Research in Autism Spectrum Disorders 117 (2024) 102461

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

journal homepage: www.elsevier.com/locate/rasd

A cross-linguistic examination of language measures in autism: A comparison between Dutch and English

Marjolein Mues[a,b,*,1], Yanru Chen[c,2], Ellen Demurie[b,3], Maide Erdogan[b,4], Sarah Schaubroeck[b,5], Helen Tager-Flusberg[c,6], Herbert Roeyers[b,7]

[a] *Brain Development Lab, Department of Psychology and Human Development, Vanderbilt University, 2201 West End Ave, Nashville, TN 37235, United States*

[b] *Research in Developmental Disorders Lab, Department of Experimental-Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium*

[c] *Center for Autism Research Excellence, Department of Psychological and Brain Sciences, Boston University, 100 Cummington Mall, Boston, MA 02215, United States*

ARTICLE INFO

Keywords:
Autism
Language
Development
Assessment modality
Cross-linguistic

ABSTRACT

Background: Language abilities are highly heterogeneous in autism. While a multimodal assessment approach is recommended to capture the language variability, this is not always possible. Therefore, it is important to gain contextual information about different language assessments to determine which assessment is most appropriate for different research questions. As most current work is based on English-speaking populations, this paper compares three language assessment modalities (standardized assessment, parent survey, and a natural language sample) between English-speaking and Dutch-speaking autistic and neurotypical children.

Method: The Mullen Scales of Early Learning, the Vineland Adaptive Behavior Scales and a naturalistic language sample were employed to measure language in 100 preschool-aged participants. Correlation analyses and mixed-effect regressions were conducted, and Bland-Altman plots were created to examine the similarity between measures.

Results: English-speaking and Dutch-speaking parents rated their children's expressive language higher than their receptive language. The best agreement between measures was for standardized language and parent report. Agreement was higher for children with low language scores. Primary language (English vs. Dutch) did not significantly affect the results, but age, nonverbal cognitive abilities and biological sex were shown to predict expressive and receptive language outcome.

Conclusions: For autistic children with low language levels, parent-reported and standardized language measures provide researchers with similar information. Depending on the available time

* Correspondence to: Department of Psychology and Human Development, Vanderbilt University, 2201 West End Ave, Nashville, TN 37235, United States.

E-mail address: marjolein.mues@vanderbilt.edu (M. Mues).

1 https://orcid.org/0000–0002-8235–1060
2 https://orcid.org/0000–0001-6419–7189
3 https://orcid.org/0000–0001-6419–7189
4 https://orcid.org/0000–0002-4190–7472
5 https://orcid.org/0000–0001-5094–2395
6 https://orcid.org/0000–0002-8768–5414
7 https://orcid.org/0000–0002-4895–583X

https://doi.org/10.1016/j.rasd.2024.102461
Received 18 January 2024; Received in revised form 18 June 2024; Accepted 30 July 2024
Available online 10 August 2024

and resources, researchers may choose to use one of these methods. However, for autistic children with (above) average language abilities, multiple modalities should be considered to gain a comprehensive understanding of their language abilities across different settings. A natural language sample is of most added value next to a standardized assessment or parent report.

1. Introduction

Language abilities are often impaired in children with autism spectrum disorder, henceforth autism, a neurodevelopmental condition characterized by challenges in social communication and social interaction, and restricted and repetitive behaviors and interests (American Psychiatric Association, 2022). However, not all autistic children experience language difficulties, and there is great heterogeneity across the spectrum (Baird & Norbury, 2016; Kjelgaard & Tager-Flusberg, 2001; Pickles et al., 2014; pp. 9, 1226). Language level, together with adaptive functioning and IQ, is one of the best-known predictors of short- and long-term developmental outcomes of autistic children (Billstedt et al., 2007; Howlin et al., 2004; Kover et al., 2016; Magiati et al., 2014; Mayo et al., 2013; Sevaslidou et al., 2019), and language difficulties in childhood are associated with decreased levels of adaptive functioning in later life for autistic individuals (Howlin et al., 2013). Therefore, it is important to examine the language abilities in autistic populations using the most appropriate measures.

Much of our scientific knowledge about child and language development is based on work examining monolingual English-speaking, mostly North American populations (Arnett, 2008; Blasi et al., 2022; Legare & Harris, 2016). This potentially limits our understanding of language abilities in general, and in autism specifically, as different languages can be very diverse in their structural properties (e.g., syntax, vocabulary) and pragmatic rules, which are especially challenging for autistic children. Moreover, different linguistic and cultural settings may have an effect on how language is assessed. Natural language sampling is a common method in clinical practice to assess a child's expressive language abilities, but how these samples are obtained (e.g., parent-child interactions) may be influenced by language and/or cultural differences. For example, preliminary research showed that English-speaking children mirror their parents' language use more in comparison to Spanish-speaking children who produced less spontaneous speech (Aguado-Orea, 2022). Parents who speak a language other than English also seem to show less responsive parenting behaviors during a parent-child interaction compared to English speaking participants (Bennetts et al., 2016), which, in turn, may also influence children's language in return. Moreover, parent-reported language, another common method for assessing children's language abilities, may be influenced by cultural differences on what parents perceive as typical and atypical language development.

In this paper, we compare an American English-speaking sample to a Flemish Dutch-speaking sample. Given that one main goal of this study was to examine potential differences in language assessments in autism within linguistically similar environments, we chose Dutch for the cross-linguistic comparison due to its linguistic similarity to English. For example, both Dutch and English have some overlap in vocabulary, as they have shared roots. There are many English loanwords in Dutch and vice versa (e.g., the English "cookie" came from the Dutch "koekje"). English and Dutch also have the same four basic verb tenses (i.e., present simple, past simple, present perfect, and past perfect), although there are stricter rules for determining tense in English compared to Dutch, especially with regards to the present perfect (Veenstra & Knooihuizen, 2021). Another important difference between Dutch and English is that English uses the Subject-Verb-Object (SVO) order as the basic sentence structure, while Dutch is technically a Subject-Object-Verb language (although it masquerades as an SVO language in sentences with only one verb; Koster, 1975). Moreover, apart from linguistic differences, research shows that there are also cultural differences between (Belgian-)Dutch-speaking and (American-)English speaking people (Boiger et al., 2013; Eigenhuis et al., 2015). Taken together, English and Dutch are closely related, yet distinct languages, such that a cross-linguistic comparison between the two can contribute to a more comprehensive understanding of language abilities in autistic children.

The language measures used in this cross-linguistic comparison were the Mullen Scales of Early Learning (MSEL; Mullen, 1995) that was administered by a trained researcher, the parent-report Vineland Adaptive Behavior Scales, 3rd edition (VABS-3; Sparrow et al., 2016), and a natural language sample based on parent-child interaction. Direct assessment measures, such as the MSEL, have the benefit that a trained professional can assess child language in a standardized manner across the sample and that scores can be compared to a normative sample. However, it can be challenging to capture some children's language abilities with standardized tests (Kasari et al., 2013) due to behavioral, motivational or attention difficulties (e.g., Akshoomoff, 2006; Kasari et al., 2013; Koegel et al., 1997). Moreover, standardized assessments often show limited variability in scores among children with severe language difficulties due to a floor effect (Chen et al., 2023; Kasari et al., 2013).

Parent-rated measures do not have most of these drawbacks, as the evaluation is based on language in a naturalistic setting and over a longer time period. Several studies showed that parent report and standardized assessments were correlated in English-speaking autistic and language-delayed samples (e.g., Garibaldi et al., 2021; Luyster et al., 2008; Ring & Fenson, 2000). These correlations were stronger for expressive language than receptive language (Garibaldi et al., 2021; Ring & Fenson, 2000). Some studies have compared scores on the VABS-3 and the MSEL in an English-speaking sample by converting scores to the same format (e.g., age equivalent scores; see, for example, Luyster et al., 2008). Results showed that parent-reported expressive language abilities of autistic children were consistently lower than scores on standardized tests (Luyster et al., 2008; Miller et al., 2017). For receptive language, the opposite was true, such that parent-rated scores were higher than standardized test scores (Luyster et al., 2008; Miller et al., 2017). It may be that parents underestimate their child's expressive language level and overestimate their child's comprehension abilities, but a test-effect is likely also present as the VABS-3 includes more items gauging both language and communication abilities while the MSEL focuses solely on language (Luyster et al., 2008). Previous research has shown that agreement between different measures is stronger in children with lower language scores while more variability is present in those with higher language scores for both neurotypical and language-delayed populations (Bennetts et al., 2016; Garibaldi et al., 2021).

A third way to characterize language abilities is to use a natural language sample. Children's language during daily interactions, such as a play-based caregiver-child interaction (Kover et al., 2014), can be recorded, transcribed and further analyzed to gain a more ecologically valid representation of children's expressive language abilities, given that the data are collected in a natural setting. Although analyzing a natural language sample can be time-consuming, it also provides the opportunity to capture more fine-grained linguistic data. Previous studies with English-speaking autistic children showed that standardized language assessments and natural language samples provided comparable assessments of language skills as they are intercorrelated (Barokova et al., 2020; Condouris et al., 2003; Thomas et al., 2022). However, autistic children sometimes show more impairments in natural language measures than in standardized assessments (see also Condouris et al., 2003).

In general, a multi-method approach is recommended to assess language abilities in autistic populations (Tager-Flusberg et al., 2009), but this may not always be possible due to limited time and resources. Therefore, it is important to identify how standardized language assessments, parent-reported language and natural language samples relate to each other across multiple languages, which can help the researchers select the most appropriate measure.

2. The present study

In the current study, we aim to investigate how different language assessment modalities relate to each other in meaningful ways reflecting the heterogeneous language profiles of autistic children, and whether these modalities capture language similarly or differently in autism across English and Dutch. As most previous work has focused on English-speaking samples, a cross-linguistic comparison with a closely related but distinct language can give us insight into how these measures may capture language differently in autistic children who speak different languages.

Our first aim was to investigate if scores of different language assessment modalities for receptive language and expressive language converge in our sample of autistic and neurotypical children and whether this would be the same for the English-speaking and Dutch-speaking samples. Given that there has not been a cross-linguistic comparison between English and Dutch in autism research, we do not have an a priori hypotheses. Second, previous research in English-speaking samples has shown a high correlation between parent reports and standardized language assessments with some differences between the two, specifically that parents underestimate expressive language and over-estimate receptive language. We hypothesize that we will replicate this finding in both our English- and Dutch-speaking groups (Luyster et al., 2008; Miller et al., 2017; Ring & Fenson, 2000). Third, in line with previous work on the relationship between natural language samples and standardized assessments of expressive language, we expect a significant positive correlation between natural language samples and standardized assessments of expressive language in our sample as well, at least in those children who speak English (Condouris et al., 2003; Thomas et al., 2022).

Table 1
Participant characteristics.

	Autistic participants		Neurotypical participants		Group differences	
	English-speaking	Dutch-speaking	English-speaking	Dutch-speaking	Language	Group
Sample size	25	25	25	25		
Biological sex	4f, 21m[a]	7f, 18m[a,b]	11f, 14m[b]	7f, 18m[a,b]	$X^2(1) = .05$	$X^2(1) = 2.38$
Age in months	43.28[a]	44.60[a]	44.28[a]	41.88[a]	$t(95.60) = -.34$	$t(96.44) = .54$
ADOS-2 overall total score[1]	15.32					
SRS total score[2]		91.13				
Nonverbal cognition[3]	30.74[a]	35.21[a]	49.9[b]	43.46[c]	$t(93.03) = -.38$	$t(88.38) = -6.25$***
Race	2 more than one, 4 Asian, 1 Black, 15 white, 3 missing	4 more than one, 1 Asian, 2 Black, 18 white	7 more than one, 4 Asian, 14 white	1 more than one, 24 white		

Note: Mean scores per group.
[1] ADOS scores were only available for English-speaking autistic participants as the covid-19 pandemic made it impossible to reliably administer the ADOS-2 to Dutch-speaking participants.
[2] SRS total scores were collected to give a characterization of the autism traits in the Dutch-speaking autistic participants.
[3] Age equivalent scores.
[a b c] Letters give an indication of group differences across group and language. Significant differences were present for groups with different letters. No group differences were present for groups with the same letter. For example, biological sex distribution differed between English-speaking autistic and neurotypical participants but not between any of the other group comparisons. Group differences across groups (irrespective of language) and languages (irrespective of group) are shown in the two right columns.

Significance level for group differences: $p \le .05$*, $p \le .01$**, $p \le .001$***

3. Methods

3.1. Participants

Participants included in this study were drawn from two projects. English-speaking participants were drawn from the Training Early Social Language in Autism (TESLA) project at Boston University's Center for Autism Research Excellence, and Dutch-speaking participants were drawn from the Preschool Brain Imaging and Behavior Project (PIP) of the AIMS-2-TRIALS consortium. PIP is a larger European study and consists of five data acquisition sites. For this project, only data collected at Ghent University, Belgium was employed. For inclusion in both studies, autistic participants needed to have a confirmed DSM-5 diagnosis of autism spectrum disorder. Neurotypical participants could not have a first-degree relative with an autism diagnosis. Participants could not have uncorrected vision or hearing difficulties and had to speak English (for TESLA) and Dutch (for PIP) as their primary language. For PIP additional inclusion criteria applied as MRI was a part of the larger study (i.e., no MRI contraindications).

In total, the TESLA study included 89 participants ($n = 52$ autistic children, $n = 37$ neurotypical children) between the ages of 22 and 60 months old. PIP included 74 participants ($n = 38$ autistic children, $n = 36$ neurotypical children) between the ages of 31 and 54 months old. After prioritizing participants with no missing data, we matched TESLA participants with PIP participants based on mean length of utterance, as our research questions are focused on language abilities. Moreover, as language is highly variable at the age of our participants (e.g., Pickles et al., 2014, pp. 12269), and given the differences in age range between TESLA and PIP, we also stratified the sample based on chronological age. This resulted in language- and age-matched samples. The final sample included 100 children (25 English-speaking autistic and 25 neurotypical participants from the TESLA project, and 25 Dutch-speaking autistic and 25 neurotypical participants from PIP). The English-speaking participants' autism status was verified by meeting the autism spectrum cut-off on the Autism Diagnostic Observation Schedule-2 (Lord et al., 2012), and all the Dutch-speaking participants had an official diagnostic report of autism spectrum disorder. The TESLA sample consisted of monolingual English-speaking participants. The PIP sample included one bilingual child that was exposed to Spanish in addition to Dutch. Gender ratio did not differ between TESLA and PIP samples, but there was a difference between English-speaking autistis and non-autistic children in TESLA. All participant characteristics can be found in Table 1.

Written informed consent was obtained from participants' legal guardian prior to their participation in either study. All experimental protocols and procedures were approved by the ethical committees of Boston University and Ghent University/Ghent University Hospital respectively.

3.2. Measures

3.2.1. Language assessments

3.2.1.1. Mullen scales of early learning. The MSEL is a test of general cognitive abilities that includes a receptive and an expressive language scale used in this study (Mullen, 1995). An example of a receptive language item of the Mullen is to give the child instructions with items in front of them, such as: "Here are a ball and a book. I would like you to do something for me. Give me the ball and then close the book". An example of an expressive language items is making verbal analogies. For example, the researcher says "John is a boy; Susan is a …?" where the child should then finish the sentence by saying "girl". The test was administered by a researcher or a clinician (Mullen, 1995). A summary of participants' scores per site can be found in Table 2. The MSEL is translated into Dutch but has not been normed on a Dutch sample and therefore American norms were used.

3.2.1.2. Vineland adaptive behavior scales. The VABS-3 (Sparrow et al., 2016) Comprehensive Parent/Caregiver questionnaire (English-speaking participants) or Comprehensive Interview (Dutch-speaking participants) was used to reflect children's receptive and expressive communication skills in this study. The VABS Parent/Caregiver questionnaire and the interview form have similar enough

Table 2
Median language scores per group.

	Autistic participants		Neurotypical participants	
	English	Dutch	English	Dutch
MSEL receptive language[1]	25 (29.2, 20.2)	34 (30.3, 14.9)	53 (52.8, 12.0)	39 (41.6, 8.3)
MSEL expressive language[1]	22 (27.4, 19.5)	31 (30.4, 13.7)	51 (51.7, 14.9)	43 (45.5, 8.5)
VABS−3 receptive language[1]	33 (28.9, 13.5)	25 (25.1, 11.6)	44 (45.1, 7.6)	42 (44.0, 13.8)
VABS−3 expressive language[1]	42 (45.2, 32.6)	26 (25.5, 12.7)	45 (51.8, 21.1)	50 (51.8, 15.0)
Mean length of utterance[2]	1.64 (2.1, 1.2)	2.43 (2.3, 1.1)	3.57 (3.7, 0.5)	3.72 (3.8, 0.6)
Number of total words[2]	57 (118.8, 146.6)	148 (125.7, 96.2)	320 (314.8, 97.2)	264 (265.2, 91.3)
Number of different words[2]	33 (47.5, 46.9)	55 (46.4, 30.7)	114 (118.4, 27.9)	97 (99.6, 18.2)

Median scores (mean scores, standard deviation)
[1]Age equivalent scores
[2]Raw scores
Abbreviations: MSEL = Mullen Scales of Early Learning; VABS-3 = Vineland Adaptive Behavior Scales - Third Edition
Note: TESLA ad PIP samples were matched based on chronological age and expressive language.

content to warrant a combination for analyses (Sparrow et al., 2016). Parents responded to questions regarding their child's daily communication skills using a scale of 0 ((almost) never present), 1 (sometimes present), 2 (always/mostly present) or don't know. Separate scores for receptive and expressive language were reported. The VABS-3 is translated into Dutch but has not been normed on a Dutch sample and therefore American norms were used. An overview of participant scores is shown in Table 2.

3.2.1.3. Natural language samples. Natural language samples at both sites were derived from a play session between the child and one of their main caregivers (mostly mothers, $n = 90$ mothers out of 100 participants). Sessions took place at the respective university and were videorecorded. Children and caregivers were given a standardized set of toys to play with at both sites, which included different kinds of toys. For Dutch-speaking participants, the set included building blocks, a children's book, a doll, a play tea set, miniature cars and a stuffed animal, while the English-speaking participants received a play set that did not include a book, cars, or a stuffed animal but included miniature dinosaur shapes sorters, baby dolls, miniature utensils, and a basketball.

Play sessions took between 10 and 20 min per session. The first ten minutes of each session after the researchers left the room were transcribed. The video recordings of the child and their caregiver were transcribed verbatim and divided into utterances separated by breath pauses. All intelligible utterances were transcribed. The English videos were transcribed by the first author and the Dutch videos were transcribed by the first author, two graduate students in clinical psychology and one student in speech and language pathology. All transcribers are native speakers of Dutch. The first author who coded the English videos is bilingual Dutch and English-speaking and has extensive experience with the English language. Students who helped with transcription were blind to the diagnosis of the child, but the first author was not. 20 % of the Dutch recordings were transcribed by all transcribers. Inter-rater reliability of MLU between the transcribers as measured with intraclass correlation was.93.

From the transcriptions, the mean length of utterance in morphemes (MLU), the total number of words (TNW) and the number of different words based on lemma (NDW) were calculated. For the transcriptions of the English videos, this was done through the software program Systematic Analysis of Language Transcripts (SALT) (Miller & Iglesias, 2015). As this software is not available for Dutch, processing of the Dutch transcriptions was done manually by the first author based on SALT conventions (Miller & Iglesias, 2015).

3.3. Nonverbal cognition

Nonverbal cognitive abilities were calculated to characterize the sample using the MSEL (Mullen, 1995). Scores were calculated as follows: [average (MSEL visual reception age equivalent score + MSEL fine motor skills age equivalent score) / chronological age in months] * 100 to generate a composite score. An overview of mean scores per group can be found in Table 1.

3.4. Statistical analyses

We first examined the language data of all groups and assessed within-sample differences between standardized and parent-reported language using the paired-sample Wilcoxon test. We then employed non-parametric Spearman's correlations between all variables. Non-parametric tests were used because of the limited sample size and as most variables in the different groups did not follow a Gaussian distribution as indicated by q-q plots and the Shapiro-Wilk test for normality.

Next, Bland-Altman plots were created for each pair of language comparisons using the blandr package in R (Datta, 2017). To compare scores on the same scale, raw scores of all language measures were transformed into z-scores prior to the creation of the plots. In order to reduce the number of tests, participants' MLU, TNW and NDW z-scores were combined into one composite score representing spontaneous expressive language by adding the z-scores and dividing this number by three, giving all values equal weight (Andrade, 2021). Bland-Altman plots can be used to visualize differences in measurements between two techniques to determine their similarity (Giavarina, 2015). The plots show three lines: a horizontal line in the middle of the chart, representing the mean of difference between test scores of both methods, also referred to as the bias between the instruments, and two lines indicating the lower limit of agreement (LL) and upper limit of agreement (UL). The LL and UL are derived from the mean of difference minus/plus 1.96 times the standard deviation of the difference between assessments. The LL and UL are surrounded by 95 % confidence intervals. If the line representing the mean of differences is close to zero, the average difference in measurements is small. If 95 % of the data points in the plot are scattered between the limits of agreement, this suggests a good level of agreement between the two methods (Giavarina, 2015). If they are further away from zero, outside of the LL and UL, this indicates weak agreement (Karun & Puranik, 2021).

In a final step, the data was transformed from wide to long format to execute mixed-model regression analyses with a per subject random intercept. Language score was the dependent variable, type of test, language (English versus Dutch) and group status (autistic or neurotypical) were the independent variables. Nonverbal cognition, biological sex and age were included as control variables. To minimize data comparisons, the composite score for natural language was used. Z-scores were used here since the language measures are each on a different scale.

Interaction effects between the independent variables were assessed using model comparison with the Akaike Information Criterion (AIC), which estimates the quality of each model relative to another model taking into account the trade-off between model fit and complexity (Akaike, 1998). The model with the lowest AIC was chosen as the most parsimonious if the difference in AIC was at least two points (Burnham & Anderson, 2004). Interaction effects were included in the "final model" if they significantly contributed to the model fit as shown by the AIC. Corrections for multiple comparisons for all analyses were done using the False Discovery Rate (FDR). All data analyses were conducted in R (R Core Team, 2022).

4. Results

4.1. Missing data

There was no missing data for the natural language samples and the standardized language assessments. However, the parent-reported VABS-3 was not completed for 28 English-speaking participants (n autistic = 13, n neurotypical = 15), resulting in a limited sample size of 22 English-speaking participants with VABS-3 data (n autistic = 12, n neurotypical = 10).

4.2. Description of language abilities

After examining the natural language sample, we found that seven of the included English-speaking autistic participants and four Dutch-speaking autistic participants were considered minimally verbal (i.e., produced less than twenty functional words spontaneously (Kasari et al., 2013)). None of the neurotypical children were considered minimally verbal. The ratio verbal/minimally verbal in the autistic group did not differ significantly between languages ($X^2(1) = 1.05$, $p = .31$).

All language scores per group per language are shown in Table 2. As can be seen in Table 2, VABS-3 expressive language scores were higher than VABS-3 receptive language scores for both languages. A paired Wilcoxon test confirmed that this effect was significant for the whole sample ($V = 1719$, $p < .01$) and for the English-speaking autistic group and Dutch-speaking neurotypical group (English-speaking autistic group: $V = 71$, $p = .01$; Dutch-speaking neurotypical group: $V = 244$, $p < .05$). When examining the MSEL scores, a significant effect was found only in the Dutch-speaking neurotypical group ($V = 268$, $p < .001$), indicating higher language scores for expressive than receptive language. Compared to each other, VABS-3 expressive language abilities were significantly higher than MSEL expressive language abilities for the English-speaking autistic sample (TESLA: $V = 74$, $p < .01$) and the Dutch-speaking neurotypical sample (PIP: $V = 247.5$, $p < .05$). The reverse was true for the English-speaking neurotypical sample and Dutch-speaking autistic samples, which showed higher expressive language scores on the Mullen (TESLA: $V = 0$, $p < .01$; PIP: $V = 51.5$, $p = .01$). For receptive language, only the autistic groups showed a significant effect, indicating higher scores for receptive scores on the MSEL than the VABS-3 (English-speaking: $V = 0$, $p < .001$; Dutch-speaking: $V = 60$, $p < .01$).

Comparisons between all language scores per group per language (English versus Dutch) are shown in Fig. 1 (expressive language) and Fig. 2 (receptive language).

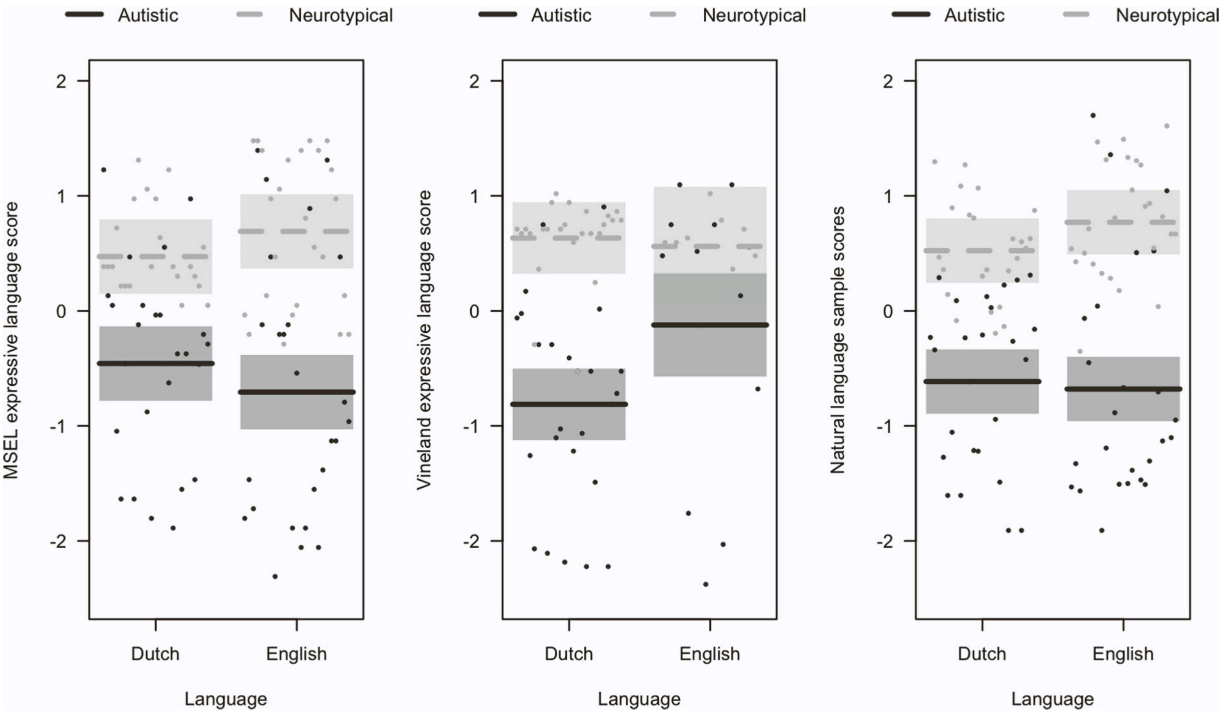

Fig. 1. Expressive language methods per group per site. Note: Figure shows the mean-centered language scores per participant, as well as the group mean standard deviation per language measure. The average score for the whole sample lies at zero. A negative score thus indicates a score that is lower than the whole sample average, while a positive score indicates a score that is higher than the whole sample average.

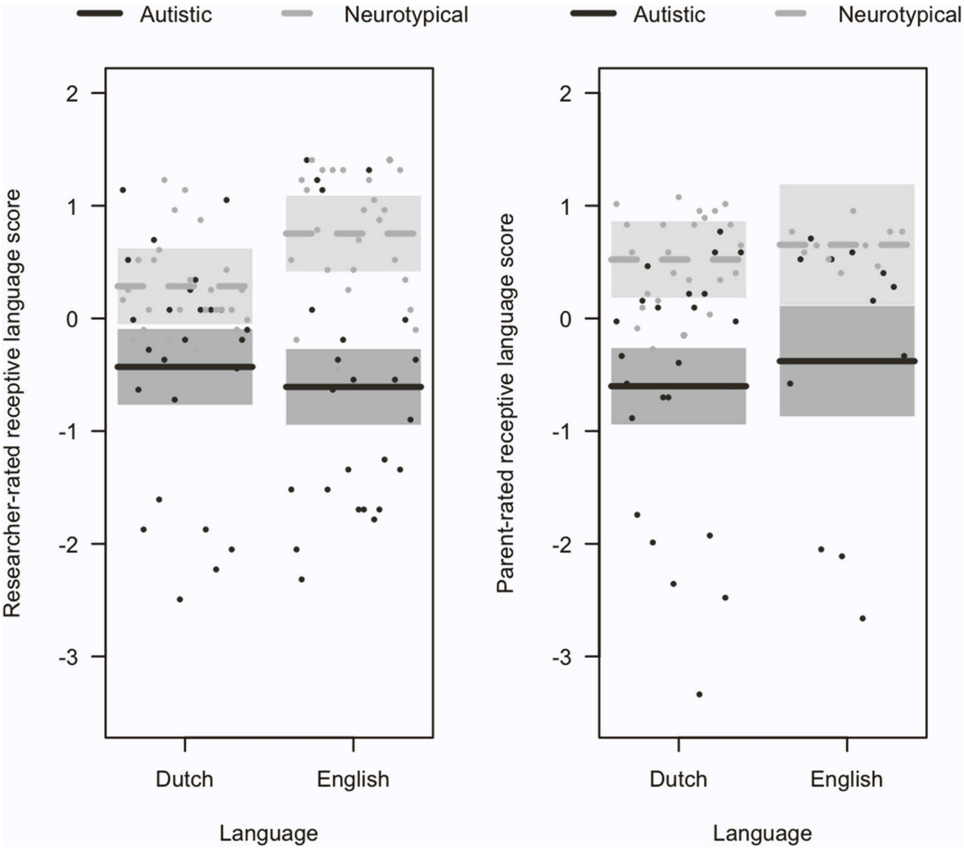

Fig. 2. Receptive language methods per group per site. Note: Figure shows the mean and standard deviations per language measure in z-scores, meaning that the whole sample average lies at zero.

4.3. Correlation analyses

Spearman correlation analyses were executed between all language measures. For the autistic groups in both languages, all language measures significantly correlated together, while this was not the case for the neurotypical group. All correlations coefficients and their significance are shown in Table 3 (autistic sample) and Table 4 (neurotypical sample).

4.4. Bland-Altman plots

Bland-Altman plots were created separately for the two sites for all pairs of language assessment types for receptive and expressive language, i.e., VABS-3 versus MSEL expressive language; MSEL versus natural language sample; VABS-3 versus natural language sample; VABS-3 versus MSEL receptive language. Given the limited sample sizes, plots are collapsed across neurotypical and autistic children per language. The plots are shown in Figs. 3 to 10.

4.4.1. Expressive language

4.4.1.1. MSEL versus VABS-3. Plots comparing MSEL and VABS-3 expressive language scores are shown in Figs. 3 and 4. Note that the sample size of the English-speaking participants for the VABS-3 is limited, resulting in bigger confidence bands (i.e., the red and green areas). In both plots, the mean difference between measurements is small (close to the zero line), and the data points mostly lie within the LL and UL. There is however a cone-shaped pattern, with data points on the left sides of the figures remaining closer to the zero line, while data points on the right sides show more variability. This heteroskedasticity shows that the difference is proportional to the size of the measurement, i.e., variability increases when the measurements increase in size, thus in this case with higher language scores, while agreement remains high for lower language scores.

4.4.1.2. MSEL versus natural language. Plots examining the agreement between the MSEL and natural language are shown in Figs. 5 and 6. For both plots, the mean of difference line is very close to the zero line, indicating minimal differences in the averages of the measurements. However, for both the English-speaking and Dutch-speaking participants, two data points fall outside of the UL and LL

Table 3
Spearman correlations between language measures - autistic sample.

	English	Dutch	English	Dutch	English	Dutch	English	Dutch	English	Dutch	English	Dutch
	1	1	2	2	3	3	4	4	5	5	6	6
1. MSEL EL												
2. MSEL RL	**.94** ***	**.95** ***										
3. VABS−3 EL	.80 * *	**.87** ***	**.86** ***	**.87** ***								
4. VABS−3 RL	.61 *	**.82** ***	.70 * *	**.88** ***	**.93** ***	**.86** ***						
5. MLU	**.88** ***	**.85** ***	**.83** ***	**.84** ***	.79 * *	**.78** ***	.69 *	**.75** ***				
6. TNW	**.72** ***	**.71** ***	**.66** ***	**.71** ***	.70 *	**.70** ***	.71 * *	**.75** ***	**.86** ***	**.82** * **		
7. NDW	**.84** ***	**.83** ***	**.83** ***	**.83** ***	**.85** ***	**.81** ***	**.82** * **	**.80** ***	**.91** ***	**.94** * **	**.94** * **	**.91** * **

Significance level: $p \leq .05$ * , $p \leq .01$ * *, $p \leq .001$ * **
Correlation coefficients in **bold** survived FDR corrections for multiple comparisons
Abbreviations: MSEL = Mullen Scales of Early Learning; EL = expressive language; RL = receptive language; VABS-3 = Vineland Adaptive Behavior Scales - Third Edition; MLU = mean length of utterance; TNW = total number of words, NDW = number of different words

Table 4
Spearman correlations between language measures - neurotypical sample.

	English	Dutch	English	Dutch	English	Dutch	English	Dutch	English	Dutch	English	Dutch
	1	1	2	2	3	3	4	4	5	5	6	6
1. MSEL EL												
2. MSEL RL	**.81** ***	**.85** ***										
3. VABS−3 EL	**.90** ***	.55 **	.56	.39 *								
4. VABS−3 RL	.70 *	.02	-.05	-.07	.68 *	.44 *						
5. MLU	.03	.44 *	.26	.33	.46	.54 *	.08	-.08				
6. TNW	-.15	-.13	.10	-.26	.15	.13	-.16	-.23	.43 *	.39 *		
7. NDW	.12	.39 *	.28	.28	.52	.18	.14	-.15	**.64** ***	.48 * *	**.80** ***	.56 * *

Significance level: $p \leq .05$ * , $p \leq .01$ * *, $p \leq .001$ * **
Correlation coefficients in **bold** survived FDR corrections for multiple comparisons
Abbreviations: MSEL = Mullen Scales of Early Learning; EL = expressive language; RL = receptive language; VABS-3 = Vineland Adaptive Behavior Scales - Third Edition; MLU = mean length of utterance; TNW = total number of words, NDW = number of different words

confidence intervals, thus failing to keep 95 % of data points within the limits of agreement, indicating less agreement between the MSEL and natural language samples. Here too, heteroskedasticity is present in the plots, with data points showing more variability for higher scores.

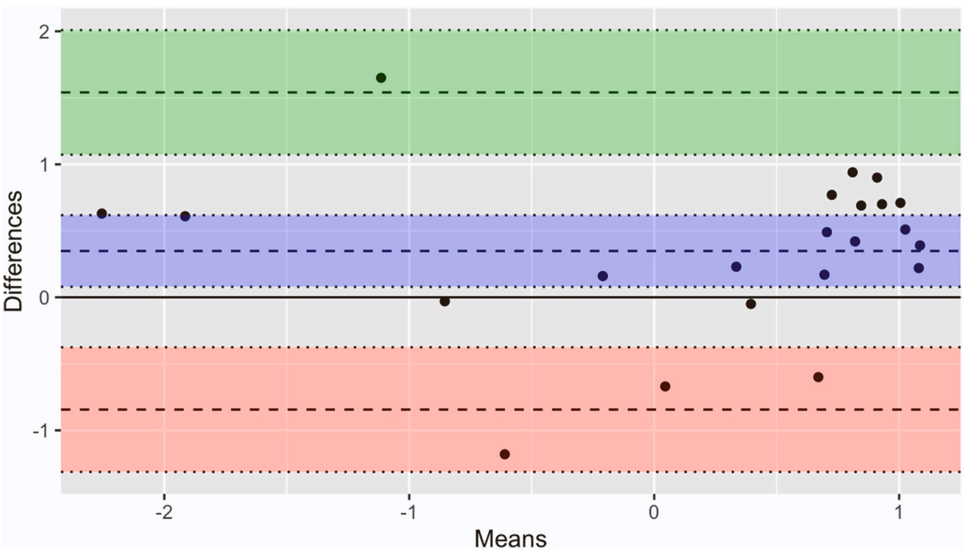

Fig. 3. Bland-Altman plot comparing MSEL and VABS-3 expressive language – English-speaking participants. Note: The y-axis shows the differences between the two measurements while the x-asis shows the averages of the two measurements. The solid horizontal line in the plot represents the mean difference between the two measurements and the dotted lines represent the ± 1.96 standard deviations of the differences.

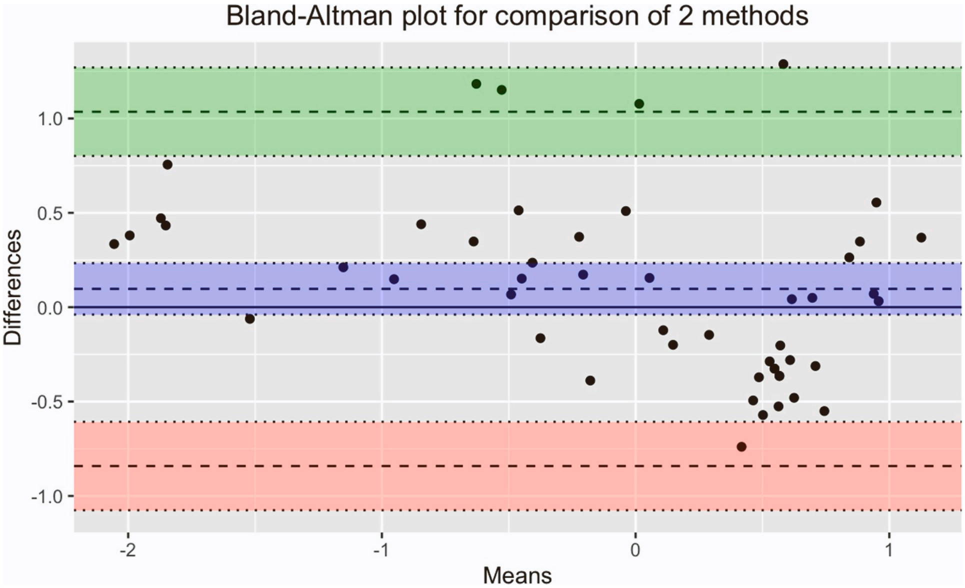

Fig. 4. Bland-Altman plot comparing MSEL and VABS-3 expressive language – Dutch-speaking participants.

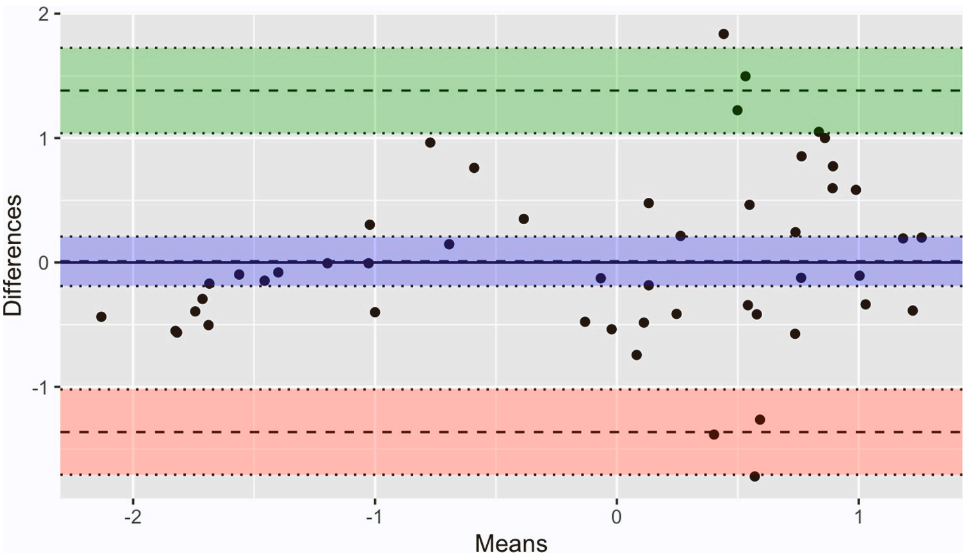

Fig. 5. Bland-Altman plot comparing MSEL and natural expressive language – English-speaking participants.

4.4.1.3. VABS-3 versus natural language. Figs. 7 and 8 depict the agreement between VABS-3 expressive language and natural language. The plots indicate small to minimal average differences in measurements, as data points for both languages lie mostly within the LL and UL, except for one data point in the Dutch-speaking sample. There is a cone-shaped pattern, similar to the previous plots, indicating the presence of heteroskedasticity, with data points showing more variability for higher scores.

4.5. Receptive language

For receptive language, only the MSEL and VABS-3 language were compared, as there is no measure of natural receptive language. Bland-Altman plots per site are shown in Figs. 9 and 10. Similar to most plots for expressive language, the mean difference line is close to zero and all data points fall within the LL and UL, except one data point for the Dutch-speaking sample. However, a pattern of heteroskedasticity is present with around the mean scores for language (i.e., x-axis score of 0).

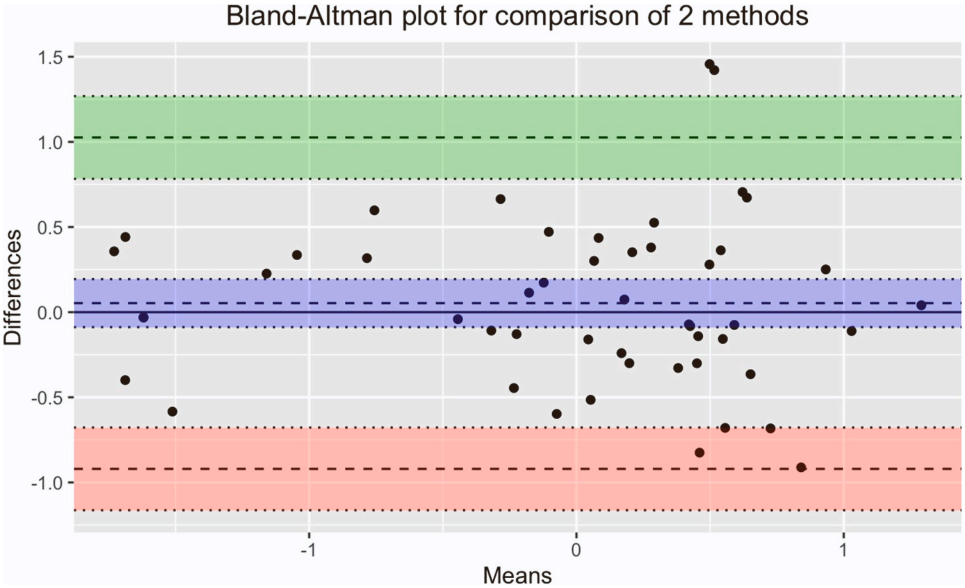

Fig. 6. Bland-Altman plot comparing MSEL and natural expressive language – Dutch-speaking participants.

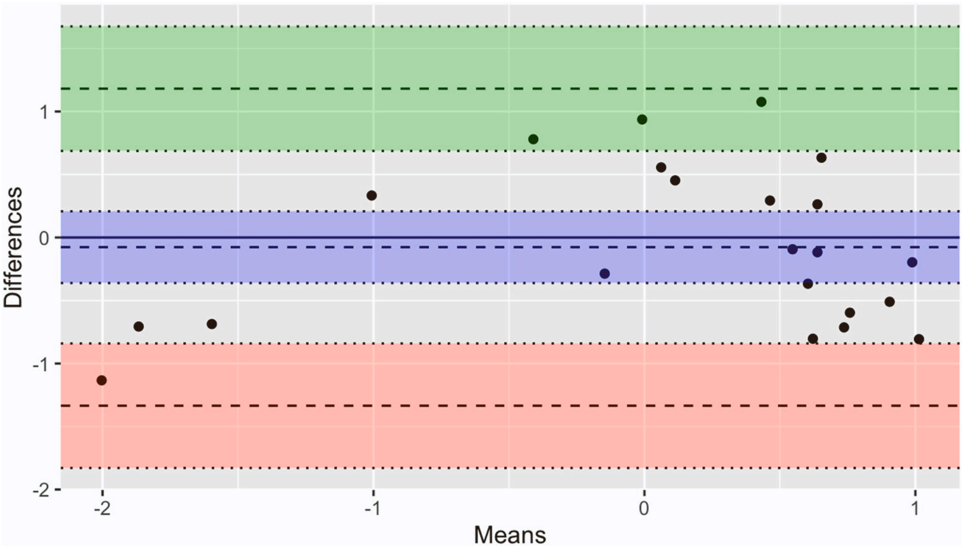

Fig. 7. Bland-Altman plot comparing VABS-3 and natural expressive language - English-speaking participants.

4.6. *Regression analyses*

4.6.1. *Testing of mixed-effects assumptions*

The assumptions for mixed-effects regression were assessed for selected models. The variance inflation factor (VIF) was examined to assess collinearity between numerical variables. All VIFs were below 10, suggesting no collinearity in the data (largest VIF was 2.91) (Vittinghoff et al., 2012). Autocorrelation was assessed with the acf function in R (R Core Team, 2022) and was not detected. Heteroskedasticity was not present based on visual inspection of residual plots and residuals of the selected model followed a Gaussian distribution in both models.

4.7. *Expressive language*

After model comparison using AIC and testing all interaction effects, the most parsimonious/ best fit model for expressive language included the pre-specified independent variables (i.e., group status (autistic or neurotypical), type of test, language (English versus

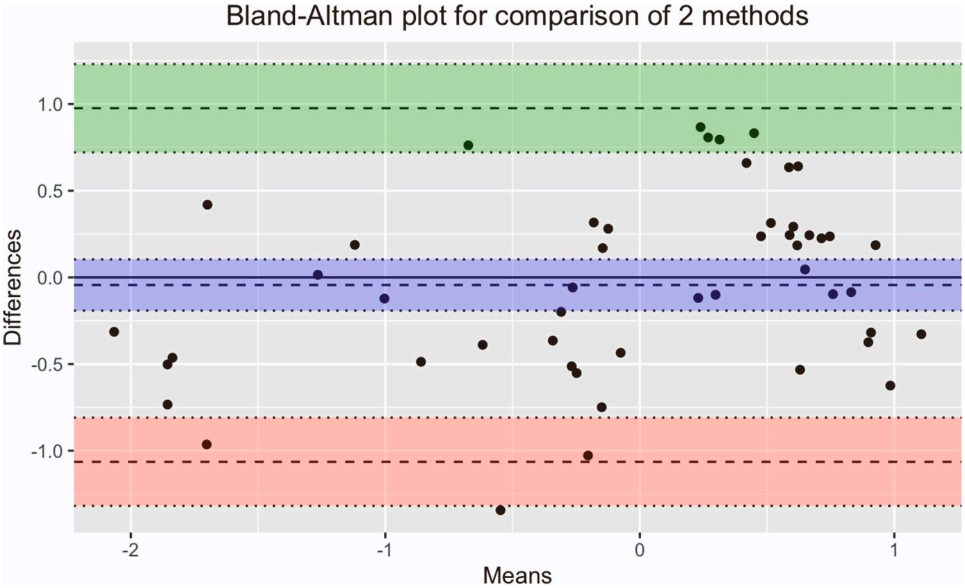

Fig. 8. Bland-Altman plot comparing VABS-3 and natural expressive language - Dutch-speaking participants.

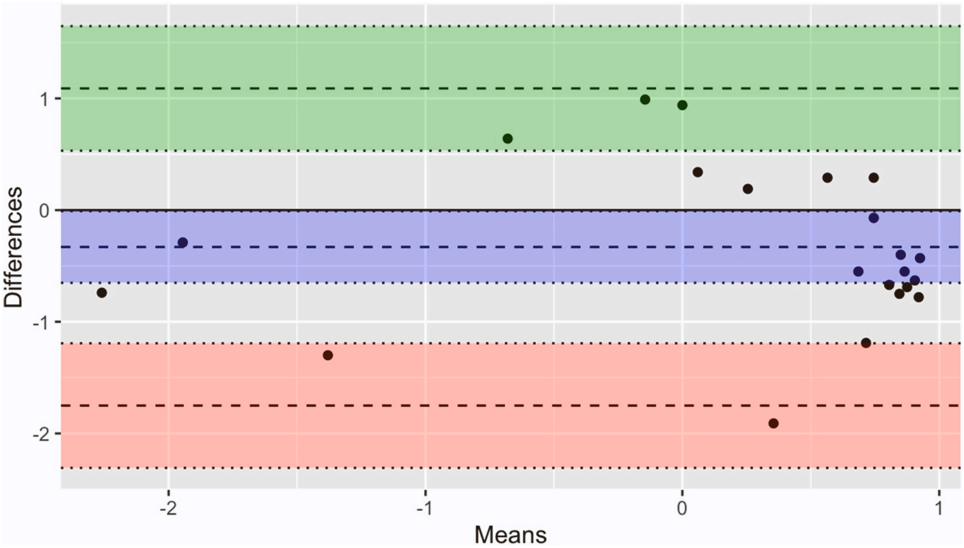

Fig. 9. Bland-Altman plot comparing MSEL and VABS-3 receptive language - English-speaking participants.

Dutch), nonverbal cognition, biological sex and age) (R^2 total model =.67, AIC = 458.91, comparison AICs of competing models > 10). Significant main effects were present for group, with higher scores for the neurotypical participants (b =.49, SE =.15, $p < .001$), age, with language scores increasing with age (b =.04, SE =.006, $p < .001$), nonverbal cognition, with higher nonverbal cognitive abilities resulting in higher language scores (b =.59, SE =.07, $p < .001$), and biological sex, with boys achieving significantly lower language scores than girls ($b = -.32$, SE =.11, $p = .01$). Interaction effects were tested and were non-significant.

4.8. *Receptive language*

The final best fit model selected for receptive language was the same as for expressive language (R^2 total model =.69, AIC = 317.25, comparison AICs of competing models > 5). Interestingly, main effects were present for age, with scores getting higher as age increases (b =.05, SE =.01, $p < .001$), nonverbal cognition, with higher nonverbal cognition resulting in higher language scores (b =.73, SE =.07, $p < .001$) and biological sex, with boys achieving lower scores than girls ($b = -.15$, SE =.05, $p = .01$), but no main effect was present for group.

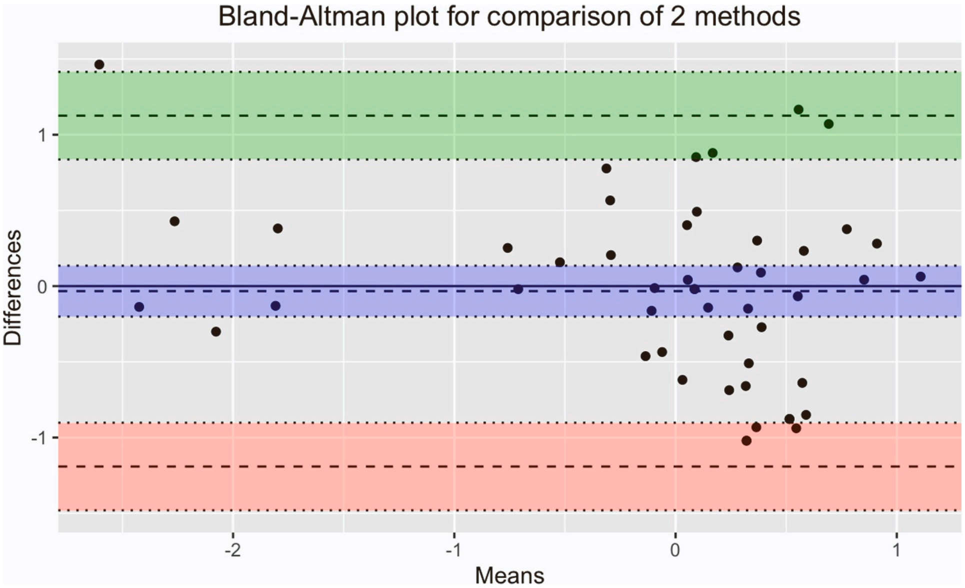

Fig. 10. Bland-Altman plot comparing MSEL and VABS-3 receptive language - Dutch-speaking participants.

5. Discussion

In this paper, we aimed to examine different language assessments and their relationship to each other in English-speaking and Dutch-speaking autistic children. Overall, findings in this study showed that there is substantial heterogeneity in language abilities within the autistic samples in both languages, which is in line with previous results (e.g., Baird & Norbury, 2016; Pickles et al., 2014, pp. 12269). Our results also suggest that there is a strong agreement between parent-reported language and standardized language assessment while there is less agreement between these variables and variables derived from a natural language sample (i.e., MLU, total number of words, number of different words). This may suggest that a multimethod approach of language assessment can be beneficial in capturing the different facets of autistic children's language abilities.

Previous work in exclusively English-speaking autistic populations has shown that the three language measurement modalities correlate highly with each other (Condouris et al., 2003; Luyster et al., 2008; Ring & Fenson, 2000; Thomas et al., 2022), but that parents tend to overestimate their child's receptive language and underestimate expressive language abilities compared to standardized assessment (Luyster et al., 2008; Miller et al., 2017; Ring & Fenson, 2000). In this study, we also found high correlations between the three expressive language measures for the autistic samples, both English-speaking and Dutch-speaking. In contrast, we did not find the same degree of correlation for the neurotypical participants, although the MSEL and VABS-3 scores of expressive language mostly correlated highly with each other, as did the measures of natural language (MLU, TNW, NDW) amongst each other. The hypothesized effect of parental underestimation of expressive language abilities, based on earlier research (Luyster et al., 2008; Miller et al., 2017) was not present in our data. On the contrary, our results show that parents across both the English-speaking and Dutch-speaking sites rated their autistic child's expressive language abilities higher than their receptive language abilities. Such an effect was also present for the standardized language measure in Dutch-speaking neurotypical children. The contrasting findings between our and previous studies may be due to the different ages of the participants, as previous work included children around 30 months (Luyster et al., 2008; Ring & Fenson, 2000), while the average age in our study was 44 months.

Moreover, unlike previous results (Luyster et al., 2008), we found that parents of autistic children in both languages and parents of English-speaking neurotypical children rated their child's expressive language abilities significantly higher than scores on the MSEL indicated. The reverse effect was found for receptive language, but only for the autistic groups. Parents may rate their child's expressive language higher because they may misinterpret echolalia as genuine language production (Cohn et al., 2023), though this would not be judged in this way by researchers and would not be considered in the scores achieved on a standardized test. It could also be the case that autistic children have more expressive communication with their caregiver than with an unfamiliar examiner (Barokova et al., 2020). Another important point to mention here is that the MSEL assess specifically structural linguistic abilities, while the VABS-3 assesses more general expressive and receptive communication abilities, which could influence the results.

A second research question in this study examined how different language measures converge cross-linguistically. To this end, we created Bland-Altman plots for each pair of language assessments for both languages. The results show that the different language assessments converge relatively well with each other, which is in line with our correlation analyses. However, for most comparisons, heteroskedasticity was present in the plot indicating that the variability increases with higher language scores. In other words, language measures are mostly in agreement for those with lower language scores but differ more for higher language scores. This is the case for both the English-speaking and Dutch-speaking children and is in line with previous studies examining agreement between standardized language measures and parent reports in neurotypical and language-delayed children (Bennetts et al., 2016; Garibaldi et al., 2021). However, it should be noted that there were also more participants present at the right end of the scale and the higher agreement for those with lower language scores can be caused by floor effects of minimally verbal participants. The data from the natural language sample correlated least well with the other measures, indicating that there is the least amount of overlap in the information between the natural language sample and the other measures. This may indicate that a natural language sample taps into a different linguistic variety that is not captured by standardized assessments and vice versa. Adding such a sample may therefore provide researchers and clinicians with information regarding an additional facet of the language abilities in autistic children.

Lastly, we examined the effect of type of test (MSEL, VABS-3, and natural language), group and language on language abilities simultaneously, while controlling for age and nonverbal cognitive abilities with mixed-effects regression. These results showed that autistic children scored significantly lower on all expressive language assessments than neurotypical children on a group-level. Age, nonverbal cognition and a female biological sex were positively related to both expressive and receptive language. No main effect of language or type of test was significant, indicating that children's language scores were not significantly impacted by the choice of language assessment or whether they spoke English or Dutch. Interestingly, no significant effect of group (autistic versus neurotypical) was present for receptive language, although Table 2 shows lower scores for autistic participants. It appears that group status did not add unique variance once age, nonverbal cognition and biological sex were taken into account, showing the importance of including these variables in data analyses.

Taken together, our results did not identify sole effects of English-speaking versus Dutch-speaking on the language scores of our participants, likely reflecting the closeness between English and Dutch, and indicating that the three language measure modalities (i.e., standardized measures, parent reports, and natural language samples) are similarly appropriate to capture language abilities between English-speaking and Dutch-speaking autistic children. Moreover, for the autistic participants, standardized, parent-reported and natural language assessments were in good agreement with each other, although each with their own specific nuances (e.g., significantly higher scores on expressive language for parent reported language) that may impact the choice of test type depending on potential research questions. Especially agreements between the standardized test and parent report were quite high. This has further implications for clinical practice as it means that parent report, such as parent questionnaires alone may already provide a relatively good estimate when screening autistic children for potential language difficulties, which may be especially beneficial for communities with limited resources. Our results also show that, similar to previous work in other populations (Bennetts et al., 2016; Garibaldi et al., 2021), convergence between language methods is highest for children with lower scores. The finding indicates that, if one language measurement modality is unfeasible due to limited resources or study format (e.g., online study), utilizing another measurement modality will capture similar language profiles for autistic children with low language levels. On the other hand, less convergence between measurement modalities in autistic children with higher language levels reiterates the importance of using multiple sources of information to fully capture the language abilities of those with (above) average competence. In addition to a standardized test or parent report, getting a natural language sample would provide most unique information about spontaneous language in natural settings for verbally fluent autistic children.

Given that our work focuses on linguistically similar languages (i.e., English vs Dutch), future work should assess languages that are further apart from each other. Studying more diverse languages will enrich our understandings of how to choose and incorporate the most appropriate language measures for autistic individuals with culturally- and linguistically diverse backgrounds. Future research could also include different populations, such as bilingual children, as more than half of the world's population is bilingual (Hopkins et al., 2017), to examine if language abilities of monolingual and bilingual autistic children are revealed differently across different assessment modalities.

5.1. Strengths and limitations

This study has multiple strengths. First, we adopted a cross-linguistic approach to examine our research questions so generalize our results beyond the English language. A second strength is the multimodal language assessment, including natural language sample, which requires specific training on transcribing the spontaneous language and therefore can be a time-consuming method, but can give valuable, ecologically valid insights into children's language abilities (Barokova & Tager-Flusberg, 2020).

Several limitations of the current study provided directions for future studies. First, although the total sample size was 100, dividing this number into four equal groups resulted in a limited sample size per group. The sample was especially limited for parent-reported language data of the TESLA site as some data were missing. Future cross-linguistic comparison research should consider including a larger sample size to increase statistical power for analyses. We also did not keep track of effects of potential language interventions that children received. For example, English-speaking participants may have received different interventions than Dutch-speaking participants, which could have influenced their language scores. Next, although the coder for the natural language variables in Dutch is well-qualified and experienced in working with this type of data, the codings would have ideally been done using SALT software rather than manually. This was however not possible as SALT software is not available in Dutch. To mitigate this fact however, the coder coded some videos twice with high reliability. Moreover, due to restrictions because of the covid-19 pandemic, ADOS-2 scores could not reliably be acquired for the Dutch-speaking autistic participants. Although we did collect SRS-2 scores for this sample to gain more information about the level of autistic traits, the lack of reliable ADOS scores makes it impossible to compare the two samples on the level of autism traits. Lastly, the MSEL and VABS-3 standardized assessments are translated in Dutch but are not validated or normed on a Dutch sample. Therefore, we used the same American norms for all groups. Although these norms were only used to acquire age equivalent scores to assess group differences and other analyses were done using raw scores transformed into z-scores, this may have had an effect on the results.

6. Conclusion

This study conducted a cross-linguistic comparison of how different language assessments capture language abilities in autistic and neurotypical preschoolers. Our findings showed a good convergence between language abilities measured by standardized measure, parent report, and natural language sample, with a better agreement between assessments for autistic children with lower language scores. As expected, autistic children obtained significantly lower language scores than neurotypical children at the group-level across all three assessment modalities. Within the autism groups, there was great individual variability and language scores ranged from clinical scores to very good structural language capacities. Significant differences were found with respect to parent-rated and standardized assessments of language, with parents rating their child's expressive language abilities significantly higher, and their receptive language significantly lower than researchers did. We found no differences between the English-speaking and Dutch-speaking sites, likely resembling the closeness between the two languages. Future work should assess languages that are more linguistically distinct, especially in non-Western populations.

CRediT authorship contribution statement

Marjolein Mues: Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization. **Yanru Chen:** Writing – review & editing, Supervision, Project administration, Investigation, Data curation, Conceptualization. **Ellen Demurie:** Writing – review & editing, Supervision, Project administration, Investigation, Data curation. **Maide Erdogan:** Writing – review & editing, Investigation, Data curation. **Sarah Schaubroeck:** Writing – review & editing, Investigation, Data curation. **Helen Tager-Flusberg:** Writing – review & editing, Supervision, Funding acquisition. **Herbert Roeyers:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declared no conflicts of interest.

Data availability

The data that has been used is confidential.

Acknowledgements

We would like to thank all the families participating in PIP in Europe and TESLA in the United States as well as all those involved in the projects. We would like to specially thank Manon Krol and Jennifer Leanno for their work on the TESLA study and Laura Kiekens, Lara Demanet and Kevser Kaymak for their help with transcribing the Dutch language samples. This work has received support from the NIH/NIDCD grant n° P01HD064653) and EU/EFPIA/SFARI/Autistica/AUTISM SPEAKS Innovative Medicines Initiative 2 Joint Undertaking (AIMS-2-TRIALS grant n° 777394). Any views expressed are those of the author(s) and not necessarily those of the funders.

References

Aguado-Orea, J. (2022). Estimations of child linguistic productivity controlling for vocabulary and sample size. *Frontiers in Psychology, 13*, Article 996610. https://doi.org/10.3389/fpsyg.2022.996610

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer New York. https://doi.org/10.1007/978-1-4612-1694-0_15.

Akshoomoff, N. (2006). Use of the mullen scales of early learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychology, 12*(4-5), 269–277. https://doi.org/10.1080/09297040500473714

American Psychiatric Association. (2022). Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR (5th ed., text rev.). https://doi.org/https://doi.org/10.1176/appi.books.9780890425787.

Andrade, C. (2021). Z Scores, standard scores, and composite test scores explained. *Indian Journal of Psychological Medicine, 43*(6), 555–557. https://doi.org/10.1177/02537176211046525

Arnett, J. J. (2008). The neglected 95%: Why american psychology needs to become less american. *American Psychologist, 63*(7), 602–614.

Baird, G., & Norbury, C. F. (2016). Social (pragmatic) communication disorders and autism spectrum disorder. *Archives of Disease in Childhood, 101*(8), 745–751. https://doi.org/10.1136/archdischild-2014-306944

Barokova, M., & Tager-Flusberg, H. (2020). Commentary: Measuring language change through natural language samples. *Journal of Autism and Developmental Disorders, 50*(7), 2287–2306. https://doi.org/10.1007/s10803-018-3628-4

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

Bennetts, S. K., Mensah, F. K., Westrupp, E. M., Hackworth, N. J., & Reilly, S. (2016). The agreement between parent-reported and directly measured child language and parenting behaviors. *Frontiers in Psychology, 7*, 1710. https://doi.org/10.3389/fpsyg.2016.01710

Billstedt, E., Gillberg, C. I., & Gillberg, C. (2007). Autism in adults: Symptom patterns and early childhood predictors. Use of the DISCO in a community sample followed from childhood. *Journal of Child Psychology and Psychiatry, 48*(11), 1102–1110. https://doi.org/10.1111/j.1469-7610.2007.01774.x

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences, 26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Boiger, M., Deyne, S. D., & Mesquita, B. (2013). Emotions in "the world": Cultural practices, products, and meanings of anger and shame in two individualist cultures. *Front Psychol, 4*, 867. https://doi.org/10.3389/fpsyg.2013.00867

Burnham, K.P., & Anderson, D.R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection [doi:10.1177/0049124104268644]. Sociological Methods & Research, 33(2), 261–304. https://doi.org/10.1177/0049124104268644.

Chen, Y., Siles, B., & Tager-Flusberg, H. (2023). Receptive language and receptive-expressive discrepancy in minimally verbal autistic children and adolescents. *Autism Research*, 1–13. https://doi.org/10.1002/aur.3079

Cohn, E. G., McVilly, K. R., & Harrison, M. J. (2023). Echolalia as defined by parent communication partners. *Autism and Developmental Language Impairments, 8*, Article 23969415231151846. https://doi.org/10.1177/23969415231151846

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

R. Core Team. (2022). R: A Language and Environment for Statistical Computing. In R Foundation for Statistical Computing. http://www.r-project.org.

Datta, D. (2017). blandr: a Bland-Altman Method Comparison package for R. *Zenodo*. https://doi.org/https://doi.org/10.5281/zenodo.824514.

Eigenhuis, A., Kamphuis, J. H., & Noordhof, A. (2015). Personality differences between the united states and the netherlands:the influence of violations of measurement invariance. *Journal of Cross-Cultural Psychology, 46*(4), 549–564. https://doi.org/10.1177/0022022115570671

Garibaldi, A., Venkatesh, L., Bhat, J. S., & Boominathan, P. (2021). Relationship between parental report of language skills and children's performance among 3-year-olds: Implications for screening language among preschoolers. *International Journal of Pediatric Otorhinolaryngology, 151*, Article 110943. https://doi.org/10.1016/j.ijporl.2021.110943

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochem Med (Zagreb), 25*(2), 141–151. https://doi.org/10.11613/bm.2015.015

Hopkins, B., Geangu, E., & Linkenauger, S. (2017). The Cambridge encyclopedia of child development. Cambridge University Press.

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

Howlin, P., Moss, P., Savage, S., & Rutter, M. (2013). Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(6), 572–581. https://doi.org/10.1016/j.jaac.2013.02.017

Karun, K. M., & Puranik, A. (2021). BA.plot: An R function for Bland-Altman analysis. *Clinical Epidemiology and Global Health, 12*, Article 100831. https://doi.org/10.1016/j.cegh.2021.100831

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. https://doi.org/10.1002/aur.1334

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

Koegel, L. K., Koegel, R. L., & Smith, A. (1997). Variables related to differences in standardized test outcomes for children with autism. *Journal of Autism and Developmental Disorders, 27*(3), 233–243. https://doi.org/10.1023/A:1025894213424

Koster, J. (1975). Dutch as an SOV language. *Linguistic Analysis, 1*(2), 111–136.

Kover, S. T., Davidson, M. M., Sindberg, H. A., & Ellis Weismer, S. (2014). Use of the ADOS for assessing spontaneous expressive language in young children with ASD: a comparison of sampling contexts. *Journal of Speech, Language and hearing research: JSLHR, 57*(6), 2221–2233. https://doi.org/10.1044/2014_JSLHR-L-13-0330

Kover, S. T., Edmunds, S. R., & Weismer, S. E. (2016). Brief report: Ages of language milestones as predictors of developmental trajectories in young children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 46*(7), 2501–2507. https://doi.org/10.1007/s10803-016-2756-y

Legare, C. H., & Harris, P. L. (2016). The ontogeny of cultural learning. *Child Development, 87*(3), 633–642. https://doi.org/10.1111/cdev.12542

Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., & Bishop, S.L. (2012). Autism Diagnostic Observation Schedule - Second Edition (ADOS-2). Western Psychological Services.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(8), 1426–1438. https://doi.org/10.1007/s10803-007-0510-1

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

Mayo, J., Chlebowski, C., Fein, D. A., & Eigsti, I. M. (2013). Age of first words predicts cognitive ability and adaptive skills in children with ASD [Article]. *Journal of Autism and Developmental Disorders, 43*(2), 253–264. https://doi.org/10.1007/s10803-012-1558-0

Miller, J.F., & Iglesias, A. (2015). *Systematic Analysis of Language Transcripts*. [Computer program]. SALT Software, LLC.

Miller, L. E., Perkins, K. A., Dai, Y. G., & Fein, D. A. (2017). Comparison of parent report and direct assessment of child skills in toddlers [Article]. *Research in Autism Spectrum Disorders, 41-42*, 57–65. https://doi.org/10.1016/j.rasd.2017.08.002

Mullen, E.M. (1995). Mullen Scales of Early Learning. American Guidance Service, Inc.

Pickles, A., Anderson, D. K., & Lord, C. (2014). Heterogeneity and plasticity in the development of language: A 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry, 55*(12), 1354–1362. https://doi.org/10.1111/jc

Ring, E. D., & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language, 20*, 141–159. https://doi.org/10.1177/014272370002005902

Sevaslidou, I., Chatzidimitriou, C., & Abatzoglou, G. (2019). The long-term outcomes of a cohort of adolescents and adults from Greece with autism spectrum disorder. *Annals of General Psychiatry, 18*(1), 1–6. https://doi.org/10.1186/s12991-019-0250-6

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

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., Rice, M., Stoel-Gammon, C., Wetherby, A., & Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language and Hearing Research: JSLHR, 52*(3), 643–652. https://doi.org/10.1044/1092-4388(2009/08-0136)

Thomas, R. P., Wittke, K., Blume, J., Mastergeorge, A. M., & Naigles, L. (2022). Predicting language in children with ASD using spontaneous language samples and standardized measures. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05691-z

Veenstra, J., & Knooihuizen, R. (2021). English and the use of the simple past in Dutch. *Taal & Tongval, 73*(1), 81–101.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2012). Regression methods in. *Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. New York: Springer,. https://books.google.be/books?id=boMUM6N8V4wC.