Research in Autism Spectrum Disorders 6 (2012) 1265–1272

Contents lists available at SciVerse ScienceDirect

Research in Autism Spectrum Disorders

Journal homepage: http://ees.elsevier.com/RASD/default.asp

Validation of an interview-based rating scale developed in Japan for pervasive developmental disorders

Hiroyuki Ito [a,*], Iori Tani [b], Ryoji Yukihiro [c], Jun Adachi [d], Koichi Hara [e], Megumi Ogasawara [f], Masahiko Inoue [g], Yoko Kamio [h], Kazuhiko Nakamura [i], Tokio Uchiyama [j], Hironobu Ichikawa [k], Toshiro Sugiyama [l], Taku Hagiwara [d], Masatsugu Tsujii [m]

[a] Research Center for Child Mental Development, Hamamatsu University School of Medicine, Japan
[b] Faculty of Humanities, Tokaigakuen University, Japan
[c] Faculty of Human and Cultural Studies, Kyoto Gakuen University, Japan
[d] Hokkaido University of Education, Japan
[e] Faculty of Integrated Arts and Sciences, University of Tokushima, Japan
[f] School of Education, Tokyo Gakugei University, Japan
[g] Graduate School of Medical Sciences, Tottori University, Japan
[h] Department of Child and Adolescent Mental Health, National Center of Neurology and Psychiatry, Japan
[i] Department of Psychiatry and Neurology, Hamamatsu University School of Medicine, Japan
[j] Faculty of Human Development and Culture, Fukushima University, Japan
[k] Tokyo Metropolitan Children's Medical Center, Japan
[l] Department of Child and Adolescent Psychiatry, Hamamatsu University School of Medicine, Japan
[m] School of Contemporary Sociology, Chukyo University, Japan

ARTICLE INFO

Article history:
Received 27 July 2011
Received in revised form 1 April 2012
Accepted 2 April 2012

Keywords:
PDD
Autism spectrum disorders
ADI-R
Psychometrics

ABSTRACT

The pervasive developmental disorders (PDDs) Autism Society Japan Rating Scale (PARS), an interview-based instrument for evaluating PDDs, has been developed in Japan with the aim of providing a method that (1) can be used to evaluate PDD symptoms and related support needs and (2) is simpler and easier than the currently used "gold standard" instruments such as the Autism Diagnostic Interview-Revised (ADI-R). We examined the reliability and validity of PARS on the basis of data from 572 participants (277 PDD patients and 295 nonclinical controls). Inter-rater reliability was sufficient at both the item and scale level. Factor analysis extracted four subscales, for which internal consistency was found to be high. The sub and total scores of PARS showed correlations with the domain and total scores of ADI-R, in line with theoretical prediction, indicating the convergent validity of PARS. A receiver operating characteristic analysis showed that PARS has good discriminative validity in differentiating between PDD patients and nonclinical controls, regardless of intellectual capacity. Considering that PARS can be easily implemented by professionals with appropriate knowledge regarding PDDs, PARS may be superior to the existing instruments in terms of cost performance.

1. Introduction

Over the course of many years, several instruments have been developed for the diagnosis, evaluation, and screening of pervasive development disorders (PDD). In recent years, the Autism Diagnostic Interview-Revised (ADI-R; Le Couteur et al., 1989; Lord, Rutter, & Le Couteur, 1994) has been broadly accepted as a standardized interview-based diagnostic instrument

* Corresponding author at: Research Center for Child Mental Development, Hamamatsu University School of Medicine, 1-20-1, Handayama, Higashi-ku, Hamamatsu 431-3192, Japan. Tel.: +81 53 435 2331; fax: +81 53 435 2291.
E-mail addresses: ito_hiroyuki@pd5.so-net.ne.jp, ito_h@hama-med.ac.jp (H. Ito).

1750-9467/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.rasd.2012.04.002

for PDD. The Autism Diagnostic Observational Schedule (ADOS; Lord et al., 2000, 1989) is also widely used as an observation-based diagnostic instrument. These instruments have a high level of discriminative validity with respect to the differentiation of PDD from non-PDD and are useful in reaching a definitive diagnosis; however, their implementation requires special training and significant time, leading to the development of numerous simpler evaluation scales in recent years.

The Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001), which has been broadly accepted as a screening instrument, is a unique tool that comprises a combination of questionnaires, telephone interviews, and structured follow-up interviews. Although it is a highly useful tool, its use is limited to toddlers because it was developed with the aim of early identification of PDD. In countries such as Japan and other Asian countries lacking the medical and governmental services for PDD that exist in the United States and Europe, it is believed that many people with undiagnosed PDD exist in a broad age group. In fact, Kawamura, Takahashi, and Ishii (2008) reported that in Toyota City, Japan, where a new systematic PDD screening system has been implemented, there were 11 times more detections of PDD compared with that observed in a survey done 20 years ago. However, few regions in the world have an adequate PDD detection system of this kind. Considering this, the development of a simple and practical evaluation scale that can be applied to a wide age group is an important and pressing issue.

The Autism Spectrum Screening Questionnaire (ASSQ; Ehlers, Gillberg, & Wing, 1999), Autism Screening Questionnaire (ASQ; Berument, Rutter, Lord, Pickles, & Bailey, 1999), and Social Responsiveness Scale (Constantino et al., 2003) have been developed as PDD evaluation scales that can be applied to a relatively broad age group. As all of these evaluation tools are in the format of a questionnaire that can be evaluated by parents or teachers, they have the advantage of being fairly easy to implement. However, in most cases, parents lack the specialized knowledge needed to understand PDD, so the standards for rating individual items can vary greatly depending on the individual conducting the evaluation, possibly leading to a deterioration of the reliability of evaluation results. Furthermore, though teachers generally have more PDD-related knowledge than do parents, they have less specific knowledge of each individual child; hence, their evaluations tend to be less reliable than those of parents. In practice, the sensitivity (true positive rate) and specificity (one minus false positive rate) of the ASSQ in distinguishing PDD and non-PDD was .91 and .77, respectively, for the parent evaluation and .90 and .58, respectively, for the teacher evaluation (Ehlers et al., 1999). Considering that the sensitivity and specificity of the ADI-R were 1.00 and .90, respectively (Lord et al., 1997), the level of accuracy of the ASSQ in distinguishing PDD from non-PDD was insufficient in the hands of both parents and teachers. Furthermore, in a simultaneous comparison conducted by Charman et al. (2007), sensitivity and specificity in identifying autistic spectrum disorders was .86 and .78, respectively, for the ASQ and .78 and .67, respectively, for the SRS, thereby indicating its insufficient precision in practical use.

To resolve this dilemma between accuracy and simplicity, the PDDs Autism Society Japan Rating Scale (PARS) has been developed in Japan as an instrument for evaluating PDDs (Adachi et al., 2006; Kamio et al., 2006; Tsujii et al., 2006). This scale was developed with the aim of providing an instrument that is simpler to use than the ADI-R and ADOS; is applicable to any age group, unlike the M-CHAT; and has better reliability and validity than questionnaire scales such as the ASSQ and ASQ. While PARS uses an interview format similar to ADI-R, the procedures, which are briefly summarized in the manual, can be implemented after simple training. Furthermore, because the criteria for rating each item is clearly defined in PARS, a more reliable and valid evaluation is possible than with questionnaire scales. In order to ease the rating process and shorten the evaluation time, the evaluator assigns values at three levels—none (0 points), somewhat apparent (1 point), and apparent (2 points)—for the 34 items listed as typical behavioral symptoms of PDD. This innovation ensures that the time required to implement PARS is kept to 30–90 min, depending on the interviewer's proficiency and the target's age and symptoms.

There is no international literature on the psychometric properties of PARS, although PARS is now widely used in Japan. This study examined the reliability and validity of PARS and involved a study population of 628 test subjects that included 302 people with PDD and 326 people without PDD. Specifically, we evaluated the inter-rater reliability, factor structure, internal consistency, correlation with the ADI-R, and the ability to distinguish subjects with PDD from a nonclinical sample.

2. Methods

2.1. PARS

The PARS instrument has been developed (Adachi et al., 2006; Kamio et al., 2006; Tsujii et al., 2006) and published (PARS Committee, 2008) in Japan. It involves the evaluation of PDD symptoms through a semi-structured interview conducted with a parent or family member of the subject as the target. This tool can be used to assess not only the risk of PDD but also the need for support pertaining to administrative and medical services. PARS comprises both an evaluation of symptoms when they were most pronounced during infancy (named the peak symptoms scale) and an evaluation of current symptoms (named the current symptoms scale). The former is used mainly to an assessment of PDD risk, and the latter is mainly used in assessment of actual support needs. The peak symptoms scale, which comprises 34 items, is the same for subjects of all age groups, whereas the current symptoms scale, which comprises 57 items, has 3 versions targeting different age groups: preschoolers, primary schoolers, and adolescents/adults. This study reports on data obtained from the peak symptoms scale.

The PARS peak symptoms scale comprises 34 items that describe the characteristic behavioral symptoms of PDDs during the preschooler phase. The items were selected by a panel of eight child psychiatrists and a developmental clinical psychotherapist who were specialized in autism research and clinical practice with more than 10 years of expertise. They compiled behavioral characteristics shown by children with PDD and classified them into eight categories—Interpersonal

Relationship, Communication, Restricted Interests, Stereotyped Behavior, Resistance, Hypersensitivity, Clumsiness, and other complications. From these, 34 items relating to symptoms that are specific to PDD, as well as items relating to nonspecific symptoms with high need for either clinical or administrative support, were selected. Twenty-two out of the 34 items corresponded to diagnostic features for PDD in the *Diagnostic and Statistical Manual 4th Edition, Text Revision* (DSM-IV-TR; American Psychiatric Association, 2000), and 8 corresponded to associated features. Symptoms described in the remaining four items (items 15, 27, 28, and 32) were not listed in the DSM-IV-TR, but since they are often present in PDD children seen in everyday clinical experience, they were included in the scales.

The evaluation of each item in PARS is based on a 30-page manual (PARS Committee, 2008). This manual includes detailed explanations of the questioning and rating standards for each item. For example, for item 1 of the peak symptoms scale (not making eye contact), a sample question "has the child ever had difficulty making eye contact?" is presented, and the rating standards are listed in detail: "0: made eye contact always," "1: had some difficulty making eye contact (made eye contact when requesting or showing interest in something but not otherwise; sometimes made eye contact and sometimes did not; made eye contact only with the parents but not with others)," and "2: rarely made eye contact (did not make eye contact with parents; avoided eye contact)." In this way, evaluation based on subjective criteria of the interviewer is avoided, and a more objective evaluation is possible.

2.2. Sample

The 572 subjects of the main sample comprised two broad groups: a PDD group made up of 277 subjects and a nonclinical control group made up of 295 subjects (Table 1).

Participants in the PDD group were diagnosed as having PDD or subordinate disorders based on the DSM-IV by experienced psychiatrists of medical and educational facilities in 28 areas throughout Japan. The diagnoses were made by integrating data from parental interviews; developmental and medical information; records provided by parents, other caregivers, and teachers; and direct observations of and interactions with the children. Subjects were referred to the facilities due to developmental concerns and randomly recruited for the study by examiners belonging to the facilities. Among these, 175 subjects underwent full-scale IQ tests using intelligence scales such as the Wechsler (Japanese WISC-III Publication Committee, 1998; Shinagawa, Kobayashi, Fujita, & Maekawa, 1990), Binet (Tanaka Institute for Educational Research, 2003), and K-ABC scales (Kaufman, Nadeen, & Kaufman, 1993). Of the 175 subjects, 51 were considered mentally retarded (IQ < 70), while 118 were not (IQ ≥ 70). To evaluate the correlation between PARS and the ADI-R, an ADI-R interview was additionally administered to 74 subjects (mean age = 14.0 years; SD = 3.6; range = 7–24 years; mean IQ = 86.2; SD = 24.7; range = 40–135) from the PDD group.

Table 1
Characteristics of the main sample.

	Age			IQ			Gender		
	M[a]	*SD*[b]	Range	*M*	*SD*	Range	Male	Female	Total
All age groups									
PDD[c] group	12.5	5.8	3–39	81.6	29.2	19–142	233	44	277
Without MR[d] (IQ[e] ≥ 70)	12.7	5.5	4–39	97.2	16.8	70–142	105	13	118
With MR (IQ < 70)	12.3	4.9	5–31	43.6	15.7	18–69	44	13	57
IQ unknown	12.4	6.3	3–32	–	–	–	84	18	102
Nonclinical control group	10.8	7.6	3–38	–	–	–	153	142	295
Preschoolers (age, 3–6 years)									
PDD group	5.1	1.0	3–6	74.1	24.5	22–121	27	12	39
Without MR (IQ ≥ 70)	5.4	0.8	4–6	87.7	13.6	70–121	9	5	14
With MR (IQ < 70)	5.9	0.4	5–6	47.0	17.8	22–68	3	3	6
IQ unknown	4.5	1.0	3–6	–	–	–	15	4	19
Nonclinical control group	4.8	1.0	3–6	–	–	–	69	63	132
Primary schoolers (age, 6–12 years)									
PDD group	9.9	1.8	6–12	80.9	31.9	18–140	94	15	109
Without MR (IQ ≥ 70)	10.2	1.7	7–12	99.6	16.2	71–140	46	5	51
With MR (IQ < 70)	9.2	2.0	6–12	40.5	13.5	18–65	16	5	21
IQ unknown	10.0	1.7	7–12	–	–	–	32	5	37
Nonclinical control group	9.2	1.8	6–12	–	–	–	34	33	67
Adolescents and adults (age, 12–39 years)									
PDD group	17.3	5.2	12–39	77.4	31.2	19–142	112	17	129
Without MR (IQ ≥ 70)	17.1	5.5	12–39	97.9	16.9	70–142	50	3	53
With MR (IQ < 70)	15.9	3.7	12–31	44.9	16.7	19–69	25	5	30
IQ unknown	17.8	5.0	12–32	–	–	–	37	9	46
Nonclinical control group	20.1	6.0	13–38	–	–	–	50	46	96

[a] Mean.
[b] Standard deviation.
[c] Pervasive development disorders.
[d] Mental retardation.
[e] Intelligence quotient.

Participants in the nonclinical control group were recruited from the local communities by individual examiners at locations such as schools, daycare centers, universities, offices, parents' circles, and neighborhood organizations. Individuals were excluded from the nonclinical control group if they had a clinical diagnosis of any psychiatric disease. IQs were not recorded for the nonclinical control group because they did not have histories of any psychiatric problems or special needs education and were considered to have normal intellectual ability.

Furthermore, separate from the main sample, data from 56 participants (mean age = 9.2 years; SD = 5.8; range = 3–26 years) diagnosed as having PDD by experienced psychiatrists were analysed to evaluate the inter-rater reliability of PARS.

The protocol of this study was approved by the institutional review board of Hamamatsu University School of Medicine.

2.3. Procedure

Psychiatrists, clinical psychologists, and graduate students involved in the service for developmental disorders administered the PARS interview by referring to the manual. They had undergone a brief training, which had the following agenda: (a) a lecture on psychiatric features of individuals with PDD; (b) instructions on the rating criterion of each item of PARS; and (c) open completion, scoring, and discussion of the interview. They conducted the PARS interview with the informants (many of whom were parents) after obtaining the appropriate informed consent. The interviewers were not completely blind to the probands' diagnosis because some of them recruited participants themselves. For some participants, an additional ADI-R interview was implemented by Japanese interviewers who had undergone a three-day long ADI-R training workshop in the United States to learn the implementation and scoring methods of ADI-R (Lord et al., 1994). They created a Japanese translation of the ADI-R and received permission from the original author and the publisher to use it through a validation process based on Japanese sample (Tsuchiya et al., submitted for publication). The ADI-R generates algorithm scores for each of the three subdomains; (a) qualitative impairments in reciprocal social behavior; (b) delays in language development; and (c) restricted range of interest and/or stereotypic behaviors. The item composition of the subdomain of delays in language development differs depending on whether or not a subject can use language. We implemented ADI-R only for subjects who can use language.

For the sample used for evaluation of inter-rater reliability, PARS was administered independently to each informant by two interviewers (one experienced specialist and one less experienced trainee).

2.4. Statistical analyses

A comprehensive examination of the reliability and validity of PARS was conducted in five steps. First, to consider the inter-rater reliability of PARS, the correlation coefficient between the scores recorded by the two interviewers of the same subject was calculated. Second, to examine the factor structure of PARS, exploratory factor analysis (mean-adjusted weight least-square estimation with promax rotation) was performed based on the PDD group data, and four subscales were extracted. As the score for each item was considered as an ordered categorical variable of three values, factor analysis was carried out using the polychoric correlation coefficient (see Holgado-Tello, Chacon-Moscoso, Barbero-Garcia, & Vila-Abad, 2010). Third, the α coefficient was calculated based on data of the PDD group to examine the internal consistency of the overall scale and four subscales. Fourth, to examine convergent validity, correlation of PARS scores with the ADI-R algorithm scores was considered using Pearson's coefficient.

Fifth, to consider how well PARS distinguishes between PDD and non-PDD, *t*-tests and receiver operating characteristic (ROC) analysis (Swets, 1988) were performed. ROC analysis plots the curve (ROC curve) of the true positive rate (sensitivity) vs. the false positive rate (one minus specificity) as the discrimination cutoff value is varied. The larger the area under the ROC curve (AUC), the higher the discriminative power of the scale. In general, sensitivity and specificity are in a trade-off relationship, and the two cannot be simultaneously maximized. In the present study, the cutoff value was set at the point where the sum of sensitivity and specificity was the largest, and sensitivity and specificity for that point were reported. Further analysis including the presence of mental retardation (MR) as a variable was conducted to consider whether the discriminative power of PARS is influenced by IQ level.

Before initiating the abovementioned analyses, we examined the difference in the scale scores for the 3 age groups because previous studies (Adachi et al., 2006; Kamio et al., 2006; Tsujii et al., 2006) have examined the scale properties of the PARS separately for each age group. One-way ANOVA showed that the total PARS score did not significantly differ for the 3 age groups, both in the PDD group, $F(2, 280) = .41$, $p = .66$, and in the control group, $F(2, 315) = 2.49$, $p = .08$. Therefore, we decided to perform the analyses without any distinction between the age groups.

Significance levels of statistical tests were set at 5% and 1%. Mplus (Muthén & Muthén, 1998–2007) was used for factor analysis, and SPSS 15.0J (SPSS Inc., 2006) was used for other analyses.

3. Results

3.1. Inter-rater reliability

Spearman's rank correlation coefficients between the scores of two interviewers were significant for all items ($p < .05$ in item 27; $p < .01$ in remaining items), with an average value of .68 (SD = .11). For the total score, the Pearson's correlation coefficient between the scores of the interviewers was $r = .78$ ($p < .01$).

Table 2
Corrected item-total correlations and factor loadings.

No.	Item	I-T corr.[a]	Factor loading			
			F1 SC[b]	F2 SD[c]	F3 SB[d]	F4 RI[e]
5	Does not communicate interest by pointing	.70	**.83**	.17	.01	−.20
6	Verbal development is delayed	.71	**.82**	−.29	.00	.09
7	Conversation does not continue	.79	**.81**	−.22	.03	.29
4	Does not bring items to show	.67	**.79**	.16	.08	−.23
1	Does not make eye contact	.74	**.69**	−.01	.06	.04
2	Is not interested in other children	.74	**.62**	.23	−.02	−.05
9	Does not play with other children	.79	**.57**	.08	.15	.06
3	Does not look back when name is called	.70	**.53**	.02	.20	.06
28	Becomes unstable bringing back to unpleasant memories	.53	−.20	**.82**	−.06	−.01
26	Becomes confused when everyday situations or routines changes	.69	.06	**.67**	−.12	.06
33	Suddenly cries or becomes upset	.60	.12	**.62**	.02	.05
32	Is very scared over nothing	.54	−.10	**.60**	−.06	.18
34	Show self-injurious action like banging head on wall or chewing hands	.46	.01	**.41**	.26	−.15
27	Cannot maintain personal independence due to disrupted lifestyle	.41	−.17	**.40**	.25	−.19
30	Disturbed by particular sounds	.63	−.03	**.37**	.19	.21
24	Does not like to be touched	.58	.14	**.37**	.20	.10
31	Is either insensitive or oversensitive to pain, heat, etc.	.62	−.15	**.36**	.28	.03
20	Does not like to be held	.56	.18	**.25**	.16	.17
22	Turns pages or crumples paper repeatedly in the same way	.54	−.03	−.14	**.67**	.23
19	Eats or swallows nonfood items	.37	.00	−.05	**.66**	−.22
14	Likes watching things that revolve	.59	.03	−.05	**.66**	.13
18	Is hyperactive and may go anywhere if left unattended	.62	.05	−.20	**.65**	−.02
17	Walks on tiptoes	.47	−.01	−.01	**.60**	−.18
23	Moves entire or part of the body repeatedly in the same pattern	.56	.03	.07	**.54**	.06
12	Becomes immersed in sensory play	.61	.15	−.05	**.51**	.06
15	Looks at things from the corner of eye or from extremely close	.62	.15	−.03	**.48**	.23
11	Repeats the words of commercials, etc.	.61	−.08	−.06	.00	**.81**
10	Parrot-like repetition stands out	.68	.37	−.10	−.08	**.68**
13	Loves road signs, logos, numbers, and letters	.59	−.13	.09	.06	**.60**
8	Speaks only one way to say what he/she wants	.70	.09	.04	−.06	**.51**
21	Repeatedly watches specific scenes of videos	.62	−.11	.15	.14	**.49**
25	Persistently asks the same question	.48	−.28	.19	.00	**.38**
16	Becomes immersed lining up toys and bottles	.61	.05	.21	.03	**.34**
29	Extremely unbalanced diet, eats very few food items	.57	.03	.18	.11	**.24**
			Interfactor correlations			
			F1	F2	F3	F4
		F2	.25			
		F3	.45	.50	–	
		F4	.27	.42	.33	–

Bold loadings indicate grouping in sub-scales.
[a] Corrected item-total correlation.
[b] Social Communication.
[c] Sensitivity/Difficulty.
[d] Stereotyped Behavior.
[e] Restricted Interests.

3.2. Factor structure and internal consistency

Table 2 shows the corrected item-total correlation for each item and the results of factor analysis. Based on a scree plot (9.25, 3.76, 2.36, 2.02, 1.68, 1.62, . . .) that showed a leveling-off of eigenvalues after the fourth factor (cf. Cattell, 1966) and perceived interpretability, a four-factor solution was employed. The four factors explained 42.27% of the variability of the total score, and each factor was named in decreasing order according to the factor loading of the items grouped in the factor, starting with Social Communication, Sensitivity/Difficulty, Stereotyped Behavior, and Restricted Interests. The α coefficient based on data of the PDD group was .84 for the communication scale (8 items), .74 for the sensitivity/difficulty scale (10 items), .72 for the stereotyped behavior scale (8 items), and .70 for the Restricted Interests scale (8 items). The α coefficient for all scales was .86. All of the individual item-to-total score correlations were positive and mainly substantial, in the range of .37–.79 (29 of the 34 exceeding .50). The mean values for each subscale and the total score for each group are shown in Table 3.

3.3. Correlation with the ADI-R

The correlation of PARS subscores and total score with ADI-R domain scores and total score is shown in Table 4. The score of Qualitative Abnormalities in Reciprocal Social Interaction in ADI-R showed moderate correlation with the score of Social

Table 3
Means and standard deviations of PARS total score and subscores.

	Social Communication		Sensitivity/ Difficulty		Stereotyped Behavior		Restricted Interest		Total sore	
	M[a]	SD[b]	M	SD	M	SD	M	SD	M	SD
PDD[c] group	10.03	4.62	7.36	4.61	6.12	4.02	7.96	4.09	31.46	12.52
Without MR[d] (IQ ≥ 70)	8.83	4.37	7.04	4.99	5.46	3.68	8.11	4.23	29.45	13.00
With MR (IQ < 70)	12.66	3.18	7.83	4.39	8.21	4.12	8.45	4.17	37.14	11.55
Nonclinical control group	0.38	1.19	0.43	1.05	0.54	1.07	0.88	1.50	2.23	3.64

[a] Mean.
[b] Standard deviation.
[c] Pervasive development disorders.
[d] Mental retardation.

Table 4
Correlations between the ADI-R and PARS.

PARS	ADI-R			
	Social Interaction[a]	Communication[b]	Stereotyped Behavior[c]	Total score
Social Communication	.48**	.43**	.07	.48**
Sensitivity/Difficulty	.17	.03	.37**	.20
Stereotyped Behavior	.03	.27*	.42**	.25*
Restricted Interest	.07	.10	.41**	.19
Total score	.27**	.31**	.46**	.41**

[a] Qualitative abnormalities in reciprocal social interaction.
[b] Qualitative abnormalities in communication.
[c] Restricted, repetitive, and stereotyped patterns of behavior.
* $p < .05$.
** $p < .01$.

Communication in PARS. Furthermore, the score of Qualitative Abnormalities in Communication in the ADI-R showed moderate correlation with the score of Social Communication in PARS, and weak correlation with the score of Stereotyped Behavior and the total score in PARS. The score of Restricted, Repetitive, and Stereotyped Patterns of Behavior in the ADI-R showed weak correlation with the score of Sensitivity/Difficulty in PARS and moderate correlations with the score of Stereotyped Behavior and Restricted Interests and the total score in PARS. The total score of the ADI-R showed a moderate correlation with the score of Social Communication and the total score in PARS and a weak correlation with Stereotyped Behavior.

3.4. Discriminative validity

Table 5 and Fig. 1 shows the results of the *t*-test and ROC analysis between the PDD groups (whole group and without MR group) and the nonclinical control group. Three main points can be concluded from the table and figure. First, PARS shows high discriminative power even when the presence or absence of MR is controlled. Second, for either comparison, the total score has more discriminative power than the subscores. This is a general trend seen in other evaluation instruments such as

Table 5
Discriminative validity of the total and subscores of PARS.

	t[a]	AUC[b]	Cutoff point	Sensitivity	Specificity
PDD vs. nonclinical control					
Social Communication	33.9	.973	3	.929	.959
Sensitivity/Difficulty	24.6	.961	2	.921	.902
Stereotyped Behaviors	22.5	.928	2	.896	.851
Restricted Interests	27.2	.953	3	.875	.902
Total score	37.6	.991	9	.975	.956
PDD without MR vs. nonclinical control					
Social Communication	20.8	.964	3	.908	.959
Sensitivity/Difficulty	14.3	.949	2	.882	.902
Stereotyped Behaviors	14.4	.921	2	.882	.851
Restricted Interests	18.2	.952	3	.882	.902
Total score	22.5	.990	9	.975	.956

[a] All *t* values are significant at the 1% level.
[b] Area under the curve.

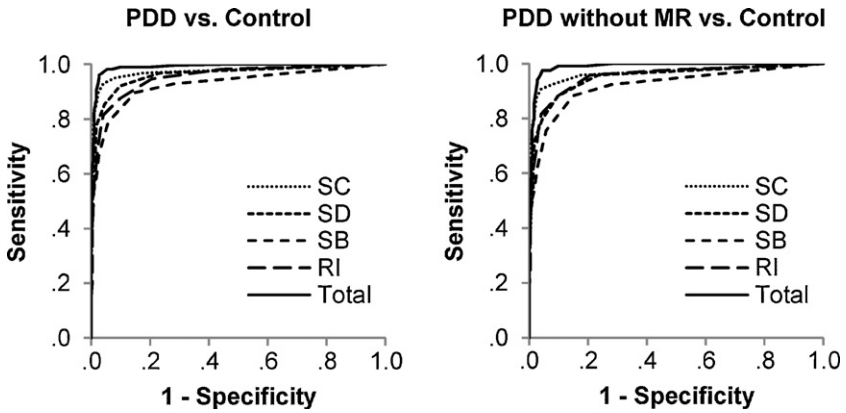

Fig. 1. Receiver operating characteristic curves for discrimination between normal control group and whole PDD (left) and PDD without MR group (right). SC, Social Communication; SD, Sensitivity/Difficulty; SB, Stereotyped Behaviors; RI, Restricted Interests.

the ADI-R (Lord et al., 1997) and ASQ (Berument et al., 1999). Third, the desired cutoff values are not affected by the presence or absence of MR.

4. Discussion

The objective of this study was to validate PARS, a scale developed for (1) the evaluation of PDD symptoms in a simpler manner than "gold standard" instruments, such as the ADI-R and ADOS, and (2) more objective evaluation than questionnaire scales, such as the ASSQ and ASQ. As long as the interviewer has a certain level of expertise pertaining to PDDs, PARS can be used after brief training and can be administered in an hour on an average by simplifying and structuring the interview procedure as much as possible and by using simple and clear terms in the manual. In this study, we administered PARS to individuals with PDD and nonclinical controls in order to examine its reliability and validity.

The rating scores recorded by two different interviewers of the same subject showed a sufficient correlation for individual items as well as for the overall score, demonstrating the inter-rater reliability of PARS. The developers of questionnaire scales have often criticized the form of the interview method, stating "the severity of each assessed behavior is rated by the interviewer 'second-hand' on the basis of the parent's answers" (Constantino et al., 2003). This criticism is based on the belief that the interview process produces random or systematic measurement error due to its "second-hand" nature. However, the PARS interview's high inter-rater reliability indicates that it produces little random error, probably because of each item's clearly defined rating criteria. We believe that a semi-structured interview conducted by specialists in treatment of developmental disorders will provide a more accurate measurement than a questionnaire scale based on the subjective judgments of people who lack specialized knowledge, as long as rating criteria are clearly defined and sufficient inter-rater reliability of the evaluation instrument is maintained.

Factor analysis extracted four subscales: Social Communication, Sensitivity/Difficulty, Stereotyped Behaviors, and Restricted Interests. The Social Communication scale corresponds to the "reciprocal social interaction skills" and "communication skills" criteria of the DSM-IV-TR (American Psychiatric Association, 2000), and the Stereotyped Behavior scale and the Restricted Interests scales correspond to the DSM-IV-TR's "presence of stereotyped behavior, interests, and activities." While there is no clear correspondence of the Sensitivity/Difficulty scale with the DSM-IV-TR criteria, it addresses many peripheral symptoms such as sensory over-responsibility and problematic behavior, which are thought to be important in practical support for PDD patients. Through these four scales, PARS not only covers core PDD symptoms but also covers a wide variety of peripheral symptoms. Each subscale and the overall scale showed an α coefficient greater than .70, which demonstrated sufficient internal consistency.

Correlation with the ADI-R clearly duplicated the correspondence relationships with DSM-IV stated above, demonstrating the convergent validity of PARS. Furthermore, the Sensitivity/Difficulty scale showed a correlation with the ADI-R's Restricted, Repetitive, and Stereotyped Patterns of Behavior domain. This might show that the limited interest or fixation on specific things or objects may be the root cause of peripheral symptoms included in the Sensitivity/Difficulty scale.

Through the ROC analysis of the ability of PARS to distinguish between PDD and non-PDD, PARS showed high discriminative power regardless of the intellectual capacity of the patient. The total score demonstrated a higher discriminative power than the subscores, similar to the case with the ADI-R (Lord et al., 1997) and ASQ (Berument et al., 1999). Considering its ease of implementation, PARS may be superior to the ADI-R or ADOS in terms of cost performance. Furthermore, the ROC analysis indicated that the selected cutoff value of PARS is relatively stable regardless of the intellectual capacity of the patient. The fact that a fixed cutoff level can be employed regardless of the nature of the interview subjects is considerably important in terms of convenience and utility in practical use.

One limitation of the study is that the interviewers were not completely blind to the probands' diagnosis. This factor might have a positive influence on the result of discriminative power analysis. Thus, the conclusion about our measurement technique's discriminative power is limited. However, it is unlikely that this problem systematically affects the result of our other analyses (i.e., factor analysis, reliability analysis, and correlation analysis), because the lack of blindness might uniformly raise the score of the PDD group and lower the score of the control group. Such uniform changes do not affect these kinds of analyses.

Finally, we discuss future issues. First, although this study examined the discriminative power of PARS in differentiating between PDD patients and the general population, there is a need to examine its discriminative power in other developmental disorders, such as attention deficit hyperactivity disorder, which shows somewhat similar symptoms to PDD (Hattori et al., 2006), or in other mental disorders, including schizophrenia, depression, and anxiety disorder, which often occur together with PDD. Second, the effectiveness of PARS in distinguishing subordinate diagnoses of PDD, which was not included among the objectives of this study, also needs to be considered. By appropriately combining the four subscales extracted in the factor analysis, PARS might be able to distinguish among subordinate diagnoses. We believe this is also an important issue with respect to the versatility of PARS. Third, an English version needs to be developed if PARS is to be used internationally. Currently, PARS is published in Japan and is being used by many clinical and research institutions (Yamada et al., 2007), but it cannot be used overseas as the Japanese version is the only one that exists. Since PARS is simpler than the ADI-R or ADOS and has sufficient reliability and validity, it can be an extremely useful instrument worldwide.

Acknowledgement

This study was supported by a grant from the Foundation for Children's Future.

References

Adachi, J., Yukihiro, R., Inoue, M., Uchiyama, T., Kamio, Y., & Kurita, H. (2006). Reliability and validity of the childhood part of the PARS (PDD-Autism Society Japan Rating Scale). *Rinsho Seishin Igaku (Clinical Psychiatry), 35*, 1119–1126.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., Text Revision) (DSM-IV-TR)*. Washington, DC: American Psychiatric Association.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry, 175*, 444–451.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., et al. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *British Journal of Psychiatry, 191*, 554–559.

Constantino, J. N., Davis, A. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., et al. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders, 33*, 427–433.

Ehlers, S., Gillberg, C., & Wing, L. (1999). A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *Journal of Autism and Developmental Disorders, 29*, 129–141.

Hattori, J., Ogino, T., Abiru, K., Nakano, K., Oka, M., & Ohtsuka, Y. (2006). Are pervasive developmental disorders and attention-deficit/hyperactivity disorder distinct disorders? *Brain & Development, 28*, 371–374.

Holgado-Tello, F. P., Chacon-Moscoso, S., Barbero-Garcia, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity, 44*, 153–166.

Japanese WISC-III Publication Committee. (1998). *Nihonban WISCIII chinou kensahou* (Japanese Wechsler Intelligence Scale for Children, 3rd ed.). Tokyo: Nihon Bunka Kagakusha.

Kamio, Y., Yukihiro, R., Adachi, J., Ichikawa, H., Inoue, M., Uchiyama, T., et al. (2006). Reliability and validity of the pervasive developmental disorders (PDD) Autism Society Japan rating scale: A behavior checklist for adolescents and adults with PDDs. *Seishin Igaku (Psychiatry), 48*, 495–505.

Kaufman, Nadeen, & Kaufman, (1993). *K-ABC Shinri Kyoiku Asesument Batteri [Kaufman Assessment Battery for Children]*. Tokyo: Maruzen Meitsu.

Kawamura, Y., Takahashi, O., & Ishii, T. (2008). Reevaluating the incidence of pervasive developmental disorders: Impact of elevated rates of detection through implementation of an integrated system of screening in Toyota, Japan. *Psychiatry and Clinical Neurosciences, 62*, 152–159.

Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., et al. (1989). Autism diagnostic interview: A standardized investigator-based instrument. *Journal of Autism and Developmental Disorders, 19*, 363–387.

Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folstein, S., et al. (1997). Diagnosing autism: Analyses of data from the autism diagnostic interview. *Journal of Autism and Developmental Disorders, 27*, 501–517.

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 communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*, 205–223.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., et al. (1989). Autism diagnostic observation schedule. A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19*, 185–212.

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

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Pervasive Developmental Disorders Autism Society Japan Rating Scale (PARS) Committee. (2008). *Kouhansei Hattatsu Syogai Nihon Jiheisyo Kyokai Hyotei Syakudo. [Pervasive Developmental Disorders Autism Society Japan Rating Scale]*. Tokyo: Spectrum Publishing Co.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*, 131–144.

Shinagawa, F., Kobayashi, S., Fujita, K., & Maekawa, H. (1990). *WAIS-R Seijin Chinou Kensahou: Nihonban. [Japanese Wechsler Adult Intelligence Scale-Revised]*. Tokyo: Nihon Bunka Kagakusha.

SPSS Inc.. (2006). *SPSS base 15.0 user's guide*. Chicago, IL: SPSS Inc.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*, 1285–1293.

Tanaka Institute for Educational Research. (2003). *Tanaka-Binet Chinou Kensa V*. (Tanaka-Binet intelligence scale, 5th ed.). Tokyo: Taken Shuppan.

Tsuchiya, K., Matsumoto, J., Yagi, A., Inada, N., Kuroda, M., Inokuchi, E., et al. Reliability and validity of autism diagnostic interview – Revised – Japanese version, submitted for publication.

Tsujii, M., Yukihiro, R., Adachi, J., Ichikawa, H., Inoue, M., & Uchiyama, T. (2006). Reliability and validity of the infant part of the PARS (PDD-Autism Society Japan rating scale). *Rinsho Seishin Igaku (Clinical Psychiatry), 35*, 1119–1126.

Yamada, A., Suzuki, M., Kato, M., Suzuki, M., Tanaka, S., Shindo, T., et al. (2007). Emotional distress and its correlates among parents of children with pervasive developmental disorders. *Psychiatry and Clinical Neurosciences, 61*, 651–657.