ORIGINAL PAPER

Evaluation of the ADOS Revised Algorithm: The Applicability in 558 Dutch Children and Adolescents

**Annelies de Bildt · Sjoerd Sytema ·
Natasja D. J. van Lang · Ruud B. Minderaa ·
Herman van Engeland · Maretha V. de Jonge**

Abstract The revised ADOS algorithms, proposed by Gotham et al. (J Autism Dev Disord 37:613–627, 2007), were investigated in an independent sample of 558 Dutch children (modules 1, 2 and 3). The revised algorithms lead to better balanced sensitivity and specificity for modules 2 and 3, without losing efficiency of the classification. Including the restricted repetitive behaviour domain in the algorithm contributes to a clinical ASD classification in modules 2 and 3. For module 1, the results indicate less improvement, probably due to the low-functioning population. In most groups, the advantages of the revised algorithms are achieved without losing the strength of the original algorithm.

Keywords ADOS · Autism · ASD · Diagnosis

A. de Bildt · R. B. Minderaa
Department of Child and Adolescent Psychiatry, University
Medical Centre Groningen, Groningen, The Netherlands

A. de Bildt (✉)
Accare, UCKJP Groningen, P.O. Box 660, 9700 AR Groningen,
The Netherlands
e-mail: a.de.bildt@accare.nl

S. Sytema
Department of Psychiatry, University Medical Centre
Groningen, Groningen, The Netherlands

N. D. J. van Lang
Department of Child and Adolescent Psychiatry Curium, Leiden
University Medical Center, Leiden, The Netherlands

H. van Engeland · M. V. de Jonge
Department of Child and Adolescent Psychiatry, University
Medical Centre Utrecht, Utrecht, The Netherlands

Introduction

The autism diagnostic observation schedule (ADOS; Lord et al. 1999) is a widely used and valuable instrument as a tool for diagnosing autism spectrum disorders (ASDs). The ADOS consists of four modules, each for a separate developmental or language level of functioning. Each module contains different tasks, and all of them are intended to provide the examiner with information on social, communicative, play and stereotyped behavior. Based on the ADOS algorithm a classification can be given of autistic disorder (AD), autism spectrum disorder (not being AD) or non-ASD.

Ongoing research showed that sensitivity and specificity of the original algorithm may be more related to cognitive and verbal level of functioning and chronological age, than seemed to be the case in the initial publications on the ADOS (Lord et al. 2000; Bishop and Frazier Norbury 2002; Joseph et al. 2002; de Bildt et al. 2004). Additionally, comparing the different modules based on the algorithm is complicated by the fact that the number and content of items are not totally comparable between the modules.

Recently, Gotham et al. (2007) published revised algorithms for modules 1 through 3 of the ADOS. The aims were to create more homogeneous algorithms over the various levels of development in order to make a first step towards a '*calibrated metric of severity of autism, as independent as possible from current language levels*' and to improve the sensitivity and specificity of the ADOS classifications and their balance and thus to improve the diagnostic validity of the instrument. This more homogeneous algorithm was achieved by organizing the same number of items (i.e., 14 items) in the algorithms of each module, with similar content of the items per module. This increases comparability between the algorithms over the

various modules and therefore between the various levels of development the respective modules are meant for. Additionally, the revised algorithm applies to smaller, more specific cells, in developmental level and age. The algorithms are therefore more specific for each subgroup, and this increases the sensitivity and specificity of the algorithm. Another change is the two domains included in the revised algorithm: a domain 'social affect' (SA) and one 'restricted repetitive behaviors (RRB)'. The SA domain is included based on the fact that one factor was found to underlie the social and communication items in previous research with the ADOS (Robertson et al. 1999; Lord et al. 1999, 2000). It is a combination of 10 items from the former 'social' and 'communication' domains, yet per module without three items (two communication, one social) and one or two new (social) items. The RRB domain was added based on findings that such items may be important for diagnostic stability, as reported by Lord et al. (2006). This domain combines three items from the RRB section of the ADOS and one language item for each module. Gotham et al. (2007) give a clear and complete overview of the items in the new algorithm per module (p. 619).

In their study, the revised algorithm led to an improved specificity for non-autism ASD in lower functioning children, except for children under non-verbal mental ages under 15 months. The sensitivity for non-autism ASD remained relatively low in all groups. Adding the RRB domain did not contribute to distinguishing AD from pervasive developmental disorder-not otherwise specified (PDD-NOS), yet did contribute to the discrimination between PDD-NOS and non-spectrum cases. The authors strongly recommended replication of their revised algorithms in other populations.

Overton et al. (2008) explored the revised algorithm in 26 Hispanic children, administered with modules 1–3. In module 1 ($n = 14$), the accuracy of the revised algorithm was slightly increased. Classifications in module 2 ($n = 3$) and 3 ($n = 8$) remained unchanged after applying the revised algorithm. As mentioned by the authors, the sample is small and selective.

Gotham et al. (2008) replicated the revised algorithms in 1,282 children in the US. The results were comparable to the original study of Gotham et al. (2007). Comparability between the modules was increased. Additionally, predictive value of the ADOS for AD was increased, also increasing the validity of an autism diagnosis beyond the ADI-R. Besides, age and verbal IQ effects on the ADOS total scores were decreased.

Based on the research from Gotham et al. it can be concluded that the revised algorithms increase the comparability between the algorithms of the various modules in the various age or developmental groups. A certain score of a young child on module 2 has the same meaning as the same score for an older child on module 3. It also leads to a classification more independent from age and verbal IQ. Importantly, achieving this comparability and higher independence from developmental level has not been at the expense of the sensitivity and specificity.

In the current paper, the revised algorithms were applied to 558 children administered with modules 1, 2 or 3. The aim of the current paper was to investigate the revised algorithm in these children: how well does the revised algorithm add to a clinical classification of ASD or non-spectrum disorder?

Methods

Participants

ADOS administrations (modules 1–3) of 558 children were reevaluated with the revised algorithms of Gotham et al. (2007). All ADOS administrations had taken place as part of two large studies in the Netherlands concerning the genetics of ASDs. These studies included children referred for child-psychiatric problems/ASD and children from an epidemiological study of ASD in mental retardation (population based; De Bildt et al. 2005). This means that not all, and especially not all low-functioning participants from the current study were referred for problems in the autism spectrum, yet they were all evaluated by experienced clinicians. The majority of the children completed a diagnostic evaluation based on DSM-IV-TR criteria at the University of Utrecht ASD clinic, or at the University of Groningen Child and Adolescent Psychiatry Clinic/Autism Clinic (ATN). The others completed a standardized classification procedure in the epidemiological study (see for more detail De Bildt et al. 2004, 2005).

In order to enhance comparability with the studies of Gotham et al. the age was 12 years or younger for modules 1 and 2, and up to 16 years for module 3. Gotham et al. their sample into smaller and more homogeneous groups, analyzed separately. Following this division in the current sample, this resulted in $n = 99$ for module 1, Some Words; $n = 124$ for module 2, 5 and Older; and $n = 335$ for module 3. Because very few participants had no words at all at the time of ADOS-administration, a 'no-words' group as described by Gotham et al. could not be included in the current study. Additionally, children younger than 5 could not be included for module 2 due to the small number of administrations in this age range. Thus, the current study only addresses three of the five divisions of the new algorithm.

The age range for the total sample is 13–198 months. The sample contained 193 (34.6%) children with a clinical classification of AD, 221 (39.6%) with non-autism ASD,

Table 1 Characteristics of the participants

		M1 SW				M2 5 and Older				M3			
		N	Mean	SD	Range	N	Mean	SD	Range	N	Mean	SD	Range
Autism	Age	61	71.16	30.50	13–144	40	88.64	26.09	60–155	92	122.52	30.24	68–196
	IQ	45	60.20	26.22	5–114	34	77.71	22.06	34–120	88	90.55	21.55	20–129
	ADI S	59	20.54	5.82	7–29	38	19.87	5.72	4–30	90	20.72	5.50	8–29
	ADI C–V	39	14.97	4.54	2–25	38	15.39	4.83	4–25	90	16.78	4.70	3–25
	ADI C-NV	20	11.95	2.16	7–14	–	–	–	–	–	–	–	–
	ADI RR	59	5.39	2.26	0–11	38	6.18	2.62	1–12	90	6.24	2.56	1–12
	ADOS S	61	10.67	2.77	1–14	40	8.78	3.42	0–14	92	8.97	2.91	1–14
	ADOS C	61	5.34	2.03	0–9	40	5.08	2.26	0–10	92	3.75	1.72	0–7
	ADOS SA	61	15.93	4.01	2–20	40	11.60	4.41	0–19	92	11.46	4.74	0–19
	ADOS RR	61	2.84	1.60	0–6	40	2.18	1.57	0–6	92	1.98	1.61	0–6
ASD	Age	14	68.28	20.64	40–111	34	88.01	22.48	61–150	173	121.44	30.24	47–198
	IQ	13	67.77	29.14	30–127	28	68.96	23.44	32–128	162	96.27	23.42	34–154
	ADI S	13	19.08	4.91	10–26	28	17.89	5.65	8–27	161	17.04	6.40	0–29
	ADI C–V	11	15.36	3.72	7–19	28	13.07	4.58	6–22	161	13.44	5.01	0–26
	ADI C-NV	2	11.50	.71	11–12	–	–	–	–	–	–	–	–
	ADI RR	13	5.85	2.12	2–9	34	4.79	2.62	1–10	161	4.49	2.53	0–12
	ADOS S	14	7.71	3.38	2–13	34	5.56	3.48	0–13	173	6.60	3.34	0–14
	ADOS C	14	3.00	1.57	0–7	34	3.32	1.98	0–8	173	2.50	1.68	0–7
	ADOS SA	14	10.57	3.78	4–18	34	7.21	4.64	0–18	172	7.88	4.64	0–19
	ADOS RR	14	2.50	1.91	0–6	34	1.35	1.63	0–6	173	1.11	1.37	0–8
Non-ASD	Age	24	98.52	34.92	37–155	50	97.40	26.23	63–155	70	132.36	26.76	77–196
	IQ	17	30.12	12.82	12–58	40	56.75	20.11	27–113	61	67.39	26.93	27–129
	ADI S	20	11.40	6.32	3–28	39	12.72	6.90	1–27	51	11.57	7.81	0–26
	ADI C–V	10	8.20	5.59	3–22	37	8.24	4.66	1–20	51	9.12	5.83	0–20
	ADI C-NV	10	8.80	3.65	3–14	2	8.00	5.66	4–12	–	–	–	–
	ADI RR	20	2.25	2.34	0–7	39	3.31	3.00	0–12	51	2.96	2.91	0–9
	ADOS S	24	4.21	3.59	0–14	50	4.32	3.11	0–13	70	4.46	2.92	0–12
	ADOS C	24	1.92	2.00	0–7	50	2.28	1.69	0–9	70	1.59	1.51	0–5
	ADOS SA	24	6.50	5.45	0–19	50	5.58	3.75	0–17	70	4.83	4.12	0–16
	ADOS RR	24	1.00	1.18	0–4	50	.80	1.07	0–4	70	.60	1.03	0–4

M1 SW module 1 some words, *M2* module 2, *M3* module 3, *Age* age in months, *IQ* IQ based on various tests, *ADI S* ADI-R social total, *ADI C–V* ADI-R communication total for verbal children, *ADI C-NV* ADI-R communication total for non-verbal children, *ADI RR* ADI-R restricted, repetitive behaviors total, *ADOS S* ADOS social total, *ADOS C* ADOS communication total, *ADOS SA* ADOS social affect (revised algorithm), *ADOS RR* ADOS restricted, repetitive behaviors (revised algorithm)

and 144 (25.8%) with a non-spectrum classification. Of these 144 the majority had MR (110, 76.4%; most of them from the epidemiological study), 10 ADHD (with or without ODD), 4 a language disorder, 2 selective mutism, 5 anxiety, 2 ODD, 1 motor coordination disorder, 6 were unclear, 3 had no psychiatric disorder, and 1 had another psychiatric disorder. In all modules the majority were boys: 77.8% in module 1, 79% in module 2 and 81.5% in module 3. In Table 1, characteristics of the participants are presented by module and diagnostic group. Of all participants only one administration was included in this study.

In general, the current sample is older, contains more clinical non-autism ASD-classifications and lower functioning cases (especially non-spectrum) than the samples of Gotham et al. (2007, 2008). For all modules, IQ is significantly lower in the non-spectrum group than in the A(S)D groups. For module 1, Some words and module 3, children in the non-spectrum group are significantly older than in the A(S)D groups.

Instruments

The ADOS was administered by trained psychologists or psychiatrists who fulfilled requirements of reliability and administration of the ADOS in research (Lord et al. 1999). In the current study the original algorithm (Communication

and Social Domains; Comm-Soc) was applied in addition to the two revised algorithms: SA (Social Affect, combining items from the Communication and Social Domains) and SARRB (Social Affect combined with Restricted Repetitive Behaviors, combining items from the Communication, Social and Repetitive restricted behaviors domains). IQ's were available for 488 (87.5%) of the children, based on standardized intelligence tests. In module 3 and 2 the majority of cases were tested with WISC-R (Wechsler 1974; Vander Steene et al. 1986), WPPSI-R (Wechsler 1989; Vander Steene and Bos 1997) or RAVEN progressive matrices (Raven 1995, 1996). Most cases in module 1 and some cases in module 2 were administered a Dutch nonverbal intelligence test, the Snijders-Oomen Niet-verbale intelligentie test-Revisie (SON-R; Snijders et al. 1996), and some in module 1 were administered the Dutch modification of the Bayley scales of Infant Development (Bayley 1969; Van der Meulen and Smrkovsky 1983).

The Autism Diagnostic Interview-Revised (ADI-R; Rutter et al. 2003) was administered to 499 children (89.4%).

Statistics

The current study aimed to replicate the findings of Gotham et al. and Overton et al. Therefore, Pearson r correlations were computed first, between the revised algorithm domains (SA, RRB and SARRB) and participant characteristics, in order to investigate whether an interrelationship between these variables existed. Second, sensitivity, specificity and efficiency of the ADOS classification were calculated for each of the applied algorithms, compared to clinical classifications of AD versus non-spectrum and non-autism ASD versus non-spectrum.

To also contribute to the further investigation of the revised algorithms, logistic regressions were applied to study the contribution of the ADOS scores on the algorithm domains to the clinical classification of ASD (including AD) versus non-spectrum. All three ADOS algorithm domains were investigated: (1) the original ADOS algorithm total score (Communication and Social; Comm-Soc), analyzed separately; (2) the SA algorithm score; and (3) the RRB algorithm score, combined in one analysis. For all analyses, a p value of <.05 was considered significant.

Results

Correlations between revised algorithm totals and participant characteristics showed no correlation of the revised algorithms with IQ or age. Correlations between the

original algorithm domains and IQ and age, showed the same pattern. The correlation between the original and revised algorithm scores is high for all modules, as expected. With respect to the ADI-R, the correlation between SA and SARRB and the Social and Communication domains of the ADI-R is highest in module 1 (see Table 2 for more detail).

In Table 3, the sensitivity, specificity and efficiency of the original and revised algorithms are presented, when applicable in relation to formerly reported values (Gotham et al. 2007, 2008; between parentheses). Efficiency represents the percentage of cases classified correctly based on the algorithm used, as compared to the clinical classification mentioned. By including this value, it is possible to see whether improvement of the balance between sensitivity and specificity (as aimed for with the revised algorithms) affects the efficiency. In the current data, efficiency generally remains stable or improves, thus the predictive value of the ADOS does not seem to lose strength with better balanced sensitivity and specificity. The greatest, yet still slight decrease in efficiency exists for module 2, 5 and older, of the SA only algorithm, comparing non-autism ASD versus non-spectrum (and then only .07). The two major differences between the proposed and original algorithms in the current study are: (1) the increased sensitivity for the SA and SARRB algorithm as compared to the original algorithm in classifying AD versus non-spectrum, with a somewhat decreased specificity; and (2) the more balanced sensitivity and specificity in module 2 (AD vs. non-spectrum).

In module 1, AD versus non-spectrum, the two revised algorithms have a slightly higher sensitivity than the original algorithm, and lower specificities. The balance between sensitivity and specificity did not improve. Comparing the algorithms for non-autism ASD versus non-spectrum, the original and the SARRB algorithms perform equally well, the SA specificity is lower.

In module 2, AD versus non-spectrum, the sensitivity of SA and SARRB is higher than for the original algorithm, with lower specificities. With respect to non-autism ASD versus non-spectrum, the sensitivity of SA is increased compared to the original algorithm, whereas the specificity is decreased, compared to the original and the SARRB algorithms. The balance between sensitivity and specificity was improved for AD versus non-spectrum.

For module 3 the revised algorithms are more sensitive for AD and non-autism ASD than the original one, although there is a slight decrease in specificity for AD and ASD.

With logistic regression the contribution of the algorithms to a clinical classification of ASD (incl. AD) versus non-spectrum was investigated. The odd's ratios (OR, presented in Table 4) express the increase or decrease of

Table 2 Correlations between revised algorithm totals and participant characteristics

			IQ	Age	ADI S	ADI C–V	ADI C–NV	ADI RR	Onset	ADOS S	ADOS C	ADOS SC	ADOS P	ADOS RR
M1 SW	SA total	r	.10	−.03	**.53**	**.59**	**.51**	.27	−.02	**.96**	**.85**	**.98**	**.59**	**.39**
		p	.38	.74	.00	.00	.00	.01	.82	.00	.00	.00	.00	.00
		N	75	99	92	92	92	92	92	99	99	99	99	99
	RRB total	r	−.08	−.10	.16	.29	**.36**	.09	.18	**.47**	**.48**	**.51**	**.35**	**.89**
		p	.52	.31	.14	.03	.00	.39	.08	.00	.00	.00	.00	.00
		N	75	99	92	60	92	92	92	99	99	99	99	99
	SA + RRB	r	.07	−.06	**.50**	**.59**	**.53**	.25	.02	**.94**	**.85**	**.97**	**.97**	**.55**
		p	.55	.59	.00	.00	.00	.01	.82	.00	.00	.00	.00	.00
		N	75	99	92	60	92	92	92	99	99	99	99	99
M2 O	SA total	r	−.03	.10	**.37**	**.39**	**.34**	.29	.20	**.95**	**.87**	**.97**	**.51**	**.30**
		p	.80	.28	.00	.00	.00	.00	.04	.00	.00	.00	.00	.00
		N	102	124	105	103	105	105	105	124	124	124	124	124
	RRB total	r	−.03	.06	.21	.28	.21	**.38**	.30	**.43**	**.46**	**.46**	.16	**.93**
		p	.78	.54	.03	.01	.03	.00	.00	.00	.00	.00	.09	.00
		N	102	124	105	103	105	105	105	124	124	124	124	124
	SA + RRB	r	−.03	.10	**.38**	**.42**	**.36**	.35	.26	**.94**	**.88**	**.96**	**.49**	**.51**
		p	.77	.27	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00
		N	102	124	105	103	105	105	105	124	124	124	124	124
M3	SA total	r	.02	−.04	.24	.23	.19	.11	.10	**.95**	**.84**	**.98**	**.34**	**.31**
		p	.76	.47	.00	.00	.00	.07	.09	.00	.00	.00	.00	.00
		N	310	334	302	302	302	302	302	334	334	334	334	334
	RRB total	r	.00	−.07	.17	.25	.13	**.38**	.07	**.37**	**.41**	**.41**	.18	**.90**
		p	.95	.21	.00	.00	.03	.00	.26	.00	.00	.00	.00	.00
		N	311	335	302	302	302	302	302	335	335	335	335	335
	SA + RRB	r	.02	−.05	.25	.27	.20	.19	.10	**.94**	**.85**	**.97**	**.35**	**.50**
		p	.76	.33	.00	.00	.00	.00	.07	.00	.00	.00	.00	.00
		N	310	334	302	302	302	302	302	334	334	334	334	334

M1 SW module 1 some words, *M2 O* module 2 5 and older, *M3* module 3, *IQ* total IQ based on various tests, *ADI S* ADI-R social total, *ADI C–V* ADI-R communication total for verbal children, *ADI C–NV* ADI-R communication total for non-verbal children, *ADI RR* ADI-R restricted, repetitive behaviors total, *ADOS S* ADOS social total, *ADOS C* ADOS communication total, *ADOS SC* ADOS social-communication total, *ADOS P* ADOS play total, *ADOS RR* ADOS restricted, repetitive behaviors total, *SA* ADOS social affect (revised algorithm), *RRB* ADOS restricted, repetitive behaviors (revised algorithm), *SA+RRB* ADOS social affect + restricted, repetitive behaviors (revised algorithm)

Bold values are significant and clinically relevant (>.3)

Table 3 Sensitivity and specificity of the ADOS classifications based on SA, SARRB and Comm-Soc algorithms, compared to the clinical classifications

	Meets Comm-Soc for ADOS AUT			Meets SA+RRB for ADOS AUT			Meets SA only for ADOS AUT		
	SE	SP	Efficiency	SE	SP	Efficiency	SE	SP	Efficiency
AD versus NS									
Mod 1 SW AD = 61 NS = 24	.82 (.88[a]/.73[b])	.88 (.96[a]/.94[b])	.84	.92 (.97[a]/.89[b])	.71 (12; .91[a]/.91[b])	.86	.90 (.91[a]/81[b])	.71 (12; .93[a]/.94[b])	.85
Mod 2 older AD = 40 NS = 50	.63 (.96[a])	.92 (.97[a])	.79	.88 (.98[a])	.76 (9; .90[a])	.81	.80 (.92[a])	.82 (9; .97[a])	.81
Mod 3 AD = 92 NS = 70	.73 (.86[a]/.72[b])	.84 (.89[a]/.96[b])	.78	.87 (.91[a]/.82[b])	.73 (9; .84[a]/.92[b])	.81	.82 (.85[a]/.79[b])	.73 (9; .87[a]/.90[b])	.80
	Meets Comm-Soc for ADOS ASD			Meets SA+RRB for ADOS ASD			Meets SA only for ADOS ASD		
	SE	SP	Efficiency	SE	SP	Efficiency	SE	SP	Efficiency
Non-autism ASD versus NS									
Mod 1 SW ASD = 14 NS = 24	.86 (.67[a]/1.00[b])	.63 (.84[a]/.80[b])	.71	.86 (.77[a]/.95[b])	.63 (8; .82[a]/.75[b])	.71	.86 (.75[a]/.75[b])	.54 (6; .79[a]/.75[b])	.66
Mod 2 older ASD = 34 NS = 50	.56 (.86[a])	.64 (.77[a])	.61	.53 (.83[a])	.62 (8; .83[a])	.58	.62 (.72[a])	.48 (6; .77[a])	.54
Mod 3 ASD = 172 NS = 70	.64 (.68[a]/.49[b])	.67 (.77[a]/.89[b])	.65	.68 (.72[a]/.60[b])	.63 (7; .76[a]/.88[b])	.67	.70 (.61[a]/.86[b])	.66 (6; .78[a]/.86[b])	.69

The absolute numbers between parentheses are the cut-offs of the various algorithms, as described by Gotham et al. (2007)

SE sensitivity, *SP* specificity, *Efficiency* cases classified correctly/total cases, *AD* clinical classification of AD, *ASD* clinical classification of non-autism ASD, including PDD-NOS and Asperger's disorder, *NS* clinical classification non-spectrum, *SA* ADOS social affect (revised algorithm), *SA+RRB* ADOS social affect + restricted, repetitive behaviors (revised algorithm), *Comm-Soc* ADOS social-communication total (original algorithm)

[a] The percentages between parentheses are the values of the studies by Gotham et al. (2007)

[b] The percentages between parentheses are the values of the studies by Gotham et al. (2008)

Table 4 Contribution of the ADOS classifications to clinical classifications, controlled for age and IQ

		ASD (include AD) versus non-spectrum Odds ratio (95% CI)
M1 SW	SA	1.43 (1.08–1.906)*
	RRB	1.61 (.69–3.76)
	Comm-Soc	1.56 (1.20–2.03)**
M2 5+	SA	1.25 (1.10–1.41)***
	RRB	1.57 (1.09–2.26)*
	Comm-Soc	1.27 (1.14–1.42)***
M3	SA	1.25 (1.14–1.37)***
	RRB	1.47 (1.04–2.06)*
	Comm-Soc	1.32 (1.21–1.45)***

SA ADOS social affect score (revised algorithm),*RRB* ADOS restricted, repetitive behaviors score (revised algorithm), *Comm-Soc* ADOS social-communication total score (original algorithm). ASD classification includes AD, PDD-NOS and Asperger's disorder

* $p < .05$, ** $p < .01$, *** $p \leq .001$

the probability of a clinical ASD classification for each additional point on the score of the algorithm domain mentioned. Six analyses were run: two for each module, one for the original algorithm score and one for the SA and RRB algorithm scores, included as separate variables in one analysis. Thus it is possible to investigate the contribution of each, taking the other into account. For all modules, the SA algorithm score and the original algorithm score contribute approximately equally to the clinical classification. The RRB domain contributes to the clinical classification in module 2 and 3, above the SA domain contribution and above the original algorithm contribution, yet does not seem to add to a clinical ASD classification in module 1, when taking SA into account.

Discussion

The current study was undertaken in order to evaluate the revised ADOS algorithms as proposed by Gotham et al. (2007). The aim was to investigate the sensitivity and specificity of the revised algorithms in an independent sample of 558 children, and to investigate their contributions to a clinical ASD classification. The independent sample consisted of children referred for child-psychiatric problems/ASD and children from an epidemiological study of ASD in mental retardation. Thus, not all participants were referred for problems in the autism spectrum, yet they were all evaluated by experienced clinicians. The sample contained children from three of the five divisions as made by Gotham: administered with module 1 (Some Words group only); module 2 (5 and Older only) and module 3. In general, the current sample was older and contained more

clinical non-autism ASD-classifications and lower functioning non-spectrum cases than the sample of Gotham et al. (2007).

It is promising that the correlation between IQ and the revised algorithms as reported by Gotham et al. in all modules (based on Verbal IQ and Non Verbal IQ; 2007) was not replicated in the current study, comparable to the replication study in the US (Gotham et al. 2008). The fact that no correlation was found with IQ or with age in all modules indicates that the revised algorithm domains seem to be independent from these variables in the current sample, which is what Gotham et al. strived for. Nevertheless, this can not be put forward as an enhancement on behalf of the new algorithms, since the same pattern was found for the original algorithm domains, with no correlations with age and IQ in either module. The comparison of data may be complicated by the fact that Gotham et al. (2007) had measured VIQ and NVIQ, whereas the current IQ measures were based on various tests, resulting in differences between the outcomes. For part of the children TIQ's were available based on verbal and nonverbal subtests, whereas others could only complete non-verbal IQ-tests. However, both correlations in Gotham's sample between VIQ and NVIQ and the ADOS algorithms were significant whereas the correlations in the current sample were not.

With respect to sensitivity and specificity, our results indicate that applying the revised algorithms improves the balance between these in module 2 and 3, without losing strength with respect to efficiency of the classification. This is comparable to the results as reported by Overton et al. (2008). The efficiency of the revised algorithms (i.e., the percentage of cases classified correctly) was better for AD than for non-autism ASD. This may be due to the fact that part of the non-spectrum children (the clinical group) were referred for developmental or behavioral problems that gave reason to investigate whether or not ASD was apparent, and therefore may have scored on the ADOS without receiving an ASD diagnosis or may not have sored on the ADOS, while receiving an ASD diagnosis. Amongst children with MR, the distinction between AD and non-autism ASD is even less clear than in a normally intelligent population, as is the discrimination between ASD and non-spectrum in some cases. It is important to keep in mind that the ADOS does not reflect the distinctions made by clinicians between AD and non-autism ASD (see also Lord et al. 2000). Additionally, even the combination of ADOS and ADI-R does not lead to a perfect discrimination between ASD and non-spectrum cases, which indicates that the actual criteria for such diagnosis are not clear enough yet (Risi et al. 2006).

The balance between sensitivity and specificity is important since a very specific instrument without

sufficient sensitivity tends to miss cases, whereas a very sensitive measure without sufficient specificity would be overinclusive. In a former study in a Dutch, low-functioning, population, the ADOS tended to be overinclusive, e.g., to have a high sensitivity without an accordingly high specificity (De Bildt et al. 2004). The currently reported balance implies that this is less the case with the revised algorithms in the current (only partially overlapping) sample. However, in module 1, Some Words of the current sample, the balance did not improve, due to a decrease in specificity when sensitivity increased.

The question is whether this balance is reasonable to aim for in this sample. Not only the population as a whole is lower functioning than the previously reported samples, also the clinically classified non-spectrum cases are lower functioning than the A(S)D cases and, in modules 1 and 3, older too. This is due to the fact that the majority of the non-spectrum cases were recruited from a population-based study of ASD in MR, not referred cases for ASD (De Bildt et al. 2005).

Based on the specific character of this population, the current sample may increase the difficulty in obtaining a satisfactory specificity, and therefore may affect the balance between sensitivity and specificity: Compared to the findings of Gotham et al. (2007, 2008), the current study reports lower values of specificity for all algorithms. As reported by Gotham et al. (2007), in the nonverbal module 1 participants with very low nonverbal mental ages ($\leq$15 months) specificity was 50 or more % lower as compared to nonverbal children administered with module 1 with mental ages of more than 15 months. Although in the current sample all children were verbal, the same issue may affect the results due to the very low IQ's reported. This may especially be the case in module 1, Some Words, where the non-spectrum cases have a mean IQ of 30.

Another interesting issue with respect to distinguishing ASD from low-functioning children with MR without ASD is the addition of the RRB domain to the SA domain for a classification based on the ADOS. The current study indicates that including the RRB domain additional to SA in the algorithm contributes significantly to the clinical ASD classifications in module 2, 5 and Older and module 3, yet does not increase correct classification of ASD in module 1, Some Words. For modules 2 and 3 these results resemble what Gotham et al. (2007, 2008) reported. For module 1, Some words, the outcome is not surprising, since RRB's are important features of low-functioning children with MR as well and may not be specific for or give lead to an ASD diagnosis in such a population.

Due to the restricted behavioural repertoire shown in low functioning children, together with the overlap with behaviours in children with ASD, it may be ambitious to strive for an instrument that is able to distinguish between those two without missing or overincluding cases. However, this should not be judged as a flaw of the algorithms or ADOS, yet is inherent to the nature of the two disorders, because of their behavioural resemblance.

Nevertheless, for (at least part of the) clinicians diagnosing ASD this is daily practice: distinguishing the specific developmental disorder ASD from a more general developmental disorder or MR. Standardized and valid instruments would be of great value in this process.

The currently reported increased balance between sensitivity and specificity in relatively low-functioning children administered with module 2 (5 and older) and module 3 indicates that in these groups, the revised algorithms do not lose any strength compared to the original one, and therefore should be preferred. Adding RRB in the algorithm increases the discriminative power. For module 1, Some Words, an even lower functioning group, the revised algorithms are less clearly preferred. Although the sensitivity increased, the specificity decreased, and the RRB domain did not add to distinguishing between ASD and MR. Further research in very low functioning children is needed in order to obtain the perfect algorithm, or perhaps even a better applicable ADOS-module. After all, the original algorithm of module 1 is not applicable to children with mental ages under 24 months, and although 15 months seems feasible with the revised algorithms, MR and ASD lead to the same behaviours in such low-functioning children.

Whether a high sensitivity (revised algorithms) or a high specificity (original algorithm) is preferred when balance between the two is not feasible (which needs to be further investigated before concluding so) depends on the question behind the administration of the particular ADOS. For research it may be important to only include definite cases of ASD, resulting in the need for a higher specificity. On the contrary, for other uses it may be important not to miss any possible ASD-case, with a need for a higher sensitivity as a result. In any case, this discussion once again emphasizes the fact that the ADOS should never be used as the only indication for whether an ASD is present, also not in low-functioning children.

Unfortunately, we were not able to reliably investigate the children with MR separately, since their number in the current sample was too small, too much distributed amongst the various modules and too little amongst various classifications within these groups. With respect to the further development of the algorithm of the ADOS, investigating it in a large (combined) group with MR will increase the knowledge on the value of the ADOS and its algorithms in this specific group.

Additionally, the number of cases that could be included in the current study may limit interpreting the findings. We could not include module 1, no words or module 2,

younger than 5. The current study therefore leaves questions open on the value of the revised algorithms of the ADOS in very young or low functioning children. Additional studies concerning these subpopulations will contribute to investigation of the value of the revised algorithms.

To conclude, our results corroborate the findings as reported by Gotham et al. (2008) that the advantages of the revised algorithm (i.e., better representation of observed diagnostic features, increased comparability between modules in algorithm content and number, and improved predictive value for autism) have not been at great expense of the sensitivity and specificity, for module 2, 5 and older and module 3. With respect to module 1, in low-functioning children, more research is needed in order to reach good balance between high sensitivity and high specificity.

References

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: Psychological Corporation.

Bishop, D. V. M., & Frazier Norbury, C. (2002). Exploring the borderlands of autistic disorder and specific language impairment: A study using standardised diagnostic instruments. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*, 917–929. doi:10.1111/1469-7610.00114.

de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., et al. (2004). Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R) and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders, 34*, 129–137. doi:10.1023/B:JADD.0000022604.22374.5f.

de Bildt, A., Sytema, S., Kraijer, D., & Minderaa, R. (2005). Prevalence of pervasive developmental disorders in children and adolescents with mental retardation. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 46*, 275–286. doi:10.1111/j.1469-7610.2004.00346.x.

Gotham, K., Risi, S., Dawson, G., Tager-Flusberg, H., Joseph, R., Carter, A., et al. (2008). A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*, 642–651. doi:10.1097/CHI.0b013e31816bffb7.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The autism diagnostic observation schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders, 37*, 613–627. doi:10.1007/s10803-006-0280-1.

Joseph, R. M., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*, 807–821. doi:10.1111/1469-7610.00092.

Lord, C., Risi, S., DiLavore, P. C., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from two to nine. *Archives of General Psychiatry, 63*, 694–701. doi:10.1001/archpsyc.63.6.694.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr, Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communicative deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*, 205–223. doi:10.1023/A:1005592401947.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *ADOS. Autism diagnostic observation schedule. Manual*. Los Angeles: WPS.

Overton, T., Fielding, C., & Garcia de Alba, R. (2008). Brief report: Exploratory analysis of the ADOS revised algorithm: Specificity and predictive value with hispanic children referred for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(6), 1166–1169.

Raven, J. C. (1995). *Colored progressive matrices sets I and II* (1996th ed.). Oxford: Oxford Psychologists Press Ltd.

Raven, J. C. (1996). *Progressive matrices: A perceptual test of intelligence. Individual form* (1996th ed.). Oxford: Oxford Psychologists Press Ltd.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*, 1094–1103. doi:10.1097/01.chi.0000227880.42780.0e.

Robertson, J. M., Tanguay, P. E., L'Ecuyer, S., Sims, A., & Waltrip, C. (1999). Domains of social communication handicap in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 738–745.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *ADI-R. Autism diagnostic interview revised. Manual*. Los Angeles: Western Psychological Services.

Snijders, J. T., Tellegen, P. J., Winkel, M., & Laros, J. A. (1996). *SON-R 2, 5–7 Snijders-Oomen Niet-verbale Intelligentietest-Revisie [SON-R 2, 5–7 Snijders-Oomen Non-verbal Intelligence test-Revised]*. Lisse, the Netherlands: Swets & Zeitlinger.

Van der Meulen, B. F., & Smrkovsky, M. (1983). *Bayley Ontwikkelingsschalen (BOS 2–30) [Bayley development scales (BOS 2–30)]*. Lisse, the Netherlands: Swets & Zeitlinger.

Vander Steene, G., & Bos, A. (1997). *Wechsler Preschool and Primary Scale of Intelligence-Revised. Vlaams-Nederlandse Aanpassing*. Lisse, the Netherlands: Swets & Zeitlinger.

Vander Steene, G., van Haasen, P. P., De Bruyn, E. E. J., Coetsier, P., Pijl, Y. J., & Poortinga, Y. H. (1986). *Wechsler intelligence scale for children-revised. Nederlandstalige Uitgave*. Lisse, the Netherlands: Swets & Zeitlinger.

Wechsler, D. (1974). *Wechsler intelligence scale for children-revised*. New York: Psychological Corporation.

Wechsler, D. (1989). *Wechsler preschool and primary scale of intelligence-revised*. New York: Psychological Corporation.