Chapter 2

Measuring pervasive developmental disorders in children and adolescents with mental retardation

A comparison of two screening instruments (ABC and PDD-MRS) used in a study of the total population with mental retardation from a designated area

Annelies de Bildt
Sjoerd Sytema
Cees Ketelaars
Dirk Kraijer
Fred Volkmar
Ruud Minderaa

In press, Journal of Autism and Developmental Disorders

Abstract

The performance of two screening instruments for pervasive developmental disorders was studied in the total population of children and adolescents with mental retardation between 4 and 18 years (N=1059) in Friesland, a northern province of the Netherlands. Parents completed the ABC, staff completed the PDD-MRS. The screening instruments were related to the ADI-R and ADOS for 184 participants. The agreement between ABC and PDD-MRS was fair (κ=.24). The ABC had a better criterion related validity compared to the ADI-R, the PDD-MRS compared to the ADOS. However, related to the clinical classification, both instruments performed equally well. Concluding, the ABC and PDD-MRS partially identify the same cases related to external criteria. In addition, each instrument has its own contribution. Both instruments are valuable in detecting children with mental retardation who are at high risk for pervasive developmental disorders.

2.1 Introduction

Pervasive developmental disorders are characterized by severe and pervasive impairments in reciprocal social interaction, verbal or non-verbal communication and stereotyped behavior, interests and activities. Yet, in the years since Kanner first published on autism (Kanner, 1943), the definition changed between childhood psychosis and autism, or infantile autism, or early childhood autism. Later, subgroups were made and every study used its own criteria to include or exclude participants. In the present study the concept of pervasive developmental disorders contains both Autistic Disorder (AD) and Pervasive Developmental Disorders-Not Otherwise Specified (PDD-NOS), depending on the definition and criteria of each of the instruments used.

It is well known that mental retardation occurs very frequently in children with pervasive developmental disorders (Wing & Gould, 1979; Wing, 1981; Rutter, 1983; Wing, 1993; Bryson, 1996, 1997). Standard estimates of 70-90% of children with a pervasive developmental disorder have IQ's of 70 or lower (DeMyer et al., 1974; Steffenburg & Gillberg, 1986). Conversely, it is not sufficiently clear how many children with mental retardation have a pervasive developmental disorder.

Differentiating between mental retardation with and without a pervasive developmental disorder can be complicated, especially in low functioning children, since impairments in social functioning are apparent in both children with mental retardation and in children with a pervasive developmental disorder. Thus, in children with mental retardation and a pervasive developmental disorder, the low mental age and the pervasive developmental disorder may account for an overlap in behavior (Wing, Gould, Yeates, & Brierley, 1977; DiLavore, Lord, & Rutter, 1995; Wing, 1997; Kraijer, 1997; Towbin, 1997). A differentiation within the spectrum of pervasive developmental disorders between AD and PDD-NOS in children with mental retardation is even more complicated, if possible at all. However, for the approach of children with mental retardation and a pervasive developmental disorder, the recognition of the presence of a co-morbid pervasive developmental disorder is of great importance, much more so than the differentiation within the spectrum of pervasive developmental disorders (Siegel, 1996; Kraijer, 1997).

From both clinical experiences and research, it is well known that co-occurrence of a pervasive developmental disorder in mental retardation causes many additional problems for children and their environment (Tsai, 1996; Kraijer, 1997). Bryson

(1996) mentioned significant behavior and additional psychiatric problems in 50% of the people with a pervasive developmental disorder. Tsai (1996) reviewed various studies with high percentages of additional behavioral problems. The co-morbid problems vary from separation anxiety, self-injurious behavior, general anxiety or fears, hyperactivity, compulsions/rituals, depression, irritability/agitation, sleep problems, tics, excessive masturbation and rumination (Kraijer, 1997). Children with a pervasive developmental disorder show marked problems in adaptive behavior as well (Volkmar et al., 1987; Carter et al., 1998). The additional problems have major clinical implications with respect to treatment for children and adolescents with a co-morbid pervasive developmental disorder.

Identification of a pervasive developmental disorder is a central factor to provide specific services and interventions for those affected, and to raise the awareness of health professionals and educational authorities about autism and other severe developmental disorders (Bryson, 1996). For identification it is necessary to be able to measure or screen for pervasive developmental disorders with instruments, to collect information on the behavior in a standardized way. Information on the behavior of a child is then collected in the same way for each child, and can be compared to information collected on other children with a pervasive developmental disorder. However, instruments are only valuable if they measure pervasive developmental disorders validly and reliably. Therefore it is important to study the utility of instruments for pervasive developmental disorders in the population with mental retardation.

In this paper, two screening instruments will be evaluated, based on the results of a total population-based screening of children and adolescents with mental retardation in Friesland, a northern province in the Netherlands. The population was administered with two standardized screening instruments: the Autism Behavior Checklist (Krug, Arick, & Almond, 1980) and the Scale of Pervasive Developmental Disorder in Mentally Retarded persons (Kraijer, 1997).

The objective of this paper was to compare the performance of the two screening instruments in a population with mental retardation in two ways: First, we investigated the two measures in relationship to each other. Second, we tested the validity of the screening instruments related to the diagnostic instruments Autism Diagnostic Interview-Revised (Lord, Rutter, & Le Couteur, 1994), Autism Diagnostic Observation Schedule (Lord, Rutter, DiLavore, & Risi, 1998), and to the clinical

classification (APA, 2000), in order to investigate the value of the screening instruments to detect children at high risk for a classification of a pervasive developmental disorder by these three diagnostic tools. For clinical practice the clinical classification is the gold standard, often founded with the combination of ADI-R and ADOS. This combination is considered the gold standard for research. However, administering these instruments, or going through the diagnostic process, is very time consuming both for parents and clinicians, and should be restricted to children at high risk, to prevent unnecessary burden for parents. Therefore it would be useful to have screening instruments that detect children who are at high risk to be classified as having a pervasive developmental disorder by the diagnostic instruments or the clinical classification.

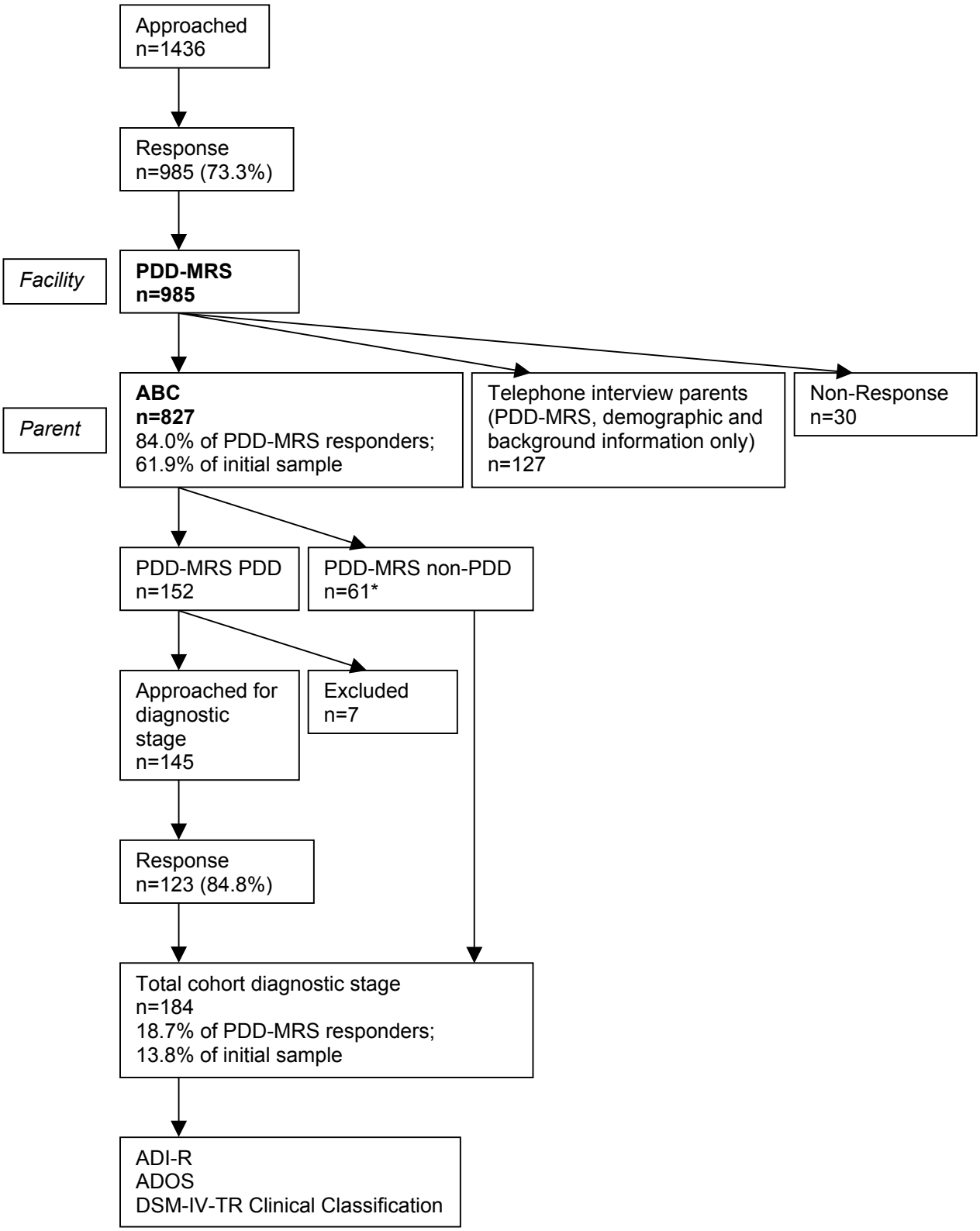
2.2 Method

The design of this study involved screening all children and adolescents with mental retardation between 4 and 18 years of age in Friesland. The screening instruments attempted to identify participants with a pervasive developmental disorder (PDD-MRS) or AD (ABC). Positive screens (PDD-MRS positive cases) were further evaluated with the two diagnostic instruments and a clinical classification. A random selection of negative screens was also evaluated to determine false negative rates. A graphic overview of the study is presented in figure 2.1.

Participants

The participants for this study were recruited from Friesland, a northern province of the Netherlands. At the time of the study, the total population was about 618,000, including approximately 120,000 children between 4 and 18 years old. All 1436 children and adolescents between 4 and 18 years, known to facilities for children and adolescents (suspected) with mental retardation (schools, day-care facilities and institutions) were approached. All levels of mental retardation were included. No participants were excluded based on etiology of mental retardation, presence of sensory or motor impairments, or co-morbid psychiatric disorders or behavior problems. With this procedure, 1059 (671 males, 388 females) participated, a response rate of 73.7%.

Figure 2.1 Overview of sampling procedure



** As a control group, these participants were randomly selected from two subgroups based on their ABC scores: below or above cut-off; all have PDD-MRS scores below cut-off.*

However, 7% of the children and adolescents who responded appeared to have no mental retardation, or to be above 18 or under 4 years old. If we hypothesize this same proportion in the non-response group, a corrected initial sample would be 1336 children between 4 and 18 years with mental retardation, with a response of 985. In this study, only participants with mental retardation, between 4 and 18 years, of whom we had information on both screening instruments were included (n=827, response rate of our corrected initial sample of 61.9%). The characteristics of the cohort are presented in table 2.1.

Table 2.1 Characteristics of the cohort

		All respondents with ABC and PDD-MRS scores		**All respondents who were clinically classified as: PDD**		**Non-PDD**	
		n	***%***	***n***	***%***	***n***	***%***
Sex	***Male***	521	63.0	59	62.1	50	56.2
	Female	306	37.0	36	37.9	39	43.8
Age	***< 12***	437	52.8	42	44.2	47	52.8
	≥ 12	390	47.2	53	55.8	42	47.2
Level of MR	***Profound***	80	9.7	29	30.5	11	12.4
	Severe	102	12.3	24	25.3	29	32.6
	Moderate	185	22.4	17	17.9	15	16.9
	Mild	460	55.6	25	26.3	34	38.2
Total		**827**		**95**		**89**	

For this study, the participants were assigned to the four levels of mental retardation: profound, severe, moderate or mild. In 83.6% of the cases, this classification was based on information from intelligence tests or developmental tests obtained by the facility. In most cases these were standardized tests, e.g. Wechsler Intelligence Scale for Children-Revised, WISC-R (Wechsler, 1974; Vander Steene et al., 1986), Wechsler Preschool and Primary Scale for Intelligence-Revised, WPPSI-R (Wechsler, 1989; Vander Steene & Bos, 1997), Non-verbal intelligence tests, e.g. Snijders-Oomen Niet-verbale intelligentie test-Revisie, SON-R (Snijders, Tellegen, Winkel, & Laros, 1996), and the Dutch modification of the Bayley Scales of Infant Development (Bayley, 1969; Van der Meulen & Smrkovsky, 1983). The other participants were assigned to one of these categories based on the Vineland

Adaptive Behavior Scales-Survey Form, VABS (Sparrow, Balla, & Cicchetti, 1984), administered as part of the study and on clinical review of functioning.

For the diagnostic stage of the study, 184 children and adolescents were selected: 123 participants identified as PDD by the PDD-MRS, and 61 controls, identified as non-PDD by the PDD-MRS. The controls were randomly selected from two groups, e.g. AD or non-AD by the ABC. The characteristics of the diagnostic cohort are also presented in table 2.1.

Instruments

Scale of Pervasive Developmental Disorder in Mentally Retarded persons (PDD-MRS)

The PDD-MRS is a Dutch instrument for the spectrum of pervasive developmental disorders based on the DSM-III-R, including AD and PDD-NOS, but with no differentiation between them. It is well studied and widely used since 1990 in the care of people with mental retardation in the Netherlands and Belgium (Kraijer, 1997). The instrument was developed for use with children and adults with mental retardation. It is a 12-item questionnaire completed by clinicians, with dichotomous items on the three aspects of pervasive developmental disorders: communication, social behavior and stereotyped behavior. Weighted factors 1, 2, or 3 are assigned to the item scores and the maximum score is 19. Psychometric qualities were tested in a large population (n=1230), including all levels of mental retardation. Sensitivity (92.3%) and specificity (92.4%) were excellent, compared to a DSM-III-R diagnosis made by a clinician. Scores on the PDD-MRS are divided into three categories: a PDD category (scores of 10 and more), a doubtful PDD/non-PDD category (scores between 7 and 9) and a non-PDD category (scores of 6 or less). In this study, school psychologists or teachers completed the PDD-MRS on 985 participants.

Autism Behavior Checklist (ABC)

At the time of this study, the ABC (Krug et al., 1980) was the only available standardized internationally applied instrument for AD, suitable for screening a large population. The ABC was included with the criteria as given by Oswald and Volkmar (1991). They proposed a cut-off of 58 and higher for AD and below 58 for non-AD, based on their study of sensitivity and specificity for the weighted sum scores,

compared to a clinical DSM-III diagnosis of AD. Parents completed the ABC on 827 of the participants who were also administered with the PDD-MRS.

Autism Diagnostic Interview-Revised (ADI-R)

The ADI-R is a standardized investigator-based interview that aims to provide data on the behavior of a child or young adult to discriminate between AD and non-AD (Le Couteur et al., 1989; Lord et al., 1994). The ADI-R focuses on the three aspects of autism, based on the DSM-IV and ICD-10. The ADI-R is conducted in an interview with parents or caregivers and is applicable for mental ages from about 18 months into adulthood (Lord et al., 1994).

The ADI-R was administered (and audiotaped) with parents of the 184 participants in the diagnostic stage of the study, as outlined in figure 2.1. Seven interviewers were trained in administering and scoring the interview. All interviewers had reached 80% reliability in scoring the ADI-R as required, before they started interviewing parents.

Autism Diagnostic Observation Schedule (ADOS)

The ADOS is a semi-structured observational instrument, developed for children, adolescents and adults who may have a pervasive developmental disorder, based on the DSM-IV (Lord et al., 1998). Scores on the ADOS are divided into three categories: AD, PDD-NOS and non-PDD. The instrument is based on an earlier version of the Autism Diagnostic Observation Schedule (ADOS, (Lord et al., 1989)) and the Pre-Linguistic Autism Diagnostic Observation Schedule (PL-ADOS, (DiLavore et al., 1995)). The assessment consists of various standardized situations, in which certain behavior (social, communicative, play or stereotyped) is expected to be elicited. The ADOS consists of four modules, each applicable for children, adolescents or adults of different levels of language and development. Interrater reliability, internal consistency, test-retest reliability and diagnostic validity are reported to be high, on item, domain and classification levels for autism and non-spectrum diagnoses (Lord et al., 2000).

Three trained examiners administered the ADOS with the 184 participants of the diagnostic stage of the study. Each examiner had reached 80% reliability, as required, before administering the ADOS. All (videotaped) observations took place in the school, day-care facility or institution. To investigate the interrelationship between

the screening instruments and the ADOS, the screening instruments were compared to both the ADOS AD category and the ADOS PDD category (including both AD and PDD-NOS).

Clinical classification

Clinical classifications were assigned by four experienced clinicians, two board certified child and adolescent psychiatrists (R.B.M. and C.E.J.K.), one clinical and developmental psychologist (D.W.K.) and one resident (E.J.M.). When a case was very difficult to assess, consensus classification was made through reviewing and discussing the available information.

The clinical classification was made according to DSM-IV-TR criteria, based on parent information, collected with the ADI-R and observation of the child on video, during the ADOS. The clinicians were blind for the outcome on the algorithms of the ADI-R and the ADOS. Each combination of two clinicians classified ten children in common. They were unaware of which child, which other clinician or outcome of the other classification. In order to measure the level of agreement of classification (AD, PDD-NOS, non-PDD) between clinicians, a weighted kappa was calculated. The weights used were 1 for exact agreement, .5 if one rater scored AD and the other PDD-NOS and 0 in all other cases. The percentage of agreement found was 81.2% and the weighted kappa coefficient was .66 (sd .13). Both the percentage of agreement and the weighted kappa values are considered good according to the criteria of Cicchetti (2001), that combine the criteria reported earlier by Cicchetti and Sparrow (1981) for weighted kappa values and the criteria reported by Cicchetti, Volkmar, Klin and Showalter (1995) for percentages of agreement.

To examine the contribution of both screening instruments to the clinical classification, the screening instruments were compared to both the AD classification of the clinicians and the broader PDD classification (including both AD and PDD-NOS).

Statistics

The agreement between the screening instruments was evaluated using Pearson's correlation and kappa statistic, the latter interpreted with the criteria proposed by Landis and Koch: 0-.2 slight, .21-.4 fair, .41-.6 moderate, .61-.8 substantial, .81-1 almost perfect (Landis & Koch, 1977).

To investigate the multivariate relationship between caseness on the ADI-R, ADOS and the clinical classification on one hand, and the screening instruments on the other, a Logistic Regression was applied. Age, sex and level of mental retardation were included in this analysis as well (SPSS Inc., 1999).

The sensitivity and specificity of the screening instruments compared to ADOS and ADI-R was studied with Receiver Operating Characteristic (ROC) analysis (Stata Corporation, 2001), on 184 participants. With this method sensitivity and specificity can be studied over the whole range of scores of the screening instruments instead of the defined cut-off only. The overall measure of agreement between criterion and screening instruments is the area under the curve (AuC). This is an estimate of the probability that a randomly chosen participant with the disorder (according to the criterion) will have a higher score (on the screening instrument) at each cut-point than a randomly chosen participant who does not have the disorder (according to the criterion) (Murphy, 1990).

2.3 Results

Agreement between the PDD-MRS and the ABC

Table 2.2 presents the participants identified as PDD by the PDD-MRS and AD by the ABC in the total cohort and in the group that was clinically classified. The PDD-MRS identified 18.4% of all participants as PDD, the ABC 17.3% as AD.

Table 2.2 PDD-caseness according to PDD-MRS and ABC

	All respondents with ABC and PDD-MRS scores			**All respondents who were clinically classified as:**					
				PDD			**Non-PDD**		
	ABC +	***ABC-***	***Total***	***ABC+***	***ABC-***	***Total***	***ABC+***	***ABC-***	***Total***
PDD-MRS+	64	88	152	43	34	77	9	37	46
PDD-MRS-	79	596	675	12	6	18	11	32	43
Total	143	684	827	55	40	95	20	69	89

Although the prevalence rates of the PDD-MRS and the ABC were comparable, the instruments partially identified different participants. The PDD-MRS agreed with the ABC on caseness in 44.8% of the participants. The ABC identified 42.1% of the participants identified as cases by the PDD-MRS. Both kappa statistic and correlation between total scores showed that level of agreement was fair (κ=.24, $p<.01$;

Pearson's r=.46, p<.01). One might expect a strong association between the comparable domains of the instruments. However, the Pearson's r correlation between the domains of the different instruments was .29 for PDD-MRS 'contact' and ABC 'relating', .26 for PDD-MRS 'communication' and ABC 'language', and .33 and .48 for PDD-MRS 'behavior' and ABC 'sensory' and 'body/object use' respectively (all correlations p<.01). These values were in the same range as the total scores correlation.

*Table 2.3 Contribution of both screeners to the **AD** classification of ADOS, ADI-R and clinical classification*

		ADOS		ADI-R		Clinical classification	
		Odds ratio	***95% CI***	***Odds ratio***	***95% CI***	***Odds ratio***	***95% CI***
PDD-MRS	***Non-PDD***	1		1		1	
	PDD	4.20**	2.05-8.63	2.90**	1.38-6.13	7.90**	2.54-24.52
ABC	***Non-PDD***	1		1		1	
	PDD	1.94	.99-3.81	8.21**	4.03-16.7	5.71**	2.60-12.53
Sex	***Female***	1		1		1	
	Male	2.03*	1.04-3.97	.88	.44-1.75	1.01	.50-2.42
Age	***< 12***	1		1		1	
	≥ 12	.84	.43-1.64	.88	.44-1.74	.94	.43-2.07
Level of MR	***Mild***	1		1		1	
	Moderate	1.98	.76-5.20	.61	.22-1.69	2.59	.79-8.53
	Severe/ Profound	4.14**	1.94-8.87	1.30	.61-2.78	2.35	.86-6.40

** p<.05, ** p<.01*

Validity of the screeners compared to ADOS, ADI-R and clinical classification

The odds ratio's (OR) presented in table 2.3 and 2.4, express the increased risk of being classified as AD or PDD according to a criterion (ADI-R, ADOS or clinical classification) when a participant had a score in the clinical range of the ABC or the PDD-MRS, controlled for all the other variables included in the model. For example, a participant classified as PDD on the PDD-MRS had an elevated chance of 4.2 to be classified as PDD on the ADOS related to the participants that were not classified as PDD on the PDD-MRS. Although participants classified as AD on the ABC also had more chance (OR=1.94) to get an ADOS AD diagnosis, this difference did not reach statistical significance.

The highest probability of an AD classification on the ADI-R was found for children identified as AD by the ABC, but children identified as PDD by the PDD-MRS had an increased probability as well. With respect to the clinical classification, children identified as AD or PDD on either of the screening instruments had an almost equally high probability to be clinically classified as AD or PDD. In general, when the screening instruments were included in the logistic regression equation, sex, age and level of mental retardation had scarcely any additional influence on PDD-caseness. An exception is the ADOS, where severe/profound mental retardation and being a boy gave an elevated additional risk for AD/PDD.

*Table 2.4 Contribution of both screeners to the **PDD** classification of ADOS and clinical classification*

		ADOS		Clinical classification	
		Odds ratio	***95% CI***	***Odds ratio***	***95% CI***
PDD-MRS	***Non-PDD***	1		1	
	PDD	7.87*	3.64-17.03	4.52*	2.16-9.46
ABC	***Non-PDD***	1		1	
	PDD	2.05	.92-4.59	5.23*	2.59-10.57
Sex	***Female***	1		1	
	Male	1.38	.65-2.95	1.25	.64-2.43
Age	***< 12***	1		1	
	≥ 12	1.87	.88-3.97	1.23	.63-2.40
Level of MR	***Mild***	1		1	
	Moderate	1.39	.49-3.94	1.07	.40-2.83
	Severe/Profound	3.54*	1.55-8.20	1.33	.63-2.80

* *p<.01*

Table 2.5 shows the overall level of agreement between the screening instruments and AD- or PDD-caseness according to ADOS, ADI-R and the clinical classification. From a ROC analysis the 'area under the curve' statistic (AuC) was obtained. We tested the difference on that statistic between the screening instruments. With perfect agreement the AuC has a value of 1, and with a value of 0.5 the agreement is no more than what could be expected by chance. Note that in this analysis each possible cut-off score of the screening instruments was tested. The AuC therefore is a continuous measure of the criterion related validity. The

calculation of sensitivity and specificity, which are the other statistics presented in the table, was based on the standard cut-off scores. All AuC's presented in the table were significantly higher than what could be expected by chance. The PDD-MRS had a significantly higher AuC than the ABC when these measures were related to the ADOS PDD classification, but the reverse was true when they were related to the ADI-R AD classification. Compared to the clinical classification, both screeners were equally able to measure PDD or AD as defined by the clinical classification. These results resemble the tendency in the odds ratio's presented in table 2.3 and 2.4.

The sensitivity and specificity statistics in general show a higher sensitivity for the PDD-MRS, hence this instrument was better able to identify children as PDD who were identified as AD or PDD by the diagnostic instruments. The ABC had a higher specificity, which implies that this instrument was less inclined to identify children as AD who were not classified as AD or PDD by the diagnostic instruments.

Table 2.5 Sensitivity and specificity of PDD-MRS and ABC compared with ADOS, ADI-R and clinical classification (AD classifications and broader PDD classifications when applicable)

			ADOS		**ADI-R**	**Clinical classification**	
	Obs		***PDD***	***AD***	***AD***	***PDD***	***AD***
PDD-MRS (total)	184	Area under the Curve	.779**	.715	.680	.742	.770
		Sensitivity[1]	.800	.815	.770	.811	.917
		Specificity[1]	.648	.478	.423	.483	.419
ABC (total)	184	Area under the Curve	.618	.631	.789*	.752	.757
		Sensitivity[1]	.454	.500	.644	.579	.708
		Specificity[1]	.704	.685	.804	.775	.699

** p < .05, compared to PDD-MRS, ** p < .01, compared to ABC, [1] sensitivity and specificity based on cut-off (PDD-MRS: 10+ PDD, 9- no PDD; ABC: 58 + AD, 57- no AD)*

2.4 Discussion

This study reports on the performance of two screening instruments, e.g. ABC and PDD-MRS, in a large population of children and adolescents with mental retardation. The first objective of the study was to examine the interrelationship between the ABC and the PDD-MRS. Although the percentage of children identified as PDD by the PDD-MRS resembles the percentage of children identified as AD by the ABC (18.4% and 17.3% respectively), they partially identify different cases.

Therefore, the agreement about clinical cases and the correlation between scores on the instruments, both total and domain scores, are only moderate.

The second objective was to investigate the validity of the screening instruments related to the diagnostic instruments ADI-R, ADOS and to the clinical classification. The results indicate that the ABC is most valid compared to the ADI-R, the PDD-MRS to the ADOS. Compared to the clinical classification, both screening instruments are equally valid. The PDD-MRS has a high sensitivity and a lower specificity. The advantage of the PDD-MRS therefore is that it identifies many children as PDD, who are actually classified as such by the diagnostic instruments. However, it is inclined to be over-inclusive. The ABC has a higher specificity and a lower sensitivity, indicating the merit that it does not identify many children as AD who are classified as non-PDD by the diagnostic instruments. Yet, the ABC is inclined to miss children who are classified as AD or PDD by the diagnostic instruments.

Two factors seem to play a role in the differences in subject identification of the different screening instruments and diagnostic tools: the different underlying concepts and the different sources of information. With respect to the first, the PDD-MRS was developed for the spectrum of pervasive developmental disorders, based on the DSM-III-R. The concept of the ADOS is comparable, based on the DSM-IV, which may explain their level of agreement. The ABC and ADI-R were developed to identify the more narrow concept autism and although they arise from different times and theoretical frameworks, they seem to resemble each other fairly closely.

With respect to the source of information, the PDD-MRS is completed by school psychologists/teachers, both ABC and ADI-R are administered with parents. The ADOS involves an external observer. From clinical practice it is known that parents and teachers or psychologists may have different perceptions about symptoms of a pervasive developmental disorder of a child. The same source of information for ADI-R and ABC might increase the strength of the relationship between them. Although the ADOS and the PDD-MRS use different sources of information, their relationship is strong as well. This could be explained by the fact that the situation in which the behavior of the child is evaluated is similar, as both instruments observe the child's behavior in school or in the facility.

The present study has several limitations that should be acknowledged in order to properly interpret the findings. Two issues concern the population of our study.

One is the fact that we only included participants with mental retardation, although pervasive developmental disorders are not restricted to the mental retardation population. In general however, to investigate disorders with a low prevalence in the population the best option is to study a high-risk population. As 70-90% of the children with a pervasive developmental disorder have mental retardation, the population with mental retardation may certainly be considered a high-risk population. The other issue is, despite our large number of participants with mental retardation, we may have missed some systematically. We did not approach children/adolescents who were in regular schools only. Yet, most of the children/adolescents with mental retardation who attend regular schools, also attend special schools part of the week. In that case, they were included in our study through their special school. Other participants we may have missed were young children who were not referred to special education yet, or older adolescents who left school and were not known to another facility yet. However, these numbers will be very small.

A limitation with respect to the design of the study is the fact that we used different instruments for different sources. Therefore, it is impossible to draw firm conclusions about the underlying factor for the limited level of agreement between the screening instruments. The optimal design to study the relationship between the performance of different instruments related to their content and to the source of information would be to use the same instruments for both sources. However, the findings of Szatmari et al. (1994), who used the ABC for both parents and teachers, indicate a low correlation between the scores of both informants. This supports our idea that the source of information should be considered as an important factor when interpreting the differences in subject identification. Another limitation was the fact that the reliability of the instruments was not tested. Nevertheless, the interviewers were thoroughly trained and reached an interrater reliability of 80% before administering the ADI-R or ADOS.

From a clinical point of view, insight in the presence of pervasive developmental disorders is very important for planning of services and intervention of problems specifically related to pervasive developmental disorders in the population with mental retardation. Since co-morbid problems due to the full spectrum of pervasive developmental disorders in children with mental retardation are known to be severe, early identification, and raising the awareness of health professionals and educational authorities about identifying pervasive developmental disorders, are of

great importance. However, administering extensive diagnostic instruments such as the ADI-R or ADOS to large groups of children with mental retardation is impossible. Thus, screening instruments that detect children who are at high risk to be classified as AD or PDD by these instruments or clinically, are of great value. The present study shows that, although the PDD-MRS and the ABC measure pervasive developmental disorders differently, the instruments have their own specific contribution and show high values of validity, the ABC in comparison to the ADI-R, the PDD-MRS in comparison to the ADOS. Each of them contributes to the clinical classification.

For clinical practice, this leads to the question which screener would be most useful in identifying pervasive developmental disorders in children and adolescents with mental retardation. This depends on the aim of the screening. Screening with the ABC is very specific, but bears the risk of under-inclusiveness. The PDD-MRS detects many children at risk, yet is over-inclusive. From a clinical perspective, over-inclusiveness would be preferred from these two. Missing a child who actually has a pervasive developmental disorder, causes more problems for both the child and his/her environment, than identifying a child as having a pervasive developmental disorder, who appears to have no pervasive developmental disorder in the long term. The combination of both screeners results in double the time and double the costs, however this could be considered in specific circumstances.