [18]. Incorporating a well-matched typically developing control group – matched for chronological age and nonverbal cognitive abilities [82] – would strengthen analyses. However, selecting appropriate control groups in ASD research is complex, as researchers must decide whether the control group should be matched for chronological age, developmental age in terms of receptive or expressive language abilities, consist

of infants at low risk for ASD evaluated longitudinally [27] or include children with other neurodevelopmental disabilities [3,44,80]. Moreover, heterogeneity in ASD samples combined with homogeneity in control groups can introduce generalization biases in the interpretation of results [4]. As a result, we, like others [4,69], focused exclusively on children with ASD.

Standardized tests, while offering objectivity and psychometric reliability, also present challenges. Poor test performance may reflect difficulties beyond language ability, such as attention deficits, social interaction impairments, imitation difficulties, irregular eye contact during interaction, a particular use of language, such as echolalia, or test engagement issues [83]. Moreover, standardized assessments often rely on conformity and structured demands to ensure a valid assessment [83] that may mask the true nature of language difficulties in children with ASD [18]. Alternative paradigms designed for younger children – more accessible and less demanding than traditional tests –, such as gaze-based responses (e.g., [84]), offer promising avenues but require specialized resources and lack standardized norms.

Future research should expand sample sizes and develop streamlined, adapted protocols suitable for clinical use with preschoolers with ASD. This need is emphasized in the literature, which calls for tailored assessments to guide effective interventions [85].

Conclusions

This study demonstrated that most children with ASD aged 4 to 6 could successfully complete standardized language assessments despite their notably heterogeneous language profiles, including many minimally verbal participants. Importantly, children performed better on receptive than expressive lexicon tests, suggesting that comprehension-based assessments might better capture underlying word knowledge without confounding production difficulties. This finding challenged previous reports of more pronounced receptive language deficits and highlighted the importance of considering assessment modality when interpreting language skills in ASD. Reducing verbal demands appear to facilitate identification of underlying lexical competencies, suggesting that adapted assessment approaches could help distinguish performance limitations from genuine comprehension deficits.

Our findings also suggest that standardized tests provide more precise and detailed language profiles than developmental batteries or hetero-assessments alone. The partial but non-equivalent concordance between assessment methods underscored their complementary rather than interchangeable nature. While developmental batteries and hetero-assessments provided valuable contextual information, they could not substitute for the precision and objectivity of direct standardized testing. Given the multifaceted nature of language development in ASD [18], an integrated, multimodal assessment approach appear essential for accurately identifying both language strengths and therapeutic targets in this population.

Overall, our results support the feasibility and clinical relevance of adapting standardized language assessments for preschool-aged children with ASD. However, given the test demands intrinsic to test modalities, adapted tools are essential for distinguishing performance limitations from genuine comprehension deficits. Moreover, future research should explore alternative paradigms designed specifically for younger children – approaches that are more accessible and less demanding than traditional standardized tests –. While such paradigms offered promising avenues for assessment, they required specialized resources and lacked standardized norms.

Finally, given the critical importance of early language development for later social functioning and the unique developmental trajectories in ASD, bridging clinical practice and research through comprehensive, evidence-based language assessment protocols appear crucial for optimizing intervention outcomes during this pivotal developmental period [85].

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Supplementary file 1: Table S1: Descriptive characteristics of the participants; Supplementary file 2: Figure S1: Wright map of the Denomination-Lex 1 scoring; Figure S2: Wright map of the

Denomination-Lex 2 scoring; Figure S3: Wright map of the Designation from a cue subtest; Figure S4: Wright map of the Understanding of topological terms subtest; Figure S5: Wright map of the E.CO.S.SE; Figure S6: Wright map of the Denomination-Phono 1 scoring; Figure S7: Wright map of the Denomination-Phono 2 scoring. Figure S8: Wright map of the subtest Orofacial and lingual praxis.; Supplementary file 3: Table S1: Inter-correlation matrix of Rasch scores on language tests; Table S2: Inter-correlation matrix between scores on hetero-assessments and scores on developmental battery; Table S3: Inter-correlation matrix between Rasch scores on language tests and scores on hetero-assessments; Table S4: Inter-correlation matrix between Rasch scores on language tests and scores on the developmental battery; Supplementary file 4: Supplementary stepwise regression analysis.

Authors' contributions: J.M. contributed to the study design, led data acquisition, performed data analysis, directed statistical analysis and interpretation, and drafted the manuscript. S.F. contributed to the literature review and data acquisition, and edited the manuscript. A.E.A. supervised the conceptualization of the analytical plan and statistical analysis, contributed to data interpretation, and edited the manuscript. C.A. contributed to the study design, data interpretation, and manuscript drafting. P.C. contributed to the study design, supervised data acquisition and analysis, and contributed to the interpretation and drafting of the manuscript. All authors were involved in the preparation and editing of the manuscript and approved the final submission.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the South Mediterranean I (protocol code 2020-A02980-39 and date of approval 02 December 2020).

Informed Consent Statement: All participants' parents provided informed consent prior to study enrollment. Given that assessments were conducted as part of routine clinical practice at our institution, at least one parent per participant reviewed the written information sheet with the examiner and signed a statement of non-opposition to research participation.

Data Availability Statement: All data generated or analyzed during this study are included in this published article and its supplementary information files.

Acknowledgments: The authors would like to express their sincere gratitude to the children and their parents who participated in this study. They are also deeply thankful to all the physicians and psychologist (Drs Pascal Andreo, Céline Bartoli, Monique Brillaux-Plassard, Yves Bernaud, Charlotte Bertier, Anne Dispa, Chloé Di Meglio, Catherine Grange, Céline Henry, Sigrid Martin, Anne-Christine Orsini, Christel Persehay, Anne Riaux and Meriem Zahed, and Mrs. Virginie Maurice), and their teams who referred patients for inclusion in this research. Finally, the authors wish to acknowledge the invaluable contributions of the master's students in neuropsychology and neuroscience who assisted with data collection: Marguie Deuil, Gabriela Dias Batista Monticelli, Clélia Esquer, Océane Lemaire, Aurélie Léonian and Thomas Leporatti.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5 (Fifth Edition). American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596>

2. Bolduc M, Poirier N. La démarche et les outils d'évaluation clinique du trouble du spectre de l'autisme à l'ère du DSM-5. *Rev Psychoéduc.* 2017;46(1):73–97. <https://doi.org/10.7202/1039682ar>
3. Torrens V, Ruiz C. Language and Communication in Preschool Children with Autism and Other Developmental Disorders. *Children.* 2021;8(3):192. <https://doi.org/10.3390/children8030192>
4. Volden J, Smith IM, Szatmari P, Bryson S, Fombonne E, Mirenda P, Roberts W, Vaillancourt T, Waddell C, Zwaigenbaum L, Georgiades S, Duku E, Thompson A. Using the Preschool Language Scale, Fourth Edition to Characterize Language in Preschoolers With Autism Spectrum Disorders. *Am J Speech Lang Pathol.* 2011;20:200–208. [https://doi.org/10.1044/1058-0360\(2011/10-0035\)](https://doi.org/10.1044/1058-0360(2011/10-0035))
5. Rey V, Tardif C, Delahaie M, Thomas K, Massion J. Etude exploratoire des capacités phonologiques chez les enfants présentant un déficit de langage. *Trav Interdisc Lab Parole Lang.* 2001;20:149–168. <https://hal.science/hal-00313776v1>
6. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord.* 2001;31(1):5–17. <https://dx-doi-org/10.1023/a:1005653411471>
7. Houzel D. Les signes précoces de l'autisme et leur signification psychopathologique. In Golse B, Delion P, editors. *Autisme : état des lieux et horizons.* Paris : Erès; 2005. p. 163–174. <https://doi.org/10.3917/eres.delio.2008.02.0163>
8. Howlin P, Goode S, Hutton J, Rutter M. Adult outcome for children with autism. *J Child Psychol Psychiatry.* 2004;45(2):212–229. <https://doi.org/10.1111/j.1469-7610.2004.00215.x>
9. Baghdadli A. Troubles autistiques : du repérage des signes d'alerte à la prise en charge. *Contraste.* 2006;25:23–51. <https://doi.org/10.3917/cont.025.0023>
10. DeGiacomo A, Fombonne E. Parental recognition of developmental abnormalities in autism. *Eur Child Adolesc Psychiatry.* 1998;7(3):131–136. <https://doi.org/10.1007/s007870050058>
11. Szatmari P, Archer L, Fisman S, Streiner DL, Wilson F. Asperger's Syndrome and Autism: Differences in Behavior, Cognition, and Adaptive Functioning. *J Am Acad Child Adolesc Psychiatry.* 1995;34(12):1662–1671. <https://doi.org/10.1097/00004583-199512000-00017>
12. Howlin P. Outcomes in autism spectrum disorders. In: Volkmar FR, Paul R, Klin A, Cohen D, editors. *Handbook of Autism and Pervasive Developmental Disorders.* Vol. 1. Hoboken, NJ: John Wiley and Sons; 2005. p. 201–220. <https://doi.org/10.1002/9780470939345>
13. Preschool - Ministry of National Education, Higher Education, and Research. Available online: <https://www.education.gouv.fr/1-ecole-maternelle-11534> (accessed on 02 February 2025).
14. Atzori P, Beggiato A, Colineaux C, Humeau E, Vantalou, V. Dépistage précoce, évaluation diagnostique et prises en charge éducatives précoces de l'autisme. *J Pédiatr Puéric.* 2022;35(2):61–72. <https://doi.org/10.1016/j.jpp.2021.12.008>
15. Filipek PA, Accardo PJ, Baranek GT, Cook EH, Dawson G, Gordon B, Gravel JS, Johnson CP, Kallen RJ, Levy SE, Minschew NJ, Ozonoff S, Prizant BM, Rapin I, Rogers SJ, Stone WL, Teplin S, Tuchman RF, Volkmar FR. The Screening and Diagnosis of Autistic Spectrum Disorders. *J Autism Dev Disord.* 1999;29(6):439–484. <https://doi.org/10.1023/a:1021943802493>
16. Haute Autorité de Santé (2018). Trouble du spectre de l'autisme – Signes d'alerte, repérage, diagnostic et évaluation chez l'enfant et l'adolescent. Recommandation de bonne pratique. Available online: https://www.has-sante.fr/upload/docs/application/pdf/2018-02/trouble_du_spectre_de_lautisme_de_lenfant_et_ladolescent__recommandations.pdf
17. Rutter M, LeCouteur A, Lord C. (2003). *Autism Diagnostic Interview-Revised.* Los Angeles: Western Psychological Services.
18. Kover ST, McDuffie AS, Hagerman RJ, Abbeduto L. Receptive Vocabulary in Boys with Autism Spectrum Disorder: Cross-Sectional Developmental Trajectories. *J Autism Dev Disord.* 2013;43(11):2696–2709. <https://doi.org/10.1007/s10803-013-1823-x>
19. Belteki Z, Lumbreras R, Fico K, Haman E, Junge C. The Vocabulary of Infants with an Elevated Likelihood and Diagnosis of Autism Spectrum Disorder: A Systematic Review and Meta-Analysis of Infant Language Studies Using the CDI and MSEL. *Int J Environ Res Public Health.* 2022;19(3):1469. <https://doi.org/10.3390/ijerph19031469>

20. Friedman L. Sterling A. A Review of Language, Executive Function and Intervention in Autism Spectrum Disorder. *Semin Speech Lang.* 2019;40(4):291–304. <https://doi.org/10.1055/s-0039-1692964>
21. Garrido D. Petrova D. Watson LR. Garcia-Retamero R. Carballo G. Language and motor skills in siblings of children with autism spectrum disorder: A meta-analytic review. *Autism Res.* 2017;10(11):1737–1750. <https://doi.org/10.1002/aur.1829>
22. Posar A. Visconti P. Update about “minimally verbal” children with autism spectrum disorder. *Rev Paul Pediatr.* 2022;40:e2020158. <https://doi.org/10.1590/1984-0462/2022/40/2020158>
23. Chenausky K. Nelson III C. Tager-Flusberg H. Vocalization Rate and Consonant Production in Toddlers at High and Low Risk for Autism. *J Speech Lang Hear Res.* 2017;60(4):865–876. https://doi.org/10.1044/2016_JSLHR-S-15-0400
24. Estes A. Zwaigenbaum L. Gu H. St. John T. Paterson S. Elison JT. Hazlett H. Botteron K. Dager SR. Schultz RT. Kostopoulos P. Evans A. Dawson G. Eliason J. Alvarez S. Piven J. IBIS network. Behavioral, cognitive, and adaptive development in infants with autism spectrum disorder in the first 2 years of life. *J Neurodev Disord.* 2015;7(24):1–10. <https://doi.org/10.1186/s11689-015-9117-6>
25. Levin AR. Varcin KJ. O’Leary HM. Tager-Flusberg H. Nelson CA. (2017). EEG power at 3 months in infants at high familial risk for autism. *J Neurodev Disord.* 2017;9(34):1–13. <https://doi.org/10.1186/s11689-017-9214-9>
26. Roemer EJ. West KL. Northrup JB. Iverson JM. Word comprehension mediates the link between gesture and word production: Examining language development in infant siblings of children with Autism Spectrum Disorder. *Dev Sci.* 2019;22(3):e12767. <https://doi.org/10.1111/desc.12767>
27. Yirmiya N. Gamliel I. Shaked M. Sigman M. Cognitive and Verbal Abilities of 24- to 36-month-old Siblings of Children with Autism. *J Autism Dev Disord.* 2007;37:218–229. <https://doi.org/10.1007/s10803-006-0163-5>
28. Bussu G. Jones EJ. Charman T. Johnson MH. Buitelaar JK. Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis. *J Autism Dev Disord.* 2018;48:2418–2433. <https://doi.org/10.1007/s10803-018-3509-x>
29. Ozonoff S. Young GS. Belding A. Hill M. Hill A. Hutman T. Johnson S. Miller M. Rogers SJ. Schwichtenberg AJ. Steinfeld M. Iosif AM. (2014). The Broader Autism Phenotype in Infancy: When Does It Emerge? *J Am Acad Child Adolesc Psychiatry.* 2014;53(4):398–407. <https://doi.org/10.1016/j.jaac.2013.12.020>
30. Messinger DS. Young GS. Webb SJ. Ozonoff S. Bryson SE. Carter A. Carver L. Charman T. Chawarska K. Curtin S. Dobkins K. Hertz-Picciotto I. Hutman T. Iverson JM. Landa R. Nelson CA. Stone WL. Tager-Flusberg H. Zwaigenbaum L. Early sex differences are not autism-specific: A Baby Siblings Research Consortium (BSRC) study. *Mol Autism.* 2015;6(32):1–12. <https://doi.org/10.1186/s13229-015-0027-y>
31. Franchini M. Duku E. Armstrong V. Brian J. Bryson SE. Garon N. Roberts W. Roncadin C. Zwaigenbaum L. Smith IM. Variability in Verbal and Nonverbal Communication in Infants at Risk for Autism Spectrum Disorder: Predictors and Outcomes. *J Autism Dev Disord.* 2018;48:3417–3431. <https://doi.org/10.1007/s10803-018-3607-9>
32. Gangi DN. Ibañez LV. Messinger DS. Joint Attention Initiation with and without Positive Affect: Risk Group Differences and Associations with ASD Symptoms. *J Autism Dev Disord.* 2014;44(6):1414–1424. <https://doi.org/10.1007/s10803-013-2002-9>
33. Lazenby DC. Sideridis GD. Huntington N. Prante M. Dale PS. Curtin S. Henkel L. Iverson JM. Carver L. Dobkins K. Akshoomoff N. Tagavi D. Nelson III CA. Tager-Flusberg H. Language Differences at 12 Months in Infants Who Develop Autism Spectrum Disorder. *J Autism Dev Disord.* 2016;46(3):899–909. <https://doi.org/10.1007/s10803-015-2632-1>
34. Reynell JK. Grubber CP. (1990). *Reynell Developmental Language Scale.* Los Angeles: Western Psychological Association.
35. Wiig EH. Secord W. Semel E. (1992). *CELF-preschool: Clinical evaluation of language fundamentals—preschool version.* San Antonio, TX: Psychological Corporation.
36. Norrelgen F. Fernell E. Eriksson M. Hedvall Å. Persson C. Sjölin M. Gillberg C. Kjellmer L. Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism.* 2015;19(8):934–943. <https://doi.org/10.1177/1362361314556782>

37. Rose V, Trembath D, Keen D, Paynter J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *J Intellect Disabil Res.* 2016;60(5):464–477. <https://doi.org/10.1111/jir.12284>
38. Tager-Flusberg H, Kasari C. Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. *Autism Res.* 2013;6(6):468–478. <https://doi.org/10.1002/aur.1329>
39. Jarrold C, Boucher J, Russell J. Language Profiles in Children with Autism: Theoretical and Methodological Implications. *Autism.* 1997;1(1):57–76. <https://doi.org/10.1177/13623613970110>
40. Durrleman S, Delage H. Autism spectrum disorder and specific language impairment: overlaps in syntactic profiles. *Lang Acquis.* 2016;23(04):361–386. <https://doi.org/10.1080/10489223.2016.1179741>
41. Kjelgaard MM, Tager-Flusberg H. An Investigation of Language Impairment in Autism: Implications for Genetic Subgroups. *Lang Cogn Process.* 2001;16(2-3):287–308. <https://doi.org/10.1080/01690960042000058>
42. Modyanova N, Perovic A, Wexler K. Grammar Is Differentially Impaired in Subgroups of Autism Spectrum Disorders: Evidence from an Investigation of Tense Marking and Morphosyntax. *Front Psychol.* 2017;8:320. <https://doi.org/10.3389/fpsyg.2017.00320>
43. Roberts JA, Rice ML, Tager-Flusberg H. Tense marking in children with autism. *Appl Psycholinguist.* 2004;25(03):429–448. [https://doi.org/10.1017.S0142716404001201](https://doi.org/10.1017/S0142716404001201)
44. Thurm A, Lord C, Lee LC, Newschaffer C. Predictors of Language Acquisition in Preschool Children with Autism Spectrum Disorders. *J Autism Dev Disord.* 2007;37:1721–1734. <https://doi.org/10.1007/s10803-006-0300-1>
45. Dunn LM, Dunn DM. (2007). PPVT-4: Peabody Picture Vocabulary Test. Bloomington, MN: Pearson Assessments.
46. Williams KT. (2007). EVT-2: Expressive vocabulary test. Pearson Assessments.
47. Sparrow SS, Balla DA, Cicchetti DV. (1985). Vineland Adaptive Behavior Scales. Circle Pines, MN: American Guidance Services.
48. Hedrick DL, Prather EM, Tobin AR. (1975). Sequenced inventory of communication development. Seattle: University of Washington Press.
49. Elliott C. (1990). Manual for the Differential Abilities Scales. San Antonio, TX: Psychological Corporation.
50. Mullen EM. (1985). Manual for the Infant Mullen Scales of Early Learning. Cranston, RI.
51. Mullen EM. (1989). Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Service, Inc.
52. Butler LK, Tager-Flusberg H. Fine motor skill and expressive language in minimally verbal and verbal school-aged autistic children. *Autism Res.* 2023;16(3):630–641. <https://doi.org/10.1002/aur.2883>
53. Sparrow SS, Cicchetti DV, Balla DA. (2005). Vineland Adaptive Behavior Scales (2nd ed.). Minneapolis: NCS Pearson, Inc.
54. Schopler E, Lansing MD, Reichler RJ, Marcus LM. (2004). Psychoeducational Profile Third Edition (PEP-3). Pro-Ed ed. USA.
55. Bovet F, Danjou G, Langue J, Moretto M, Tockert E, Kern S. Les inventaires français du développement communicatif (IFCD) : un nouvel outil pour évaluer le développement communicatif du nourrisson. *Méd Enf.* 2005;25(6):327–332. Available online: <https://solidarites-sante.gouv.fr/IMG/pdf/05.ifdc-2.pdf>
56. Fenson L, Dale P, Reznick J, Thal D, Bates E, Hartung J, Pethick S, Reilly J. (1993). The MacArthur Communicative Development Inventories: User's guide and technical manual. San Diego, CA: Singular Publishing Group.
57. Coudougnan E. (2012). Le bilan orthophonique de l'enfant autiste : des recommandations à la pratique. *Rééduc Ortho.* 2012;50(249):77–90. <https://api.semanticscholar.org/CorpusID:193613155>
58. Macchi L, Herman F, Colli-Vaast L, Merle A, Danchin P. Propriétés psychométriques des tests francophones de langage oral chez l'enfant. *Etud Linguist Appl.* 2023;210(2):145–160. <https://doi.org/10.3917/ela.210.0019>
59. Dunn LM, Thériault-Whalen CM, Dunn LM. (1993). Echelle de Vocabulaire en Images Peabody : EVIP. Toronto : Psycan.
60. Coquet F, Ferrand P, Roustit J. (2009). EVALO 2-6. Ortho édition.
61. Lecocq P. (1996). L'E.CO.S.SE, une épreuve de compréhension syntaxico-sémantique. Villeneuve d'Ascq : Presses Universitaires du Septentrion.

62. Rasch G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Denmark's Paedagogiske Instiut. [https://doi.org/10.1016/S0019-9958\(61\)80061-2](https://doi.org/10.1016/S0019-9958(61)80061-2)
63. Wright BD. Stone MH. (1979). Best Test Design. Chicago: MESA Press.
64. Smith AB. Rush R. Fallowfield LJ. Velikova G. Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Med Res Methodo*. 2008;8(33):1–11. <https://doi.org/10.1186/1471-2288-8-33>
65. Bond TG, Fox CM. (2015). Applying the Rasch Model. *Fundamental Measurement in the Human Sciences*. 3rd ed. New York, NY: Routledge.
66. Fisher Jr WP. Reliability, Separation, Strata Statistics. *Rasch Meas Trans*. 1992;6(3):238. <https://www.rasch.org/rmt/rmt63i.htm>
67. Wright BD. Reliability and separation. *Rasch Meas Trans*. 1996;9:472. <https://www.rasch.org/rmt/rmt94n.htm>
68. Koegel LK. Bryan KM. Su PL. Vaidya M. Camarata S. Definitions of Nonverbal and Minimally Verbal in Research for Autism: A Systematic Review of the Literature. *J Autism Dev Disord*. 2020;50(8):2957–2972. <https://doi.org/10.1007/s10803-020-04402-w>
69. Hudry K. Leadbitter K. Temple K. Slonims V. McConachie H. Aldred C. Howlin P. Charman T. Pact Consortium. Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *Int J Lang Commun Disord*. 2010;45(6):681–690. <https://doi.org/10.3109/13682820903461493>
70. Mervis CB. Klein-Tasman BP. Methodological Issues in Group-Matching Designs: α Levels for Control Variable Comparisons and Measurement Characteristics of Control and Target Variables. *J Autism Dev Disord*. 2004;34(1):7–17. <https://doi.org/10.1023/b:jadd.0000018069.69562.b8>
71. Mervis CB. Robinson BF. Methodological issues in cross-syndrome comparisons: Matching procedures, sensitivity (Se) and specificity (Sp). *Monogr Soc Res Child Dev*. 1999;64(1):115–130. <https://doi.org/10.1111/1540-5834.00011>
72. Kasari C. Brady N. Lord C. Tager-Flusberg H. Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Res*. 2013;6(6):479–493. <https://doi.org/10.1002/aur.1334>
73. Lord C. Rutter M. DiLavore PC. Risi S. Gotham K. Bishop SL. (2012). *Autism Diagnostic Observation Schedule, (ADOS-2)*. Torrance, California: Western Psychological Services.
74. Bal VH. Katz T. Bishop SL. Krasileva K. Understanding definitions of minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *J Child Psychol Psychiatry*. 2016;57(12):1424–1433. <https://doi-org/10.1111/jcpp.12609>
75. Fenson L. Marchman VA. Thal DJ. Dale Philip Reznick JS. Bates E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore, MD: Paul H. Brookes Publishing Company.
76. DiStefano C. Tucker ZT. Jeste S. (2018). Choosing a definition of “minimally verbal”: characteristics of minimally verbal children with ASD in the Autism Genetic Resource Exchange. Abstracts of the INSAR (International Society for Autism Research) annual meeting. 2018:28. <https://insar.confex.com/insar/2018/webprogram/Paper28434.html>
77. Kaushanskaya M. Marian V. Mapping phonological information from auditory to written modality during foreign vocabulary learning. *Ann N Y Acad Sci*. 2008;1145(1):56–70. <https://doi-org/10.1196/annals.1416.008>
78. Ellis Weismer S. Lord C. Esler A. Early Language Patterns of Toddlers on the Autism Spectrum Compared to Toddlers with Developmental Delay. *J Autism Dev Disord*. 2010;40(10):1259–1273. <https://doi.org/10.1007/s10803-010-0983-1>
79. Luyster RJ. Kadlec MB. Carter A. Tager-Flusberg H. Language Assessment and Development in Toddlers with Autism Spectrum Disorders. *J Autism Dev Disord*. 2008;38(8):1426–1438. <https://doi.org/10.1007/s10803-007-0510-1>
80. Martelli ME. Gigliotti F. Giovannone F. Lentini G. Manti F. Sogos C. (2025). Developmental Patterns in Autism and Other Neurodevelopmental Disorders in Preschool Children. *Children*. 2025;12:125. <https://doi.org/10.3390/children12020125>

81. Moore HL, Couteur AL, Charman T, Green J, Parr JR, Grahame V. What is the concordance between parent- and education professional-reported adaptive functioning in autistic children using the VABS-II? *J Autism Dev Disord.* 2023;53(8):3077–3091. <https://doi.org/10.1007/s10803-022-05602-2>
82. Thomas MS, Annaz D, Ansari D, Scerif G, Jarrold C, Karmiloff-Smith A. Using developmental trajectories to understand developmental disorders *J Speech Lang Hear Res.* 2009;52(2):336 – 358. [https://doi.org/10.1044/1092-4388\(2009/07-0144\)](https://doi.org/10.1044/1092-4388(2009/07-0144))
83. Tager-Flusberg H. Strategies for Conducting Research on Language in Autism. *J Autism Dev Disord.* 2004;34(1):75–80. <https://doi.org/10.1023/b:jadd.0000018077.64617.5a>
84. Swensen LD, Kelley E, Fein D, Naigles LR. Processes of Language Acquisition in Children With Autism: Evidence from Preferential Looking. *Child Dev.* 2007 ;78 (2), 542 – 557. <https://doi.org/10.1111/j.1467-8624.2007.01022.x>
85. Park CJ, Yelland GW, Taffe JR, Gray KM. Morphological and syntactic skills in language samples of pre school aged children with autism: Atypical development? *Int J Speech Lang Pathol.* 2012;14(2):95–108. <https://doi.org/10.3109/17549507.2011.645555>
86. Professions and socio-professional categories 2003. Available online: <https://www.insee.fr/fr/metadonnees/pes2003/categorieSocioprofessionnelleAgregee/1?champRecherche=true> (accessed on 24 September 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.