



Examining Phenotypical Heterogeneity in Language Abilities in Chinese-Speaking Children with Autism: A Naturalistic Sampling Approach

Xue-Ke Song¹ · Cassandra Lee¹ · Wing-Chee So¹

Accepted: 20 May 2021 / Published online: 26 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Phenotypical heterogeneity in language abilities is a hallmark of autism but remains poorly understood. The present study collected naturalistic language samples from parent–child interactions. We quantified verbal abilities (mean length of utterance, tokens, types) of 50 Chinese-speaking children ($M = 5; 6$) and stratified subgroups based on their autism traits, IQ, and language abilities. Using hierarchical cluster analysis, four groups were identified. Group 1, the least affected group, had mild autism, the highest IQ, and the strongest verbal abilities. Group 2, the severely affected group, had the lowest IQ, most severe autism symptoms, and weakest verbal abilities. Group 3 and Group 4 displayed average levels of verbal abilities and IQ. These findings may characterize the heterogeneous profiles of verbal abilities in Chinese-speaking children.

Keywords Heterogeneity · Cluster analysis · Naturalistic sampling approach

Introduction

Autism spectrum disorder (ASD) is widely recognized as a complex, heterogeneous neurodevelopmental condition. According to the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric et al., 2013), individuals with ASD have impairments in social communication and interaction, as well as restricted and repetitive behaviors. Based on the estimation from Western countries, it affects about 1 in 54 children aged 8 (Maenner et al., 2020). The preliminary prevalence estimates of ASD in mainland China (Sun et al., 2013, 2019) is similar to that of developed countries, i.e., around 0.8–1.5% among preschool and school-aged children. Based on this estimate, in Hong Kong there are approximately over 10,000 individuals aged from 0 to 12 years diagnosed with ASD, of which over 7,000 are below six years of age. With such a large number of young children diagnosed with ASD and the significance of early intervention for these children, it is

crucial to understand phenotypical heterogeneity in language impairments among young Chinese-speaking children with ASD in Hong Kong.

DSM-5 differentiates individuals with ASD based on the level of the support (ranging from basic support to substantial to very substantial support) that should be offered. Those with more severe autism symptoms should receive more substantial support than those with milder symptoms. Researchers have generally reached consensus that there are large individual variations in cardinal symptoms, cognitive skills, and language communications among individuals with ASD (Fennell et al., 2010; Masi et al., 2017). For example, high-functioning autism individuals, who had high cognitive abilities and intact language skills, exhibited various levels of social communication impairments and restricted or repetitive behaviors (Klopper et al., 2017; Ring et al., 2008). Minimally verbal or nonverbal individuals with autism, who had severe impairments in cognitive and language abilities, showed severe deficits in their social and communication skills. However, this group of individuals were always excluded from analysis because of the difficulties of collecting their data from standardized tests (Bacon et al., 2019; Bal et al., 2016). To summarize, notable heterogeneity, spanning the entire range in each of these domains, imposes challenges for researchers to study its etiology, prognosis, diagnosis, and treatment. To address these

✉ Wing-Chee So
wingchee@cuhk.edu.hk

¹ Department of Educational Psychology Department, The Chinese University of Hong Kong, Shatin, Hong Kong, SAR, China

challenges, the present study aimed to delineate homogeneous subgroups of verbal abilities based on the autism symptoms, cognitive abilities, and language abilities in a sample of children with ASD aged three to eight. The following sections first summarized the previous research that examined heterogeneity in social communication skills based on standardized instruments followed by the studies of heterogeneity using multidimensional measures. Finally, we highlighted the importance of studying heterogeneity with naturalistic language samples.

Heterogeneity of Social Communication Impairments Based on Standardized Instruments

Recent studies have adopted data-driven approaches to examine clinically meaningful subgroups (Cholemkery et al., 2016; Georgiades et al., 2013; Hu & Steinberg, 2009). Of these studies, Georgiades et al. (2013) examined the phenotypical heterogeneity in social communication skills based on the scores in the Autism Diagnostic Interview-Revised (ADI-R) in 391 toddlers aged 38.3 months on average. They defined three subgroups by factor analysis according to the severity of social communication impairments and stereotyped behaviors. One group (Class 3) showed the most severe impairments in social communication and repetitive behavior domains, whereas the other two groups showed fewer impairments in both domains. Class 1 exhibited more social communication deficits than repetitive behaviors, while Class 2 showed reverse patterns. Similarly, Cholemkery et al. (2016) conducted cluster analysis and used ADI-R as a classification criterion to define clinically meaningful subgroups in samples with a broad age range from toddlers to adults. This study classified three subgroups based on the levels of impairment of social communication and stereotyped behaviors. One subgroup had the most severe social communication deficits and stereotyped behavior, whereas the other two subgroups exhibited moderate symptoms in these two domains (one with severe impairments in restricted behaviors but mild impairments in social interaction and the other one showing the opposite profile).

These two studies proposed that there are mixed levels of autism symptoms, with one group having mild social communication deficit and severe repetitive behavior and another group with reverse profile (severe social communication deficit but mild repetitive behavior). However, other studies did not find similar patterns and reported that severity of two domains covaried across subgroups (Wiggins et al., 2012).

Heterogeneity of Social Communication Impairments Based on Multidimensional Measures

An important breakthrough of the study conducted by Zheng et al. (2020) abandoned standardized diagnostic instruments

(e.g., ADI-R, Autism Diagnostic Observation Schedule); rather they opted for dimensional measures for multiple domains, including autism symptoms, cognitive abilities, verbal skills, and adaptive function in 188 preschoolers. Cluster analysis on principal component was conducted, resulting in an identification of three clusters. Cluster 1 was the least affected group and characterized as having milder impairments on social communication skills and repetitive behaviors, as well as advanced language and cognitive skills. Cluster 2 exhibited more severe social communication impairment and restrictive behaviors than Cluster 1, yet with comparable language and cognitive skills. Cluster 3 was the most affected group and showed more severe cardinal symptoms than Cluster 1. The cognitive and language abilities in Cluster 3 were the lowest than the other two clusters. Taken together, Zheng et al. (2020) identified two subgroups with comparable levels of autism symptoms but different cognitive and language abilities.

Klopper et al. (2017) also investigated phenotypical autism subgroups in individuals without intellectual disabilities. He reported two subgroups, which differed significantly along the dimensions of social communication impairments and restricted and repetitive behaviors, but not in receptive language and verbal comprehension ability. These results suggested that the severity of core autism symptoms may dissociate with cognitive and language abilities, therefore supporting the multi-dimensional approaches when defining homogenous subgroups.

Heterogeneity of Language Impairments Using Language Samples

Despite the tremendous efforts to define homogenous subgroups over the past decades (Beglinger & Smith, 2001; Syriopoulou-Delli & Papaefstathiou, 2020), empirically reliable subgroups have not yet been identified. Moreover, majority of previous research studied heterogeneity of language impairments using standardized or global language tests. Only very few studies collected naturalistic language samples (Bacon et al., 2019; Tek et al., 2014; Wittke et al., 2017). Of these studies, Wittke et al. (2017) investigated variations in vocabularies and grammar use among preschoolers with ASD. They manually divided participants into three subgroups (38 high verbal, 11 low verbal, and 33 minimally verbal) based on the scores of the vocabulary language tests and non-verbal IQs. Then they collected language samples from Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012). All but two participants in minimally verbal group did not generate enough speech and thus were excluded. The two children in minimally verbal group who produced sufficient tokens to transcribe and children in low verbal group were then combined and assigned to the global language impairment group. They used 10% grammatical

errors as a cut-off point to manually identify high-verbal group children into grammatical impairment group and language normal group. The global language impairment group scored low in the overall lexical ability but its grammatical skills remained intact; the grammatical impairment group showed the opposite. Grammatical impairment group was comparable to the normal language group in terms of the means for utterance level (MLU), but they showed significantly more grammatical errors than the other two groups. Their results indicated that children with ASD displayed unbalanced language skills in grammar and lexical components.

Tek et al. (2014) also looked into spontaneous language samples and examined the structural language acquisition of 12 toddlers with autism across six-time points over four months. Based on the raw scores in the expressive language test, toddlers with autism were divided into two subgroups using median split based on the scores of the standardized language test at time point 1: high verbal skills (HV) and low verbal skills (LV). The study reported that the ASD-HV group showed drastic improvements in vocabulary, grammatical complexity and production, with their development patterns similar to those of the typically development group. On the other hand, the ASD-LV group showed slight improvement in most of the language measures even at the last time point.

Although Wittke et al. (2017) and Tek et al. (2014) manually categorized subgroups and their criterion for cut-off points were controversial, these two studies shed light on the use of naturalistic language samples instead of standardized language tests to define homogenous language subgroups. While global language tests are convenient for researchers, they have disadvantages for children who are inattentive and lacking motivation. On the contrary, naturalistic language sampling can capture the nuanced aspects of linguistic deficits in details, thereby being more sensitive in evaluating language abilities than global language tests (Casenhiser et al., 2015; Chiang, 2009; Jyotishi et al., 2017; Rice et al., 2010). For instance, Casenhiser et al. (2013) conducted a social-communication-based intervention program and reported that the treatment group improved significantly in social skills manifested in the naturalistic language samples in the one-year follow-up assessment. Surprisingly, this improvement was not evident in the standardized language test. Specifically, the treatment group performed better on language measures such as the mean length of utterance (MLU), numbers of utterance and other communication aspects (Casenhiser et al., 2015).

Having said that, language samples in Wittke's study were collected in a structured assessment (ADOS-2). However, it is not an ideal method for eliciting speech from children who are minimally verbal (Kasari et al., 2013). Indeed, these children may be able to produce more vocabulary when talking

to their parents during free play (Kover et al., 2014). Kover et al. (2014) collected naturalistic language samples from different partners and contexts, and compared utterance and words tokens, as well as pragmatic language. They found that participants produced more utterances and number of conversational turns when interacting with parents than when interacting with examiners.

Overall, previous findings about the number of homogeneous groups in children with ASD regarding their language abilities are not conclusive. Additionally, to date, very few studies collected naturalistic language samples for cluster analyses. Naturalistic language sampling approach has its advantage in measuring language abilities in children with ASD. Firstly, naturalistic language sampling is more sensitive to measure the change in language abilities. Different domains of expressive language measures can be derived from naturalistic language samples, including phonology, syntax, lexicon, and communicational skills (Barokova & Tager-Flusberg, 2020). Secondly, children with ASD could produce more utterances and display various communication skills when they are freely talking to their caregivers. Therefore, the current study adopted the naturalistic language sampling approach to examine heterogeneity in language abilities in children with ASD.

We here focused on Chinese (Cantonese)-speaking children. A large portion of the aforementioned studies investigating language variations have been focusing on children with ASD from western countries. Research into the heterogeneity of language impairment in Chinese-speaking children is scarce. Only a small number of studies investigated the expressive language profiles in Chinese-speaking children with ASD. Previous language studies have primarily compared Chinese children with ASD with typically developing children in terms of their language abilities and they did not collect language samples (Su & Naigles, 2019; Yi et al., 2013; Zhou et al., 2015). Of these studies, Su et al. (2018) examined uneven expressive language development in a sample of 160 participants between 17 and 84 months old Chinese preschoolers with ASD. Parents completed the Putonghua Communicative Development Inventory-Toddler form (Tardif & Fletcher, 2008). Three subgroups (low verbal, middle verbal, and high verbal) were defined based on the total vocabulary production. The three subgroups displayed discrepancies in lexical components (e.g., the proportion and total utterance on nouns, verbs, and pronouns), syntax, and MLU. These results suggested that there are variations in language abilities in Chinese-speaking children. However, this study did not study heterogeneity from naturalistic language samples and examine the relation of heterogeneity with IQs and autism severity.

Taken together, the present study aimed to delineate homogeneous subgroups of verbal abilities in Chinese (Cantonese)-speaking children with ASD and to compare

phenotypical presentations and detailed lexical components among these subgroups. To the best of our knowledge, this study was the first one looking at the heterogeneity in language abilities in Chinese-speaking children. We here focused on some aspects of grammatical complexity (e.g., MLU) and lexical diversity (e.g., types, the number of different words, and tokens, the number of total utterances) as indicators of the language abilities. Mean length of utterance (MLU) indicates grammatic complexity of language ability (Brown, 1973) and is highly correlated with scores on the standardized measurement of grammatical development (Condouris et al., 2003). Although Scarborough et al. (1991) reported that the correlation between MLU and grammatical development becomes weaker when MLU exceeded 3.0, many studies confirmed that MLU outperformed standardized language measures in the estimation of grammatic complexity (Casenhiser et al., 2015; Jyotishi et al., 2017). Besides, the number of different words indicates lexical diversity and the richness of expressive language and is significantly correlated with the scores on standardized measurement of semantic development (Condouris et al., 2003).

Hierarchical cluster analysis was conducted to define the optimal number of clusters based on multiple measurements, including autism symptoms, IQ, language abilities. Since clustering is an unsupervised method of machine learning, we could not predict the numbers of clusters for our current data; rather we would choose the optimal number of clusters based on the elbow method. Then we characterized the generated subgroups by phenotypical measurements and detailed lexical components in terms of the types and frequencies of words children could use spontaneously with their caregivers. The findings of the current study would fill the gap in the literature by defining subgroups of children with autism based on their speech collected in the naturalistic language samples.

Methods

Participants

Participants in the current study were taken from the larger Robot for Autism Behavioral Intervention project (RABI) (So, 2020), an intervention conducted at the Chinese University of Hong Kong for Chinese-speaking individuals with autism aged 3 to 18. RABI recruits Chinese-speaking participants diagnosed with autism throughout Hong Kong. A subset of these participants ($N=59$; 52 males) contributed their language samples before intervention. Among the 59 participants, eight did not complete assessments of IQ and autism severity within the time allotted and one was absent from the parent–child interaction. Those nine children were

excluded from this study. Thus, the final sample for the current study was 50 (45 males).

Assessments

Autism Diagnostic Observation Schedule—Second Edition (ADOS-2)

ADOS-2 assesses and diagnoses ASD across age, developmental level, and language skills (Lord et al., 2012). In the present study, it was conducted by a trained professional who completed ADOS-2 Advanced/Research Training. ADOS comparison scores converted from the total raw scores and chronological ages were reported. In this study, raw scores of social affect (SA) and restricted and repetitive behavior (RRB) were also reported separately in order to better capture the distinct features of both domains.

Kaufman Brief Intelligence Test, Second Edition (KBIT-2)¹

KBIT-2 assesses both verbal and nonverbal intelligence in people from 4 through 90 years of age (Kaufman, 2004). It is composed of two separate scales. The Verbal Scale contains two kinds of items—Verbal Knowledge and Riddles—both of which assess crystallized ability (knowledge of words and their meanings). The Nonverbal Scale includes a Matrices subtest that assesses fluid thinking—the ability to solve new problems by perceiving relationships and completing analogies. Test items are free of cultural and gender bias. Children’s standardized Verbal and Nonverbal Scores, plus a composite IQ, were calculated and reported.

Parent–Child Interaction

Caregivers were invited to interact with their children for 20 min in a treatment room at the Chinese University of Hong Kong. Each time, a child was presented with a standardized set of age-appropriate toys and his/her parent was instructed to play with the child as they normally would at home. All participating children and their caregivers played with the same set of toys. These toys included a food-themed set (book, puzzle, toy), wooden trains and a police-themed set (soft toy, book). Each session was video-recorded using two cameras with high-definition zoom-in functions to capture the head and hand movements of parent and child.

¹ There was a child aged 3; 7 and his KBIT-2 standard score was 95. The findings were still the same after excluding him.

Transcriptions

The language samples were transcribed by the research assistants trained in the Codes for the Human Analysis of Transcripts (CHAT) format using Computerized Language Analysis (CLAN) software (MacWhinney, 2000). Each language sample was transcribed verbatim by one transcriber, who viewed each recording multiple times until the entire sample was transcribed. According to CHAT coding conventions, utterances or portions of utterances that could not be fully transcribed after three viewings were indicated as unintelligible. A consensus procedure was implemented in all the 50 transcripts. Once transcriptions were completed, they were proofed by a second transcriber. Transcribers viewed each other's video recordings while reading the initial transcriptions (Shriberg et al., 1984). When errors or discrepancies were discovered, transcribers would discuss among themselves until agreement was reached. Otherwise, those utterances or portions of utterances were considered unintelligible.

Coding

CLAN conventions were deployed to perform morphological analysis on the transcripts, as well as to mark syntactic errors and extract word type and token variables for all parts of speech. The grammatical category labels for the Cantonese corpus were based on the established 33 categories (e.g., modal verb, directional verb, noun, proper noun, pronoun, preposition, adjective, adverbs) (MacWhinney, 2000; Matthews & Yip, 1994). 20% of the transcriptions from 10 of the children were randomly selected and coded by a research assistant who was naïve to the objectives of the present study. The inter-rater reliabilities for the grammatical coding were measured by Cronbach's Alpha, which were 0.85 for verbs, 0.93 for nouns, 0.88 for prepositions, 0.89 for adverbs, 0.95 for adjectives, and 0.92 for pronouns. Lexical measures included both types (number of different morphemes) and tokens (total numbers of words including nouns, verbs, adjectives, adverbs, pronouns, and prepositions). The mean length of utterance (MLU, total number of words divided by total utterances), total number of types, and total number of tokens were generated using CLAN. Detailed lexical components, including the percentages of nouns, verbs, pronouns, adjectives, adverbs, and prepositions, were also calculated.

Each utterance was also coded for jargon and echolalia. Echolalia is the repetition, with similar intonation, of words or phrases that someone else has said; it can be immediate, or right after someone said it, or delayed, meaning a repetition of something heard in the past (Tager-Flusberg et al., 2005). Jargon was coded if the child used strings of non-meaningful speech with odd intonation. Utterances

containing echolalia or jargon were not included in the analyses of MLU and types and tokens.

Analyses

Data analyses were conducted by statistical software R version 4.0.0. We first calculated descriptive statistics of assessments and language parameters of children with ASD in this study. Then we examined the relationships among autism symptoms, cognitive abilities, and language abilities using Pearson correlation, and reported significance levels. To define homogeneous subgroups, hierarchical clustering algorithm and the “complete” method built into R's `hclust` function were used. The elbow method was applied to the mean of the within-cluster variances in order to determine the optimal number of clusters. A radar plot was displayed to visualize the characteristics of each subgroup. Although there are no rules of thumb about the minimal sample size for cluster analysis, the common practice is that there should be no less than 2^k cases (where k refers to the number of variables) (Dolnicar, 2002; Formann, 1984). Therefore, we chose five variables to conduct hierarchical cluster analysis: SA, RRB, overall IQ, MLU, and the number of type tokens. Then the minimal sample required should be 32, which is smaller than the current sample size (50). All numerical data were normalized before conducting cluster analysis. Finally, we calculated descriptive statistics for all measurements by subgroups and conduct ANOVA to examine the differences among subgroups. The assumptions of ANOVA, including the independence of observations, the normal distribution, and the homogeneity of variance, were examined, respectively, by using the QQ plots and the residuals versus fits plot. These assumptions were met. Tukey's HSD post-hoc test was performed to adjust for multiple comparisons.

Results

On average, each child contributed 124.96 utterances ($SD = 37.73$), which formed a reliable sample for analysis. Table 1 shows the descriptive statistics of ADOS comparison scores (ADOS), standardized IQ scores (both verbal and nonverbal), as well as language measures. Findings have shown notable variations of language abilities. MLU ranged from 1.47 to 5.17 ($M = 2.65$, $SD = 2.49$). The total number of types and tokens ranged from 24 to 224 ($M = 127.96$, $SD = 0.77$), and 36 to 900 ($M = 356.82$, $SD = 190.65$), respectively. Table 2 shows the correlation coefficients among these variables. Social affect score (SA) was highly correlated with ADOS score ($r = 0.97$), repetitive and restricted behavior (RRB) score ($r = 0.48$), overall IQ ($r = -0.47$), verbal IQ ($r = 0.41$), number of types ($r = 0.41$), and number of tokens ($r = 0.39$). MLU and number of types

Table 1 Descriptive statistics of assessments and language parameters

	Mean	Median	SD	Range
Age in months	65.26	65.50	10.33	43–100
ADOS-2				
ADOS-CS	5.5	6	2.72	1–10
SA	9.22	9	4.72	1–20
RRB	1.72	1.50	1.82	0–6
KBIT-2				
Overall IQ	94.98	93	21.60	50–147
Nonverbal IQ	93.88	96	23.05	33–142
Verbal IQ	94.76	93	21.99	55–147
Language parameters				
MLU	2.65	2.49	0.77	1.47–5.17
Token types	127.96	118.50	47.99	24–224
Token total	356.82	303.50	190.65	36–900

ADOS-2 autism diagnostic observation scale-second edition, ADOS-CS autism diagnostic observation scale-comparison scores, SA social affect, RRB restrictive and repetitive behaviors, KBIT-2 Kaufman brief intelligence test- second edition, IQ intelligence quotient, MLU mean length of utterance

also moderately correlated with overall IQ score ($r=0.31$ and 0.28 respectively).

Figure 1 shows the dendrogram illustrating the process of cases clustering, in which four-clusters were initially observed. Figure 2, the elbow method graph, shows that the four-cluster solution gained much smaller within-cluster variance than the three-cluster solution and five-cluster solution, indicating that the optimal number of subgroups was four.

The descriptive statistics for the four subgroups are shown in Table 3. Tukey’s HSD post-hoc tests were conducted to examine whether there were significant differences among the four subgroups. Figure 3 shows the radar plot that visually displays autism severity and developmental

profiles by subgroups on standardized scores. In general, there were significant main effects of subgroups on all measures including autism symptoms, $F(3,46)=27.19$, $p<0.001$, SA score, $F(3,46)=20.5$, $p<0.001$, RRB score, $F(3,46)=3.9$, $p<0.001$, overall IQ, $F(3,46)=5.672$, $p=0.002$, nonverbal IQ, $F(3,46)=3.196$, $p=0.032$, verbal IQ, $F(3,46)=3.835$, $p=0.015$, MLU, $F(3,46)=18.55$, $p<0.001$, types, $F(3,46)=29.48$, $p<0.001$, and tokens, $F(3,46)=32.38$, $p<0.001$.

Group 1 ($N=8$) was the least affected group and had relatively better language abilities than the other three subgroups. Their autism was milder than that of Group 2, $p<0.001$ and Group 4, $p<0.001$; their SA score was also lower than that of Group 2, $p<0.001$ and Group 4, $p<0.001$. Group 1 also had the highest scores on nonverbal IQ, verbal IQ, and overall IQ with the means of these scores all exceeding 110. Regarding their language profiles, Group 1 had the highest MLU (Group 1 vs. Group 2, $p<0.001$; vs. Group 3, $p<0.001$; vs. Group 4, $p=0.02$), number of types (Group 1 vs. Group 2, $p<0.001$; vs. Group 3, $p<0.001$; vs. Group 4, $p=0.004$), and number of tokens (Group 1 vs. Group 2, $p<0.001$; vs. Group 3, $p<0.001$; vs. Group 4, $p<0.001$).

In contrast with Group 1, Group 2 ($N=14$) showed the moderately severe autism symptoms and the lowest language abilities. Their ADOS comparison and SA scores were higher than those of Group 1 ($p<0.001$, and $p<0.001$, respectively) and those of Group 3 ($p<0.001$, and $p<0.001$, respectively). Their IQ scores were the lowest with the mean of the overall IQ and that of verbal IQ both around 80. However, the differences among Group 2, Group 3, and Group 4 on nonverbal IQ and verbal IQ did not reach statistically significant levels. Children in Group 2 had the lowest MLU (Group 2 vs. Group 1, $p<0.001$; vs. Group 3, $p=0.010$; vs. Group 4, $p<0.001$), number of types (Group 2 vs. Group 1, $p<0.001$; vs. Group 3, $p=0.001$; vs. Group 4, $p<0.001$), and number of tokens (Group 2 vs. Group 1, $p<0.001$; vs. Group 3, $p=0.009$; vs. Group 4, $p<0.001$).

Table 2 Correlation coefficients among variables

	ADOS-CS	SA	RRB	Overall IQ	Nonverbal IQ	Verbal IQ	MLU	Types	Tokens
ADOS-CS	1	–	–	–	–	–	–	–	–
SA	0.97***	1	–	–	–	–	–	–	–
RRB	0.68***	0.48***	1	–	–	–	–	–	–
Overall IQ	–0.32*	0.47***	–0.10	1	–	–	–	–	–
Nonverbal IQ	–0.07	–0.25	0.10	0.74***	1	–	–	–	–
Verbal IQ	–0.32*	0.41***	–0.11	0.77***	0.38**	1	–	–	–
MLU	–0.19	–0.25	–0.12	0.31*	0.24	0.29*	1	–	–
Types	–0.32*	0.41***	–0.09	0.28*	0.26	0.24	0.64***	1	–
Tokens	–0.31	0.39***	–0.10	0.23	0.30*	0.18	0.75***	0.92***	1

ADOS-CS autism diagnostic observation scale-comparison score, SA social affect, RRB restrictive and repetitive behaviors, MLU mean length of utterance

$p<0.05^*$, $p<0.01^{**}$, $p<0.001^{***}$

Fig. 1 Dendrogram of hierarchical clustering. The x-axis shows individual cases: red represents Group 1; yellow for Group 2; gray for Group 3; and blue for Group 4

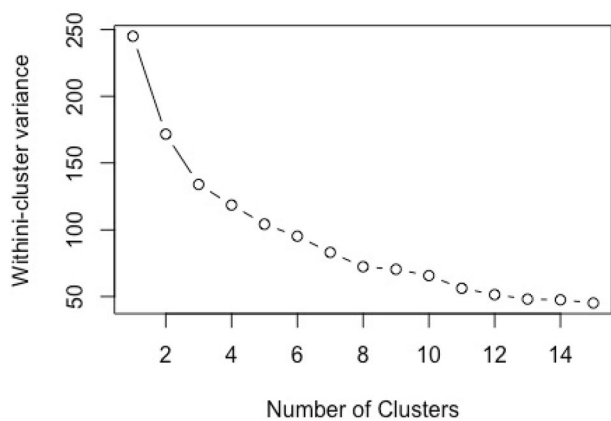
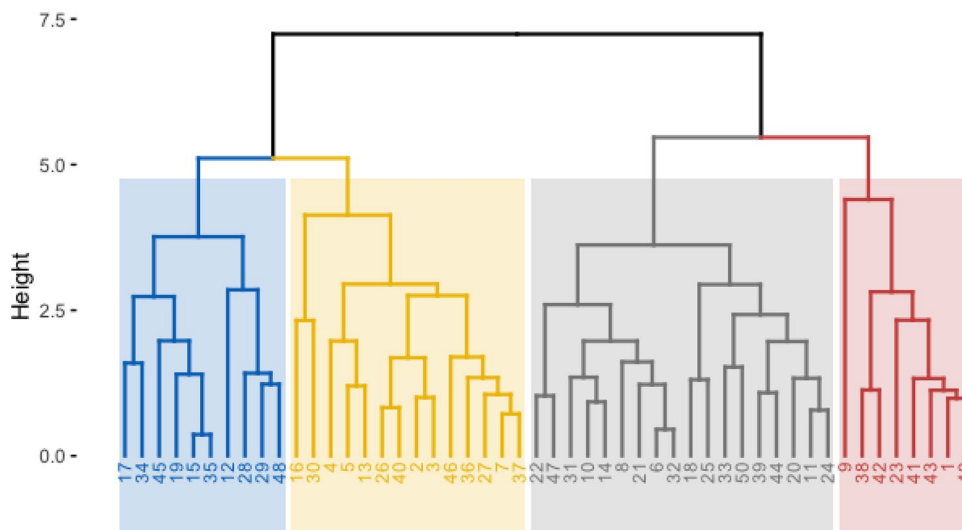


Fig. 2 Elbow method graph shows the four-cluster solution gained much smaller within-cluster variance than the three-cluster solution and five-cluster solution

Group 3 ($N=18$), the largest subgroup, showed mild autism symptoms and average levels on various language abilities. Their ADOS comparison score and SA and RRB scores were lower than those of Group 2 ($p < 0.001$, $p < 0.001$, and $p = 0.017$, respectively) and those of Group 4 ($p < 0.001$, $p < 0.001$, and $p < 0.001$, respectively). Their overall IQ was higher than that of Group 2, $p = 0.002$. Group 3 showed average levels on language abilities: their MLU, types and tokens were higher than those of Group 2 ($p = 0.010$, $p = 0.001$ and $p = 0.009$, respectively), but lower than those of Group 1 ($p = 0.001$, $p < 0.001$ and $p < 0.001$, respectively).

Group 4 ($N=10$) had the most severe autism symptoms, especially on RRB domain, and average levels on language abilities. Their ADOS comparison score and SA scores were higher than those of Group 1 ($p < 0.001$

Table 3 Z-scores of ADOS-2, IQ and language parameters in four clusters

Mean (SD)	Group 1 (N=8)	Group 2 (N=14)	Group 3 (N=18)	Group 4 (N=10)	F-value	p-value	Tukey's posthoc test ($p < 0.05$)
Age in months	63.5 (11.11)	61.86 (10.81)	67.67 (6.03)	67.1 (14.55)	1.016	0.394	/
ADOS-CS	3.38 (1.69)	7 (1.62)	3.56 (1.92)	8.6 (1.26)	27.19	<0.001	1 < 2; 1 < 4; 2 > 3; 3 < 4
SA	4.75 (2.12)	12.36 (3.82)	6.39 (3.22)	13.5 (2.8)	20.5	<0.001	1 < 2; 1 < 4; 2 > 3; 3 < 4
RRB	1.25 (1.39)	2 (1.92)	0.5 (0.79)	3.9 (1.2)	13.94	<0.001	1 < 4; 2 > 3; 2 > 4; 3 > 4
Overall IQ	111.62 (21.55)	79.93 (13.12)	101.11 (21.37)	91.7 (19.53)	5.672	0.002	1 > 2; 2 < 3
Nonverbal IQ	111 (26.86)	83.21 (14.98)	94.67 (23.24)	98.1 (16.56)	3.196	0.032	1 > 2
Verbal IQ	113 (21.45)	81.86 (16.17)	96.67 (25.27)	90.4 (19.29)	3.835	0.015	1 > 2
MLU	3.68 (0.7)	1.96 (0.35)	2.59 (0.52)	2.92 (0.64)	18.55	<0.001	1 > 2; 1 > 3; 1 > 4; 2 < 3; 2 < 4
Types	197.75 (17.04)	81.07 (30.19)	122.39 (26.55)	147.8 (37.65)	29.48	<0.001	1 > 2; 1 > 3; 1 > 4; 2 < 3; 2 < 4
Tokens	668.62 (146.6)	190.36 (91.48)	322.5 (104.64)	402.2 (119.31)	32.28	<0.001	1 > 2; 1 > 3; 1 > 4; 2 < 3; 2 < 4

ADOS-CSS autism diagnostic observation scale-comparison score, SA social affect, RRB restrictive and repetitive behaviors, MLU mean length of utterance

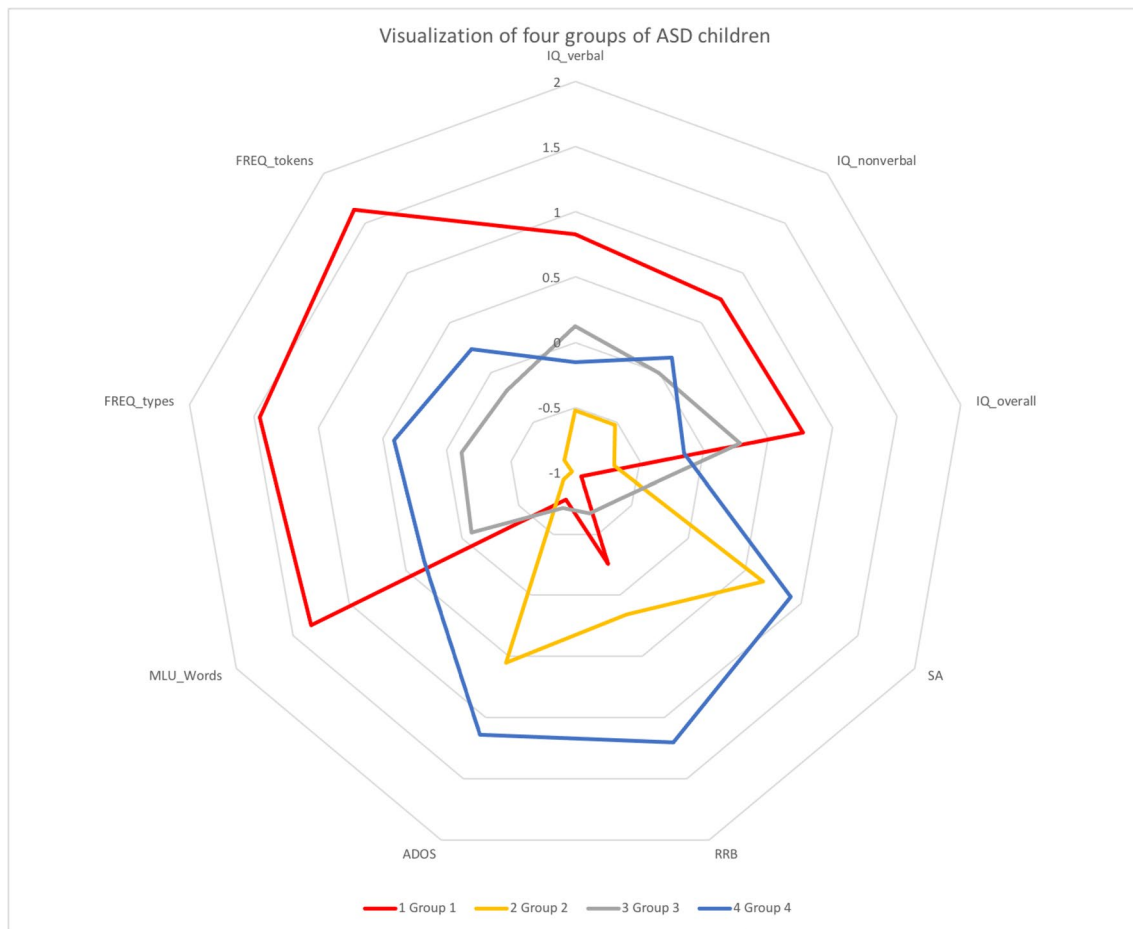


Fig. 3 Radar plot for the means of z-scores for different variables in each group: red represents Group 1; blue Group 2; green Group 3; and magenta Group 4

Table 4 Mean number of total utterances and mean percentage of different types of lexical components in four clusters

Mean (SD)	Group 1 (N=8)	Group 2 (N=14)	Group 3 (N=18)	Group 4 (N=10)	F-value	p-value	Tukey’s posthoc test (p<0.05)
Total utterance	157.75 (23.77)	122 (34.9)	129.72 (54.39)	89.3 (37.84)	6.702	<0.001	1 > 2; 1 > 3; 2 > 4
Nouns (in %)	14.69 (5.26)	19.29 (7.75)	25.99 (9.97)	21.03 (7.35)	5.024	0.004	1 < 2; 2 < 4
Verbs (in %)	19.87 (4.62)	21.38 (6.63)	21.64 (7.47)	18.81 (2.83)	0.298	0.826	/
Prepositions (in %)	0.8 (0.57)	1.36 (0.59)	1.04 (1.01)	1.53 (1.72)	0.84	0.479	/
Adjectives (in %)	2.81 (2.14)	4.36 (4.92)	3.79 (2.76)	6.38 (4.56)	1.002	0.401	/
Adverbs (in %)	7.52 (3.01)	8.31 (2.86)	4.97 (3.34)	6.32 (2.6)	2.215	0.099	/
Pronouns (in %)	9.9 (3.81)	7.54 (3.09)	5.2 (3.97)	4.32 (2.67)	4.315	0.009	1 > 2; 2 > 4

and $p < 0.001$, respectively), and Group 3 ($p < 0.001$ and $p < 0.001$, respectively), and comparable to those of Group 2. Their RRB score was the highest than that of Group 1, $p < 0.001$, Group 2, $p = 0.008$, and Group 3, $p < 0.001$. Children in Group 4 did not differ from the other three subgroups on IQ scores. Their average overall IQ was around 90. Similar to Group 1, Group 3 showed average

levels on language abilities: their MLU, types and tokens were higher than those of Group 2 ($p = 0.001$, $p < 0.001$ and $p < 0.001$, respectively), but lower than those of Group 1 ($p = 0.021$, $p = 0.004$ and $p < 0.001$, respectively).

The profiles of detailed lexical components divided by each group were shown in Table 4. Children in Group 1 produced more sentences during the parent–child interaction

with their number of total utterances being higher than that of Group 2, $p < 0.001$, and Group 3, $p = 0.04$. Verbs and nouns were the two most commonly used lexicons in all groups, contributing to approximately 40% of all lexical components. Children of all four groups used relatively few prepositions, adjectives, adverbs, and pronouns (less than 10% on each components). There were no significant differences in the percentages of verbs, $F(3, 46) = 0.298$, $p = 0.826$, prepositions, $F(3, 46) = 0.84$, $p = 0.479$, adjectives, $F(3, 46) = 1.002$, $p = 0.401$, and adverbs, $F(3, 46) = 2.215$, $p = 0.099$ among the four groups, but children in Group 1 used fewer nouns, $p = 0.008$ and more pronouns, $p = 0.041$, than those in Group 2.

Discussion

This research identified four distinct subgroups among 50 Cantonese-speaking children based on their language outputs derived from parent–child interactions, and autism severity and cognitive functioning assessed using standardized assessments. Compared to other subgroups, Group 1, the least affected group, had the highest IQ, mild autism, and strongest verbal abilities. Group 2, the most severely affected group, had the lowest IQ, most severe autism, and weakest verbal abilities. Group 3 and Group 4 displayed average levels of verbal abilities and IQ. However, Group 4 had moderately severe autism than Group 3.

Based on analyses using the CLAN program, we conducted hierarchical cluster analyses and defined four subgroups. Hierarchical cluster analyses are more objective in defining clinically meaningful subgroups than the manual stratification used in Wittke et al. study (2017). Dendrogram initially observed four subgroups, which was then verified by the elbow method. Both measures reached a consensus that four subgroups should be identified. This result is reported in some of the previous studies (Eaves et al., 1994; Hu & Steinberg, 2009; Wiggins et al., 2017). In one study, Eaves et al. (1994) identified four clinically meaningful subtypes, which showed distinct differences in behavioral and cognitive areas. The biggest group showed impairments in verbal and nonverbal social communication and mild intellectual disabilities. The second largest group had autism symptoms similar to the first group but had more severe intellectual disabilities. The other two subgroups had better intellectual functioning and mild autism. Nevertheless, other studies defined three subgroups only (Cholemkey et al., 2016; Fein et al., 1999; Georgiades et al., 2013; Wittke et al., 2017; Zheng et al., 2020). That being said, it is difficult to merge the language profiles of our subgroups with those of subgroups identified in the previous studies. This is not surprising because assessments of language abilities are different across studies. We collected naturalistic language samples

derived from spontaneous parent–child interactions while Wittke et al. (2017) collected the language samples from a structured assessment and Zheng et al. (2020) administered standardized language tests. Besides, the present study targeted children aged four to eight years of age while Zheng focused on preschoolers and Cholemkey et al. (2016) on individuals with autism ranging from toddlers to adults. Meta-analyses should be conducted in order to sort out the similarities and differences of the language profiles characterized in these studies.

Regarding children with different severity levels of autism in this study, their social affect covaried with restrictive and repetitive behaviors (see also Zheng et al., 2020). Children who had severe social communication deficits also had severe repetitive behaviors. Their autism fell along the spectrum ranging from mild (Group 1 and Group 3) to severe (Group 2 and Group 4). Only Group 1 and Group 3 showed mixed patterns, but their difference did not reach statistically significant. Yet, some previous studies found dissociations between social affect and restrictive and repetitive behaviors (Cholemkey et al., 2016; Georgiades et al., 2013). The controversial results shed lights on future studies, which should further explore the relationship between social affect domain and repetitive and restrictive domain by collecting data from different measurements and diverse samples from different cultures and age groups.

In line with the previous researches, our study also found two subgroups (Group 2 and Group 4) exhibited severe autism symptoms but had different levels on language abilities and cognitive abilities. For example, Eaves et al. (1994) identified two subgroups with similar autism symptoms, but their cognitive ability differed, and Zheng et al. (2020) also reported two subgroups with comparable levels of autism symptoms were characterized as different cognitive and language abilities. Likewise, Group 3 and Group 4 in our study, showing comparable mild language impairments, exhibited different levels of autism symptoms, which was similar to previous findings (Klopper et al., 2017). Our study also did not report a straightforward relationship between IQ and language abilities and autism symptoms in children with autism. Only Group 1 and Group 2 differed significantly on nonverbal and verbal IQ: the group with higher IQ (Group 1) had better verbal skills and milder autism symptoms than the one with lower IQ (Group 2) (Eagle et al., 2010; Fein et al., 1999). This finding is similar to those of previous studies (Ellis Weismer & Kover, 2015; Mougá et al., 2020).

Our findings shed light on the heterogeneity of language abilities with regard to the use of lexical components. To date, very few studies identified autism subgroups according to detailed lexical components (Bacon et al., 2019; Wittke et al., 2017). In general, the two most commonly used words found in the present language samples were verbs and nouns (approximately 40% of the total lexical components).

Relatively fewer prepositions, adjectives, adverbs, and pronouns (around 10%) were produced. There were no significant differences in the use of lexical components across four subgroups on the proportions of verbs, prepositions, adjectives, and adverbs produced by these children. These findings were consistent with the previous studies, which also found that children with autism used similar frequency of types of words compared to matched typically developing children (Eigsti et al., 2007; Park et al., 2012).

Yet, Group 1 used fewer nouns and more pronouns than Group 2. Group 1 also had the highest MLU, number of types, and number of tokens, in comparison to other groups. These findings may suggest that children with better language abilities in general or grammar specifically would be more capable of producing pronouns than those with weaker language skills. Understanding how children with autism produce lexical components in spontaneous conversation would provide clinicians and caregivers with valuable insights in designing interventions for improving language use in social contexts. Clinicians or therapists may design interventions for those with weaker language skills for improving their use of pronouns. However, the current study focused on the percentage of the use of pronouns and did not examine whether their pronouns were properly used. Correct use of pronouns may relate to the theory of mind understanding (Niemic, 2007). In future study, we should examine the proper use of pronouns and its relation to theory of mind understanding.

Despite the fact that our study was pioneering research into the heterogeneity of language abilities in Chinese-speaking children, it has a few limitations. First, our language samples were collected from 50 children only. Cluster analysis would benefit from a larger sample size. Related to this, while Kover et al. (2014) reported that children with ASD produce more utterances with parents than with examiners, some parents, especially those with ASD as well or having ASD traits, may have difficulties interacting with their children, which in turn may influence the children's language production. Therefore, we are currently collecting ASD children's language samples from their interactions with parents as well as teachers in order to more comprehensively and accurately evaluate these children's language abilities. Secondly, this study included children aged three to eight years of age. Heterogeneity in language abilities and autism-related traits were observed in individuals with autism along their lifespan (Fountain et al., 2012; Pickles et al., 2014). The rate and trajectories of language development are highly variable up to 6 years of age but are becoming relatively stable after 6 (Pickles et al., 2014). This study therefore included children from both variable and stable periods and yet considered the effects of developmental maturation and chronological age in the heterogeneity in language abilities.

Thirdly, our study only focused on lexical components and did not examine other aspects of language. Further studies should be conducted to evaluate other language domains including phonology, syntax, and pragmatics. We are currently investigating how children with ASD represent temporal concepts in Cantonese through using aspect markers (Tse et al., 2012). We are also studying the heterogeneity of gesture production as previous findings have shown that gesture development is delayed in individuals with autism (Bono et al., 2004; Mastrogiuseppe et al., 2015) and there are individual variations in gestural recognition and production (So et al., 2016).

Conclusion

To the best of our knowledge, our study could be the first to identify subgroups of children with autism using multiple standardized tests and naturalistic language samples derived from parent–child interactions in Hong Kong. Our language samples also successfully reflected heterogeneity of language abilities in children with different levels of autism severity and intellectual functioning. Further studies investigating the heterogeneity of language impairments in autism should be based on naturalistic language samples collected in spontaneous interactions. The results highlighted the heterogeneity of phenotypical presentations in children with autism and the significance of applying multiple measurements. Classification is an effective method for researchers to better understand the behavior and abilities for individuals with autism. Although delineating subgroups is beneficial for us to see the distinctions between different subgroups, the distinct profile of each child was covered by merging them into a more homogeneous subgroup. It is necessary to consider differences and profiles from individual levels when conducting intervention and study (Lai et al., 2013). Future research should track the developmental trajectory or etiology of autism spectrum disorder of these clusters.

Acknowledgments We acknowledge the help of our research assistants Chi-Chung Chan, Chun-Ho Cheng, Johnny Fung, Fai-Yeung Kwok, Wing-Wun Law, Shing-Hey Lee, Ying-Yi Lee, Oi-Ki Leung, Jonathan Tse, Ching-Yi Wong, Tiffany Wong, and Yan-Yan Yip with data collection and transcription. Special thanks to all of the children and their parents for their help and dedication to education. This research has been fully supported by a grant from the Innovation and Technology Fund for Better Living (“FBL”) (Project no. ITB/FBL/8005/17/P). The authors declare that they have no conflict of interest. All procedures performed in studies involving human participants were conducted in accordance with the ethical standards of the institutional research committee (Survey and Behavioral Research Ethics Reference no. SBRE-19-307) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Author Contributions XKS analyzed the data and wrote the article. CL analyzed the data. WCSO designed the study and wrote the article.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5[®]). American Psychiatric Publishing, Washington
- Bacon, E. C., Osuna, S., Courchesne, E., & Pierce, K. (2019). Naturalistic language sampling to characterize the language abilities of 3-year-olds with autism spectrum disorder. *Autism, 23*(3), 699–712.
- Bal, V. H., Katz, T., Bishop, S. L., & Krasileva, K. (2016). Understanding definitions of minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 57*(12), 1424–1433.
- Barokova, M., & Tager-Flusberg, H. (2020). Commentary: Measuring language change through natural language samples. *Journal of Autism and Developmental Disorders, 50*(7), 2287–2306.
- Beglinger, L. J., & Smith, T. H. (2001). A review of subtyping in autism and proposed dimensional classification model. *Journal of Autism and Developmental Disorders, 31*(4), 411–422.
- Bono, M. A., Daley, T., & Sigman, M. (2004). Relations among joint attention, amount of intervention and language gain in autism. *Journal of Autism and Developmental Disorders, 34*(5), 495–505.
- Brown, R. (1973). *A first language: The early stages*. Cambridge: Harvard University Press.
- Casenhiser, D. M., Binns, A., McGill, F., Morderer, O., & Shanker, S. G. (2015). Measuring and supporting language function for children with autism: Evidence from a randomized control trial of a social-interaction-based therapy. *Journal of Autism and Developmental Disorders, 45*(3), 846–857.
- Casenhiser, D. M., Shanker, S. G., & Stieben, J. (2013). Learning through interaction in children with autism: Preliminary data from a social-communication-based intervention. *Autism, 17*(2), 220–241.
- Chiang, H.-M. (2009). Differences between spontaneous and elicited expressive communication in children with autism. *Research in Autism Spectrum Disorders, 3*(1), 214–222.
- Cholemky, H., Medda, J., Lempp, T., & Freitag, C. M. (2016). Classifying autism spectrum disorders by ADI-R: Subtypes or severity gradient? *Journal of Autism and Developmental Disorders, 46*(7), 2327–2339.
- 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.
- Dolnicar, S. (2002). *A review of unquestioned standards in using cluster analysis for data-driven market segmentation*. Australia: University of Wollongong.
- Eagle, R. F., Romanczyk, R. G., & Lenzenweger, M. F. (2010). Classification of children with autism spectrum disorders: A finite mixture modeling approach to heterogeneity. *Research in Autism Spectrum Disorders, 4*(4), 772–781.
- Eaves, L. C., Ho, H. H., & Eaves, D. M. (1994). Subtypes of autism by cluster analysis. *Journal of Autism and Developmental Disorders, 24*(1), 3–22.
- Eigsti, I.-M., Bennetto, L., & Dadlani, M. B. (2007). Beyond pragmatics: Morphosyntactic development in autism. *Journal of Autism and Developmental Disorders, 37*(6), 1007–1023.
- Ellis Weismer, S., & Kover, S. T. (2015). Preschool language variation, growth, and predictors in children on the autism spectrum. *Journal of Child Psychology and Psychiatry, 56*(12), 1327–1337.
- Fein, D., Stevens, M., Dunn, M., Waterhouse, L., Allen, D., Rapin, I., & Feinstein, C. (1999). Subtypes of pervasive developmental disorder: Clinical characteristics. *Child Neuropsychology, 5*(1), 1–23.
- Fernell, E., Hedvall, Å., Norrelgen, F., Eriksson, M., Höglund-Carlsson, L., Barnevik-Olsson, M., & Gillberg, C. (2010). Developmental profiles in preschool children with autism spectrum disorders referred for intervention. *Research in Developmental Disabilities, 31*(3), 790–799.
- Formann, A. K. (1984). *Die latent-class-analyse: Einführung in Theorie und Anwendung*. Germany: Beltz.
- Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics, 129*(5), e1112–e1120.
- Georgiades, S., Szatmari, P., Boyle, M., Hanna, S., Duku, E., Zwaigenbaum, L., & Mirenda, P. (2013). Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach. *Journal of Child Psychology and Psychiatry, 54*(2), 206–215.
- Hu, V. W., & Steinberg, M. E. (2009). Novel clustering of items from the Autism diagnostic interview-revised to define phenotypes within autism spectrum disorders. *Autism Research, 2*(2), 67–77.
- Jyotishi, M., Fein, D. A., & Naigles, L. R. (2017). Didn't I just say that? Comparing parent report and spontaneous speech as indicators of grammatical development. *Research in Developmental Disabilities, 61*, 32–43.
- 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.
- Kaufman, A. S. (2004). *Kaufman brief intelligence test—second edition (KBIT-2)*. Circle Pines, MN: American Guidance Service.
- Klopper, F., Testa, R., Pantelis, C., & Skafidas, E. (2017). A cluster analysis exploration of autism spectrum disorder subgroups in children without intellectual disability. *Research in Autism Spectrum Disorders, 36*, 66–78.
- Kover, S. T., Davidson, M. M., Sindberg, H. A., & Weismer, S. E. (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, 57*(6), 2221–2233.
- Lai, M.-C., Lombardo, M. V., Chakrabarti, B., & Baron-Cohen, S. (2013). Subgrouping the autism “spectrum”: Reflections on DSM-5. *PLoS Biology, 11*(4), e1001544.
- Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule—2nd edition (ADOS-2)*. Los Angeles, CA: Western Psychological Corporation.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk transcription format and programs*. (Vol. 1). London: Psychology Press.
- Maenner, M. J., Shaw, K. A., & Baio, J. (2020). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries, 69*(4), 1.
- Masi, A., DeMayo, M. M., Glozier, N., & Guastella, A. J. (2017). An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience Bulletin, 33*(2), 183–193.

- Mastrogiuseppe, M., Capirci, O., Cuva, S., & Venuti, P. (2015). Gestural communication in children with autism spectrum disorders during mother–child interaction. *Autism, 19*(4), 469–481.
- Matthews, S., & Yip, V. (1994). *Cantonese*. London: Routledge.
- Mouga, S., Correia, B. R., Café, C., Duque, F., & Oliveira, G. (2020). Language predictors in autism spectrum disorder: Insights from neurodevelopmental profile in a longitudinal perspective. *Journal of Abnormal Child Psychology, 48*(1), 149–161.
- Niemiec, V. (2007). *Connections between theory of mind and pronoun competence: Typically-developing children and children with autism spectrum disorders* Doctoral dissertation. USA: The Ohio State University.
- Park, C. J., Yelland, G. W., Taffe, J. R., & Gray, K. M. (2012). Morphological and syntactic skills in language samples of pre school aged children with autism: Atypical development? *International Journal of Speech-Language Pathology, 14*(2), 95–108.
- 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.
- Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3–9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research, 53*(2), 333–349.
- Ring, H., Woodbury-Smith, M., Watson, P., Wheelwright, S., & Baron-Cohen, S. (2008). Clinical heterogeneity among people with high functioning autism spectrum conditions: Evidence favouring a continuous severity gradient. *Behavioral and Brain Functions, 4*(1), 11.
- Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics, 12*(1), 23–45.
- Shriberg, L. D., Kwiatkowski, J., & Hoffmann, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech Language and Hearing Research, 27*(3), 456–465.
- So, W. C. (2020). *Robot for autism behavioral intervention*. Hong Kong: Chinese University of Hong Kong.
- So, W. C., Wong, M. Y., Cabibihan, J. J., Lam, C. Y., Chan, R. Y., & Qian, H. H. (2016). Using robot animation to promote gestural skills in children with autism spectrum disorders. *Journal of Computer Assisted Learning, 32*(6), 632–646.
- Su, Y., & Naigles, L. R. (2019). Online processing of subject–verb–object order in a diverse sample of mandarin-exposed preschool children with autism spectrum disorder. *Autism Research, 12*(12), 1829–1844.
- Su, Y., Naigles, L. R., & Su, L. Y. (2018). Uneven expressive language development in mandarin-exposed preschool children with ASD: Comparing vocabulary, grammar, and the decontextualized use of language via the PCIDI-toddler form. *Journal of Autism and Developmental Disorders, 48*(10), 3432–3448.
- Sun, X., Allison, C., Matthews, F. E., Sharp, S. J., Auyeung, B., Baron-Cohen, S., & Brayne, C. (2013). Prevalence of autism in mainland China, Hong Kong and Taiwan: A systematic review and meta-analysis. *Molecular Autism, 4*(1), 7.
- Sun, X., Allison, C., Wei, L., Matthews, F. E., Auyeung, B., Wu, Y. Y., & Brayne, C. (2019). Autism prevalence in China is comparable to Western prevalence. *Molecular Autism, 10*(1), 7.
- Syriopoulou-Delli, C. K., & Papaefstathiou, E. (2020). Review of cluster analysis of phenotypic data in autism spectrum disorders: Distinct subtypes or a severity gradient model? *International Journal of Developmental Disabilities, 66*(1), 13–21.
- Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. *Handbook of Autism and Pervasive Developmental Disorders, 1*, 335–364.
- Tardif, T., & Fletcher, P. (2008). *Chinese communicative development inventories: User's guide and manual*. Beijing: Peking University Medical Press.
- Tek, S., Mesite, L., Fein, D., & Naigles, L. (2014). Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(1), 75–89.
- Tse, S. K., Li, H., & Leung, S. O. (2012). Tense and temporality: How young children express time in Cantonese. *Chinese Language and Discourse, 3*(1), 35–56.
- Wiggins, L. D., Robins, D. L., Adamson, L. B., Bakeman, R., & Henrich, C. C. (2012). Support for a dimensional view of autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders, 42*(2), 191–200.
- Wiggins, L. D., Tian, L. H., Levy, S. E., Rice, C., Lee, L. C., Schieve, L., & Thompson, W. (2017). Homogeneous subgroups of young children with autism improve phenotypic characterization in the study to explore early development. *Journal of Autism and Developmental Disorders, 47*(11), 3634–3645.
- Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., & Naigles, L. R. (2017). Grammatical language impairment in autism spectrum disorder: Exploring language phenotypes beyond standardized testing. *Frontiers in Psychology, 8*, 532.
- Yi, L., Fan, Y., Zhao, J., Huang, D., Li, Y., & Zou, X. (2013). Atypical understanding of mental terms in Chinese-speaking children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 7*(11), 1411–1417.
- Zheng, S., Hume, K. A., Able, H., Bishop, S. L., & Boyd, B. A. (2020). Exploring developmental and behavioral heterogeneity among preschoolers with ASD: A cluster analysis on principal components. *Autism Research, 13*, 796–809.
- Zhou, P., Crain, S., Gao, L., Tang, Y., & Jia, M. (2015). The use of grammatical morphemes by Mandarin-speaking children with high functioning autism. *Journal of Autism and Developmental Disorders, 45*(5), 1428–1436.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.