

# Language and Verbal Memory in Individuals with a History of Autism Spectrum Disorders Who Have Achieved Optimal Outcomes

Katherine Tyson · Elizabeth Kelley · Deborah Fein · Alyssa Orinstein ·  
Eva Troyb · Marianne Barton · Inge-Marie Eigsti · Letitia Naigles ·  
Robert T. Schultz · Michael Stevens · Molly Helt · Michael Rosenthal

Published online: 28 August 2013  
© Springer Science+Business Media New York 2013

**Abstract** Some individuals who lose their autism spectrum disorder diagnosis may continue to display subtle weaknesses in language. We examined language and verbal memory in 44 individuals with high-functioning autism (HFA), 34 individuals with “optimal outcomes” (OO) and 34 individuals with typical development (TD). The OO group scored in the average range or above on all measures and showed few differences from the TD group. The HFA group performed within the average range but showed significantly lower mean performance than the other groups on multiple language measures, even when controlling for verbal IQ. Results also indicate that OO individuals show strong language abilities in all areas tested, but that their language may show greater reliance on verbal memory.

**Keywords** Optimal outcome · Language · Recovery · Autism

## Introduction

Language and communication difficulties are a defining feature of autism spectrum disorders (ASDs) (American Psychiatric Association 2000). Many children with ASD show delayed and atypical language development (Boucher 2012; Eigsti et al. 2011). However, ASD language profiles vary considerably among individuals and among studies, with impairments ranging from global, severe impairment to subtle difficulties with pragmatic use of language in high-functioning individuals, who have cognitive abilities in the average range (Boucher 2012; Condouris et al. 2003). Verbal memory may also be an area of difficulty in ASD (Bennetto et al. 1996; Fein et al. 1996).

In spite of the prevalence of language impairment in ASD, a small, but growing body of research indicates that, through intensive early intervention, some children with ASDs make significant cognitive and language gains and even go on to lose the ASD diagnosis (Fein et al. 2013; Harris and Handleman 2000; Helt et al. 2008; Kelley et al. 2006, 2010; Sallows and Graupner 2005; Sutter et al. 2007). These children may be placed in mainstream classrooms, have cognitive and adaptive skills in the average range, and no longer meet formal criteria for an ASD. Studies of “optimal outcome” are reviewed in detail in Helt et al. (2008). Some of these studies have followed groups of children longitudinally (e.g., Harris and Handleman 2000; Lovaas 1987; McEachin et al. 1993), while others simply examined them at one time point. Kelley et al. (2006) studied a group of OO children compared to typically developing (TD) individuals, at one time point

---

K. Tyson (✉) · D. Fein · A. Orinstein · E. Troyb · M. Barton ·  
I.-M. Eigsti · L. Naigles · M. Helt  
Department of Psychology, University of Connecticut, Storrs,  
CT, USA  
e-mail: katherine.tyson@uconn.edu

E. Kelley  
Department of Psychology, Queens University,  
Kingston, ON, Canada

D. Fein  
Department of Pediatrics, University of Connecticut,  
Farmington, CT, USA

R. T. Schultz  
Center for Autism Research, Children’s Hospital of Philadelphia,  
Philadelphia, PA, USA

M. Stevens  
Institute of Living, Hartford Hospital, Hartford, CT, USA

M. Rosenthal  
Child Mind Institute, New York, NY, USA

only. Kelley et al.'s (2010) study included OO, TD, and HFA groups studied at one time point.

Helt et al. (2008) have estimated that individuals with ASD who achieve this “optimal outcome” (OO) status comprise between 3 and 25 % of individuals originally diagnosed with an ASD. Studies of individuals who lose an ASD diagnosis have incorporated varying definitions for this optimal outcome status. For instance, Sallows and Graupner (2005) described children who lost an ASD diagnosis as performing in the normal range on measures of cognitive, language, adaptive, and academic functioning. Individuals in their sample showed mild elevations on some personality and diagnostic scales measuring attention, language, social skills, and rigid play. Lovaas (1987) described outcomes following early, intensive behavioral intervention, which included achievement of cognitive functioning in the average range and placement in mainstream classrooms. Sutura et al. (2007), in a sample of 4-year-olds previously diagnosed with ASD, defined optimal outcomes as characterized by having scores within the normal range on measures of adaptive and cognitive functioning and no diagnosis of autistic disorder or PDD-NOS. In Kelley et al.'s (2006) study, individuals with OOs were diagnosed with ASD before age 5 and had IQ scores within the normal range, were placed in age-appropriate mainstream classes, and received minimal services at school. A few of these children continued to receive behavioral intervention services for autism symptoms, and performance on the ADOS was not included as a stipulation for the definition of OO. Kelley et al. (2010) defined OO similarly, but required that individuals with optimal outcomes not meet criteria for an ASD on the ADOS.

Examination of residual deficits in these individuals with OO is important for determining the degree to which symptoms have resolved and for identifying core deficits that may be resistant to amelioration. Findings of language and verbal memory impairments in HFA might suggest areas in which to look for residual deficits in OO individuals.

The literature on ASD language profiles indicates heterogeneous patterns of strengths and weaknesses. For younger and lower-functioning children, phonology and prosody tend to be difficulties (Lord and Paul 1997). In general, grammar tends to be appropriate to individuals' mental age and overall language level (e.g., Condouris et al. 2003; Goodwin et al. 2012; Jordan 1993; Lord and Paul 1997; Swensen et al. 2007). However, some individuals with ASD may use simpler syntax (Eigsti et al. 2007) and a more limited range of morphological and syntactic forms in spontaneous speech (e.g., prepositions, conjunctions, embedded sentences) than TD peers (Fein et al. 1996; Kjelgaard and Tager-Flusberg 2001; Menyuk and Quill 1985; Scarborough et al. 1991; Tek et al. 2013). Difficulties with syntactic judgments have been found even

in adolescents with HFA (Eigsti and Bennetto 2009). Semantics appear to be relatively preserved in ASD (Fein et al. 1996; Kelley et al. 2010; Kjelgaard and Tager-Flusberg 2001; Mottron 2004; Swensen et al. 2007). However, individuals with ASD may have different ways of organizing semantic categories than individuals with typical development (Dunn et al. 1996; Klinger and Dawson 2001; Naigles et al. 2013; Tek et al. 2008) and may have difficulty with specific semantic classes, such as mental state words (Kelley et al. 2006).

One question, then, concerns the degree to which these linguistic difficulties are resolved in individuals with OO. A few recent studies have examined language outcomes in children and adolescents who lost their diagnosis and achieved OO, although the children included in these studies were not required to have a history of language delay (Kelley et al. 2006, 2010). In a sample of TD and OO children ages 5–9 years old, Kelley et al. (2006) found that grammatical abilities of OO children were indistinguishable from those of their TD peers. However, these OO children did have some difficulty with pragmatic and semantic language, with weaknesses compared to TD peers on tests of categorical induction, theory of mind, mental state verb production (e.g., ‘think’ vs. ‘guess’), and narrative story production.

In a second study including children aged 8–14 years with OOs (Kelley et al. 2010), which included some of the same children as those in the Kelley et al. (2006) study when they were a few years older, weaknesses in language abilities on a number of standardized language assessments were no longer evident in OO individuals, although narrative production was not included in this study. This second study compared the OO group to an age-matched group of TD children, as well as a group of children with HFA. The OO group performed significantly better than the HFA group on receptive vocabulary, a test of figurative language, and a subtest asking participants to infer the most likely connection between statements. The OO group's performance on these tests was indistinguishable from that of the TD group. In the current study, we extend these investigations of language to older individuals with OO, ranging up to 21 years of age, and with a more rigorously defined OO group (Fein et al. 2013).

Verbal memory is another area that has been studied in ASD. In general, children with ASD have more difficulty with material of increased complexity, organization, or meaningfulness. Fein et al. (1996) found that children with HFA, compared to a group with SLI, had the least difficulty remembering digits, more with sentences, and most with stories. Similarly, Gabig (2008) found a hierarchy of verbal memory difficulty by complexity in a group of children with ASD, compared to age-matched TD children, on repeating nonwords, digits, sentences, and a story.

Children with ASD may have difficulty recalling lists of words based on meaningful clustering or other strategies (Hermelin and O'Connor 1970; Tager-Flusberg 1988). Studies of the California Verbal Learning Test (CVLT), a list-learning task, have shown mixed results with children and adults with ASD. Minshew and Goldstein (1993) tested verbal memory in high-functioning autistic individuals between ages 12 and 40 years (mean = 21) and found increased intrusion errors on both the initial and interference lists, suggesting less self-monitoring, and lower recall of the interference list, suggesting susceptibility to proactive interference. Bennetto et al. (1996), in a study of intact and impaired types of memory in ASD, administered the

CVLT-C to a sample of children with ASD with a mean verbal IQ (VIQ) of 82.32. Results suggested that individuals with ASD displayed a flatter learning curve across trials than a group of individuals with learning disorders matched on sex, age, and VIQ. Phelan et al. (2011), in contrast, found no differences on CVLT Trials 1–5 recall, short- and long-delay free recall, and long-delay recognition of the initial list, when comparing CVLT verbal memory in ASD and TD children and adolescents matched on age and IQ. As no studies to date have assessed verbal memory in OO, it is not known whether this continues to be an area of vulnerability for individuals who have achieved OO, nor whether variation in verbal memory accounts for any variation in language skills in this group.

In addition to studies of verbal memory in ASD samples, studies of verbal memory in children with specific language impairment (SLI) indicate associations with language impairment (Mainela-Arnold et al. 2012). Since there is overlap between ASD and SLI (Kjelgaard and Tager-Flusberg 2001), verbal memory impairments in SLI may have implications for ASD. Baird et al. (2010) found that children with current language impairments showed more difficulty with verbal memory performance than children with no history of language impairment, and that these difficulties were still present in children with only a past history of language difficulties, although lower than those with current language impairment. Phonological memory seems to be a particularly strong marker of SLI; children with SLI consistently show deficits in nonword repetition tasks relative to TD children (Botting and Conti-Ramsden 2001; Coady et al. 2010) and the Baird et al. (2010) study found that children whose past language impairment had resolved still showed nonword repetition deficits as large as the group with current language impairment, suggesting that phonological repetition is a strong marker of language impairment even after surface language has normalized.

Determining whether verbal memory deficits and language deficits are closely related has been a topic of debate in the child language literature (Mainela-Arnold et al.

2010; Montgomery 2002). MacDonald and Christiansen (2002) have suggested that individual differences in language processing and verbal working memory tasks are simply different components of linguistic processing, and that both originate from variations in exposure to language and to biological differences that affect processing accuracy. Although the relationship between verbal memory and other aspects of language remains unclear, correlations between aspects of verbal memory and language abilities, such as sentence comprehension, have been reported in both typical and atypical development (e.g., Schuh and Eigsti 2012). Even in individuals with HFA who have language abilities in the average range, certain individuals may have difficulty with verbal memory tasks, and these deficits may also be tied to deficits in language and non-language areas (e.g., social skills; Schuh and Eigsti 2012). In children with SLI, Mainela-Arnold et al. (2012) have also found associations between short-term memory and vocabulary skills, and between working memory and syntax ability. Baird et al. (2010) showed that degree of language impairment was correlated with the level of overall verbal memory deficit, although this was not examined separately for each group (controls, past language impairment, current language impairment).

Although the processes of working memory (in which material must be held in mind and manipulated, rather than just repeated), phonological memory, and memory for meaningful units such as words, sentences, and narratives, can be theoretically distinguished, studies do not always make this distinction (e.g., de Abreu et al. 2011; Mainela-Arnold et al. 2010; Montgomery 2002). Therefore, we cannot make strong predictions about which aspect of memory would be impaired and which aspect in particular would be related to language in the groups we studied.

In the present study, we anticipated a relationship between verbal memory in HFA and these individuals' possible language deficits. We also expected that strengths in verbal memory in OO might be associated with better performance on language measures. If individuals with OO showed any residual deficits in verbal memory, we expected that they would display similar weaknesses to those found in studies of individuals with HFA. Thus, relative to TD individuals, we expected individuals with OO to have a flatter learning curve and less semantic clustering. Since individuals with OO are at risk for attention problems (Fein et al. 2005), we also expected they might display increased rates of intrusion and perseveration errors. Given the findings of Fein et al. (1996) and Gabig (2008) we expected that if children with HFA had residual verbal memory impairment, it would be with more semantically complex material (e.g., sentence repetition worse than nonword repetition).

The current study explores relationships among a set of language skills and verbal memory in a fully-characterized group of individuals with OO, including individuals ages 8–21, compared to individuals with typical development and with current HFA diagnoses. One primary goal was to determine whether the OO group displayed any residual deficits in language functioning that might have implications for core deficits in autistic language. Because we included older individuals than in previous studies of OO, we were also able to examine some additional complex language skills in the areas of syntactic and semantic functioning.

The current study expands on previous examinations of language abilities in a sample of OO children and adolescents using some of the same measures included in Kelley et al.’s (2006) and (2010) studies. In those papers, OO was defined similarly, but without requirements of an early language delay and diagnosis of ASD before age 5; in addition, those studies used no IQ and Vineland score cutoffs. This current sample includes 17 of the same children from Kelley et al. (2006), 16 of whom were also included in Kelley et al.’s (2010) paper.

Based upon the literature on OO language profiles, we predicted that OO individuals would exhibit grammatical and semantic capabilities that were indistinguishable from TD peers, and significantly better than the HFA group. We

expected that the OO group would display a similar verbal memory profile to the TD group, but that because verbal memory difficulties seem to persist in some children with past histories of language impairment and correlate with the degree of language impairment, verbal memory might still contribute more to any residual language difficulties in the OO than in the TD group.

**Method**

**Participants**

Thirty-four individuals with a history of ASD and who have now reached OO, 44 high-functioning individuals with a current ASD diagnosis (HFA), and 34 TD peers were tested. Participants ranged from 8 years, 1 month to 21 years, 8 months. The groups were matched on age, gender, and nonverbal IQ, but were significantly different on VIQ, with the OO and TD groups having a VIQ about 7 points higher than the HFA group (See Table 1). The gender ratio of the TD and HFA groups was about 10:1 boys to girls, while the OO group was about 4:1 boys to girls. Fisher’s exact tests comparing the gender ratio of the OO group to either of the other groups or the two combined were not significant.

**Table 1** Participant characteristics

	HFA	OO	TD	$F/\chi^2$	$p$	Tukey/Games-Howell	Cohen’s $d$
N	44	34	34				
Sex	40 M; 4 F	27 M; 7 F	31 M; 3 F	$\chi^2 = 2.92$	0.23		
Age	13.9 (2.7) 8:7–20:0	12.8 (3.5) 8:1–21:2	13.9 (2.6) 9:11–21:8	1.66	0.20		
VIQ	105.4 (14.4) 81–142	112.7 (13.7) 80–137	112.0 (11.2) 93–138	3.62	0.03	HFA < OO	HFA/OO: 0.52
NVIQ	110.2 (12.8) 78–147	110.3 (15.1) 81–142	112.8 (11.3) 89–139	0.45	0.64		
ADOS-Communication	3.50 (1.42) 2–7	0.47 (0.62) 0–2	0.41 (0.56) 0–2	124.20	<0.001	HFA > OO, TD	HFA/OO: 2.68 HFA/TD: 2.77
ADOS-Socialization	6.77 (2.21) 4–13	1.09 (1.31) 0–4	0.50 (0.75) 0–2	183.75	<0.001	HFA > OO, TD	HFA/OO: 3.07 HFA/TD: 3.66
ADOS-total	10.27 (3.13) 7–19	1.56 (1.71) 0–5	0.91 (1.14) 0–4	213.10	<0.001	HFA > OO, TD	HFA/OO: 3.38 HFA/TD: 3.84

Table reports means, followed by SDs and ranges

By ADOS classification, 21 individuals in the HFA group fall into the autism spectrum category (cutoff total score of 7); 23 fall into the autistic disorder category (cutoff total score of 10)

Age and nonverbal IQ were used as control variables because of the variation in cognitive level within each group, and across the autism spectrum more generally. Our efforts to match groups based on gender were designed to control for the fact that the majority of individuals diagnosed with ASD are male. Most of the participants were evaluated in Connecticut. Six HFA participants and three OO participants were evaluated in Kingston, Ontario, Canada. Their performance did not significantly differ from the larger sample on any measure. The participants tested in Connecticut were primarily from the northeastern US. Participants were predominantly Caucasian, with three OO individuals, two HFA individuals, and three TD individuals reporting other races or ethnicities. All participants were native English speakers. There were no significant group differences in family income among groups.

The study was approved by the Institutional Review Boards of the University of Connecticut, the Institute of Living Hartford Hospital, Children's Hospital of Philadelphia, and Queens University. Recruitment was done through media outlets (newspaper stories, radio interviews), private practices, and clinic referrals. In some cases, therapists contacted parents of children known to have optimal outcomes, and in some cases, parents saw media reports and contacted the investigators. Participants were recruited for the TD group through flyers distributed at local schools, grocery stores, and community centers.

#### *Inclusion Criteria*

All participants had verbal, nonverbal, and full-scale IQ standard scores greater than 77 (within 1.5 SD of the average of 100). Additional OO criteria were:

1. Participants had a documented ASD diagnosis made by a physician or psychologist specializing in autism before the age of 5, verified in a written diagnostic report provided by parents. Early language delay (no words by 18 months or no phrases by 24 months) documented in the report was required. As a second step in confirming diagnosis, the report was edited to remove any diagnosis, summary, and recommendations but leaving descriptions of behavior. One of the co-investigators (MB), an expert in diagnosis of ASD and Director of the University of Connecticut Psychological Services Clinic, reviewed these reports, blind to early diagnosis and current group membership. In addition to potential OO participants, she reviewed 24 "foil" reports for children with non-ASD diagnoses, such as global delay or language disorder. Four potential OO participants were rejected for insufficient early documentation, and were dropped from the study. All 24 foils were correctly rejected.

2. Participants could not currently meet criteria for any ASD according to the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000) administered by a research-reliable administrator. In addition, the ADOSs of all potential OO cases were reviewed by one of three clinicians with more than 15 years of autism diagnostic experience (IME, MB, or DF) who confirmed that ADOS scores were below ASD thresholds and that in their expert clinical judgment, an ASD was not present.
3. Participants' scores on the Communication and Socialization domains of the Vineland (see below) had to be greater than 77 (within 1.5 standard deviations of the mean of 100).
4. Participants had to be fully included in regular education classrooms with no one-on-one assistance and no special education services to address autism deficits (e.g., no social skills training). However, participants could be receiving limited special education services or psychological support to address impairments not specific to ASDs, such as attention or academic difficulties.

For the HFA group:

1. Following Collaborative Programs of Excellence in Autism diagnostic guidelines (Luyster et al. 2005), participants had to meet criteria for ASD on the ADOS (both Social and Communication domains and total score) and according to best estimate clinical judgment.

For the TD group:

1. Participants could not meet criteria for any ASD at any point in their development, by parent report.
2. Participants could not have a first-degree relative with an ASD diagnosis
3. Participants could not meet current diagnostic criteria for an ASD on the ADOS, or by clinical judgment (see Table 1). There was no attempt to exclude TD children for other learning or psychiatric disorders (but see general exclusion criteria).
4. Scores on the Communication and Socialization domains of the Vineland had to be greater than 77.

#### *Exclusion Criteria*

Potential participants for any group were excluded from the study if (1) at the time of the telephone screening they exhibited symptoms of major psychopathology (e.g., active psychotic disorder) that would impede full participation, (2) they had severe visual or hearing impairments, or (3) they had a history of seizure disorder, Fragile X syndrome, or significant head trauma with loss of consciousness. Two

in the TD group and two in the HFA group were excluded because of possible seizure disorder; none were excluded for other reasons.

### Procedure

Phone screenings based on study criteria were conducted with parents of each potential participant. Those who passed screening were scheduled for an assessment. For participants under 18, parent consent and child assent was obtained prior to testing. For participants 18 and over, their informed consent was obtained. The evaluation was administered in a quiet room over the course of two or three testing sessions at one of the university labs or in the home. Testing lasted approximately 6 h. In most cases, parent interviews were conducted concurrently by a second examiner and lasted approximately 3 h for the OO and HFA groups and 1.5 h for the TD group. Participants received a monetary incentive for participation, even if the testing could not be completed.

### Diagnostic Measures

The *ADOS* (Lord et al. 2000) was used to determine whether the participants met diagnostic criteria for ASD at the time of the study. *ADOS* administrations were videotaped and five administrations per group were coded by a rater blind to group status, with high inter-rater reliability for both algorithm (86.7 %) and total items (85.7 %).

The *Vineland Adaptive Behavior Scales* (Sparrow et al. 1984) is a parent report measure used to evaluate adaptive functioning across the domains of Communication, Daily Living Skills, and Socialization.

Cognitive abilities were assessed using the *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler 1999), which is a brief measure of verbal and nonverbal cognitive abilities.

### Language Measures

To assess semantic and syntactic aspects of language, participants completed the Core Language Battery of the *Clinical Evaluation of Language Fundamentals* (CELF-IV; Semel et al. 2003). Subtests assess the ability to listen to and follow complex, orally presented instructions (Concepts and Following Directions), to create sentences using unmarked verb, subject, and object prompts (Formulated Sentences), to repeat orally presented sentences (Recalling Sentences), to explain semantic relationships between pairs of words (Word Classes), and to define words (Word Definitions). The Concepts and Following Directions subtest was administered only to participants 12 years of age and younger, and the Word Definitions subtest was

administered only to participants older than 12 years of age.

The *Peabody Picture Vocabulary Test-Third Edition* (PPVT; Dunn and Dunn 1997) was used to assess receptive vocabulary knowledge.

The Nonword Repetition subtest of the *Comprehensive Test of Phonological Processes* (CTOPP; Wagner et al. 1999) requires participants to repeat word-like nonsense words (phonological memory).

To evaluate verbal learning and memory, participants completed the *California Verbal Learning Test, 2nd Edition* (CVLT-II; Delis et al. 2000). Participants younger than 17 years of age completed the *California Verbal Learning Test, Children's Version* (CVLT-C; Delis et al. 1994). The CVLT is a widely-used neuropsychological measure that evaluates use of strategies in recalling words from two lists, with words being drawn from specific semantic categories. List A is administered five times, with learning assessed after each trial, followed by a single presentation of List B. Recall for List A is then reassessed, followed by recall cued by category. After an approximately 20 min delay, memory for List A is assessed via free recall, cued recall, and recognition. In addition to recall and recognition scores, learning characteristics such as semantic and serial clustering and errors of perseveration and intrusion are also calculated.

### Results

One-way between-group analyses of variance were conducted to examine group differences on the language measures. Tukey's HSD post hoc tests were used to determine specific between-groups differences for most measures. When Levene's test for homogeneity of variances was found to be violated, the Games-Howell post hoc test was used.

On most of the language measures, all groups (OO, HFA, and TD) earned mean scores in the average range. Data were analyzed using VIQ as a covariate to ensure that group differences on the language and verbal memory tasks were not better accounted for by group differences in overall verbal ability. Except where noted, all pairwise group differences remained significant at the  $p = 0.05$  level when VIQ was added as a covariate. One primary aim of the analyses was to look for even small differences between the OO and TD groups, to determine how truly 'typical' their language has become. To adopt the most conservative approach, allowing the detection of even small differences, we adopted no correction for multiple comparisons.

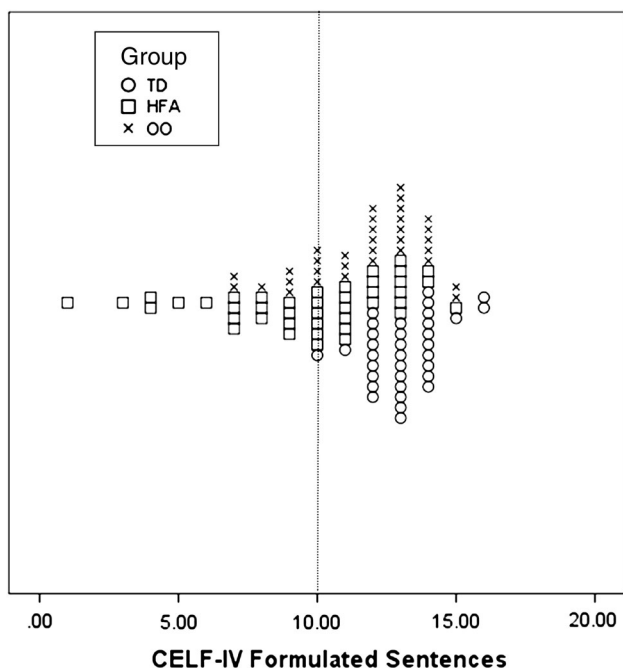
In order to examine the possibility of distinct subgroups, even in the absence of mean differences between groups,

we calculated the frequency by group of subtest scores falling below 1.5 SDs of the mean. We then conducted Chi square tests to examine if there was a significant difference in the frequency of these scores for each group. One difficulty that arises when a normal control group performs above the population mean is how to interpret group differences, when all groups' mean scores may be within the normal range. In order to better illustrate how scores within this particular sample compare to the normative samples for some of the key measurements used, we have presented scatterplots for certain scores. Figures 1, 2 and 3 show the frequency distributions for CELF-IV Formulated Sentences, CVLT List A Trials 1–5, and CVLT Discriminability scores.

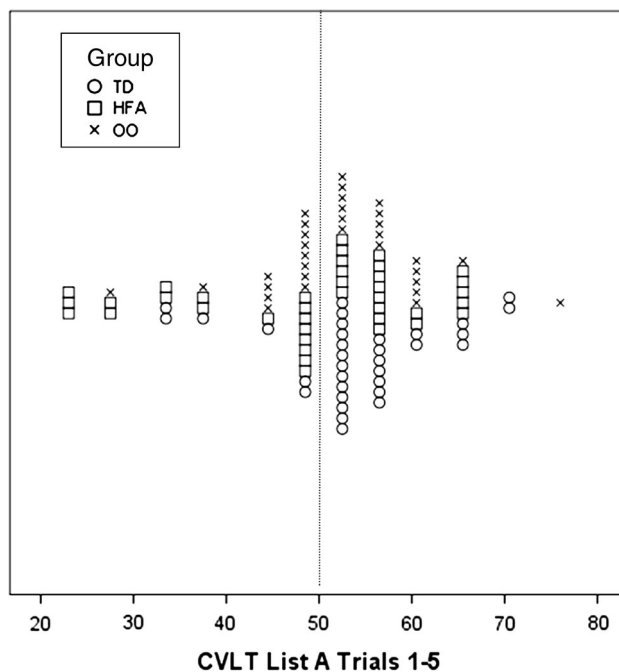
In addition, we conducted a series of multiple regression analyses to examine how aspects of verbal memory and autism symptomatology relate to language ability, as measured by the CELF Core Language Composite. Regressions were run separately for each group to see which independent language variables contributed to variance in the composite CELF-IV language score. The variables chosen as predictors were verbal memory (CVLT Trials 1–5), receptive vocabulary (PPVT), phonological memory (CTOPP nonword repetition), and Communication (ADOS Communication). A primary aim was to assess the degree to which even subtle differences in verbal memory and autism symptoms may explain group differences in language abilities.

Syntactic and Semantic Abilities

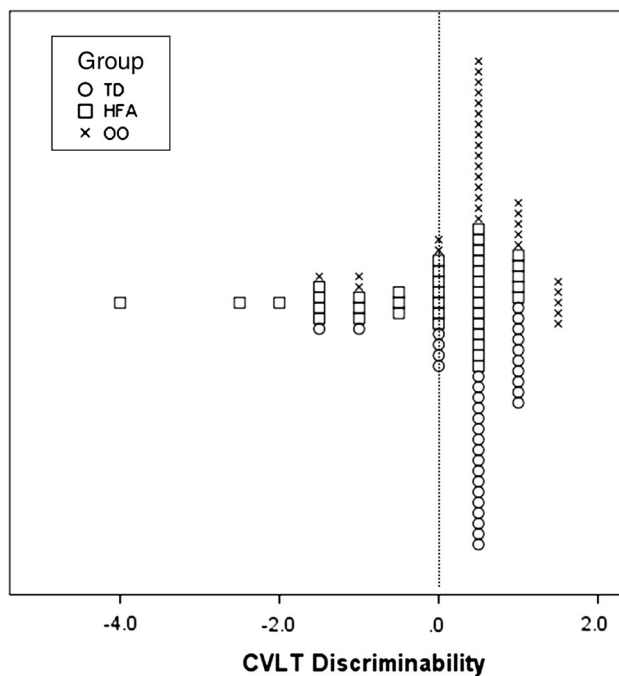
On the Core Language Composite of the CELF-IV, all three groups differed significantly from one another, with the best



**Fig. 1** Plot of CELF-IV Formulated Sentences scores by group. Note The vertical line marks the mean score of 10 for the CELF-IV normative sample



**Fig. 2** Plot of CVLT List A Trials 1–5 t-scores by group. Note The vertical line marks the mean t score of 50 for the CVLT normative sample



**Fig. 3** Plot of CELF-IV discriminability scores by group. Note The vertical line marks the mean z-score of 0 for the CVLT normative sample

performance by TD, followed by OO, followed by HFA. Post-hoc testing revealed significant differences between the HFA and OO groups at  $p = 0.03$ , between HFA and TD at  $p < 0.01$ , and between OO and TD at  $p < 0.01$ . On the Concepts and Following Directions subtest of the CELF-IV, there was a marginally significant overall difference among the groups, with a marginally significant ( $p = 0.053$ ) difference between the HFA and TD groups' performance in post hoc testing. Because this subtest is administered only to individuals twelve and younger, the sample sizes for this ANOVA were smaller than for the other subtests. Although the effect was only marginal, the Cohen's  $d$  statistic indicates a medium effect size. Only one participant in the HFA group, and none in the other groups, performed lower than one SD below the mean on this subtest.

Analysis of the remaining subtests of the CELF-IV revealed statistically significant between-group differences, though all groups had mean scores within the average range on all subtests. There was a significant difference in scaled scores on the Formulated Sentences subtest, reflecting the fact that the mean score for the TD group was significantly higher than the scores for both the OO group,  $p < 0.01$  and HFA group,  $p < 0.01$ . The OO group's scores were also significantly higher than those of the HFA group at  $p < 0.01$ . On the Recalling Sentences, Word Classes, and Word Definitions subtests, there was a significant difference in scaled scores, with the TD group scoring significantly higher than the HFA group, but the OO group not differing from either the OO or the HFA groups (see Table 2 for means and SDs).

**Table 2** Comprehensive Evaluation of Language Fundamentals, 4th Edition (CELF-IV) findings

	HFA	OO	TD	<i>F</i>	<i>p</i>	Tukey/Games-Howell	Cohen's <i>d</i>
<i>N</i>	43	33	34				
Concepts & Following Directions	10.50 (2.75) 4–13 <i>N</i> = 13 8 %	10.73 (1.67) 8–13 <i>N</i> = 16 0 %	12.00 (1.73) 9–14 <i>N</i> = 11 0 %	3.07	0.06	TD > HFA, $p = 0.053$	HFA/TD: 0.67
Formulated Sentences	9.67 (3.21) 1–15 <i>N</i> = 42 12 %	11.67 (2.19) 7–15 0 %	13.12 (1.27) 10–16 0 %	19.22	<0.01	TD > OO > HFA	HFA/OO: 0.72 HFA/TD: 1.38 OO/TD: 0.83
Recalling Sentences	9.44 (2.99) 4–15 12 %	10.27 (2.81) 4–14 0 %	11.97 (1.77) 8–15 0 %	9.01	<0.01	TD > HFA	HFA/TD: 1.01
Word Classes Total	9.88 (3.26) 1–16 10 %	11.74 (2.73) 3–18 3 %	13.06 (1.97) 9–17 0 %	12.89	<0.01	TD > HFA	HFA/TD: 1.16
Word Definitions	11.43 (2.70) 6–16 <i>N</i> = 30 0 %	13.24 (2.14) 7–16 <i>N</i> = 17 0 %	13.78 (0.99) 12–16 <i>N</i> = 23 0 %	8.72	<0.01	TD > HFA	1.12
Core Language Composite Standard Score	100.00 (13.79) 70–126 <i>N</i> = 30 0 %	108.35 (11.59) 79–126 <i>N</i> = 31 0 %	117.15 (7.06) 106–132 <i>N</i> = 33 0 %	20.78	<0.01	TD > OO > HFA	HFA/OO: 0.67 HFA/TD: 1.61 OO/TD: 0.94

Table reports means, followed by SDs, ranges, and the percentage of individuals with subtest scores lower than 5.5 (mean minus 1.5 SD). All subtest scores are scaled scores with a mean of 10 and SD of 3. Standard scores have a mean of 100 and SD of 15. The CELF-IV Core Language Composite is a standard score. *N*s are as noted at the top of the table, unless otherwise specified



For the Formulated Sentences test only, there was a significant difference among the groups in the percentage of participants with scores below 1.5 SD from the mean,  $\chi^2(2, N = 40) = 8.09, p = 0.02$ . The HFA group had 5 scores in this range, and the other groups none. All CELF-IV data are presented in Table 2.

In summary, the three groups had overall differences in core language scores, in the expected direction. The OO and TD groups also differed on the Formulated Sentences score, but not on any other subtests. Similarly, the OO and HFA group differed on performance for this subtest, but no other subtests. The TD and HFA groups' scores differed on all core language subtests.

### Receptive Vocabulary

On a measure of receptive vocabulary, the PPVT-4, there was a significant main effect of group; data are shown in Table 3. The TD group performed significantly better than the HFA,  $p < 0.01$  and OO groups,  $p = 0.05$  (Table 3). All groups' mean scores were above average.

### Phonological memory

On the Nonword Repetition subtest of the CTOPP (Table 3), there was a significant difference in mean scores (Table 3). Post-hoc tests revealed that the TD group performed significantly better than the HFA group,  $p < 0.001$  (Table 3). The OO group had an intermediate score and did not differ from the TD or HFA group.

### Verbal Learning and Memory

For the CVLT results, we have presented recall memory, learning characteristics, and recognition memory scores (Tables 4, 5, 6, 7) The Total Trials score, which assesses learning over five trials, is a  $t$ -score, while the remaining

scores are  $z$ -scores (mean of 0 and SD of 1). As was true with results for the CELF-IV, group means were all in the average range.

Within the recall scores, there was a significant difference among groups on List A Total Trials 1–5 (Table 4). Post-hoc testing showed that the TD group's performance was significantly better than that of the HFA group's,  $p = 0.05$  (Table 4). However, this finding was not significant when the ANOVA was run with VIQ as a covariate. Both with and without VIQ as a covariate, we found a significant between-group difference for List A Short-Delay Cued Recall, and post hoc testing indicated that the TD group's performance was again significantly better than that of the HFA group,  $p = 0.03$  (Table 4), with the OO group showing intermediate performance.

Among the recall error scores, there were several between-group differences. There was a significant between-group difference in the mean number of perseverations (i.e., saying a word twice or more during a single trial) (Table 4), with more perseverations in the HFA group than in the TD group,  $p = 0.03$ . Similarly, there was a significant difference among the groups for total free- and cued-recall intrusions (Table 4), with more intrusions in the HFA than in the TD group,  $p < 0.01$ .

Of the Chi square tests examining group differences in frequency of recall scores lower than 1.5 SDs from the mean, the List A Short-Delay Cued Recall yielded the only statistically significant test ( $\chi^2(2, N = 102) = 8.83, p = 0.01$ ), with the HFA group showing more low scores. Although the HFA group did have a higher frequency of intrusions and perseverations than the TD and OO groups, none of the Chi square tests comparing frequencies of below average recall error scores was statistically significant (Table 5).

No between-group differences were found for Semantic-Cluster Ratio, Serial-Cluster Ratio, or Learning Slope (Table 6). Although some individuals within each group

**Table 3** Additional language measures (PPVT, CTOPP) findings

	HFA	OO	TD	$F$	$p$	Tukey	Cohen's $d$
$N$	41	31	34				
PPVT	110.79 (14.40) 81–138 0 %	112.85 (12.78) 86–130 0 %	120.21 (10.05) 102–143 0 %	5.54	0.01	HFA < TD, OO < TD	HFA/TD: 0.76 OO/TD: 0.65
CTOPP Nonword Repetition	9.19 (2.54) 4–15 5 %	10.15 (1.87) 7–13 0 %	10.97 (2.08) 7–16 0 %	6.03	<0.01	HFA < TD	HFA/TD: 0.77

This table reports means, with standard deviations ranges, and the percentage of individuals with subtest scores lower than the mean score minus 1.5 SDs. All subtest scores are scaled scores.  $N$ s as noted at the top of the table, unless otherwise noted

**Table 4** California verbal learning test (CVLT-C or CVLT-II) recall scores

	HFA	OO	TD	<i>F</i>	<i>p</i>	Tukey/ Games-Howell	Cohen's <i>d</i>
<i>N</i>	40	31	33				
List A Total Trials 1–5	48.64 (12.13) 22–67	52.52 (8.80) 29–76	54.45 (8.67) 34–72	3.08	0.05*	HFA < TD	HFA/TD: 0.55
List B Free Recall	–0.18 (1.08) –2.5 to 2.5	0.36 (0.91) –1.0 to 2.5	0.09 (0.96) –2.0 to 2.5	2.51	0.09		
List A Short-Delay Free Recall	–0.22 (1.13) –3.0 to 2.0	–0.03 (0.88) –2.0 to 2.0	0.18 (0.75) –2.0 to 2.0	1.67	0.19		
List A Short-Delay Cued Recall	–0.18 (1.17) –3.0 to 2.0	0.13 (0.80) –1.5 to 1.5	0.42 (0.83) –1.0 to 2.0	3.49	0.03	HFA < TD	HFA/TD: 0.59
List A Long-Delay Free Recall	–0.15 (1.32) –4.0 to 2.0	0.08 (1.00) –3.0 to 2.0	0.39 (0.85) –2.0 to 2.0	2.24	0.11		
List A Long-Delay Cued Recall	–0.16 (1.35) –5.0 to 2.0	–0.03 (0.93) –2.0 to 2.0	0.36 (0.79) –2.0 to 2.0	0.48	0.62		
Perseverations (Free & Cued Recall Total)	0.23 (1.18) –1.0 to 4.5	–0.07 (0.87) –1.0 to 2.5	–0.36 (0.78) –1.0 to 1.5	3.32	0.04	HFA > TD	HFA/TD: 0.59
Free-Recall Intrusions (Total)	0.42 (1.23) –1.0 to 5.0	0.11 (0.90) –1.0 to 3.5	–0.29 (0.33) –0.50 to 0.50	5.28	0.01	HFA > TD	HFA/TD: 0.77

\* Result is non-significant when VIQ added as a covariate. List A Total Trials is a *t*-score, with *M* = 50, *SD* = 10. All other scores are *z*-scores, with *M* = 1, *SD* = 1

**Table 5** Percentage of CVLT level of recall scores lower than –1.5 SDs from the mean

	HFA	OO	TD
<i>N</i>	40	31	33
List A Total Trials 1–5	15 % (n = 6)	3 % (n = 1)	3 % (n = 1)
List B Free Recall	13 % (n = 5)	0	6 % (n = 2)
List A Short-Delay Free Recall	15 % (n = 6)	10 % (n = 3)	3 % (n = 1)
List A Short-Delay Cued Recall	20 % (n = 8)*	6 % (n = 2)	0
List A Long-Delay Free Recall	15 % (n = 6)	13 % (n = 4)	6 % (n = 2)
List A Long-Delay Cued Recall	20 % (n = 8)	16 % (n = 5)	6 % (n = 2)
Perseverations (Free & Cued Recall Total)	13 % (n = 5)	6 % (n = 2)	3 % (n = 1)
Intrusions (Free & Cued Recall Total)	13 % (n = 5)	6 % (n = 2)	0

\* A Chi square test indicated that this percentage of scores differed significantly from the percentages found in the other two groups (*p* = 0.01). No other proportions differed significantly

scored below average on these domains, there were no group differences in frequency of low scores (Table 7).

On the recognition memory portion of the CVLT, we found significant between-group differences for hits (correct endorsement of an item) and discriminability (the individual's ability to discriminate targets from distractors) (Table 6). For recognition hits, a post hoc test showed that

the OO group performed significantly better than the HFA group, *p* = 0.03 (Table 6). This finding was marginally significant when VIQ was added as a covariate. On discriminability, there was also a significant difference in mean scores, with poorer performance in the HFA group than both the TD (*p* < 0.01) and OO groups (*p* < 0.01) (Table 6). However, the HFA group was still performing

**Table 6** California verbal learning test (CVLT-C or CVLT-II) learning characteristics and recognition measures

	HFA	OO	TD	<i>F</i>	<i>p</i>	Games-Howell	Cohen's <i>d</i>
<i>N</i>	40	31	33				
Semantic-Cluster Ratio	0.01 (1.61) –2.5 to 3.5	–0.11 (1.15) –2.0 to 2.5	–0.09 (1.38) –3.0 to 4.0	0.08	0.92		
Serial-Cluster Ratio	0.17 (1.65) –1.5 to 5.0	0.05 (1.03) –1.5 to 3.5	–0.21 (1.51) –1.5 to 5.0	0.10	0.90		
Learning Slope	–0.38 (1.18) –3.0 to 2.0	–0.37 (1.15) –3.0 to 2.0	–0.24 (0.99) –2.0 to 2.0	0.17	0.85		
Correct Recognition Hits	–0.14 (1.13) –5.0 to 1.0	0.40 (0.50) –1.0 to 1.0	0.23 (0.66) –2.0 to 1.0	3.76	0.03*	HFA < OO	HFA/OO: 0.60
Discriminability	–0.18 (1.10) –4.0 to 1.0	0.55 (0.72) –1.5 to 1.5	0.49 (0.55) –1.5 to 1.0	8.15	<0.01	HFA < OO, HFA < TD	HFA/TD: 0.76 HFA/OO: 0.78
Recognition Discriminability versus Long-Delay Free Recall	–0.01 (1.05) –2.5 to 2.0	0.48 (0.85) –1.5 to 2.0	0.09 (0.63) –1.0 to 1.5	2.91	0.06*		

This table reports means, with standard deviations followed by ranges in parentheses. All scores are *z*-scores, with *M* = 1, *SD* = 1

\* Result becomes significant when VIQ is added as a covariate

**Table 7** Percentage of CVLT recognition measure and contrast scores lower than –1.5 SDs from the Mean

	HFA	OO	TD
<i>N</i>	40	31	33
Semantic-Cluster Ratio	28 % (n = 11)	16 % (n = 5)	12 % (n = 4)
Serial-Cluster Ratio	5 % (n = 2)	13 % (n = 4)	18 % (n = 6)
Learning Slope	25 % (n = 10)	26 % (n = 8)	15 % (n = 5)
Correct Recognition Hits	13 % (n = 5)	0	6 % (n = 2)
Discriminability	18 % (n = 7)	3 % (n = 1)	3 % (n = 1)
Recognition Discriminability versus Long-Delay Free Recall	8 % (n = 3)	3 % (n = 1)	0

within the average range. The OO group achieved the highest score on discriminability, suggesting that their verbal memory is more consistent with their VIQ than is the case for the other two groups. A Chi square test comparing frequencies of low scores for discriminability was significant ( $\chi^2(2, N = 102) = 6.54, p = 0.04$ ), with the HFA group having the most low scores and OO and TD groups having one participant each with a score in this range (Table 7). In addition, on the Recognition Discriminability versus Long Delay Free Recall, the OO group scored higher than the other two groups; their high score on this variable suggests that the OO group encoded and retained the list, but did relatively better on recognition than free recall, suggesting a minor weakness in retrieval without cues.

Note that the CELF-IV Recalling Sentences also involves verbal memory processes; on this measure, the HFA group was also significantly lower than the TD group, with a large effect size (see Table 2).

In summary, in the general language and verbal memory domains, individuals in the HFA group showed the poorest performance of the three groups, while still performing in the average range. Thus, for all performance domains, group means were well within the average range. The HFA group was lower than the TD group on Recalling Sentences, even when covarying VIQ. On the CVLT, on List A Short-Delay Cued Recall, the TD group performed better than the HFA group. We did not see expected differences among the groups on Semantic- or Serial-Cluster Ratio, or on Learning Slope. In an examination of recall errors, the

HFA group had higher frequencies of low scores for intrusions and perseverations than the TD and OO groups. On recognition memory, the OO group performed better than the HFA group for overall correct recognition hits, and their recognition exceeded their free recall more than the other two groups. The HFA group showed significantly weaker performance on distinguishing targets from distractors compared to the TD and OO groups.

Regression Analyses

Regression analyses were used to examine the independent contributions of phonological memory (CTOPP nonword repetition), verbal learning (the CVLT score for Trials 1–5), vocabulary (PPVT score), and symptom severity (ADOS total scores), to CELF core language ability scores. We predicted that the relationship between phonological memory and CELF-IV composite would be stronger in the HFA than the OO and TD groups, given the relationship between phonological memory and language impairment.

For the TD group, vocabulary and ADOS total score each contributed significant variability to overall CELF score; all statistics are presented in Table 8. We should note that the somewhat surprising finding for ADOS score in the TD group was driven entirely by two individuals with relatively lower CELF scores (105 and 109) and relatively higher symptom counts (though still under the diagnostic threshold; total of 5 and 5, respectively); with those two participants excluded, the ADOS was no longer a significant predictor,  $B = .158, t = .35, p = 0.73$ . For the HFA group, vocabulary contributed significant variability; there was a strong trend as well for the predictive value of verbal memory,  $p = 0.06$ . Finally, for the OO group, as for

the TD and HFA groups, vocabulary was a significant predictor; in addition, however, both verbal memory and phonological memory made independent and significant contributions.

Summary of Findings

In conclusion, the HFA group, consistent with expectations, generally performed more poorly than the OO or TD groups on measures of language and verbal memory. In particular, the HFA group was lower functioning than the TD group, even when covarying VIQ, with large effect sizes on overall CELF score and CELF subtests (Recalling Sentences, Formulating Sentences, Word Classes, Word Definitions), and with medium effect sizes on CELF Concepts and Following Directions, PPVT, CTOPP Nonword Repetition, and multiple CVLT variables. It is important to point out, though, that none of the groups' scores indicated clinically significant impairment (i.e., below average mean scores).

Our results indicate that the individuals in the OO group, many of whom previously had significant difficulties with language, are now performing well across all domains. Only three variables, Formulating Sentences, Composite CELF score (both large effect sizes), and the PPVT (medium effect size) showed OO group means lower than TD means, but OO scores were still above average on all of these measures. The OO group was not different from the TD group on any CELF or CVLT measure of verbal memory, except that their superiority on long delay recognition over free recall was greater than the other groups. Furthermore, for most core language measures, no individuals in the OO group scored below average.

**Table 8** Summary of multiple regression analyses predicting CELF Core Language Composite, ordered by group

Group	Predictor variable	B	SE B	$\beta$	$t$	$p$	Model $R^2$
TD ( $n = 31$ )	Short-term phonological memory <sup>a</sup>	.522	.517	.155	1.009	0.32	
	Verbal memory <sup>b</sup>	-.074	.128	-.082	-.58	0.57	
	ADOS total score	-1.77	.792	-.349	-2.23	0.04	
	PPVT**	.357	.107	.492	3.35	0.003	.571***
HFA ( $n = 35$ )	Short-term phonological memory	-.151	.647	-.029	-.234	0.817	
	Verbal memory <sup>‡</sup>	.251	.128	.237	1.96	0.060	
	ADOS total score	.158	.456	.047	.346	0.732	
	PPVT***	.649	.112	.740	5.77	0.000	.626***
OO ( $n = 28$ )	Short-term phonological memory**	2.650	.769	.461	3.448	0.002	
	Verbal memory*	.362	.174	.292	2.086	0.048	
	ADOS total score	-.198	.697	-.040	-.285	0.778	
	PPVT**	.330	.118	.384	2.804	0.010	.613***

<sup>‡</sup>  $p < 0.10$ ; \*  $p < 0.05$ ; \*\*  $p < 0.01$

<sup>a</sup> CTOPP nonword repetition

<sup>b</sup> CVLT Trials 1–5

Interestingly, linear regression analyses indicated that strong verbal memory and short-term phonological memory abilities seemed to make an important contribution to their language abilities, even more than in the HFA group.

## Discussion

This study aimed to add to the previous literature on language in a group of children and adolescents with a history of ASD who have achieved “optimal outcomes” (Kelley et al. 2006, 2010). We compared syntactic, semantic, and verbal memory skills of these individuals to those of individuals with TD and HFA.

Given that the OO group no longer met formal criteria for an ASD, we expected that any residual deficits in language and verbal memory would be subtle, and that the OO group would perform significantly better across language measures than the HFA group. Overall, as determined through standardized measurements, the results indicate no residual language weaknesses for the OO group and, more surprisingly, few significant language deficits in the HFA group. All group means fell well within the average range. No doubt at least some of these findings were due to the stringent inclusion criteria for the HFA group, which included both verbal and nonverbal IQ scores within the normal range, and thus resulted in a group of *very* high functioning individuals. However, they were found to be lower functioning on multiple language measures than both of the other groups, even when covarying VIQ, indicating residual language weaknesses in this HFA group.

Consistent with Kelley et al.’s (2010) findings regarding grade-school children with OOs, this older and larger sample of OO children and adolescents is performing remarkably well. On the three variables on which the OO group scored significantly lower than the TD group (CELF overall score, CELF Formulated Sentences, and PPVT), the OO mean was above average and *no* OO individual scored below average. On several CVLT learning characteristics, a small number of OO individuals scored below average, but not significantly more than in the TD group.

Although this study does not indicate that the individuals we studied have residual deficits on any language or verbal memory tests, the findings do suggest a potential need for examining subtle differences in these skills and in strategies used to complete language and memory tasks. The findings indicate that OO individuals’ symptoms have resolved almost entirely from the point of view of standardized tests, but that they may still have some differences that would perhaps be clearer in targeted psycholinguistic tasks (e.g., Naigles et al. 2013) and/or naturalistic, everyday situations, which we did not measure. Similarly, findings for the HFA group, which suggest most individuals are

performing within the average range, still indicate that multiple areas of language functioning are lower than would be expected from their VIQ, and have contributions from their autism social and communication symptoms and their verbal memory.

In this discussion, we cover comparisons of the OO and HFA group first, followed by OO versus TD comparisons, and then TD versus HFA comparisons.

There were few statistically significant differences between the OO and HFA groups. In their ability to formulate sentences and overall language skills on the CELF, OO individuals did perform significantly better than HFA individuals, most likely resulting in statistically significant findings because of the high scores of the OO group. Performance on this task was more heterogeneous in the HFA group, suggesting that some individuals with HFA may lag behind their FSIQ-matched OO peers in language tasks that assess more subtle aspects of language functioning. Differences between the HFA and OO group on the CVLT correct recognition hits and discriminability scores were likely due to the very high scores of the OO group on these verbal memory domains.

The OO group’s scores differed from the TD group across more domains than from the HFA group. The OO group’s average scores for semantic and syntactic skills were slightly lower than those of the TD group, though not so low as to result in statistically significant differences in most cases. On the Formulated Sentences and the Core Language Composite, however, there were statistically significant differences between the OO group and TD group. As with the OO versus HFA difference in these scores, these differences were likely due to the very high scores of the TD group, as all of the OO scores were average or above. An alternative explanation may be that the OO group was using slightly different strategies for formulating sentences from the TD group, such that their performance was lower than that of the TD group. These strategies may not be detectable using standardized language measures, but perhaps could be measured through targeted psycholinguistic tasks, imaging tasks or other means.

In receptive vocabulary, the TD and OO groups also differed due to the very high scores of the TD group, since both groups scored above average. These differences may be an indication of subtle processing differences that are not fully reflected in our findings, due to limitations in these standardized measures (Naigles et al. 2013). Medium to large effect sizes for all of these findings (Table 3) suggest that these results are robust and would hold were we to examine larger samples.

There were no statistically significant differences between the OO and TD groups in verbal memory. However, it is interesting to note that the OO group performed

at slightly higher levels on average on correct recognition hits and discriminability. The OO group also performed well in semantic clustering indicating that they were using good strategies for verbal working memory.

As would be expected, the TD and HFA groups differed most markedly in this sample. Significant differences between the TD and HFA group on Concepts and Following Directions, and Formulated Sentences, which contribute to the Core Language Composite may be due in part to deficits in self-monitoring and attention in the HFA individuals. Individuals in this HFA sample displayed significant difficulties with attention and self-monitoring, both as observed during testing and reported by parents (executive functioning data will be reported separately). The HFA group means on the current language measures were still well within the average range. Previous research suggesting a trajectory of ASD transitioning into ADHD may be consistent with this explanation (Fein et al. 2005). The HFA group's scores on receptive vocabulary and phonological processing were also significantly lower than those of the TD group, possibly due to the TD group's very high scores.

For verbal memory, a considerable number of individuals in the HFA group received below average scores on semantic clustering, which indicates passive reception of the words, rather than processing for meaning. As indicated by the HFA group's mean score for free recall intrusions, these individuals also appear to be doing less self-monitoring and self-correction than the TD group. Although all of the HFA group's mean verbal memory scores on the CVLT were solidly average (within .5 SD of the mean), the HFA group did produce significantly fewer correct responses than the TD group on List A Total Trials 1–5, resulting in a significantly lower T-score. The HFA group also tended to show more perseverations and intrusions than the TD group, suggesting that these skills are not quite commensurate with the high VIQs of the HFA group (though still within the average range) and may be reflecting less self-monitoring or a slight impairment in source memory.

While these results show that strong language and verbal memory skills may be part of the profile of individuals who lose an ASD diagnosis, verbal memory did not sharply differentiate the HFA and OO groups. Furthermore, our regression analysis suggested that even subtle differences in individuals' verbal memory and ASD symptom profiles may be related to differences in language outcomes.

Based upon these regression analyses, verbal memory (as measured by CVLT score) and phonological memory (nonword repetition) contributed to overall linguistic ability for only the clinical groups (OO and HFA) and more strongly for the OO group. There was no overall verbal memory or phonological memory contribution to linguistic

ability for the TD group; ADOS summed symptoms predicted language ability in the TD group only, which was a reflection of the relatively lower CELF scores of just two participants. Single word receptive vocabulary was a predictor for all groups, likely because children with bigger vocabularies are more likely to understand the lexical items in the CELF, and so can devote more resources to processing the grammatical components. The results are consistent with the possibility that the OO group made gains in language skills in part by relying more heavily on verbal memory abilities, which also played a role in language skills in HFA. This is consistent with prior research showing the critical role played by verbal memory in language acquisition (de Abreu et al. 2011; Gathercole 2006). Children who can hold sentences in short-term memory longer can analyze those sentences into their components and can better understand the organization of those sentences' components. Children with better short-term memory also might be expected to have more stable lexical representations, such that incoming words are more efficiently recognized, categorized, and parsed.

#### Limitations and Future Directions

The results of this study of language and verbal memory profiles in individuals who have achieved OO should be interpreted with some caution for the following reasons. The sample was ethnically homogeneous and included only children with very high-functioning forms of ASD; in addition, parents of individuals in all groups were invested in research participation, and completed hours of testing. Across all three groups in this study, mean cognitive scores were in the average range, which also limits generalizability of these findings to individuals with autism who have more limited cognitive and language abilities. In addition to sample limitations, the study included only standardized measures of language ability, which may not be able to capture subtle residual difficulties in functioning. For example, naturalistic conversation tasks or analysis of a spontaneously produced narrative or conversation may reveal group differences not apparent in the analyses presented here; analyses of these data are ongoing.

Future research in the language and verbal memory functioning of individuals with optimal outcomes should include prospective examination of the trajectory of language development, about which we have a somewhat limited understanding in ASD. Research should also concentrate on describing the language of these individuals in non-structured environments and tasks, which are most likely to elicit the types of difficulties we might expect in individuals with ASD or a history of ASD.

A crucial set of questions concern the reasons for which some individuals achieved OO, while some remain on the

autism spectrum, even while functioning cognitively in the normal range. We are currently examining intervention histories, and analyzing structural and functional imaging data, to try to shed light on this question.

In the current sample, individuals with OO and a clear history of autism and language delay demonstrated normal language functioning. The outcomes of these individuals are a promising area of study for future researchers and for interventionists, who should continue to elucidate the underlying reasons for such outcomes.

**Acknowledgments** The authors are very grateful to the participants and their families, to Dr. Lynn Brennan and Harriet Levin for help with recruitment, to our invaluable undergraduate research assistants, and for our Grant funding: NIH R01 MH076189.

## References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., DSM-IV-TR). Washington, DC: American Psychological Association.
- Baird, G., Dworzynski, K., Slonims, V., & Simonoff, E. (2010). Memory impairment in children with language impairment. *Developmental Medicine and Child Neurology*, *52*, 535–540.
- Bennetto, L., Pennington, B. F., & Rogers, S. J. (1996). Intact and impaired memory functions in autism. *Child Development*, *67*, 1816–1835.
- Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, *36*, 421–432.
- Boucher, J. (2012). Structural language in autistic spectrum disorder—characteristics and causes. *Journal of Child Psychology and Psychiatry*, *53*, 219–233.
- Coady, J., Evans, J. L., & Kluender, K. R. (2010). Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International Journal of Language & Communication Disorders*, *45*, 494–509.
- Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology*, *12*, 349–356.
- de Abreu, P., Gathercole, S., & Martin, R. (2011). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. *Learning and Individual Differences*, *21*, 569–574.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *California verbal learning test, children's version*. San Antonio, TX: Psychological Corporation.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California verbal learning test* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Services.
- Dunn, M., Gomes, H., & Sebastian, M. (1996). Prototypicality of responses in autistic, language-disordered, and normal children in a word fluency task. *Child Neuropsychology*, *2*, 99–108.
- Eigsti, I. M., & Bennetto, L. (2009). Grammaticality judgments in autism spectrum disorders: Deviance or delay. *Journal of Child Language*, *19*, 1–23.
- Eigsti, I.-M., Bennetto, L., & Dadlani, M. B. (2007). Beyond pragmatics: Morpho-syntactic development in autism. *Journal of Autism and Developmental Disorders*, *37*, 20–33.
- Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*, 681–691.
- Fein, D., Barton, M., Eigsti, I., Kelley, E., Naigles, L., Schultz, R. T., et al. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, *54*, 195–205.
- Fein, D., Dixon, P., Paul, J., & Levin, H. (2005). Brief report: Pervasive developmental disorder can evolve into ADHD: Case illustrations. *Journal of Autism and Developmental Disorders*, *35*, 525–534.
- Fein, D., Dunn, M., Allen, D. A., Aram, D. M., Hall, N., Morris, R., et al. (1996). Language and neuropsychological findings. In I. Rapin (Ed.), *Preschool children with inadequate communication* (pp. 123–154). London: MacKeith Press.
- Gabig, C. (2008). Verbal working memory and story retelling in school-age children with autism. *Language, Speech, and Hearing Services in Schools*, *39*, 498–511.
- Gathercole, S. E. (2006). Complexities and constraints in nonword repetition and word learning. *Applied Psycholinguistics*, *27*, 599–613.
- Goodwin, A., Fein, D., & Naigles, L. (2012). Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. *Autism Research*, *5*, 109–123.
- Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: A four-to six-year follow-up. *Journal of Autism and Developmental Disorders*, *30*, 137–142.
- Helt, M., Kelley, E., Kinsbourne, M., Pandey, J., Boorstein, H., Herbert, M., et al. (2008). Can children with autism recover? If so, how? *Neuropsychology Review*, *18*, 339–366.
- Hermelin, B., & O'Connor, N. (1970). *Psychological experiments with autistic children*. New York, NY: Pergamon.
- Jordan, R. (1993). The nature of the linguistic and communication difficulties of children with autism. In D. J. Messer & G. T. Turner (Eds.), *Critical influences on child language acquisition and development* (pp. 229–249). New York: St. Martin's Press.
- Kelley, E., Naigles, L., & Fein, D. (2010). An in-depth examination of optimal outcome children with a history of autism spectrum disorders. *Research in Autism Spectrum Disorders*, *4*, 526–538.
- Kelley, E., Paul, J. J., Fein, D., & Naigles, L. (2006). Residual language deficits in optimal outcome children with a history of autism. *Journal of Autism and Developmental Disorders*, *36*, 807–828.
- Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, *16*, 807–828.
- Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology*, *13*, 111–124.
- Lord, C., & Paul, R. (1997). Language and communication in autism. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 195–225). New York: Wiley Press.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*, 205–223.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, *55*, 3–9.

- Luyster, R., Richler, J., Risi, S., et al. (2005). Early regression in social communication in autism spectrum disorders: A CPEA Study. *Developmental Neuropsychology*, *27*, 311–336.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- Mainela-Arnold, E., Evans, J. L., & Coady, J. (2010). Beyond capacity limitations II: Effects of lexical processes on word recall in verbal working memory tasks in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *53*, 1656–1672.
- Mainela-Arnold, E., Misra, M., Miller, C., Poll, G., & Park, J. S. (2012). Investigating sentence processing and language segmentation in explaining children's performance on a sentence span task. *International Journal of Language and Communication Disorders*, *47*, 166–175.
- McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal on Mental Retardation*, *97*, 359–372.
- Menyuk, P., & Quill, K. (1985). Semantic problems in autistic children. In E. Schopler & G. B. Mesibov (Eds.), *Communication problems in autism* (pp. 127–146). New York: Plenum Press.
- Minshew, N. J., & Goldstein, G. (1993). Is autism an amnesic disorder? Evidence from the California verbal learning test. *Neuropsychology*, *7*, 209–216.
- Montgomery, J. W. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology*, *11*, 77–91.
- Mottron, L. (2004). Matching strategies in cognitive research in individuals with high-functioning autism: Current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders*, *34*, 19–27.
- Naigles, L. R., Kelley, E., Troyb, E., & Fein, D. (2013). Residual difficulties with categorical induction in children with a history of autism. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-012-1754-y.
- Phelan, H. L., Filliter, J. H., & Johnson, S. A. (2011). Brief report: Memory performance on the California verbal learning test-children's version in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*, 518–523.
- Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation*, *110*, 417–438.
- Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-delayed groups. *Psycholinguistics*, *12*, 23–45.
- Schuh, J. M., & Eigsti, I. M. (2012). Working memory, language skills, and autism symptomatology. *Behavioral Sciences*, *2*, 207–218.
- Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: The Psychological Corporation.
- Sparrow, S., Balla, D., & Cicchetti, D. (1984). *The Vineland Adaptive Behavior Scales*. Circle Pine, MN: American Guidance Services.
- Sutera, S., Pandey, J., Esser, E., Rosenthal, M. A., Wilson, L. B., Barton, M., et al. (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*, 98–107.
- Swensen, L., Kelley, E., Fein, D., & Naigles, L. (2007). Children with autism display typical language learning characteristics: Evidence from preferential looking. *Child Development*, *78*, 542–557.
- Tager-Flusberg, H. (1988). On the nature of a language acquisition disorder: The example of autism. In F. S. Kessel (Ed.), *The development of language and language researchers: Essays in Honor of Roger Brown* (pp. 249–267). Hillsdale, NJ: Lawrence Erlbaum.
- Tek, S., Jaffery, G., Fein, D., & Naigles, L. R. (2008). Do children with autism show a shape bias in word learning? *Autism Research*, *1*, 202–215.
- Tek, S., Mesite, L., Fein, D., & Naigles, L. R. (2013). Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-013-1853-4.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processes*. Austin, TX: PRO-ED, Inc.
- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio: Harcourt Assessment.