Empirical Article

Clinical Psychological Science
1–19

DOI: 10.1177/2167702618809353
www.psychologicalscience.org/CPS
SAGE

Quantitative Analysis of Heterogeneity in Academic Achievement of Children With Autism

Lang Chen[1], **Daniel A. Abrams**[1], **Miriam Rosenberg-Lee**[1,2], **Teresa Iuculano**[1], **Holly N. Wakeman**[1,3], **Sandhya Prathap**[1], **Tianwen Chen**[1], **and Vinod Menon**[1,4,5]

[1]Department of Psychiatry and Behavioral Sciences, Stanford University; [2]Department of Psychology, Rutgers University; [3]Department of Psychology, University of Colorado at Boulder; [4]Department of Neurology and Neurological Sciences, Stanford University; and [5]Symbolic Systems Program in Neuroscience, Stanford University

Abstract

Autism spectrum disorders (ASD) represent a quintessential example of a clinical population with diverse symptom presentations and marked variation in cognitive abilities. However, the extensive literature lacks rigorous quantitative procedures for characterizing heterogeneity of cognitive abilities in these individuals. Here we employ novel clustering and cross-validation procedures to investigate the stability of heterogeneous patterns of cognitive abilities in reading and math in a relatively large sample (N = 114) of children with ASD and matched controls (N = 96). Our analysis revealed a unique profile of heterogeneity in ASD, consisting of a low-achieving subgroup with poor math skills compared with reading and a high-achieving subgroup who showed superior math skills compared with reading. Verbal and central executive working memory skills further differentiated these subgroups. Findings provide insights into distinct profiles of academic achievement in children with ASD, with implications for educational practice and intervention, and provide a novel framework for quantifying heterogeneity in the disorder.

Keywords

autism, academic achievement, childhood development, individual differences, working memory

Received 10/17/17; Revision accepted 8/6/18

Individuals with neurodevelopmental disorders, including autism spectrum disorders (ASD), often display a wide range of symptom presentations, cognitive abilities, and behavioral profiles (Seelaar, Rohrer, Pijnenburg, Fox, & van Swieten, 2011; Smellie, 2006; Zetusky, Jankovic, & Pirozzolo, 1985). Characterizing these heterogeneous patterns represents a crucial question: Identification of distinct subgroups may provide important theoretical information regarding how underlying clinical impairments relate to prominent manifestations of disorders and may also provide critical information to facilitate targeted remediation of clinical symptoms within a specific subgroup.

ASD represents a quintessential example of a clinical population with diverse symptom presentations and marked variation in cognitive and behavioral abilities. For example, heterogeneity of symptom presentations in ASD is well established and includes a high degree of variability in multiple core symptom domains, including social communication (Georgiades et al., 2013; Hu & Steinberg, 2009), language function (Kjelgaard & Tager-Flusberg, 2001), and repetitive and restricted behaviors (Fountain, Winter, & Bearman, 2012; Gotham, Pickles, & Lord, 2009, 2012). Significant heterogeneity in ASD has also been shown in other behavioral domains, including motor (Jansiewicz et al., 2006), cognitive (Norbury & Nation, 2011), and clinical (Ring,

Corresponding Authors:
Lang Chen, Department of Psychiatry & Behavioral Sciences, 401 Quarry Road, Stanford, CA 94305-5719
E-mail: lchen32@stanford.edu

Vinod Menon, Department of Psychiatry & Behavioral Sciences, 401 Quarry Road, Stanford, CA 94305-5719
E-mail: menon@stanford.edu

Woodbury-Smith, Watson, Wheelwright, & Baron-Cohen, 2008). Together, mounting evidence supports the hypothesis that heterogeneity is a prominent feature of ASD measured across a wide range of domains, and therefore, a critical goal for autism research is to characterize this heterogeneity.

Academic achievement is an important domain in which children with ASD have shown considerable heterogeneity (Assouline, Nicpon, & Dockery, 2012; Baron-Cohen, Wheelwright, Burtenshaw, & Hobson, 2007; Cash, 1999; Chiang & Lin, 2007; H. Kim & Cameron, 2016). Math and reading skills are two key areas of academic achievement in children with ASD, and characterizing heterogeneity in these areas is an important question for several reasons. First, children with ASD are increasingly included in mainstream classrooms, and therefore, acquiring these foundational academic skills is critical for keeping up with their peers. Second, developing educational interventions for children with ASD (e.g., Barnett & Cleary, 2015) requires a thorough understanding of academic achievement profiles within this population. Third, proficiency in these primary academic areas has practical implications for future independent living and professional careers (Newman et al., 2011; Troyb et al., 2014).

In the context of academic achievement, results from previous studies of reading have been equivocal regarding consistent and robust patterns of abilities in children with ASD. Studies of word reading have revealed inconsistent results, with reports of both preserved and deficient skills depending on the cutoffs and comparisons used (Jones et al., 2009; Mayes & Calhoun, 2003; Nation, Clarke, Wright, & Williams, 2006; Troyb et al., 2014; Wei, Christiano, Yu, Wagner, & Spiker, 2015). In one study, a small sample (N = 32) of children and adolescents with ASD aged 6 to 16 showed average word reading skills when reading abilities were measured across the entire group and compared with standardized scores (Nation et al., 2006). While the authors of this work noted significant variability in word reading ability within this sample, they did not quantify this aspect of the distribution. Additionally, two studies examined word reading ability relative to full-scale IQ (FSIQ) in children and adolescents with ASD over an extended age range (6–18 years old; Jones et al., 2009; S. H. Kim, Bal, & Lord, 2017). These studies revealed that 75% to 99% of affected individuals showed word reading ability that was comparable or higher than predicted by FSIQ, suggesting preserved reading abilities in the majority of children and adolescents with ASD. However, group mean word reading ability in individuals with ASD is often below average both in 3- and 7-year-old children (low IQ group = 81 in Mayes & Calhoun, 2003) and older children and adolescents ages 6 to 18 years (S. H. Kim et al., 2017, low IQ groups = 68.5 and 64.1; Jones et al., 2009, low IQ group = 85.2), and therefore, the use of deviance scores may mask severe word reading impairments in a substantial proportion of affected individuals. While these studies have suggested preserved word reading ability in individuals with ASD, a recent study revealed weaknesses in word reading ability in over half of the children with ASD ages 6 to 9 years (Wei et al., 2015). Specifically, results showed that two distinct subgroups of children with ASD, including "lower-achieving" and "hypercalculic" children, who accounted for 52% of the sample, had reading abilities that fell below the average range.

A second aspect of reading abilities that has shown inconsistencies in the literature involves the relationship between word reading and reading comprehension abilities. Specifically, some studies have shown that single word decoding skills are generally preserved or even fall into the superior range in individuals with ASD, whereas reading comprehension skills are frequently impaired in these children (Jones et al., 2009; Nation et al., 2006). For example, in a sample of children and adolescents with ASD, group mean reading comprehension standardized scores were 14 points lower than word reading scores, and half of the individuals who showed normal word reading also revealed impaired reading comprehension (Nation et al., 2006). However, this pattern of abilities across different aspects of reading has not always been replicated. For example, a recent study revealed that substantial discrepancies (> 10 points) between word reading and reading comprehension scores were present in only a small subgroup of children with ASD (Wei et al., 2015). Similar findings of comparable word reading and reading comprehension abilities have been demonstrated in an early study with a group of 42 children and adolescents with autism (Mayes & Calhoun, 2003). Together, previous studies of reading abilities in individuals with ASD have failed to converge on a consistent pattern of abilities associated with word reading and reading comprehension and suggest that systematic and rigorous quantitative approaches to characterizing heterogeneity of reading abilities may be needed to address this important aspect of academic learning.

In contrast to reading, the study of mathematical achievement is an emerging area of academic research in children with ASD; however, results from initial studies examining patterns of abilities in the math domain have been inconsistent. Building on anecdotal (Baron-Cohen et al., 2007) and experimental accounts of savant skills in ASD (Howlin, Goode, Hutton, & Rutter, 2009), several studies have identified a subgroup of children with ASD who show relative strengths in math abilities (Chiang & Lin, 2007; Jones et al., 2009; Wei et al., 2015) that is distinct from a large, majority subgroup that shows average math abilities. More recent studies of individuals with ASD, however, have shown a high degree of variability in math skills and suggest that weaknesses in math skills may be more prominent than giftedness in this population (Keen, Webster, & Ridley, 2016; Oswald et al., 2016; Titeca, Roeyers, Loeys, Ceulemans, & Desoete, 2015). A critical gap in this literature is an understanding of whether children with ASD show a unitary pattern of strengths and weaknesses across multiple mathematics subtests, including calculation and problem solving/reasoning, or whether heterogeneity in the ASD population is manifested by distinct subgroups of children with consistent profiles of mathematical abilities.

The lack of converging results in previous studies of both reading and mathematical abilities in children with ASD may be affected by the inconsistency of methodological approaches that have been used to address heterogeneity of academic abilities in these individuals. In one approach, heterogeneity in ASD has been examined by identifying the proportion of children with ASD who have either exceptional or impaired abilities in math or reading. The rationale for this approach is that academic "outliers," including both gifted children (e.g., hyperlexic and hypercalculic) as well as children with impairments (e.g., dyslexic and dyscalculic), may be disproportionately represented in the ASD population and characterizing the prevalence of these children highlights an important aspect of heterogeneity (Estes, Rivera, Bryan, Cali, & Dawson, 2011). Studies using this approach have reported a range of inconsistent results. For example, one study reported that 72% of adolescents with ASD had at least one measure of academic achievement that was discrepant from their general intellectual ability, including 14% with enhanced math abilities and 7% with lower than expected math skills (Jones et al., 2009). A second study reported that 4% of children with ASD had enhanced math abilities, whereas 22% of these children had a mathematical learning impairment (Oswald et al., 2016). A longitudinal study that also applied this approach revealed that the discrepancy rate of academic achievement measures varies from 1% to 69%, depending on the age, measure, and general cognitive abilities (i.e., IQ) of affected groups (S. H. Kim et al., 2017).

A second approach for characterizing heterogeneity of academic abilities in ASD involves using data-driven clustering methods for identifying subgroups of children who have similar academic learning profiles. The rationale for this approach is that by using an unbiased, data-driven method, distinct subgroups of children can be identified on the basis of multiple academic measures, thereby providing a more detailed account of abilities and disabilities across a range of academic areas. One previous study employed this approach with a sample of children with ASD and identified four distinct achievement profiles, including two small groups with isolated strengths in a particular academic area (i.e., hyperlexia and hypercalculia) as well as two larger groups who showed either a high or a low achievement profile across both reading and math domains (Wei et al., 2015).

However, extant studies using both approaches have critical weaknesses for properly characterizing heterogeneous patterns of academic skills in ASD, which may have contributed to the lack of converging evidence in previous studies. A limitation of the cutoff approach is that because of its arbitrary nature, it is unable to identify distinct subgroups of children with unique profiles across multiple academic measures. Additionally, a limitation of the data-driven clustering approach is that cross-validation procedures have not been used, so it is unknown whether the identified subgroups are stable and characteristic of the greater population of children with ASD. Many studies (Jones et al., 2009; Mayes & Calhoun, 2003; for a review, see Whitby & Mancil, 2009) have also used participants spanning a wide age range, which can be problematic for a neurodevelopmental disorder such as ASD. Moreover, most previous studies have not included an age- and IQ-matched control group to examine whether heterogeneous patterns of academic achievement identified in children with ASD are unique to this population. Finally, little is known regarding sources of heterogeneity in academic achievement in children with ASD. One longitudinal study identified environmental (e.g., schooling type, parent participation in intervention) and cognitive factors (e.g., IQ) that contribute to individual differences in academic achievement. However, additional cognitive factors, such as working memory, that are important academic skill acquisitions were not examined and may have further contributed to heterogeneous patterns of abilities in ASD. For example, previous research has shown that academic achievement in typically developing (TD) children is strongly associated with working memory (Alloway & Passolunghi, 2011; Bull & Scerif, 2001; Meyer, Salimpoor, Wu, Geary, & Menon, 2010), with visuospatial and verbal working memory predicting math and reading achievement, respectively. Importantly, it is unknown whether cognitive measures such as working memory or clinical symptoms of autism and behavioral-affective traits might account for heterogeneous patterns of academic achievement in children with ASD.

Our study had three major goals. The first goal was to apply rigorous clustering procedures and cross-validation to identify distinct subgroups within a well-characterized cohort of children with ASD on the basis of academic abilities. We employed an analytic protocol that involved: (a) using the Gap Index to determine whether heterogeneity exists in this population

(Tibshirani, Walther, & Hastie, 2001), (b) identifying the optimal number of clusters on the basis of the aggregation of 30 interval validation indices from the NbClust R package (Charrad, Ghazzali, Boiteau, & Niknafs, 2014), and (c) computing cross-validation measures to determine the reliability and consistency of our optimal clustering solution. Importantly, the rigorous and data-driven clustering methods used in the current study enable specific predictions regarding the nature of the distributions of academic achievement in children with ASD. Specifically, if the clustering methods identify a single group, this would strongly suggest a continuum between lower and higher achieving children across academic domains. Alternatively, if these methods identify multiple clusters of children, this would strongly suggest distinct subgroups of children with ASD who are distinguished by specific features of their academic achievement profiles. The second goal of our study was to investigate whether distinct profiles of academic achievement identified in children with ASD are unique to these children. We therefore compared clustering results identified in children with ASD with results identified in a group of well-matched TD children. The third goal of the study was to examine unique factors that contribute to heterogeneous patterns of academic achievement in children with ASD, with a focus on core ASD-related symptoms, behavioral-affective traits, and cognitive factors, particularly distinct aspects of working memory. These analyses were designed to test four primary hypotheses: (a) A rigorous data-driven approach would reveal distinct patterns of academic achievement in ASD; (b) subgroups of children with ASD would be more strongly distinguished by math than reading abilities (Jones et al., 2009); (c) clinical symptoms of ASD, behavioral-affective traits, and cognitive factors would significantly differ across distinct ASD subgroups and contribute to heterogeneity of academic achievement in these individuals; and (d) consistent with previous reports in TD children (Alloway & Passolunghi, 2011; Bull & Scerif, 2001; Meyer et al., 2010), visuospatial and verbal working memory would be associated with individual differences in math and reading abilities, respectively.

Method

This study was approved by Stanford University's Institutional Review Board. Parental consent and children's assent were obtained, and children were paid for their participation in the study.

Participants

Participants were identified from our database of children who previously participated in studies examining the development of mathematical, social, and language abilities in children with ASD and their IQ- and age-matched controls (see Table S1 in the Supplemental Material available online). All these children were recruited locally from schools and clinics in the San Francisco Bay area. Children with ASD were recruited from diverse socioeconomic and ethnic backgrounds: 41.5% were White, 13.6% were Asian, 2.5% were African American, 11% were identified as Other, and 31.4% were not specified (details in Supplemental Material). We also gathered information about the socioeconomic status (SES) of the family from the ASD and control samples on the basis of annual household income. For the 68.6% of the ASD population that reported the data, their annual income ranged from less than $10,000 to over $200,000 (details in Supplemental Material). Unfortunately, no information about their school placement or special programs was available.

As a means of identifying subpopulations of children with ASD that span a wide range of academic and cognitive abilities as well as maximizing the sample size for our clustering analyses, we assembled an inclusive sample of children with ASD. Specifically, we included all children with ASD who (a) were male, (b) were under the age of 13, (c) had a previous community diagnosis of autism, and (d) had completed all sections of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) and the second edition of the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2001). The rationale for only including males in this study was to include a sample as heterogeneous as possible but not further complicated by gender given the 4 to 1 ratio in ASD. These criteria yielded a sample of 114 boys with ASD between the ages of 7 and 12 years old. An age- and FSIQ-matched control group of 96 children without a diagnosis of ASD was selected from a larger group of 218 participants using parametrical matching on the basis of means of age and FSIQ (Table S1 in the Supplemental Material). Both ASD and control groups showed a wide range of FSIQ (ASD, 67–150; TD, 77–143; see Table S1 in the Supplemental Material for participant demographic information) measured with the WASI (Wechsler, 1999), and there was no difference in group means ($t = 1.00$, $p = .316$) or variance (Bartlett's K2 = 0.92, $p = .337$).

Behavioral assessments

Assessment of math abilities. Participants' mathematical abilities were determined using subtests of the WIAT-II (Wechsler, 2001). Specifically, the Numerical Operations (NO) subtest measures the ability to identify, write, and count numbers; produce numbers; and solve written calculation problems and simple equations. The Mathematical Reasoning (MR) subtest measures the ability to count, identify geometric shapes, and solve single-step and multistep word problems. To characterize subgroup profiles and perform multiple regression analysis, we computed a single measure of mathematical skills for each participant, which we called the math composite score, by combining the NO and MR subtests scores of the WIAT-II, as suggested in the WIAT-II procedures (Wechsler, 2001).

Assessment of reading abilities. Participants' reading ability was determined using the Word Reading (WR) and Reading Comprehension (RC) subtests of the WIAT-II. The WR subtest assesses pre-reading and decoding skills as well as the ability to read words from a list. The RC subtest assesses reading comprehension skills by asking the children to read passages of increasing difficulty levels and then answer multiple-choice questions about their content. A reading composite score was computed in the same way as the math composite score for a comparison.

Assessments of both math and reading abilities included in this study consist of two subtests, one of which tests more concrete aspects of domain knowledge (i.e., NO and WR) and another that tests more complex and applied aspects of domain knowledge (i.e., MR and RC). Therefore, in the data analysis, these assessments allow for a 2 × 2 analysis of variance (ANOVA) with factors domain (math vs. reading) and complexity (low vs. high).

Intelligence measure. All participants were administered the WASI (Wechsler, 1999), which provides Verbal, Performance, and FSIQ values. FSIQ was used for matching children with ASD to control participants on overall cognitive ability.

Autism symptom measures. The Autism Diagnostic Interview–Revised (ADI-R; Lord, Rutter, & Couteur, 1994) and Module 3 of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000) were used to confirm a diagnosis of ASD. The subscale scores on these measures were also used to examine whether specific autism symptoms contribute to heterogeneity in academic achievement. For the ADI-R, the Social Interaction, Communication and Language, Restricted and Repeated Behaviors, and Developmental Abnormalities subscales were included in subgroup comparison analyses. For the ADOS, the Communication, Restricted and Repetitive Behaviors (RRB), Social, and Imagination/Creativity subscales were included in subgroup comparison analyses. Calibrated Severity Score (CSS) for each child with ASD was also computed according to Gotham et al. (2009).

Social and behavioral difficulties. The Child Behavior Checklist for Ages 6 to 18 (CBCL/6–18; Achenbach & Rescorla, 2001) was administered to the parents or guardians of a total of 199 participants (111 ASD, 88 control) to characterize social and behavioral problems in the child participants. The CBCL is a well-validated standardized measure that is widely used to characterize social and behavioral problems in children. This assessment has 113 items and characterizes whether children are currently exhibiting or have exhibited within the past 6 months specific behavioral and emotional problems or traits. The CBCL includes eight subscales for empirically based symptom measures and six *Diagnostic and Statistical Manual of Mental Disorders*–oriented subscales (Achenbach & Rescorla, 2001). The full list of subscales is listed in Table S5 in the Supplemental Material.

Working memory. Because participants were pooled from different studies, either the Working Memory Test Battery–for Children (WMBT-C; Pickering & Gathercole, 2001) or the Automated Working Memory Assessment (AWMA; Alloway, Gathercole, Kirkwood, & Elliott, 2008) was administered to 179 participants (93 ASD and 86 control) to characterize the different components of participants' working memory as proposed by Baddeley (2003). The WMTB-C and AWMA share three common subtests that assess distinct aspects of working memory, including: (a) the Digital Recall subtest, which measures verbal/phonological working memory; (b) the Block Recall subtest, which measures visuospatial working memory; and (c) the Backwards Digital Recall subtest, which measures the central executive component of working memory.

Clustering and cross-validation measures

To examine distinct profiles (i.e., subgroups) of academic achievement in ASD, we performed hierarchical clustering analysis, an unbiased and data-driven approach with rigorous validation procedures (Fig. 1). The input to this analysis included four measures from the WIAT-II: Numerical Operations, Math Reasoning, Word Reading, and Reading Comprehension. We used hierarchical clustering with Euclidean distance and complete-linkage criterion because of their robustness and wide use in the behavioral literature (Murtagh & Contreras, 2012; Szekely & Rizzo, 2005; Ward, 1963).

Examining the presence of heterogeneity: the Gap index. To examine whether the ASD group is best characterized as a homogenous population on the basis of academic achievement or, alternatively, consists of multiple subgroups, we employed the Gap Index (Tibshirani et al., 2001). Specifically, the Gap Index quantifies the probability that an *N*-cluster solution ($N > 1$) characterizes the data compared with a one-cluster solution[1] based on a uniform distribution. To perform this analysis, we employed a subsampling procedure that randomly

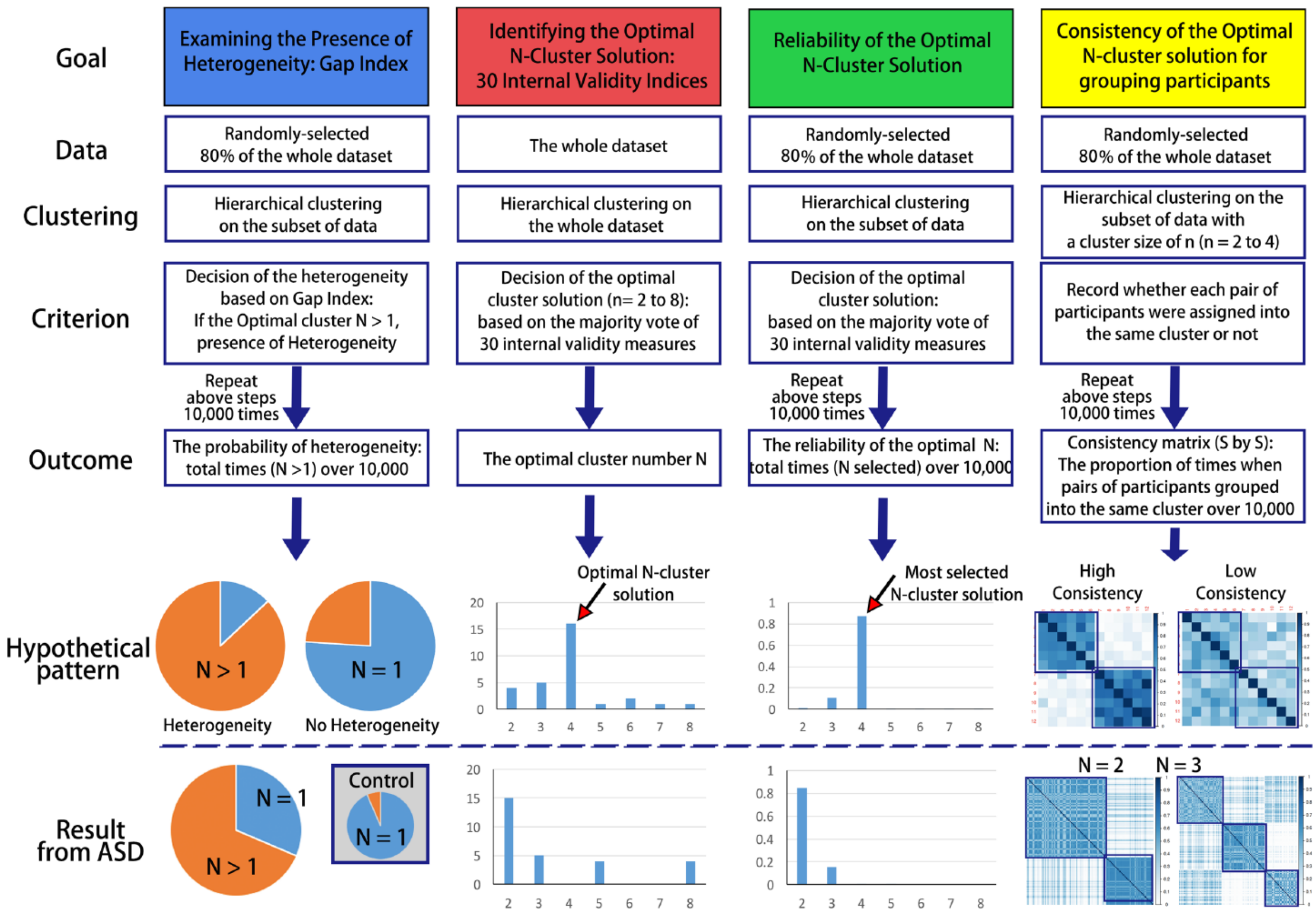

Fig. 1. Scheme of a general approach to characterize heterogeneity with clustering methods and rigorous validation measures. Each column represents different components of the analysis that separately examine the presence, identification, reliability, and consistency of heterogeneity within a specified population. The rows of the figure represent details of each component of the analysis, including hypothetical patterns of results that could emerge from the analysis (second to bottom row) as well as actual results of the analysis of children with autism spectrum disorders (ASD; bottom row). Results show there was a high probability (69%~81% dominance) that the ASD sample exhibits heterogeneity on the basis of their academic achievement scores (bottom row, left-most column), and this pattern was not observed in the age- and Full Scale IQ–matched control group (bottom row, in the shaded box; more details of the results from the control group, please see Fig. S1 in the Supplemental Material). The two-cluster solution was the optimal solution for the ASD sample (bottom row, second column from left). Reliability measures further revealed that the two-cluster solution was the dominant solution in 84.61% of the permutations (bottom row, second column from right) and that when a two-cluster solution was chosen in the ASD sample, pairs of individuals within the same cluster for the whole group were consistently assigned into the same cluster (bottom row, right-most column).

selected 80% of the data from the ASD group 10,000 times. In each subsample, hierarchical clustering was performed using the same distance and linkage metrics, and the optimal *N*-cluster solution was selected using the Gap Index implemented in the R package cluster (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2012). Across 10,000 iterations of this algorithm, we computed the proportion of times when the *N*-cluster solution ($N > 1$) was recommend by the Gap Index as the optimal solution and used it as the probability of a clustering solution favoring heterogeneity. Additionally, because the Gap Index considers the within-dispersion of clustering, we also carried out the same procedure with other algorithms, including hierarchical clustering with Ward's method and nonhierarchical clustering, namely, *K*-means.

Identifying the optimal* N*-cluster solution: 30 interval validation indices. To determine the optimal number of clusters in the ASD data, we evaluated hierarchical clustering using multiple internal validity measures. Specifically, we varied the number of clusters from two to eight, and the optimal number of clusters was determined on the basis of the majority vote of 30 indices of internal validity measures (NbClust package in R; Charrad et al., 2014).

Reliability of the optimal* N*-cluster solution. Further analyses were performed to evaluate the reliability and consistency of the optimal *N*-cluster solution. The first analysis determined whether the optimal number of clusters could be reliably selected within subsets of the data. Similar to the Gap Index analysis, subsampling was performed to evaluate reliability of the optimal *N*-cluster solution: For each subsample, 80% of the data was randomly selected, and hierarchical clustering with the NbClust package was performed. This procedure was repeated 10,000 times. A dominance score was then calculated for each cluster solution (n = 2–8) by dividing the total number of times recommended as the optimal solution over the total number of subsamples. Therefore, a higher dominance score indicates higher reliability and consistency of its corresponding number of clusters.

Consistency of the optimal* N*-cluster solution for grouping participants. A second analysis was performed to examine how consistently pairs of individuals were assigned into the same cluster when a subset of the whole sample was used for the hierarchical clustering analysis. For each *N*-cluster solution between two and four (the range of n was narrowed here on the basis of previous results showing that number of clusters > 4 was unlikely), hierarchical clustering was performed with 10,000 repetitions by randomly selecting 80% of the data. After 10,000 repetitions, a consistency matrix of S-by-S dimensions (with S as the number of participants) was computed, and each element represented the proportion of cases in which a pair of individuals was assigned to the same cluster over the 10,000 iterations (i.e., the consistency value). Based on this consistency matrix, a hierarchical clustering analysis with a fixed number of clusters (*N* corresponds to the cluster size used to produce the consistency matrix; from two to four) was conducted to cluster participants, and we examined whether the same clustering structure from the *N*-cluster solution based on the whole-sample data could be recovered. If the individual pairs from the same cluster were assigned into the same cluster consistently, the outcome of hierarchical clustering should resemble the clustering outcome based on the whole sample data. Hit rate and false alarm (FA) rate were then computed to quantify the similarity between the clustering outcomes based on consistency matrix and whole-sample data. For each cluster in every *N*-cluster solution, the cluster label from the outcome on the consistency matrix was compared with the cluster labels from the whole sample. Hypothetically, the optimal *N*-cluster solution should have a high hit rate and a low FA rate. In addition, adjusted Rand Index based on multiple methods (Steinley, 2004) was employed to compare the labeling consistency between the grouping based on the whole sample and the grouping based on subsampling consistency matrix of different *N*-cluster solutions. The confidence interval and standard error were estimated for the adjusted Rand Index (Steinley, Brusco, & Hubert, 2016).

Charactering Distinct Subgroup Profiles of Academic Achievement in ASD

ANOVAs were used to characterize distinct profiles of academic achievement among the subgroups of children with ASD. This analysis focused on group differences related to overall academic achievement between academic domains (i.e., math vs. reading). The deviance of academic achievement scores from FSIQ scores, defined here as the deviance score, was computed to assess whether individual achievement scores in both math and reading could be predicted by IQ.

Exploring Sources of Heterogeneity in ASD: Logistic and Multiple Regression Analysis

Logistic regression analysis was used to explore sources of heterogeneity within the subgroups identified in the clustering analysis. The goal of this analysis was to examine whether behavioral features of children with ASD, including core ASD symptoms, behavioral-affective traits, and cognitive abilities, could predict academically based subgroup membership when the effect of age was covaried out. This analysis allowed us to uncover underlying factors that contribute to distinct profiles of academic achievement in the ASD subgroups. Given the subgroup differences (statistical significance level for two-group difference testing was adjusted by Bonferroni correction to protect against family-wise Type I errors), the behavioral measures used in the logistic regression analysis included three working memory measures (i.e., verbal, visuospatial, and central executive).

Multiple regression was employed to examine how behavioral predictors could uniquely account for variance in specific academic domains, including math and reading, across individuals if they could predict group membership. This analysis provided details about the unique contribution of each predictor to the individual variances in domain-specific achievement. As we did in the logistic regression analysis, we examined the relationship of different components of working memory abilities with math and reading skills using multiple regression analysis. This analysis focused on the ASD group and covaried out the effects of age and FSIQ to assess independent effects of working memory components.

Finally, we conducted additional analysis to evaluate the stability and predictability of the effects of working memory on math and reading skills. In each of 10,000 repetitions/iterations, we first split the entire ASD sample into five subsets. The aforementioned multiple regression analysis was then performed on a pool of four subsets (a training set), and the remaining subset served as the testing set. The *p* value of each working memory measure (verbal, visuospatial, and central executive) was then recorded from the regression model based on the training set. A predicted math score for each individual in the testing set was computed by multiplying the beta value from the regression model, based on the training set, with the corresponding working memory (WM) score from the testing set ($\text{Math}_{\text{predicted}} = \beta_{\text{training}} \times \text{WM}_{\text{test}}$). This procedure resulted in three predicted math values for verbal, visuospatial, and central executive working memory measures, respectively, for each individual in the testing set. This process was repeated five times with each of the five subsets serving as the testing set, and therefore, five regression models were established with five training sets. If the median value of the five *p* values of a working memory measure were smaller than .05, that working memory measure was then counted once on the stability index. The correlation of predicted math values, based on each working memory measure and actual math values across all individuals, was computed, and if this correlation were significant, that working memory measure was then counted once on the predictability index. After 10,000 repetitions, a *p* value of stability for each working memory measure could be computed as 1 minus the proportion of total stability counts over the 10,000 repetitions, and a *p* value of predictability for each WM measure could be computed as 1 minus the proportion of total predictability counts over the 10,000 repetitions. The same analysis of working memory was also conducted in the control group.

Results

Cognitive and academic abilities in children with ASD

On average, children with ASD ($n = 114$) performed in the normal to above normal range across multiple measures of cognitive abilities and academic achievement (Table S1 in the Supplemental Material). This result shows that as a group, our sample of children with ASD has normal math, reading, and cognitive abilities, which is consistent with the literature on children with ASD (Hiniker, Rosenberg-Lee, & Menon, 2016; Iuculano et al., 2014). Relative to their age- and FSIQ-matched peers in the control group, the ASD group was comparable with matched controls on performance IQ (PIQ), measures of WM, and word reading (Table S1 in the Supplemental Material). The ASD group was also comparable with the control group in terms of race/ethnicity and socioeconomic status (details in the Supplemental Material). However, the ASD group showed marginally lower verbal IQ scores and significantly lower scores on Numerical Operations, Mathematical Reasoning, and Reading Comprehension compared with the matched controls (all *p*s < .05; Table S1 in the Supplemental Material).

Optimal number of subgroups in the ASD

Hierarchical clustering: Gap Index. The first goal of the analysis was to examine whether children with ASD were more likely to cluster into a single group or multiple groups on the basis of measures of academic achievement. Results from the subsamping analysis showed that the Gap Index recommended a cluster number larger than one ($N > 1$) in 6,860 out of 10,000 permutations. In other words, there was a high probability (68.60% dominance) that the ASD sample exhibits heterogeneity on the basis of academic achievement scores (Fig. 1). The same pattern of results was observed when the clustering analysis was performed using different algorithms for minimizing within-cluster dispersion (78.6%~80.5% dominance, see Supplemental Material). Moreover, the $N > 1$ solution in children with ASD was in stark contrast to the results from the control group, in which the one-cluster solution was dominant in 9,354 out of 10,000 iterations (93.54% dominance; Fig. S1 in the Supplemental Material). On the basis of the Gap Index, we conclude that the ASD group was likely to have distinct academic achievement subgroups but their age- and IQ-matched control peers were not. These results provide evidence that, on the basis of the Gap Index, children with ASD show greater heterogeneity of academic achievement compared with the control group, who showed a highly homogeneous pattern of academic achievement. Furthermore, results do not support the hypothesis that children with ASD fall on a continuum of abilities across domains; rather, results suggest that these children form distinct subgroups based on specific features of their academic achievement profiles.

Identifying optimal cluster number: interval validation measures. Consistent with results from the Gap Index, results from the hierarchical clustering analysis based on internal validity measures suggested that the two-cluster solution was the optimal solution for the ASD sample. The dendrogram produced by this analysis shows two distinct clusters that consist of a comparable number of individuals (Fig. 2a). Results also showed that 50% of 30 validation indices identified the two-cluster

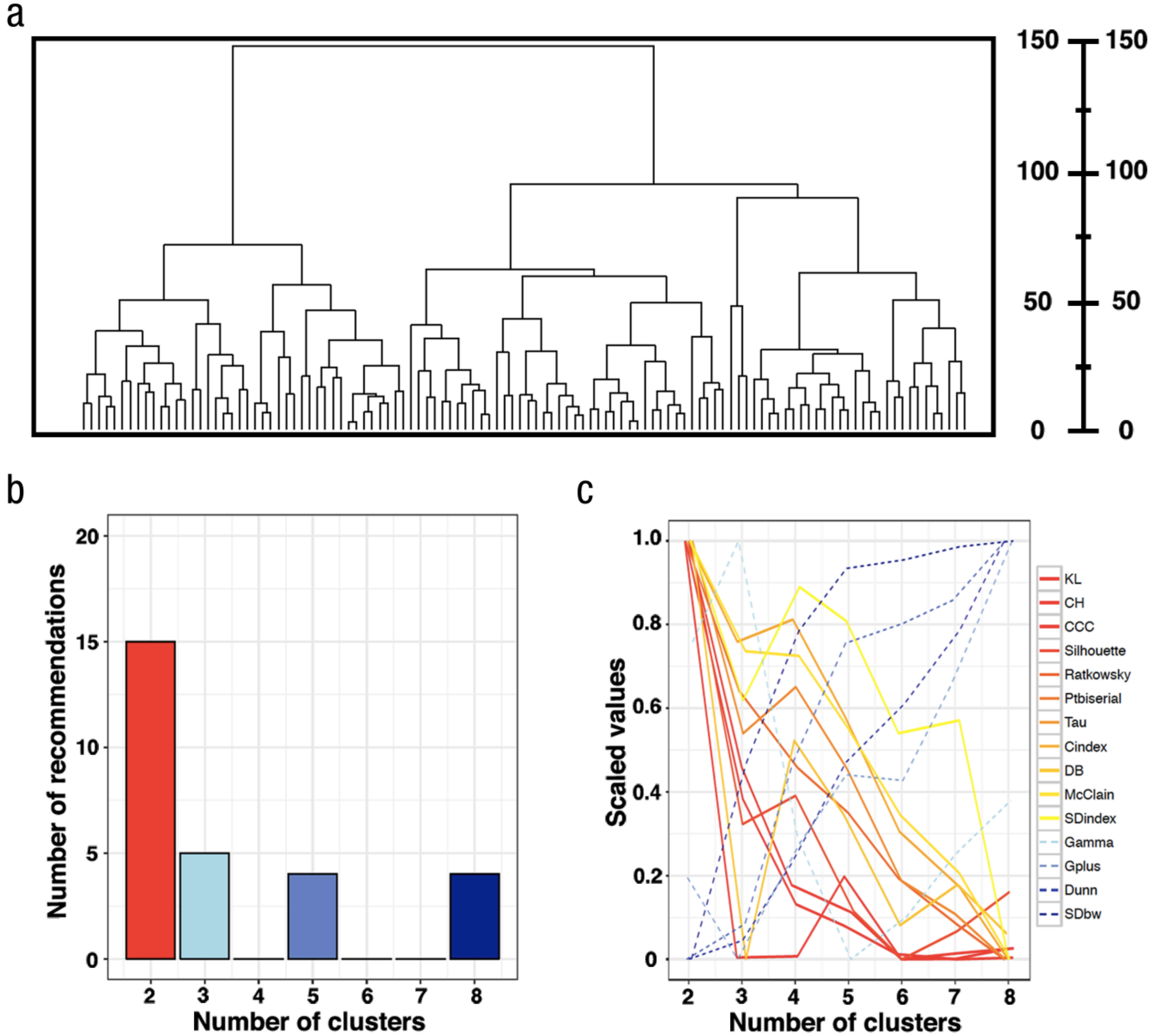

Fig. 2. Hierarchical clustering and validations for the optimal number of clusters in children with autism spectrum disorders (ASD). (a) Dendrogram from hierarchical clustering of the ASD sample (n = 114). (b) Number of indices (out of 30) from the NbClust package in R that recommended the N-cluster (2~8) solutions in ASD. (c) Scaled values of 15 commonly used indices for choosing number of clusters in ASD; these indices used either maximum or minimum values to select the optimal number of clusters, and the scaled values were computed by (original value − minimum)/range for those using maximum and 1 − (original value − minimum)/range for those using minimum as selecting criterion. Warm colors were used for indices recommending two-cluster solution, and cold colors were used for indices recommending other N-cluster solutions.

solution as the optimal one ($p < .001$), whereas the remaining indices were distributed across three, five, or eight clusters (Fig. 2b). For those indices that recommended two-cluster solutions, the scaled critical values of the two-cluster solution were also markedly higher than those of other cluster solutions (Fig. 2c).

Since the Gap Index showed that the control group formed a single homogeneous group with a high probability (92.68%), the subsequent analyses described in the Results section focused on characterizing heterogeneity within the ASD group. Additional comparative analyses examining differences with a two-cluster solution in the control group are reported in the Supplemental Material (Fig. S2).

Reliability of the optimal* N*-cluster solution in ASD: subsampling analysis. Reliability analysis of the ASD clustering results using 10,000 permutations showed that the two-cluster solution was the dominant solution in 84.61% of the permutations. This finding further highlighted the stability of the hierarchical clustering results for the ASD group.

Consistency of optimal* N*-cluster solution for grouping participants: cross-validation among pairs of participants. We further examined the robustness of the hierarchical clustering results by performing a consistency analysis, which examined whether pairs of individuals were consistently assigned into the same cluster when sampling with replacement. Results showed that when a two-cluster solution was chosen in the ASD sample, pairs of individuals within the same cluster for the whole group were consistently assigned into the same cluster (Fig. 1). Over the 10,000 permutations, consistency values for pairs in the same cluster were significantly higher than those values of pairs from different clusters, $t(6,439) = 117.55$, $p < .001$. A perfect consistency was observed in both Cluster 1 and Cluster 2 (hit rate = 1.00, FA rate = 0.00). When the cluster solution increased to three or more, individual pairs were less consistently assigned into the same clusters, as shown with lower hit rate and higher FA rate (Table S4 in the Supplemental Material). The adjusted Rand Index (Steinley, 2004; Steinley et al., 2016) further confirmed that the two-cluster solution provided better labeling consistency compared with three- and four-cluster solutions (see Supplemental Material).

Distinct profiles of academic achievement in ASD

The hierarchical clustering analysis identified two subgroups in ASD, one with 42 (36.8%) participants and the other with 72 (63.2%) participants, and the next goal of the analysis was to examine what aspects of academic achievement differed between these ASD subgroups. Results showed that the two ASD subgroups showed large and significant discrepancies (with IQ) on all academic achievement measures (Fig. 3a and Table S5 in the Supplemental Material; all *p*s < .01). Hereafter, we refer to these two subgroups as low academic achieving ASD (LA-ASD) and high academic achieving ASD (HA-ASD). Note that *low* and *high achieving* are relative terms in this context because all group-averaged achievement scores are within the normal range (i.e., within 2 *SD* from the normative mean of 100). The large overall discrepancy on measures of academic achievement is consistent with the significant FSIQ difference between LA-ASD and HA-ASD subgroups, 94.21 versus 115.65, $t(112) = -7.54$, $p < .001$. More strikingly, beyond the overall achievement discrepancy between the LA-ASD and HA-ASD subgroups, performance on math and reading measures was markedly different. Specifically, the LA-ASD group showed lower math skills compared with reading skills, whereas the HA-ASD group showed higher math skills compared with reading skills. This pattern is confirmed by a significant two-way interaction of subgroup and achievement domain, $F(1, 112) = 40.63$, $p < .001$.

As illustrated in Figure 3a, scores on the two math subscales for both the LA-ASD and HA-ASD groups were comparable, but scores on the two reading measures were not. To explore this observation, we examined a two-way interaction between academic domain and complexity of the academic measure (low complexity: Numerical Operations and Word Reading; high complexity: Math Reasoning and Reading Comprehension). The effect was significant in both subgroups, $F(1, 41) = 18.20$, $p < .001$ in LA-ASD, and $F(1, 71) = 18.00$, $p < .001$ in HA-ASD, and was largely driven by a significant difference between the reading measures (lower scores on Reading Comprehension than Word Reading). In both the LA-ASD and HA-ASD groups, the difference between Word Reading and Reading Comprehension was significant, LA-ASD $t(41) = -3.67$, $p < .001$; HA-ASD $t(61) = -5.11$, $p < .001$. However, the difference between Numerical Operations and Math Reasoning was significant only in the LA-ASD group, LA-ASD $t(41) = 2.22$, $p < .05$; HA-ASD $t(71) = 0.85$, $p = .40$. Given the statistical similarity of Numerical Operations and Math Reasoning, we computed a math composite score by averaging these two measures for each individual, and these values were used in additional analysis.

A notable feature of ASD subgroups' academic profiles is the relationship between math and reading achievement and FSIQ. Whereas Word Reading and Reading Comprehension scores for HA-ASD and LA-ASD subgroups were within 1 *SD* of the age-normed FSIQ

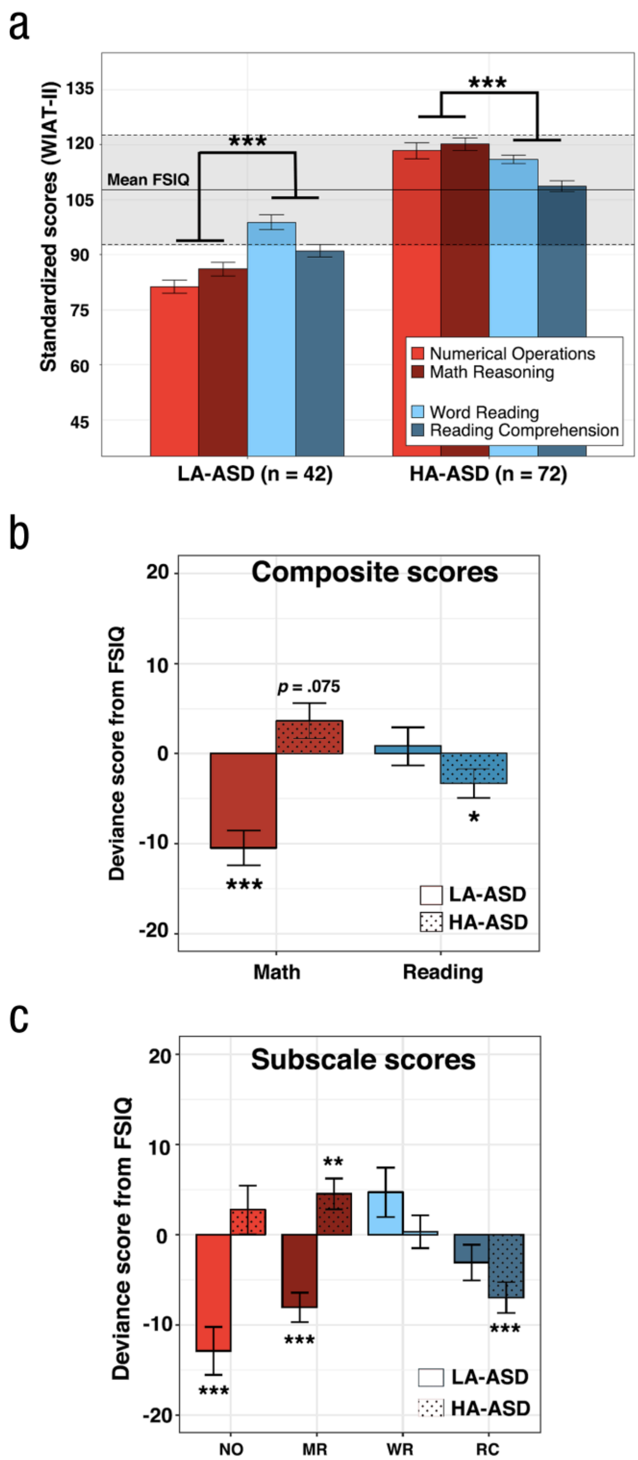

Fig. 3. Distinct subgroup profiles of academic achievement in the children with autism spectrum disorders (ASD) group. (a) Mean standardized scores of math and reading measures from second edition of the Wechsler Individual Achievement Test (WIAT-II) used for hierarchical clustering of low-achieving and high-achieving subgroups in the ASD sample. The gray shaded area in the panel indicates ±15 around the mean standardized full-scale IQ (FSIQ) scores (solid line) measured in the entire ASD sample. Significant domain (math vs. reading) differences are marked with asterisks. Deviance scores, measured in relation to FSIQ,

Fig. 3. *(continued)*
for math and reading (b) composite scores and (c) subscale scores in low-achieving (LA-ASD; solid color bars) and high-achieving (HA-ASD; dotted color bars) subgroups of ASD. NO = Numerical Operations; MR = Math Reasoning; WR = Word Reading; RC = Reading Comprehension. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

(15 points; gray horizontal bar in Fig. 3a), this pattern was not evident for the two math measures. Rather, Numerical Operations and Math Reasoning scores in the two ASD subgroups showed a striking pattern of discrepancy with FSIQ. Specifically, math scores in the LA-ASD group were below 1 *SD* of average FSIQ, whereas scores in the HA-ASD group were approximately 0.5 *SD* greater than the average FSIQ. To explore the relationship between math and reading abilities and FSIQ in the two ASD subgroups, we computed deviance scores that reflect the difference between scores on academic achievement measures and FSIQ for each participant. Results showed that math composite scores in the LA-ASD subgroup were significantly less than zero, $t(41) = -5.47$, $p < .001$, whereas math composite scores in the HA-ASD group were marginally significant greater than zero, $t(71) = 1.81$, $p = .075$, and the effect was driven by a significantly above-zero deviance score on Math Reasoning, $t(71) = 2.63$, $p = .01$ (Fig. 3b and 3c). In contrast, in the LA-ASD subgroup, deviance scores for reading skills measured by Word Reading and Reading Comprehension were not significantly different from zero ($p > .05$), and in the HA-ASD group, only the Reading Comprehension deviance score was significant below zero, $t(71) = -4.10$, $p < .001$ (Fig. 3c).

ASD symptomatology and behavioral-affective traits did not account for individual differences in academic skills in children with ASD

The next goal of the analysis was to examine whether ASD symptomatology, behavioral abilities, and cognitive factors contributed to individual differences in academic skills in the ASD subgroups. To examine this question, we compared these behavioral measures between the ASD subgroups, and results showed that neither were there differences between the two subgroups on symptom severity scores for the calibrated severity scores of the ADOS (Table S6 in the Supplemental Material), $t(85) = 0.84$, $p = .40$, nor on ADOS or ADI-R subscale scores. Moreover, behavioral-affective traits measured by the CBCL were comparable between HA-ASD and LA-ASD subgroups (Table S6 in the Supplemental Material). Together, results indicate that math achievement differences between these subgroups are unrelated to ASD symptomatology or behavioral and affective traits.

Verbal and central executive working memory predict ASD subgroup membership

In contrast to measures of symptom severity and behavioral-affective function, results showed that the LA-ASD and HA-ASD subgroups differed significantly on all three measures of WM (Table S5 in the Supplemental Material; $p < .01$), and performance on WM tests was significantly correlated with both the Numerical Operations and Math Reasoning subtest scores across the entire ASD sample (Table S2 in the Supplemental Material). To examine the role of WM in subgroup membership, the next goal of the analysis was to examine how different aspects of WM are related to subgroup membership and individual differences in academic achievement in children with ASD. We employed logistic regression to assess the effects of the different WM components on subgroup members, and results showed that after controlling for the effect of age, subgroup membership in ASD was significantly predicted by verbal WM ($z = 2.50$, $p < .05$) and central executive component of WM ($z = 2.21$, $p < .05$; see Table S8 in the Supplemental Material). Specifically, greater scores in verbal and central executive WM were associated with an increased likelihood for being clustered into the HA-ASD subgroup. In contrast, subgroup membership in the control sample was significantly predicted by all three WM components (ps < .05), indicating that individuals with high WM scores were more likely to be grouped into the HA-control subgroup.

Verbal and central executive working memory predict individual differences in math and reading skills in ASD

Given the prominent discrepancy between the two ASD subgroups with regards to WM measures, we assessed the independent contributions of WM measures to individual differences in math and reading skills across the ASD sample. Results (Table 1) from multiple regression analysis showed that after controlling for the effects of age and FSIQ, individual differences in the math composite score were predicted by verbal ($\beta = 0.23$, $t = 2.62$, $p = .01$) and central executive WM ($\beta = 0.26$, $t = 2.76$, $p < .01$) and were marginally predicted by visuospatial WM ($\beta = 0.18$, $t = 1.90$, $p = .06$). Furthermore, permutation analysis identified verbal and central executive working memory measures as stable and predictive of math skills in ASD but not visuospatial WM (see Table 1). Multiple regression and permutation analyses were then performed on reading scores in children with ASD and showed that verbal WM reliably predicted individual differences in Word Reading in children with ASD ($\beta = 0.23$, $t = 3.36$, $p < .001$); however, no WM measures were predictive of individual differences in Reading Comprehension scores in these children (see Table S9 in the Supplemental Material).

In the control sample, our analysis showed that only visuospatial WM, but not verbal or central executive WM, was a stable and predictive measure of math composite score ($\beta = 0.34$, $t = 4.51$, $p < .001$; see Table 1). In contrast, none of the WM measures uniquely contributed to individual differences in Word Reading or Reading Comprehension in the control sample (see Tables S9 and S10 in the Supplemental Material).

Discussion

Academic achievement is an important domain in which children with ASD have shown considerable heterogeneity, and previous studies of both reading and mathematical skills in children with ASD have failed to converge on a consistent pattern of abilities and deficits in these individuals. In this study, we examined and characterized distinct subgroups of children with ASD on the basis of key aspects of academic achievement and the cognitive factors that contribute to this heterogeneity. To address inconsistencies in the ASD academic achievement literature, we describe and apply a comprehensive framework and methods, employing a rigorous cross-validation approach, to examine a key facet of heterogeneity: the clustering of behavioral features between individuals. Our analyses revealed that: (a) Children with ASD clustered into two highly distinct and reliable subgroups, LA-ASD and HA-ASD, on the basis of their academic achievement; (b) LA-ASD and HA-ASD subgroups showed pronounced differences in achievement, with marked discrepancies in math but not reading abilities; (c) math composite scores in the LA-ASD group were below the normal range and lower than predicted by IQ, characteristics that were not identified in the HA-ASD or control subgroups; (d) reading comprehension and word reading were within the normal range and commensurate with IQ in both ASD-subgroups; however, reading comprehension scores in the HA-ASD group were markedly lower than word reading; and (e) verbal and central executive working memory abilities provided unique and reliable predictions of individual differences in math achievement in children with ASD. Together, these results are the first to comprehensively characterize distinct subgroups of children with ASD on the basis of their academic achievement and provide a new level of detail regarding academic achievement in children with ASD. Crucially, this work provides a novel methodological framework for examining heterogeneity in ASD and more generally, other clinical populations.

Table 1. Multiple Regression of Working Memory Measures Predicts Individual Differences in Math Scores in Children With Autism Spectrum Disorder (ASD) and Control Children

	ASD Sample (n = 93)				Control Sample (n = 86)			
Multiple regression for math	Estimates	*SE*	*z*	*p* value	Estimates	*SE*	*z*	*p* value
	Adjusted R^2 = .547				Adjusted R^2 = .566			
Intercept	26.06	13.707	1.90	.103	–11.36	13.013	–0.87	.385
Age	–2.81	1.011	–2.78	.007**	3.11	1.320	2.36	.021*
Wechsler Abbreviated Scale of Intelligence-Full Scale IQ	0.40	0.100	3.97	< .001***	0.36	0.090	4.05	< .001***
Working memory (WM)								
Verbal WM	0.23	0.089	2.62	.010*	–0.02	0.072	–0.23	.822
Visuospatial WM	0.18	0.094	1.90	.060	0.31	0.072	4.31	< .001***
Central executive WM	0.26	0.095	2.76	.007**	0.29	0.084	3.41	.001**

*$p < .05$. **$p < .01$. **$p < .001$.

Distinct profiles of academic abilities in LA-ASD and HA-ASD subgroups

Results from our clustering approach revealed two distinct subgroups of children with ASD, LA-ASD and HA-ASD, based on math and reading abilities. Unlike results from control children, these results are not consistent with the hypothesis that children with ASD fall on a continuum of academic abilities; rather, results suggest distinct subgroups of children with ASD who are distinguished by specific features of their academic achievement profiles. Two striking characteristics of these subgroups include a large discrepancy in overall achievement scores across academic domains and significant IQ differences between the subgroups. Most notably, the LA-ASD subgroup showed significantly lower math scores compared with reading scores, and math scores in this subgroup were significantly lower than IQ. These discrepancies were specific to math skills in children with ASD because reading scores in the LA-ASD subgroup were more consistent with IQ. Moreover, this pattern of poor math scores in the LA-ASD subgroup was not evident in the age- and FSIQ-matched LA-control subgroup, who showed comparable math and reading scores, both of which were within the normal range of scores and were comparable with IQ (Fig. S3 in the Supplemental Material). These findings are not consistent with results reported by S.H. Kim et al. (2017), which showed that IQ underestimated reading and arithmetic skills in a low-IQ group; however, it is important to note that the low-IQ group in that study had considerably lower IQs (M = 55 at age 9 and M = 41.1 at age 18) than those reported for the LA-ASD group in the current study (M = 98.75). In comparison, patterns of academic achievement in high-achieving children with ASD and controls are consistent with a previous report in adolescents with ASD and matched controls (Oswald et al., 2016) and show that math scores in high-achieving subgroups are both greater than reading scores and FSIQ. Together, these results suggest that weak math abilities are an important and distinguishing factor for academic achievement in a large subpopulation of children with ASD.

Math abilities in LA-ASD and HA-ASD subgroups. Results point to a unique contribution of math difficulties in the LA-ASD subgroup, which accounts for nearly 40% of the sample of children with ASD. These results are similar to but more pronounced than previous reports showing that a smaller percentage of children with ASD, ranging from 6% to 22%, have weaknesses in math abilities (Jones et al., 2009; Oswald et al., 2016). A methodological limitation of previous studies is that they have relied on arbitrary cutoff scores for defining subgroups,

Table 2. Cross-Validation of Working Memory Measures Predicts Individual Differences in Math Scores in Children with Autism Spectrum Disorder and Control Children

Cross-validation analysis for math	Stability	Predictability	Stability	Predictability
Verbal WM	0.012*	0.021*	1.000	0.945
Visuospatial WM	0.977	0.394	< 0.001***	0.004**
Central executive WM	0.007**	0.019*	< 0.001***	0.068

*$p < .05$. **$p < .01$. **$p < .001$.

whereas the current study has used advanced statistical methods and a data-driven and cross-validated approach. Only one previous study employed a data-driven approach to identify subgroups of children with ASD on the basis of academic abilities (Wei et al., 2015); however, there are a number of important differences between the methods, results, and interpretation of these previous results and those reported here. First, since this previous study did not employ cross-validation in their analysis pipeline, it is unknown how reliable or stable these results are. Second, this previous study employed a longitudinal design, and clustering was only performed at the first time point in which children were 6 to 9 years old (M = 7.5 years old), which is younger on average than the current sample (M = 9.5 years old). Results from this previous study showed high- and low-achieving groups of children with autism, which is similar to the current results. However, in contrast to the current results, this previous study also identified hypercalculic and hyperlexic subgroups of children who demonstrated isolated and superior abilities in calculation and reading, respectively. Importantly, the math advantage evident in the hypercalculic group was not sustained at subsequent time points in this previous work. Together, findings support the hypothesis that hypercalculic and hyperlexic subgroups of children with ASD may be more prevalent at younger ages but may not persist as mathematical and reading materials increase in difficulty and complexity.

Our findings highlight a number of important points regarding math difficulties in individuals with ASD, which were evident in a large subgroup of children in the current study. A critical consideration here is that math achievement is closely associated with academic and professional success and is an important component of achieving independent living (Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Peters et al., 2013; Rodríguez et al., 2013). Our analysis indicates that approximately 40% of the sampled children with ASD show specific learning difficulties in math that are overall not accompanied by reading difficulties. Importantly, math skills are generally not a primary focus of educational programs for children with ASD, which often target the remediation of language skills, including reading (Barnett & Cleary, 2015). Results from the current study emphasize the need for a new focus on educational interventions to serve the substantial subgroup of children with ASD who present impairments in this cognitive domain.

We also identified a subgroup of children with ASD, accounting for approximately 60% of the ASD sample, who showed significantly greater strength in math skills (especially, math reasoning) relative to reading and FSIQ. This result is consistent with previous anecdotal and empirical evidence showing that some children with ASD may have preserved, or even gifted, skills in the math domain (Assouline et al., 2012; Cash, 1999; Oswald et al., 2016); however, this hypothesis has not always been supported in the literature (Oswald et al., 2016). Our findings provide strong evidence that math skills are an area of specific strength in a subgroup of children with ASD, and therefore, gaining a better understanding of the learning profiles and neurocognitive mechanisms underlying strengths in math in individuals with ASD represents an important area for future investigation (Iuculano et al., 2014).

Reading abilities in LA-ASD and HA-ASD subgroups. While word reading and reading comprehension abilities in the LA-ASD and HA-ASD subgroups were both within the normal range of abilities (i.e., mean standard scores > 90), different patterns of performance were evident for these reading measures. Specifically, reading comprehension scores were lower than word reading scores for both HA-ASD and LA-ASD groups (Fig. 3a). Moreover, whereas word reading abilities were commensurate with FSIQ in both LA-ASD and HA-ASD subgroups, reading comprehension scores in the HA-ASD subgroup were significantly lower than those predicted by FSIQ, a finding that was absent in the LA-ASD subgroups (Fig. 3c). The discrepancy between reading comprehension and FSIQ was also evident in the HA-control subgroup, suggesting that this pattern is not unique to the high-achieving children with ASD and may reflect a general profile of academic abilities in high-achieving individuals.

We identified a specific weakness for reading comprehension in the HA-ASD subgroup. Converging results with previously published studies (Bartak & Rutter, 1973; Jones et al., 2009; Nation et al., 2006) strongly suggests that difficulties with reading comprehension are a reliable feature of the ASD academic profile for many children with ASD and are not specific to children with lower cognitive abilities (Troyb et al., 2014). One possible explanation for these difficulties is that compared with Word Reading, which is a relatively concrete task associated with word decoding and recognition, Reading Comprehension is a relatively abstract task that requires an understanding of social situations and interpersonal information as well as information outside the range of interests of the participant and therefore may present a particular challenge to children with ASD. Results suggest that reading- and language-related interventions for children with ASD are appropriately tailored to the IQ level of the individual, with a particular focus on reading comprehension for high-IQ children with ASD. Moreover, unlike previous studies that revealed hyperlexia in a subgroup of children with ASD (Jones et al., 2009; S. H. Kim et al., 2017; Wei et al., 2015), results from the current study did not identify a subgroup with superior word reading ability that is better than predicted by FSIQ. The current results suggest that hyperlexia may not be a general characteristic feature of children with ASD.

Sources of heterogeneity in academic profiles of children with ASD

ASD symptomology, behavior, and IQ. A final goal of the study was to examine whether aspects of ASD symptomatology or behavior differ between the academic subgroups identified in the clustering analysis. Surprisingly, results did not reveal any significant differences between the HA-ASD and LA-ASD subgroups on clinical symptom severity and behavioral difficulties, and the only differences to emerge were on cognitive abilities, including IQ and working memory. From one perspective, this may not be a surprising result: Mean FSIQ was more than 1 *SD* greater for the HA-ASD subgroup compared with LA-ASD, and this finding alone may be sufficient to explain academic achievement differences between subgroups. From another perspective, however, the fact that none of the other symptom-related or behavioral measures were different between subgroups may be considered a surprising and interesting finding. This result suggests that the prominent social communication difficulties and repetitive and restricted behaviors that are the hallmarks of childhood autism do not appear to play a dominant role in children's ability to acquire math and reading skills. This conclusion is surprising given that acquiring reading and math skills in a classroom or home setting is often a highly social process that relies on many of the skills that are most challenging for children with ASD, including joint attention and reciprocal interactions.

Working memory. WM plays an important role in children's academic achievement (Meyer et al., 2010; Wu et al., 2017). To the best of our knowledge, previous studies have not examined the role of WM in the heterogeneity of academic skills in children with ASD, and here we show that verbal and central executive WM predicted HA-ASD and LA-ASD subgroup membership. These WM components also explained significant individual differences in math skills in children with ASD taken as a group. Our finding that visuospatial WM is not a reliable predictor of composite math scores in the ASD group is surprising in the context of the extant literature showing a strong link between this aspect of WM and math skills in TD children (Alloway & Passolunghi, 2011; Ashkenazi, Black, Abrams, Hoeft, & Menon, 2013; Bull & Scerif, 2001; Meyer et al., 2010) and contrasts with results from the control sample who, consistent with this extensive literature, showed that visuospatial working memory was the only reliable WM factor to predict individual differences in math achievement. An important consideration is that multiple regression results showed that visuospatial working memory was a significant predictor of composite math scores in the ASD group; however, this component of WM had a reduced impact compared with verbal and central executive WM.

Finally, our analysis also revealed that verbal WM explained unique variance in word reading achievement in children with ASD; however, this effect was not robust according to cross-validation analysis and was not evident in TD children. The association of verbal WM with word recognition skills has been documented in the literature (Dufva, Niemi, & Voeten, 2001; Gottardo, Stanovich, & Siegel, 1996), but the nature of this association is unclear. One possibility is that shared variance in verbal WM with word recognition skills reflects overlapping cognitive processes involved in phonological coding. Thus, verbal and central executive WM function represent a primary bottleneck for the LA-ASD group. Specifically, an inability to process and manipulate verbal information and direct attentional resources may contribute to processing deficits in math and to a lesser extent, in reading.

A novel approach for comprehensively characterizing heterogeneity on the basis of clinical and behavioral measures

A major goal of the current work was to describe an approach and method that can be used to comprehensively characterize heterogeneity on the basis of clinical, cognitive, and behavioral attributes in clinical populations. This approach first consists of rigorous clustering methods that identify subgroups of individuals on the basis of their behavioral profile, which represents a methodological improvement over previous studies that have used arbitrary cutoff criteria to identify subgroups (Jones et al., 2009; Oswald et al., 2016). Next, this approach capitalizes on advanced statistical methods to validate the optimal cluster solution by using permutation testing and the Gap Index, which identifies the probability of the presence of subgroups (Tibshirani et al., 2001).

Our approach provides several advantages compared with previous studies. First, the data-driven approach described here is an improvement over previous methods that force the sample into predetermined numbers of subgroups without first demonstrating the existence of subgroups (Jones et al., 2009; Oswald et al., 2016; Wei et al., 2015). Second, the current approach uses cross-validation rather than qualitative criteria (Wei et al., 2015), which often require post hoc interpretations of the clustering results as a means of justifying their validity. As a result, it is unknown whether these previous methods provide replicable or reliable outcomes. Third, the current study incorporates a matched control group for comparison with the clinical group to provide insights into whether results from the experimental ASD group represented unique clustering results that were specific to this population or alternatively, whether these were more general patterns that were also characteristic of matched control children. Results showed unique aspects of academic profiles in the ASD subgroups that were not identified in the control subgroup, suggesting specificity of these findings to children with ASD. Together, the current approach and methods enable a new level of reliability and robustness in the identification of subgroups based on behavioral measures and may be particularly important for assessing heterogeneity in clinical populations such as ASD. Importantly, this approach and analysis may serve as a template for future studies as a means of establishing a more thorough methodology for addressing complex behavioral profiles in clinical populations.

Future Directions and Implications

Our study provides new insights into distinct academic profiles in children with ASD and has important educational and clinical implications. From an educational perspective, results showing that children with ASD consistently cluster into either a low or high performance group, each of which has distinct strengths and weaknesses in reading and math domains, suggest that these profiles may serve as a useful guide for considering tailored educational programs that can both build on children's academic strengths and fortify areas of weakness. For example, if educators are tasked with designing and overseeing a mainstreamed educational curriculum for a child with ASD with a high IQ (IQ > ~110), being cognizant of the fact that this child is likely to show significant strengths in math and word reading and weaknesses in reading comprehension may provide useful information for initially placing this child in the appropriate classes and dedicating appropriate resources for the child's education. Similarly, when considering the educational needs for a child with average IQ (~100), it may be useful and informative to these educators to know that this child is more likely to show significant deficits in math abilities but reading abilities that are commensurate with IQ. Given the limited resources of many schools and the difficult task that educators face in providing high-quality education for the diverse population of children in their classrooms, having an understanding of these subgroups of children with ASD may serve as an important tool for the initial placement of these children in mainstream classrooms. Another implication of our findings is that even children with average or above intelligence with ASD should have regular academic evaluations, particularly if they are struggling at school, because stereotypes that all children with ASD are good with numbers and computers do not hold. Future work in the area of educational research may focus on designing educational curricula that address the strengths and weaknesses of the two distinct profiles of children with ASD identified in the current work and may consider not only reading and math instruction but also the strengthening of cognitive skills, such as working memory, that support academic learning.

From a clinical perspective, results from the current study provide new information by showing that autism symptom severity is not related to educational abilities in children across symptom domains. One implication of this result is that it suggests that it may be beneficial for clinicians to consider profiles of academic and cognitive abilities in conjunction with autistic symptoms when designing appropriate interventions for children with ASD, for example, an important direction for ASD intervention, such as pivotal response training (Koegel & Koegel, 2006; Koegel, Koegel, & Brookman, 2003), which builds on a child's interest and strengths and uses motivational teaching strategies to promote verbal interactions (Dawson et al., 2010; Mundy & Stella, 2000). If a clinician were designing an intervention for an affected child with a high IQ, it may be instructive for that clinician to consider engaging the child with math given that the child is likely to show a high aptitude for math and may therefore enjoy math games. Importantly, math may provide an important avenue for verbal engagement for these high IQ children. Alternatively, if a clinician were designing an intervention for an affected child with average IQ (i.e., LA-ASD children), it may be useful for the clinician to consider engaging the child in ways that do not relate to math skills, which the child likely struggles with and may not serve to promote verbal interactions. It is hoped that future clinical research examines whether profiles of academic abilities contribute to designing effective remediation programs for children with ASD by helping to identify areas for verbal engagement and interaction in these children.

Conclusions

In conclusion, we have examined the heterogeneity of academic achievement profiles in children with ASD using a rigorous quantitative approach and methods for identifying distinct subgroups within clinical populations. Results showed that children with ASD clustered into two distinct subgroups on the basis of math and reading skills: an LA-ASD subgroup characterized by markedly lower math scores that were both below the normal range of abilities and scores predicted by FSIQ and an HA-ASD subgroup characterized by math skills that were both greater than scores predicted by FSIQ and reading skills. Importantly, these subgroup differences were unrelated to ASD symptoms or other behavioral-affective traits. Regression analysis further revealed that subgroup membership and variability in math skills was associated with verbal and central executive working memory capacities in children with ASD, which is in contrast to the strong relationship between visuospatial working memory and math function shown by controls. Our findings not only describe a novel and rigorous approach for exploring heterogeneity in clinical populations but provide new insights into the distinct profiles of academic skills in children with ASD, which have important implications for educational practice and intervention programs.

Action Editor

Kenneth J. Sher served as action editor for this article.

Author Contributions

L. Chen and V. Menon designed the research; H. N. Wakeman, S. Prathap, M. Rosenberg-Lee, T. Iuculano, and L. Chen performed the research; L. Chen, H. N. Wakeman, and T. Chen analyzed the data; L. Chen, D. A. Abrams, and V. Menon wrote the manuscript, and all authors contributed to editing it. All the authors approved the final manuscript for submission.

ORCID iD

Lang Chen https://orcid.org/0000-0002-2118-5601

Acknowledgments

We thank the reviewers for their valuable feedback.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Grant HD059205 (to V. Menon), National Institute of Mental Health Grants MH084164 (to V. Menon) and MH102428 (to D. A. Abrams), and National Center for Advancing Translational Sciences Grant UL1-TR001085 (to the Stanford Child Health Research Institute, L. Chen, and V. Menon).

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/[DOI]

Note

1. The uppercase *N* is used to denote a specific number of the cluster size in the analysis, and lower case *n* is used for the range of cluster size used in each analysis.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for ASEBA School-Age Forms and Profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth, and Families.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, *28*, 725–734. doi:10.1080/01443410802243828

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*, 133–137. doi:10.1016/j.lindif.2010.09.013

Ashkenazi, S., Black, J. M., Abrams, D. A., Hoeft, F., & Menon, V. (2013). Neurobiological underpinnings of math and reading learning disabilities. *Journal of Learning Disabilities*, *46*, 549–569.

Assouline, S. G., Nicpon, M. F., & Dockery, L. (2012). Predicting the academic achievement of gifted students with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *42*, 1781–1789. doi:10.1007/s10803-011-1403-x

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*, 189–208. doi:10.1016/S0021-9924(03)00019-4

Barnett, J. E. H., & Cleary, S. (2015). Review of evidence-based mathematics interventions for students with autism spectrum disorders. *Education and Training in Autism and Developmental Disabilities*, *50*, 172–185.

Baron-Cohen, S., Wheelwright, S., Burtenshaw, A., & Hobson, E. (2007). Mathematical talent is linked to autism. *Human Nature*, *18*(2), 125–131. doi:10.1007/s12110-007-9014-0

Bartak, L., & Rutter, M. (1973). Special educational treatment of autistic children: A comparative study? 1. Design of study and characteristics of units. *Journal of Child Psychology and Psychiatry*, *14*, 161–179. doi:10.1111/j.1469-7610.1973.tb01185.x

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science: A Journal of the American Psychological Society / APS*, *11*, 474–480. doi:10.1111/1467-9280.00291

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*, 273–293. doi:10.1207/s15326942dn1903_3

Cash, A. B. (1999). A profile of gifted individuals with autism: The twice-exceptional learner. *Roeper Review*, *22*, 22–27. doi:10.1080/02783199909553993

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust : An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6), 1–36. doi:10.18637/jss.v061.i06

Chiang, H.-M., & Lin, Y.-H. (2007). Mathematical ability of students with Asperger syndrome and high-functioning autism: A review of literature. *Autism: The International Journal of Research and Practice*, *11*, 547–556. doi:10.1177/1362361307083259

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., & Donaldson, A. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The early start Denver model. *Pediatrics*, *125*(1). doi:10.1542/peds.2009-0958

Dufva, M., Niemi, P., & Voeten, M. J. M. (2001). The role of phonological memory, word recognition, and comprehension skills in reading development: From preschool to Grade 2. *Reading and Writing An Interdisciplinary Journal*, *14*, 91–117. doi:10.1023/A:1008186801932

Estes, A., Rivera, V., Bryan, M., Cali, P., & Dawson, G. (2011). Discrepancies between academic achievement and intellectual ability in higher-functioning school-aged children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*, 1044–1052. doi:10.1007/s10803-010-1127-3

Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics*, *129*, e1112–e1120. doi:10.1542/peds.2011-1601

Georgiades, S., Szatmari, P., Boyle, M., Hanna, S., Duku, E., Zwaigenbaum, L., . . . Thompson, A. (2013). Investigating phenotypic heterogeneity in children with autism spectrum disorder: A factor mixture modeling approach. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*, 206–215. doi:10.1111/j.1469-7610.2012.02588.x

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*, 693–705. doi:10.1007/s10803-008-0674-3

Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, *130*, e1278–e1284. doi:10.1542/peds.2011-3668

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology*, *63*, 563–582. doi:10.1006/jecp.1996.0062

Hiniker, A., Rosenberg-Lee, M., & Menon, V. (2016). Distinctive role of symbolic number sense in mediating the mathematical abilities of children with autism. *Journal of Autism and Developmental Disorders*, *46*, 1268–1281.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2009). Savant skills in autism: psychometric approaches and parental reports. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1359–1367. doi:10.1098/rstb.2008.0328

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*, 67–77. doi:10.1002/aur.72

Iuculano, T., Rosenberg-Lee, M., Supekar, K., Lynch, C. J., Khouzam, A., Phillips, J., . . . Menon, V. (2014). Brain organization underlying superior mathematical abilities in children with autism. *Biological Psychiatry*, *75*, 223–230. doi:10.1016/j.biopsych.2013.06.018

Jansiewicz, E. M., Goldberg, M. C., Newschaffer, C. J., Denckla, M. B., Landa, R., & Mostofsky, S. H. (2006). Motor signs distinguish children with high functioning autism and Asperger's syndrome from controls. *Journal of Autism and Developmental Disorders*, *36*, 613–621. doi:10.1007/s10803-006-0109-y

Jones, C. R. G., Happé, F., Golden, H., Marsden, A. J. S., Tregay, J., Simonoff, E., . . . Charman, T. (2009). Reading and arithmetic in adolescents with autism spectrum disorders: Peaks and dips in attainment. *Neuropsychology*, *23*, 718–728.

Keen, D., Webster, A., & Ridley, G. (2016). How well are children with autism spectrum disorder doing academically at school? An overview of the literature. *Autism*, *20*, 276–294. doi:10.1177/1362361315580962

Kim, H., & Cameron, C. E. (2016). Implications of visuospatial skills and executive functions for learning mathematics: Evidence from children with autism and Williams Syndrome. *AERA Open*, *2*(4). doi:10.1177/2332858416675124

Kim, S. H., Bal, V. H., & Lord, C. (2017). Longitudinal follow-up of academic achievement in children with autism from age 2 to 18. *Journal of Child Psychology and Psychiatry*, *59*, 258–267. doi:10.1111/jcpp.12808

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

Koegel, R. L., & Koegel, L. K. (2006). The basis of pivotal response treatment. In R. L. Koegel, D. Openden, R. M. Fredeen, & L. K. Koegel (Eds.), *Pivotal response treatments for autism: Communication, social, and academic development* (pp. 3–30). Baltimore, MD: Paul H. Brookes Publishing.

Koegel, R. L., Koegel, L. K., & Brookman, L. I. (2003). Empirically supported pivotal response interventions for children with autism. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 341–357). New York, NY: Guilford.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). Autism Diagnostic Observation Schedule (ADOS). *Journal of Autism and Developmental Disorders*, *30*(3), 205–23. doi:10.1007/BF02211841

Lord, C., Rutter, M., & Couteur, A. L. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659–685. doi:10.1007/BF02172145

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2012). Cluster: Cluster analysis basics and extensions. *R Package Version*, *1*, 56.

Mayes, S., & Calhoun, S. (2003). Analysis of WISC-III, Stanford-Binet: IV, and academic achievement test scores in children with autism. *Journal of Autism and Developmental Disorders*, *33*, 229–241. doi:10.1023/a:1024462719081

Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences*, *20*, 101–109. doi:10.1016/j.lindif.2009.08.004

Mundy, P., & Stella, J. (2000). Joint attention, social orienting, and nonverbal communication in autism. In A. M. Wetherby & B. M. Prizant (Eds.), *Autism spectrum disorders: A transactional developmental perspective* (pp. 55–77). Baltimore, MD: Paul H. Brookes Publishing.

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*, 86–97. doi:10.1002/widm.53

Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *36*, 911–919. doi:10.1007/s10803-006-0130-1

Newman, L., Wagner, M., Knokey, A-M., Marder, C., Nagle, K., Shaver, D., & Wei, X. (2011). *The post-high school outcomes of young adults with disabilities up to 8 years after high school: A report from the National Longitudinal Transition Study-2 (NLTS2)* (NCSER 2011-3005). Washington, DC: National Center for Special Education Research.

Norbury, C., & Nation, K. (2011). Understanding variability in reading comprehension in adolescents with autism spectrum disorders: Interactions with language status and decoding skill. *Scientific Studies of Reading*, *15*, 191–210.

Oswald, T. M., Beck, J. S., Iosif, A. M., Mccauley, J. B., Gilhooly, L. J., Matter, J. C., & Solomon, M. (2016). Clinical and cognitive characteristics associated with mathematics problem solving in adolescents with autism spectrum disorder. *Autism Research*, *9*, 480–490. doi:10.1002/aur.1524

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2013). Numeracy and decision making numeracy. *Psychological Science*, *17*, 407–413. doi:10.1111/j.1467-9280.2006.01720.x

Pickering, S. J., & Gathercole, S. E. (2001). *Working memory test battery for children (WMTB-C)*. London, England: Psychological Corporation.

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*, 11.

Rodríguez, V., Andrade, A. D., García-Retamero, R., Anam, R., Rodríguez, R., Lisigurski, M., . . . Ruiz, J. G. (2013). Health literacy, numeracy, and graphical literacy among veterans in primary care and their effect on shared decision making and trust in physicians. *Journal of Health Communication, 18* (Suppl. 1), 273–289. doi:10.1080/10810730.2013.829137

Seelaar, H., Rohrer, J. D., Pijnenburg, Y. A-L., Fox, N. C., & van Swieten, J. C. (2011). Clinical, genetic and pathological heterogeneity of frontotemporal dementia: A review. *Journal of Neurology, Neurosurgery, and Psychiatry*, *82*, 476–486. doi:10.1136/jnnp.2010.212225

Smellie, W. S. A. (2006). Testing pitfalls and summary of guidance in lipid management. *BMJ*, *333*(7558), 83–86. doi:10.1136/bmj.333.7558.83

Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods*, *9*, 386–396. doi:10.1037/1082-989X.9.3.386

Steinley, D., Brusco, M. J., & Hubert, L. (2016). The variance of the adjusted Rand index. *Psychological Methods*, *21*, 261–272. doi:10.1037/met0000049

Szekely, G. J., & Rizzo, M. L. (2005). Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. *Journal of Classification*, *22*, 151–183.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*, 411–423. doi:10.1111/1467-9868.00293

Titeca, D., Roeyers, H., Loeys, T., Ceulemans, A., & Desoete, A. (2015). Mathematical abilities in elementary school children with autism spectrum disorder. *Infant and Child Development*, *24*, 606–623. doi:10.1002/icd.1909

Troyb, E., Orinstein, A., Tyson, K., Helt, M., Eigsti, I.-M., Stevens, M., & Fein, D. (2014). Academic abilities in children and adolescents with a history of autism spectrum disorders who have achieved optimal outcomes. *Autism*, *18*, 233–243. doi:10.1177/1362361312473519

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236–244. doi:10.1080/01621459.1963.10500845

Wechsler, D. (1999). *Wechsler Abbreviated Intelligence Scale*. San Antonio, TX: The Psychological Corp.

Wechsler, D. (2001). *The Wechsler Individual Achievement Test Second Edition (WIAT-II)*. San Antonio, TX: The Psychological Corp.

Wei, X., Christiano, E. R., Yu, J. W., Wagner, M., & Spiker, D. (2015). Reading and math achievement profiles and longitudinal growth trajectories of children with an autism spectrum disorder. *Autism*, *19*, 200–210. doi:10.1177/1362361313516549

Whitby, P., & Mancil, G. (2009). Academic achievement profiles of children with high functioning autism and Asperger syndrome: A review of the literature. *Education and Training in Developmental Disabilities*, *44*, 551–560.

Wu, S. S., Chen, L., Battista, C., Smith Watts, A. K., Willcutt, E. G., & Menon, V. (2017). Distinct influences of affective and cognitive factors on children's non-verbal and verbal mathematical abilities. *Cognition*, *166*, 118–129. doi:10.1016/j.cognition.2017.05.016

Zetusky, W. J., Jankovic, J., & Pirozzolo, F. J. (1985). The heterogeneity of Parkinson's disease: Clinical and prognostic implications. *Neurology*, *35*, 522–526.