ORIGINAL PAPER

Effectiveness of Using the Modified Checklist for Autism in Toddlers in Two-Stage Screening of Autism Spectrum Disorder at the 18-Month Health Check-Up in Japan

Yoko Kamio · Naoko Inada · Tomonori Koyama · Eiko Inokuchi · Kenji Tsuchiya · Miho Kuroda

Abstract To determine whether the Modified Checklist for Autism in Toddlers (M-CHAT) in conjunction with the routine 18-month health check-up identifies Japanese toddlers with autism spectrum disorder (ASD). Two-stage screening using the M-CHAT was conducted with 1,851 children attending the check-up. Final ASD diagnosis was confirmed at age ≥3 years. Screening identified 20/51 children with ASD: 12/20 true positives were developmentally delayed, whereas 16/22 false negatives were high-functioning. Sensitivity was 0.476, specificity 0.986, positive predictive value 0.455, and likelihood ratio 33.4 for children with ASD. With a few modifications, M-CHAT screening successfully detected toddlers with ASD with and without developmental delay and is a promising screening tool to complement existing community surveillance.

Keywords Autism spectrum disorder · Community-based surveillance · Early detection · Modified Checklist for Autism in Toddlers (M-CHAT) · Screening

Introduction

Autism spectrum disorders (ASDs) are lifelong developmental disorders and the earliest symptoms start to manifest overtly from the age of 1 year onwards. Since early educational intervention can optimize long-term prognosis (Kamio et al. 2013; Rogers and Vismara 2008), early detection and diagnosis are crucial. The American Academy of Pediatrics (AAP) recommends that in addition to broad developmental screening at 9, 18, and 24 months, all children receive autism-specific screening at 18 and 24 months of age, and it cautions against a "wait-and-see" approach for children with suspected ASD (Johnson and Myers 2007). Although many screening tools are available for children aged 18 months and older (Johnson and Myers 2007), several issues such as the optimal age for screening, general developmental surveillance versus standardized autism-specific screening, and barriers to standardized screening remain to be answered by a series of longitudinal studies (Barton et al. 2008; Charman et al. 2001). Moreover, most screening tools have been evaluated in clinical samples referred for specialized assessment (Allen et al. 2007; Eaves et al. 2006) or in a mixture of clinical and population-based samples (Robins et al. 2001); only a few have been examined in total population studies (Baird et al. 2000; Dietz et al. 2006; Pandey et al. 2008; Robins 2008). Also, parents who do not suspect their child to have ASD

Y. Kamio (✉) · N. Inada · T. Koyama · E. Inokuchi · M. Kuroda
Department of Child and Adolescent Mental Health, National Institute of Mental Health, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8553, Japan
e-mail: kamio@ncnp.go.jp

N. Inada
e-mail: nainada-tky@umin.ac.jp

T. Koyama
e-mail: tomok-tky@umin.ac.jp

E. Inokuchi
e-mail: eiko_i@osa.att.ne.jp

M. Kuroda
e-mail: pr6m-krd@asahi-net.or.jp

K. Tsuchiya
Research Center for Child Mental Development, United Graduate School of Child Development, School of Medicine, Hamamatsu University, Hamamatsu, Japan
e-mail: tsuchiya@hama-med.ac.jp

may respond to the same screening questions differently from those who do suspect it, and the results of screening should be interpreted cautiously if screening tools are used outside the setting in which their psychometric properties are known to apply (Gray et al. 2008).

Among the autism screening tools available, the Checklist for Autism in Toddlers (CHAT) (Baron-Cohen et al. 1992) was the first. In a total population study (n = 16,235) with follow-up from age 18 months up to 7 years (Baird et al. 2000), two-stage CHAT screening of 18-month-old children identified 10 of 94 children with Pervasive Developmental Disorders (PDDs) using the high-risk threshold, showing a sensitivity of 0.106, a specificity of 1.00, and a positive predictive value (PPV) of 0.833. In another study, two-stage screening of 31,724 children aged 14–15 months using the Early Screening of Autistic Traits Questionnaire (ESAT) identified 18 children who were diagnosed with ASD at an average age of 23.3 months, giving a PPV of 0.25 (Dietz et al. 2006). The Modified Checklist for Autism in Toddlers (M-CHAT) was developed as a more sensitive alternative to the CHAT (Robins et al. 2001) and has been extensively validated (Chlebowski et al. 2013; Pandey et al. 2008; Robins 2008; Kleinman et al. 2008), although its psychometric properties confirmed through long-term follow-up were determined for a combined clinical and low-risk sample (Kleinman et al. 2008). Against this background, the present study evaluated the utility of M-CHAT screening for Japanese toddlers in primary health settings. We targeted children aged 18 months for practical reasons: all Japanese children have a regular general health check-up at 18 months of age, as stipulated by the Maternal and Child Health Act, and the attendance rate is over 90 % (Mothers' & Children's Health & Welfare Association 2007).

Methods

Catchment Area

The catchment area was the suburbs of Fukuoka City, one of the biggest cities in Japan. Its total population is 93,093 according to the 2003 administrative register. The 2000 national census shows that 74 % of the working population is employed in manufacturing with the remainder working in the commerce, service, agriculture, forestry, or fishery sectors.

Participants

From April 2004 to March 2007, 2,141 children (95.4 % of the 2,245 total population cohort) attended the routine 18-month health check-up at a local health center. Written informed consent to participate in this study was obtained

from the parents of 2,113 children (consent rate = 98.7 %). Exclusion of 262 children without any follow-up data after age 3 left 1,851 children (87.6 %) for the subsequent analyses (Table 1). The 262 children excluded and the remaining 1,851 children were not significantly different in terms of sex ratio, mean age at M-CHAT screening, or screening results.

Screening Tool

Children were screened using the Japanese version of the Modified Checklist for Autism in Toddlers (M-CHAT-JV). Its high mother-father and test–retest reliability as well as concurrent and discriminant validity for Japanese toddlers have been reported (Inada et al. 2011). The majority of the Japanese general population aged 18 months has been confirmed to manifest all of the preverbal social behaviors screened by the M-CHAT-JV (Inada et al. 2010).

Because the original M-CHAT was intended to target children aged 2 (Robins et al. 2001), we assumed that the threshold might miss some children aged 18 months in a non-selected population. A preliminary analysis of data from the first one hundred 18-month-old children showed that the total 3 criteria used in the original study (Robins et al. 2001) still worked to identify possible cases (n = 7), but the critical 2 criteria identified only one in 100 children and missed 6 of 7 possible cases. In light of this, we modified the original threshold by defining 10 items as our key item set (comprising the original 6 items and newly added items 6, 20, 21, and 23) and lowered the threshold for the first-stage screening by replacing the original first-stage threshold of "any 3 from the total 23 or any 2 from the critical set criteria" with "any 3 from the total 23 or any 1 from the critical set criteria". For the second-stage screening, we adopted the original threshold, namely a total of 3 or any 2 from the critical set criteria.

Procedure: Screening and Follow-Up

1. *Screening using the M-CHAT* (Fig. 2) Our two-stage screening consisted at the first stage of administering the M-CHAT-JV at 18 months of age (any 3 from the total 23 or any 1 from the critical set criteria) and at the second stage of a follow-up telephone interview (FUI) at 19–20 months of age (any 3 from the total 23 or any 2 from the critical set criteria). The FUI followed a translated script with specific examples in which all failed items were reviewed with a parent in accordance with the original procedure (Robins et al. 2001). When reviewing the failed responses with the parents, trained interviewers did not use the term 'fail' and attempted not to cause anxiety or distress for the parents. They also offered feedback or advice when necessary. Parents were provided concrete examples of the target behaviors

Table 1 Characteristics of participants

	Participants classified as having ASD (n = 51) n (%), mean (SD), range	Total participants (n = 1851) n (%), mean (SD), range
Sex ratio, M : F	35 : 16	942 : 909
Age at M-CHAT-JV (months)	18.6 (0.6) 18–21	18.7 (0.6) 17–26
M-CHAT-JV total (failed items)	4.1 (3.2) 0–13	1.0 (1.4) 0–13
M-CHAT-JV critical 10 (failed items)	2.3 (2.2) 0–8	0.3 (0.9) 0–8
Number of evaluations	1.9 (0.8) 1–3	-
Age at final evaluation	50.6 (14.2) 33–73	-
IQ/DQ[a]	80.1 (26.7) 20–134	-
Developmental delay[b]	26 (51.0 %)	-
	Participants diagnosed with ASD by the research team (n = 34)	
AD : other ASD, (boys)	16 (14) : 18 (11)	
No. of evaluations	2.3 (0.6) 1–3	
Age at final evaluation (months)	49.4 (11.5) 33–73	
CARS total scores	34.0 (4.7) 24.5–44.5	
ADI-R toddler total scores[c]	25.5 (7.5) 11–39	
ADOS (a) + (b) total scores[d]	13.4 (3.8) 9–23	
IQ/DQ	82.1 (28.1) 20–134	
≥85	17	
70–84	4	
50–69	8	
35–49	4	
<35	1	

AD autistic disorder, *ADI-R* the Autism Diagnostic Interview-Revised, *ADOS* the Autism Diagnostic Observation Schedule, *ASD* autism spectrum disorder, *CARS* the Childhood Autism Rating Scale, *M-CHAT-JV* the Japanese version of the Modified Checklist for Autism in Toddlers

[a] 43 of 51 participants were assessed by standardized intellectual/developmental tests

[b] In addition to the 43 participants with IQ/DQ data, 8 participants were clinically judged on the presence of developmental delay

[c] 30 participants were evaluated using the ADI-R

[d] 19 participants were evaluated using the ADOS

in order to help our judgment of their responses. If the child continued to fail the M-CHAT-JV after the FUI, the family was told that their child was not doing some things that were important for social communication at this age and an evaluation was recommended (Fig. 1).

2. *Diagnostic evaluation at age 2* Screen positives were invited for diagnostic evaluation at age 2. Evaluations were conducted by the research team consisting of child psychiatrists, licensed psychologists, and primary care nurses who were already familiar with the children with special needs. The evaluation instruments included the Japanese versions of the Childhood Autism Rating Scale (CARS) (Kurita et al. 1989; Schopler et al. 1988), the Autism Diagnostic Interview-Revised (ADI-R) (Tsuchiya et al. 2012; Lord et al. 2000), and the Autism Diagnostic Observation Schedule (ADOS) (Lord et al. 1994). Children who were evaluated at age 2 were invited for full evaluation at ages 3, 4, and 5, irrespective of the diagnosis at this age.

3. *Routine 3-year health check-up* Children at age 3 received a routine health check-up including pediatric examination and parental interview by primary care nurses. Parental interviews were conducted based on a checklist containing autism-specific items derived from the ADI-R. The items included in the checklist comprised 10 social domain items, 8 communication domain items, and 2 repetitive or restricted behavior items. Among the 20 items in total, 7 items were picked up from the conventional checklist used for the routine health check-up at age 3 and 13 items were modified from the ADI-R items and newly added.

The social domain items inquire about eye contact, facial expression, nodding as yes, interest in peers, attracting adults' attention, point following, showing as joint attention, play with mother, play with peers, and social reference. The communication domain items ask about imitating what mother does, pretend play by himself/

Fig. 1 Study design
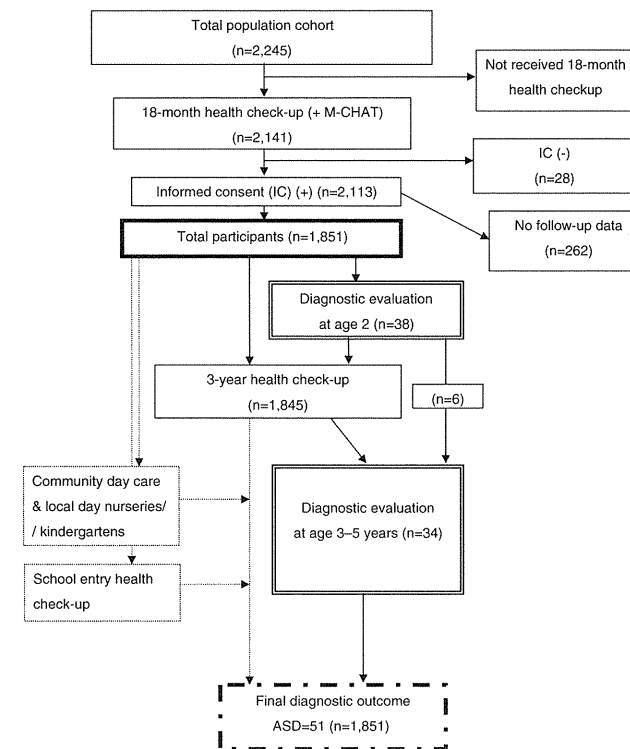
herself, pretend play with others, saying only words, saying his/her name, speaking 2-word sentences, understanding what he/she is said, and using why or what questions. The repetitive or restricted behavior domain items ask about being upset when a routine is broken or when in new environment, and stereotyped movement.

In a pilot study of 39 consecutive children who received the 3-year health check-up, failing more than 3 social or communication items produced a sensitivity of 0.857 and a specificity of 0.400 (Kamio et al. unpublished). Therefore, in the present study, this threshold in combination with behavioral observation by the primary health professional was used to detect false negative children at age 3. Among 1,830 children whose item records had no missing data, 2.24 % (41/1,830) failed more than 3 items, suggesting that the second screening at age 3 may be helpful for detecting false negatives.

The 20-item autism-specific checklist used was created in order to follow up as many false negatives as we could at

age 3. That is, children who were suspected of having ASD at age 3 based on the parental interview using the checklist or on behavioral observation during the medical examination were invited, along with screen-positive children, for full follow-up evaluation including the CARS, ADI-R, or ADOS at ages 3, 4, and 5.

4. *Community day care and local day nurseries/kindergartens* More than 90 % of the participating children went to local day nurseries or kindergartens during preschoolerhood, and children with special needs were referred to community day care centers. The research team members (primary care nurses) regularly visited these centers to monitor, consult on, and obtain clinical information about the children with special needs during preschoolerhood.

5. *School entry health check-up* Children at age 5 received a health check-up before school entry. For children with developmental concerns, detailed

🖉 Springer

interviews were conducted with the children and parents using an interview-based instrument, the Pervasive Developmental Disorders Autism Society Japan Rating Scale (Ito et al. 2012), and an IQ assessment was conducted by our research team.

Because diagnostic judgments by experienced clinicians are considered to be the "gold standard" for autism diagnosis (Volkmar et al. 2005), final diagnosis was decided according to DSM-IV-TR (American Psychiatric Association 2000) on the basis of all available information obtained after age 3 by the research team. IQs/DQs were assessed by different measures depending on mental age, using the Tanaka-Binet Intelligence Scale V for children, the Enjoji's Analytical Developmental Test under age 4, or the Japanese version of the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) at age 5.

Clinical measures were compared by group with the use of ANOVA and the Bonferroni multiple comparison test. The proportion of boys versus girls, developmental delay versus high-functioning, and the presence/absence of the targeted problems were compared with use of the Chi square test. Statistical analysis was performed using SPSS software. The protocol of this study was approved by the Ethics Committee of the National Center of Neurology and Psychiatry. This study was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Results

Throughout the screening and surveillance process of the 1,851 children, we identified 51 children with ASD: 20 screen positives, 22 screen negatives, and 9 non responders (i.e., children who needed FUI but were missed among the attrition group) (Figs. 1, 2). Thirty-four children were directly evaluated by the research team (minimum ASD). Sixteen were diagnosed with autistic disorder (AD). Table 1 outlines their demographic and diagnostic characteristics. In addition, 17 children were clinically judged by the research team to have ASD on the basis of available information, such as that from local clinicians, which brought the total number of children with ASD up to 51 (maximum ASD).

Prevalence rate was estimated to be 0.0184 (95 % confidence interval [CI] 0.0123–0.0245), and 0.0276 (95 % CI 0.0201–0.03501) for minimum and maximum ASD, respectively. The boy/girl ratio of 2.8 and 2.2 and proportion of developmental delay of 38.2 and 52.9 % in the 34 and 51 children with ASD, respectively, were in parallel with the latest reported figures (Kim et al. 2011), indicating the representativeness of this sample. Regarding AD, the prevalence rate was estimated as 0.0086 (95 % CI 0.0044–0.0129).

The sensitivity, specificity, PPV, and likelihood ratio (LR) for maximum ASD, minimum ASD, and AD through both the first-stage screening and the entire two-stage screening are shown in Table 2. Calculations for the two-stage screening including FUI were based on 1,727 children after excluding 124 FUI non-responders. Re-screening with FUI improved the specificity, PPV, and LR but reduced the sensitivity for maximum and minimum ASD and AD. Since probability is influenced by prevalence of the disorder studied, we calculated the posttest probability assuming that a prevalence rate of 2.5 % for all ASDs according to Bayes' theorem, giving a posttest probability of 0.47 and 0.51 for maximum and minimum ASD, respectively. These figures mean that almost one in every two screen positives will subsequently be diagnosed with ASD.

Among 319 screen positives at the first stage who needed FUI, only 195 were followed (response rate 61 %). One-hundred twenty-four non-responders (NR) had a significantly lower mean total M-CHAT-JV score (mean 2.81 ± 1.85) than the 195 responders (mean 3.35 ± 2.15) ($t = 2.32$, $p < 0.05$) and included significantly more girls (50 vs. 37 %) ($\chi^2 = 2.32$, $p < 0.05$), while neither group differed significantly in regard to age at M-CHAT-JV, critical items, or the proportion of nonverbal children at 18 months of age. Of the 124 NR, 9 were identified as having ASD before they were evaluated by our research team, 5 of whom had sought professional help regarding language delay.

The true positives (TP, n = 20), false positives (FP, n = 24), false negatives (FN, n = 22), and true negatives (TN, n = 1661) were compared according to demographic and diagnostic characteristics (Table 3). Although TP had significantly higher M-CHAT-JV total and critical scores than FP, FN, and TN ($ps < 0.001$), TP could not be discriminated from FP or FN by either sex ratio, maternal age at childbirth, perinatal problems, mother's feeling of difficulty with child rearing at 18 months, or mother's concerns about the child's emotional or behavioral difficulties at 3 years. A comparison between TP and FN revealed that CARS, ADI-R, and ADOS scores at 3 years or older did not significantly differ between TP and FN, but there were significantly more children with developmental delay among TP (60 vs. 27 %, $p < 0.05$). As for the 24 FP cases, mothers of 22 children reported finding child-rearing difficult on the routine 18-month health check-up questionnaire, and those of 12 children expressed some concern about their child's emotional or behavioral difficulties on the routine 3-year health check-up questionnaire. Although there were not necessarily objective records available to support their reports at or above 3 years of age, one boy had a DQ of 61 at 2, and 3 boys were clinically judged as having mild developmental delay at the 3-year pediatric check-up. In addition, the research team evaluations
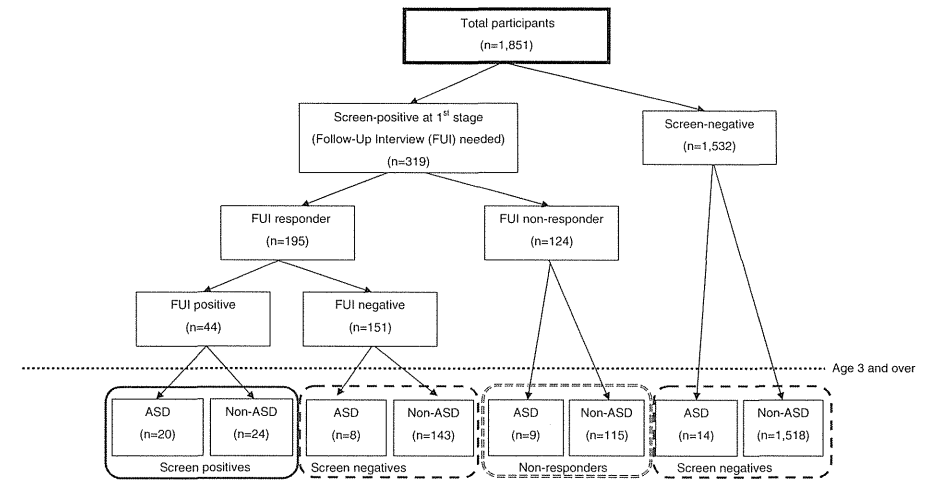
Fig. 2 Results of screening. Non-responders are children who needed a follow-up telephone interview but were missed among the attrition group

Table 2 Psychometric properties of the M-CHAT-JV screening

		Sensitivity 95 % CI	Specificity 95 % CI	PPV 95 % CI	Likelihood ratio 95 % CI
Maximum ASD	1st	0.725 (37/51)	0.843 (1,518/1,800)	0.116 (37/319)	4.631
		0.594–0.828	0.840–0.846	0.095–0.132	3.703–5.382
	1st + 2nd	0.476 (20/42)	0.986 (1,661/1,685)	0.455 (20/44)	33.433
		0.351–0.596	0.983–0.989	0.335–0.569	20.228–52.908
Minimum ASD	1st	0.765 (26/34)	0.839 (1,524/1,817)	0.082 (26/319)	4.742
		0.603–0.875	0.836–0.841	0.064–0.093	3.670–5.497
	1st + 2nd	0.613 (19/31)	0.985 (1,671/1,696)	0.432 (19/44)	41.579
		0.457–0.748	0.982–0.988	0.322–0.527	25.967–60.921
AD	1st	0.625 (10/16)	0.832 (1,526/1,835)	0.031 (10/309)	3.712
		0.388–0.815	0.830–0.833	0.019–0.041	2.276–4.885
	1st + 2nd	0.500 (7/14)	0.978 (1,676/1,713)	0.159 (7/44)	23.149
		0.274–0.725	0.977–0.980	0.087–0.231	11.695–36.670

"Maximum ASD" referred to 51 children who were classified as autism spectrum disorder (ASD) based on available information

"Minimum ASD" referred to 34 children who were directly evaluated and diagnosed as ASD by the research team. "AD" referred to 16 children who were directly evaluated and diagnosed as autistic disorder by the research team

PPV positive predictive value

confirmed two subthreshold ASD cases: one girl was diagnosed with ASD at both age 2 (IQ 68) and 3 (IQ 89), but at age 4 (IQ 123) the symptoms no longer met the diagnostic criteria. Another boy was a floppy infant with autistic features at age 2, and subsequently motor developmental delay became apparent with reduced autistic symptoms.

Discussion

This study aimed to examine prospectively the utility of an autism-specific screening in conjunction with community developmental surveillance for a non-selected Japanese population. Two-stage screening with the M-CHAT-JV identified 20 of 51 children with ASD across all intellectual

Table 3 Comparison of demographic and diagnostic characteristics: true positive, false positive, false negative, and true negative

	True positive		False positive		False negative		True negative		F	χ^2	p	
	n	M (SD)	n	M (SD)	n	M (SD)	n	M (SD)				
Sex ratio (M : F)	20	14 : 6	24	16 : 8	22	15 : 7	1,661	835 : 826		8.2	<0.05	TP > FP > FN > TN
										0.1	>0.05ᵃ	TP > FP > FN > TN
Age at M-CHAT-JV (months)	20	18.6 (0.5)	24	18.5 (0.5)	22	18.7 (0.7)	1,661	18.7 (0.6)	0.5		>0.05	
M-CHAT-JV total	20	6.3 (3.0)	24	4.5 (2.0)	22	1.7 (1.6)	1,661	0.8 (1.0)	238.0		<0.001	
M-CHAT-JV critical10	20	3.5 (2.1)	24	2.5 (1.2)	22	0.8 (1.2)	1,661	0.1 (0.5)	322.0		<0.001	
Gestational age (weeks)	18	38.2 (2.1)	23	39.0 (1.7)	22	39.1 (1.4)	1,604	39.0 (1.6)	1.7		>0.05	
Birth weight (g)	18	2949.5 (517.7)	23	3130.3 (389.5)	22	3029.6 (425.2)	1,647	3049.3 (419.4)	0.9		>0.05	
Maternal age at childbirth	20	30.2 (5.1)	24	30.8 (4.6)	22	32.8 (4.1)	1,653	30.1 (4.7)	1.8		>0.05	
Perinatal problems (present: absent)	19	4 : 15	24	6 : 18	22	6 : 14	1,597	426 : 1171		0.5	>0.05	
Mother's feeling of difficulty with child-rearing at 18 months (present: absent)	20	19 : 1	23	22 : 1	22	21 : 1	1,659	1633 : 26		3.5	>0.05	
Mother's concerns about the child's emotional or behavioral difficulties at 3 years (present: absent)	14	8 : 6	20	12 : 8	22	7 : 15	1,646	1120 : 526		14.2	<0.001	TP, FN > FP
										3.9	>0.05ᵃ	TP > FP
CARS total	20	34.2 (5.7)	10	23.7 (5.1)	12	32.8 (4.0)	4	26.4 (1.9)	14.5		<0.001ᵇ	TP, FN > FP
ADI-R toddler total	17	26.9 (9.0)	6	16.5 (7.1)	12	23.0 (4.9)	4	15.5 (9.8)	4.3		<0.05ᵇ	TP > FP
ADOS (a) + (b) total	9	13.7 (3.7)	2	1.0 (0.0)	10	12.7 (3.9)	3	6.7 (3.8)	9.9		<0.01ᵇ	TP, FN > FP
IQ/DQ	20	73.3 (27.7)	11	91.1 (15.6)	18	88.3 (26.0)	5	78.2 (6.7)	0.9		>0.05	
≥85		6		8		11		1				
70–84		2		2		3		4				
50–69		9		1		1		0				
35–49		2		0		3		0				
<35		1		0		0		0				
Developmental delay (%)		12/20 (60 %)		4/24 (17 %)		6/22 (27 %)				8.9	<0.005	TP > FP
										4.6	<0.05	TP > FN

ᵃ Three groups of TP, FN, and FP were compared using the Chi square test

ᵇ Three groups of TP, FN, and FP were compared using ANOVA

functioning levels. This indicates that the autism-specific screening at 18 months of age in primary health settings is feasible and useful when combined with community-based surveillance for preschoolers.

The controversial issue regarding the age of screening was partly answered in this study. Our findings indicate that the age of 18 months can be applied with acceptable predictive values, better than those in the earlier pioneering work (Baird et al. 2000). A possible explanation for why the M-CHAT-JV screening could identify children with ASD at this age is that the M-CHAT items might represent age-specific social development such as joint attention and pretend play that few typically developing children lack at 18 months (Inada et al. 2010; Oosterling et al. 2010), and that it could detect nonverbal social maldevelopment even in children with high-functioning ASD (HFASD). In the present study, only 30 % of 20 detected children with ASD had IQ at or above 85 and the 60 % had IQ/DQ below 70 (see Table 3). We found that the proportion of children with IQ/DQ below 70 was significantly greater among true-positive children than false-negative children, although the severity of autistic symptoms assessed by the CARS, ADI-R, or ADOS at 3 years did not differ between them. This finding suggests that the parent-report M-CHAT-JV screening measure at 18 months was more sensitive to low-functioning ASD than to high-functioning ASD, similar to earlier studies with unselected/low-risk children (Pandey et al. 2008; Kleinman et al. 2008; Baron-Cohen et al. 1996) in which detected children were mainly developmentally retarded. If the reduced sensitivity to high-functioning ASD is partly due to a lack of parental awareness, in addition to the parent-report M-CHAT-JV questionnaire, it could be possible to improve sensitivity by direct observation of some of its items by primary health nurses. In order to examine this hypothesis, a prospective study is currently underway to compare the sensitivity of the parent-report M-CHAT alone with that of the M-CHAT plus direct observation.

We recognize that we could not evaluate all screen-positive children directly, but we did instead clinically judge children who were not directly evaluated based on the information available from community surveillance. Since early detection of ASD should be economically balanced with existing surveillance procedures (Charman et al. 2002), in the absence of any better alternative screen, we recommend enhancing community developmental surveillance by supplementing it with the M-CHAT screen-rescreen procedure. Although a one-point screening model may be cost-effective, we conclude that a comprehensive model comprising repetitive screening and subsequent community surveillance will be more appropriate, considering the various developmental trajectories of children with ASD (Fernell et al. 2010; Robins et al. 2001). An

advantage of the time lag associated with the screen-rescreen procedure might be that it gives parents time to pay attention to their child's ongoing social development. To answer definitively the issue about the optimal age of screening, more empirical studies are needed and the merits and demerits for each screening procedure should be determined based on long-term follow-up data.

Our results indicated that there were at least twice as many children with HFASD missed (n = 8) as those detected (n = 16) at screening, which is consistent with Kleinman et al. (2008). In general, parents seem to be unaware of reduced social development in their child with HFASD. However, there is the possibility that these missed children show a different developmental trajectory in the very early years from that of the detected children.

Many clinicians will likely be concerned at the high screen-positive rate at the first stage of screening (17 %) because parents of children who were incorrectly suspected of having ASD might suffer unnecessary distress. This high rate might be related to the high attrition rate of 39 % (124/319) between the two stages. Since we could not systematically investigate the attrition group (the non-responders), details of the referral pattern for children with ASD who were screen positive at the first stage but who were later missed are not clear in this study. If we raise the first-stage screening threshold to approach the original one (any 3 from the total 23 or any 2 from the critical set criteria), this reduces screen positive cases (n = 39), and as a result slightly increases the PPV from 0.455 up to 0.462 (18/39) but also reduces the sensitivity from 0.476 up to 0.439. Closer inspection reveals that mothers of the majority of the false-positive children actually had been concerned about their child-rearing by age 3, and through evaluations, several children were confirmed to have problems in either cognitive, language, social, or motor domains even though the symptoms did not meet the diagnostic criteria for ASD. These findings could suggest that the false-positive cases in our study might have neurodevelopmental symptoms that extend beyond those of ASD, which are in common with those seen in many children referred to clinics (Gillberg 2010). Following this thought further, the M-CHAT screening at 18 months may be sensitive to children with mild but overlapping neurodevelopmental problems in multiple domains to some degree. This issue should be investigated in future studies using a comprehensive neurodevelopmental assessment tool.

Two major limitations exist in the current study. First, although efforts were made in cooperation with local day nurseries and clinicians to identify missed screen-positive and ASD-suspected screen-negative cases, the attrition rate was high and community-based developmental surveillance was not then sufficient in itself to monitor all children. The final diagnosis of 17 ASD cases was made based on such

indirect information. There is also the possibility that we missed a subset of children with ASD, particularly those with milder autistic symptoms, average intelligence, or girls, for whom diagnosis of ASD tends to be delayed (Mandell et al. 2005; Shattuck et al. 2009). As a result, the sensitivity and specificity of the M-CHAT-JV that we calculated based on these results can only be considered estimates of their upper bounds. Second, although various standardized instruments were used for case ascertainment of strictly defined ASD cases, the most standard ones such as the ADOS and ADI-R were not available in Japan at the beginning of this study. The total prevalence rate in our study is similar to the latest figure available from a study using strict scientific methodology (Charman et al. 2002), which indicates the quality of case ascertainment in our study.

In summary, two-stage autism-specific screening using the M-CHAT with some modification of the threshold could effectively identify Japanese children with ASD, even HFASD. We would like to emphasize that not only screening but also continual community-based developmental surveillance is necessary for detecting children with ASD. Such enhancement of multidisciplinary community assessment should result in promoting the development of children with ASD and improve their quality of life (Kamio et al. 2013).

References

Allen, C. W., Silove, N., Williams, K., & Hutchins, P. (2007). Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders, 37*, 1272–1278.

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

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S., et al. (2000). A screening instrument for autism at 18 months of age: A 6-year follow-up study. *Journal of American Academy of Child and Adolescent Psychiatry, 39*, 694–702.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry, 161*, 839–843.

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., et al. (1996). Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry, 168*, 158–163.

Barton, M., Chlebowski, C., & Fein, D. (2008). Screening for autism spectrum disorders in young children: A review. *Seishin Hoken Kenkyu (Journal of Mental Health), 54*, 7–28.

Charman, T., Baron-Cohen, S., Baird, G., Cox, A., Swettenhan, J., Wheelwright, S., et al. (2002). Is 18 months too early for the CHAT? *Journal of American Academy of Child and Adolescent Psychiatry, 41*, 235–236.

Charman, T., Baron-Cohen, S., Baird, G., Cox, A., Wheelwright, S., Swettenhan, J., et al. (2001). Commentary: The modified checklist for autism in toddlers. *Journal of Autism and Developmental Disorders, 31*, 145–148.

Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics, 131*, e1121–e1127. doi:10.1542/peds. 2012-1525.

Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14–15 months. II: Population screening with the early screening of autistic traits questionnaire (ESAT). Design and general findings. *Journal of Autism and Developmental Disorders, 36*, 713–722.

Eaves, L. C., Wingert, H., & Ho, H. H. (2006). Screening for autism: Agreement with diagnosis. *Autism, 10*, 229–242.

Fernell, E., Hedvall, A., Norrelgen, F., Eriksson, M., Höglund-Carlsson, L., Barnevik-Olsson, M., et al. (2010). Developmental profiles in preschool children with autism spectrum disorders referred for intervention. *Research in Developmental Disabilities, 31*, 790–799.

Gillberg, C. (2010). The ESSENCE in child psychiatry: Early symptomatic syndromes eliciting neurodevelopmental clinical examinations. *Research in Developmental Disabilities, 31*, 1543–1551.

Gray, K. M., Tonge, B. J., Sweeney, D. J., & Einfeld, S. L. (2008). Screening for autism in young children with developmental delay: An evaluation of the developmental behaviour checklist: Early screen. *Journal of Autism and Developmental Disorders, 38*, 1003–1010.

Inada, N., Kamio, Y., & Koyama, T. (2010). Developmental chronology of preverbal social behaviors in infancy using the M-CHAT: Baseline for early detection of atypical social development. *Research in Autism Spectrum Disorders, 4*, 605–611.

Inada, N., Koyama, T., Inokuchi, E., Kuroda, M., & Kamio, Y. (2011). Reliability and validity of the Japanese version of the modified checklist for autism in toddlers (M-CHAT). *Research in Autism Spectrum Disorders, 5*, 330–336.

Ito, H., Tani, I., Yukihiro, R., Adachi, J., Hara, K., Ogasawara, M., et al. (2012). Validation of an interview-based rating scale developed in Japan for pervasive developmental disorders. *Research in Autism Spectrum Disorders, 6*, 1265–1272.

Johnson, C. P., & Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*, 1183–1215.

Kamio, Y., Inada, N., & Koyama, T. (2013). A nationwide survey on quality of life and associated factors of adults with high-functioning autism spectrum disorders. *Autism, 17*, 16–27.

Kim, Y. S., Leventhal, B. L., Koh, Y. J., Fombonne, E., Laska, E., Lim, E. C., et al. (2011). Prevalence of autism spectrum disorders in a total population sample. *The American Journal of Psychiatry, 168*, 904–912.

Kleinman, J. M., Robins, D. L., Ventola, P. E., Pandey, J., Boorstein, H. C., Esser, E. L., et al. (2008). The modified checklist for autism in toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*, 827–839.

Kurita, H., Miyake, Y., & Katsuno, K. (1989). Reliability and validity of the childhood autism rating scale-Tokyo version (CARS-TV). *Journal of Autism and Developmental Disorders, 19*, 389–396.

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., & 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.

Mandell, D. S., Novak, M. M., & Zubritsky, C. D. (2005). Factors associated with age of diagnosis among children with autism spectrum. *Pediatrics, 116*, 1480–1486.

Mothers' & Children's Health & Welfare Association. (2007) *Maternal and child health statistics of Japan (in Japanese)*. Tokyo.

Oosterling, I. J., Wensing, M., Swinkels, S. H., van der Gaag, R. J., Visser, J. C., Woudenberg, T., et al. (2010). Advancing early detection of autism spectrum disorder by applying an integrated two-stage screening approach. *Journal of Child Psychology and Psychiatry, 51*, 250–258.

Pandey, J., Verbalis, A., Robins, D. L., Boorstein, H., Klin, A. M., Babitz, T., et al. (2008). Screening for autism in older and younger toddlers with the modified checklist for autism in toddlers. *Autism, 12*, 513–535.

Robins, D. L. (2008). Screening for autism spectrum disorders in primary care settings. *Autism, 12*, 537–556.

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.

Rogers, S. J., & Vismara, L. A. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child and Adolescent Psychology, 37*, 8–38.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The childhood autism rating scale (CARS)*. Los Angeles: Western Psychological Services.

Shattuck, P. T., Durkin, M. D., Maenner, M., Newschaffer, C., Mandell, D. S., Wiggins, L., et al. (2009). The timing of identification among children with an autism spectrum disorder: Findings from a population-based surveillance study. *Journal of American Academy of Child and Adolescent Psychiatry, 48*, 474–483.

Tsuchiya, K. J., Matsumoto, K., Yagi, A., Inada, N., Kuroda, M., Inokuchi, E., et al. (2012). Reliability and validity of autism diagnostic interview-revised, Japanese version. *Journal of Autism and Developmental Disorders,.* doi:10.1007/s10803-012-1606-9.

Volkmar, F., Chawarska, K., & Klin, A. (2005). Autism in infancy and early childhood. *Annual Review of Psychology, 56*, 315–336.

RESEARCH

Open Access

Normative data and psychometric properties of the strengths and difficulties questionnaire among Japanese school-aged children

Aiko Moriwaki and Yoko Kamio*

Abstract

Background: Although child mental health problems are among the most important worldwide issues, development of culturally acceptable mental health services to serve the clinical needs of children and their families is especially lacking in regions outside Europe and North America. The Strengths and Difficulties Questionnaire (SDQ), which was developed in the United Kingdom and is now one of the most widely used measurement tools for screening child psychiatric symptoms, has been translated into Japanese, but culturally calibrated norms for Japanese schoolchildren have yet to be established. To this end, we examined the applicability of the Japanese versions of the parent and teacher SDQs by establishing norms and extending validation of its psychometric properties to a large nationwide sample, as well as to a smaller clinical sample.

Methods: The Japanese versions of the SDQ were completed by parents and teachers of schoolchildren aged 7 to 15 years attending mainstream classes in primary or secondary schools in Japan. Data were analyzed to describe the population distribution and gender/age effects by informant, cut-off scores according to banding, factor structure, cross-scale correlations, and internal consistency for 24,519 parent ratings and 7,977 teacher ratings from a large nationwide sample. Inter-rater and test-retest reliabilities and convergent and divergent validities were confirmed for a smaller validation sample (total *n* = 128) consisting of a clinical sample with any mental disorder and community children without any diagnoses.

Results: Means, standard deviations, and banding of normative data for this Japanese child population were obtained. Gender/age effects were significant for both parent and teacher ratings. The original five-factor structure was replicated, and strong cross-scale correlations and internal reliability were shown across all SDQ subscales for this population. Inter-rater agreement was satisfactory, test-retest reliability was excellent, and convergent and divergent validities were satisfactory for the validation sample, with some differences between informants.

Conclusions: This study provides evidence that the Japanese version of the SDQ is a useful instrument for parents and teachers as well as for research purposes. Our findings also emphasize the importance of establishing culturally calibrated norms and boundaries for the instrument's use.

Keywords: Child mental health, Questionnaire, Reliability, Validity, Normative banding, Strengths and difficulties questionnaire

* Correspondence: kamio@ncnp.go.jp
Department of Child and Adolescent Mental Health, National Institute of
Mental Health, National Center of Neurology and Psychiatry, 4-1-1
Ogawa-Higashi, Kodaira, Tokyo 187-8553, Japan

Background

Mental health problems affect 10-20% of children and adolescents worldwide [1], and substantial evidence indicates continuity in psychopathology from childhood into adulthood [2-4]. Despite heightened public concern in Japan for childhood mental health problems [5-7], many of these children remain unidentified and have no access to professional support due to various barriers including an insufficient specialized community health service system and parents or school teachers having inadequate knowledge of and stigma against child mental health problems. Recognizing this urgency, the Japanese Ministry of Health, Labour and Welfare has provided basic training opportunities for primary health professionals and promoted multidisciplinary work in the community since 2008. In addition, in 2009, the Ministry of Education, Culture, Sports, Science and Technology revised the School Health Act to strengthen the role that school personnel play in the early identification of children with mental health problems.

To support such initiatives, we need to develop reliable and valid measurement tools of psychopathological symptoms in Japanese children. At present, among the various questionnaires available for measuring mental health problems in children and adolescents, the Child Behavioral Checklist (CBCL) [8] has long been viewed as the "gold standard" because of its comprehensive nature. Although the CBCL is a solid instrument for conducting in-depth assessment, the 25-item Strengths and Difficulties Questionnaire (SDQ) [9] may be more suitable for screening purposes. The SDQ was created by Goodman by adding items on concentration, peer relations, and social competence to the established Rutter questionnaires. Because the SDQ measures not only behavioral problems but also the strengths of children and adolescents aged 4–16 years [10], parents and teachers can easily complete it. Furthermore, authorized translations of the SDQ are available free of charge [11]; http://www.sdqinfo.com. Due to its ease of use, the SDQ has now been translated into more than 75 languages and extensively validated in clinical and community samples [12-25]. These prior studies revealed that population-specific SDQ norms vary widely across countries.

To the best of our knowledge, only one study has examined the Japanese version of the SDQ. That study analyzed parent ratings in a community sample of 2,899 children aged 4–12 years [18] and found a gender effect on parent ratings, showed cut-off scores according to score banding, and confirmed its five-factor structure and satisfactory internal consistencies. However, given the value of having multiple informants reporting on children's mental health problems especially for psychological assessment [26,27], we must examine whether its psychometric properties differ by rater. Also, to evaluate

clinical usefulness, we need to examine it in a psychiatric clinical population as well as in a community population. The urgency to enhance school mental health care necessitates establishing culturally calibrated norms for Japanese schoolchildren based on a nationwide sample rather than on data from a restricted local area. Therefore, this study examined the applicability of the Japanese version of the SDQs for parents and teachers by establishing norms and cut-offs according to bandings and extending validation of its psychometric properties to a large, nationwide, and representative sample as well as a smaller clinical sample.

Methods

This cross-sectional epidemiological study investigated the score distribution with gender and age effects, factor structure, reliability, and validity of the Japanese versions of the parent and teacher SDQs.

Participants and data collection

Participants comprised a large-sized sample recruited from primary and secondary schools (normative sample) and a small-sized sample (validation sample) that was locally recruited. The schools were recruited countrywide with assistance from the Japanese Ministry of Education, Culture, Sports, Science, Technology and local government boards of education. We did not include private schools, national schools, or schools for handicapped children. Data were collected between December 2009 and March 2010 at the end of the Japanese school year to ensure that teachers knew their students well.

Normative sample

The parent SDQ to be completed at home was distributed to all parents of schoolchildren (aged 7–15 years) attending mainstream classes in 148 primary schools and 71 secondary schools in the 10 geographical areas making up Japan, with a letter from the investigators and school principals informing them about the study. From the parents of 87,548 children, 25,779 returned questionnaires to the investigators (29.4% response rate). Among these schools, 142 primary schools and 69 secondary schools (2,769 classes) agreed to participate in the teacher rating portion of the study. First, parents were informed about the study with a letter from the investigators and school principals. Second, among schoolchildren whose parents gave written consent, classroom teachers chose 4 children (2 boys, 2 girls) per class using a predetermined rule. In classes where less than 4 parents gave consent, teachers were asked to complete the questionnaire for all children whose parents who consented. We received 8,272 questionnaires rated by 2,183 teachers (78.8% response rate; 2,183/2,769). Among all questionnaires returned, we excluded 1,260 parent

ratings (4.9%) and 295 teacher ratings (3.6%) with one or more missing answers, leaving 24,519 parent ratings (12,472 boys, 12,047 girls) and 7,977 teacher ratings (4,010 boys, 3,967 girls). Each of 9 grade levels comprised a minimum of 815 parent ratings and 302 teacher ratings for each gender (Table 1). The parent SDQ was rated by mothers (91.1%), fathers (7.6%), both parents (0.7%), and others (0.6%). The ratio of raters did not differ significantly between boys and girls ($\chi^2 = 1.27$, *ns*) or by age ($\chi^2 = 2.11$, *ns*). Therefore, the parent SDQ data rated by different raters were combined and analyzed in subsequent analyses.

Validation sample

Participants were recruited from research volunteers with or without mental disorders, local schools, or a local pediatric outpatient clinic specializing in neurodevelopmental disorders. Participants totaled 128 children aged 6 to 16 years, of which 73 had any psychiatric diagnosis and 55 had no diagnosis (19 typically developing, 29 from community schools). Psychiatric diagnoses given by child psychiatrists or developmental pediatricians were autism spectrum disorder ($n = 47$), attention-deficit/hyperactivity disorder ($n = 23$), anxiety disorder ($n = 2$), specific phobia ($n = 14$), social phobia ($n = 4$), obsessive-compulsive disorder ($n = 1$), adjustment disorder ($n = 2$), tic disorders ($n = 5$), and others ($n = 7$). Thirteen of 73 children with any mental disorder had more than one diagnosis. Parent ratings were obtained for 108 children (69 clinical), and teacher ratings were obtained for 75 children (42 clinical). To examine inter-rater reliability, we used data from 63 participants rated by both parent and teacher at almost the same time. We collected retest data from the parents of 34 children 14 to 137 days later, and teachers of 18 children

10 to 107 days later (practical limitations precluded a shorter collection interval).

Measures

Strengths and difficulties questionnaire

The SDQ is a 25-item questionnaire assessing child psychopathology and positive strengths of children and adolescents. Twenty-five items are classified into five subscales, four difficulties subscales (emotional symptoms, conduct problems, hyperactivity/inattention, peer problems) and one subscale on prosocial behavior. Each item is scored on a 3-point scale (0 = not true, 1 = somewhat true, 2 = certainly true). Each subscale score ranges from 0 to 10, and four difficulties subscale scores add up to a total difficulties score (range 0–40); higher difficulties scores indicate more difficulties, whereas the prosocial subscale score is reversely coded. The authorized Japanese translations of the SDQ [28] were used in this study.

Child behavioral checklist

The CBCL, a 113-item questionnaire assessing child psychopathology, comprises eight subscales (withdrawal problems, somatic complaints, anxious/depressed, social problems, thought problems, attention problems, delinquent behavior, aggressive behavior) [8]. After each item is scored on a 3-point scale, eight individual subscale scores, an internalizing score (withdrawal problems, somatic complaints, and anxious/depressed subscales), an externalizing score (delinquent and aggressive behavior subscales), and a total score can be calculated. The Japanese version was shown to be valid and reliable [29,30] and to have an 8-syndrome structure [31]. In this study, 46 parents and 29 teachers of primary schoolchildren in the validation sample completed the CBCL for Ages 4–18 (CBCL/4-18) and the Teacher Rating Form (TRF), respectively.

ADHD-rating scale-IV

The ADHD-Rating Scale-IV (ADHD-RS) is an 18-item questionnaire assessing symptom frequency characterized by attention deficit/hyperactivity disorder in children and adolescents [32]. Each item is scored on a 4-point scale, and inattention (sum of odd-numbered items), hyperactivity-impulsivity (sum of even-numbered items), and total score (sum of all items) can be calculated. The Japanese versions of the ADHD-RS home and school forms were shown to be valid, reliable, and to have a two-factor structure [33,34]. In this study, 41 parents and 43 teachers of primary schoolchildren completed the home form and school form, respectively.

Ethical considerations

The study protocol was approved by the Ethics Committee of the National Center of Neurology and Psychiatry, Japan,

Table 1 Number of children in the normative sample by gender and grade

Grade	SDQ parent ratings ($n = 24,519$)				SDQ teacher ratings ($n = 7,977$)			
	Boys	%	Girls	%	Boys	%	Girls	%
1	1,792	14.4	1,633	13.6	526	13.1	519	13.1
2	1,662	13.3	1,514	12.6	547	13.6	540	13.6
3	1,526	12.2	1,541	12.8	481	12.0	485	12.2
4	1,479	11.9	1,506	12.5	509	12.7	506	12.8
5	1,562	12.5	1,382	11.5	499	12.4	478	12.0
6	1,321	10.6	1,334	11.1	484	12.1	486	12.3
7	1,162	9.3	1,186	9.8	346	8.6	343	8.6
8	1,100	8.8	1,136	9.4	316	7.9	307	7.7
9	868	7.0	815	6.8	302	7.5	303	7.6
Total	12,472		12,047		4,010		3,967	

Note. *SDQ*, strengths and difficulties questionnaire. Most grade 1 participants were 7 years old at the time of the survey.

and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. We obtained written informed consent to participate in this study from the caregivers of each child participant.

Statistical analysis

Because the SDQ score distribution in the normative sample was significantly different from a normal distribution (Shapiro-Wilk and Kolmogorov-Smirnov tests, both $p < .01$), subsequent statistical analyses employed non-parametric tests. To examine gender effects, we used the Mann–Whitney U-test to compare scale scores between boys and girls. To examine age effects, we used the Kruskal-Wallis test and post-hoc Mann-Whitney's comparisons with Bonferroni correction on the scale scores of three age groups (7–9, 10–12, 13–15 years). We conducted exploratory factor analysis (EFA) with varimax rotation and confirmatory factor analysis (CFA) on the normative sample to confirm the five-factor model. On the normative sample, we calculated internal consistency for the total difficulties score and each subscale score, and we assessed cross-scale correlations between the five scales using Spearman's rank correlations. Inter-rater and test-retest reliabilities and convergent and divergent validities were assessed using Spearman's rank correlations on the validation sample. We also examined temporal stability using a repeated-measures Wilcoxon signed-rank test on scores rated on two occasions for a smaller validation sample. All statistical

analysis was performed with SPSS version 17.0 and AMOS version 10.0.

Results

Population distribution, and gender and age effects

Table 2 shows the means and standard deviations of parent- and teacher-rated SDQ scores in the normative sample, and also gender and age effects on the SDQ scores. Gender effects were significant for both parent and teacher ratings on total difficulties and all five subscale scores (total difficulties: $U = 67,710,000, 5,796,000$; emotional symptoms: $U = 70,330,000, 7,782,000$; conduct problems: $U = 69,980,000, 6,558,000$; hyperactivity/inattention: $U = 61,150,000, 5,180,000$; peer problems: $U = 73,270,000, 7,140,000$; prosocial behavior: $U = 67,710,000, 5,796,000$ [for parent and teacher ratings, respectively, $p < 0.001$ for all except teacher-rated emotional symptoms, $p < 0.05$ for teacher-rated emotional symptoms]). Parent ratings showed that boys scored significantly higher than girls on total difficulties and on the conduct problems, hyperactivity/inattention, and peer problems subscales, whereas girls scored significantly higher than boys on the emotional symptoms and prosocial behavior subscales. However, the effect sizes (r) of these gender differences were negligible. Teacher ratings, on the other hand, showed that boys scored significantly higher than girls on total difficulties and on all of the difficulties subscales, whereas girls scored significantly higher than boys on the prosocial behavior subscale. The effect sizes (r) of gender differences of teacher ratings on total difficulties and on

Table 2 Mean scores of parent- and teacher-rated SDQs and gender and age effects

SDQ	Boys		Girls		Gender effect (p, r)	7-9 years		10-12 years		13-15 years		Age effect (p, Cramer's *V*)
	M	(SD)	M	(SD)		M	(SD)	M	(SD)	M	(SD)	
Parent ratings	($n = 12,472$)		($n = 12,047$)			($n = 9,968$)		($n = 8,584$)		($n = 6,267$)		
Total difficulties	8.02	(5.26)	7.11	(4.76)	†	8.39	(5.09)	7.20	(4.94)	6.82	(4.94)	a† b† c†, 0.15
Emotional symptoms	1.31	(1.67)	1.49	(1.76)	†	1.59	(1.77)	1.33	(1.67)	1.21	(1.68)	a† b† c†, 0.11
Conduct problems	1.92	(1.59)	1.70	(1.43)	†	2.01	(1.57)	1.74	(1.50)	1.62	(1.43)	a† b† c†, 0.12
Hyperactivity/inattention	3.23	(2.30)	2.49	(1.98)	†	3.27	(2.26)	2.69	(2.13)	2.49	(2.00)	a† b† c†, 0.16
Peer problems	1.55	(1.69)	1.42	(1.50)	†	1.52	(1.57)	1.44	(1.58)	1.51	(1.68)	a†
Prosocial behavior	5.80	(2.15)	6.50	(2.08)	†	6.18	(2.10)	6.26	(2.15)	5.91	(2.20)	a† b† c†
Teacher ratings	($n = 4,010$)		($n = 3,967$)			($n = 3,098$)		($n = 2,962$)		($n = 1,917$)		
Total difficulties	6.37	(5.80)	3.95	(4.50)	†, 0.24	5.74	(5.70)	4.94	(5.22)	4.58	(4.79)	a† c†
Emotional symptoms	0.82	(1.48)	0.77	(1.42)	‡	0.93	(1.55)	0.76	(1.44)	0.64	(1.23)	a† b† c†
Conduct problems	1.20	(1.68)	0.68	(1.22)	‡	1.06	(1.61)	0.90	(1.45)	0.81	(1.35)	a† c†
Hyperactivity/inattention	2.89	(2.67)	1.37	(1.76)	†, 0.31	2.46	(2.60)	2.01	(2.32)	1.79	(2.04)	a† c†
Peer problems	1.47	(1.86)	1.13	(1.56)	†	1.30	(1.71)	1.28	(1.75)	1.34	(1.73)	
Prosocial behavior	5.73	(2.74)	7.14	(2.49)	†, 0.26	6.47	(2.68)	6.48	(2.70)	6.28	(2.76)	c†

Note. *SDQ*, strengths and difficulties questionnaire. Age bands 7–9 years, 10–12 years, 13–15 years correspond to grades 1–3, 4–6, 7–9, respectively.
Age effect: [a]7–9 yrs > 10–12 yrs, [b]10–12 yrs > 13–15 yrs, [c]7–9 yrs > 13–15 yrs. †$p < 0.05$, ‡$p < 0.001$.

hyperactivity/inattention and prosocial behavior subscale scores were small (0.24-0.31), although the rest were negligible (Table 2).

Age effects were also significant for both parent and teacher ratings except for the teacher-rated peer problem subscale. As for parent ratings, total difficulties and all subscale scores were significantly different by age band (total difficulties: $\chi^2 = 568.33$; emotional symptoms: $\chi^2 = 307.30$; conduct problems: $\chi^2 = 323.96$; hyperactivity/inattention: $\chi^2 = 586.60$; peer problems: $\chi^2 = 19.26$; prosocial behavior: $\chi^2 = 88.62$ [all $p < 0.001$]). Differences by age band were similar but diminished for teacher ratings (total difficulties: $\chi^2 = 51.75$; emotional symptoms: $\chi^2 = 59.14$; conduct problems: $\chi^2 = 18.69$; hyperactivity/inattention: $\chi^2 = 71.61$, all $p < 0.001$; peer problems: $\chi^2 = 5.64$, ns; prosocial behavior: $\chi^2 = 6.77$, $p < 0.05$). Post hoc comparisons between three age bands indicated that SDQ scores tended to be higher in younger children, as shown in Table 2. The effect size (Cramer's V) of age effects was small for parent-rated total difficulties, emotional symptoms, conduct problems, and hyperactivity/inattention subscale scores, although negligible for all teacher-rated scores.

Normative banding and cut-off score

Because gender or age effects were consistently observed for the total difficulties scores (Table 2), score ranges of the three bands (clinical, borderline, normal) were determined for the total difficulties scores by gender and age group (7–9, 10–12, 13–15 years) (Table 3). According to Goodman's original work [10], the highest 10th percentile of the normative sample is defined as the "clinical" range, the next 10th percentile as the "borderline" range, and the remaining 80th percentile as the "normal" range. Although discrete scores made it impossible to divide the sample into exact percentiles, as Table 3 shows, nearly 10%, 10%, and 80% of the children were in the clinical, borderline, and normal bands.

Factor analysis

Table 4 shows rotated factor loadings for a five-factor EFA performed on parent- and teacher-rated SDQ scores with a rearranged item order. Only five factors had eigenvalues greater than 1.00, consistent with the original study [14] and the previous Japanese study [18]. EFA revealed that the five factors accounted for 33.03% and 55.22% of total variance of parent and teacher ratings, respectively, and most items loaded moderately to strongly onto their predicted factors. Communality values for teacher ratings were generally fair, at over 0.40 for 23 of 25 items, whereas only 7 of 25 items exceeded 0.40 for parent ratings. Parent- and teacher-rated item 7 ("obedient") and teacher-rated item 14 ("popular") loaded onto the prosocial factor more strongly than onto the predicted factor. The loading of parent-rated item 10 ("fidgety") onto the emotional factor was also higher than that onto the predicted factor.

Furthermore, CFA results lend support to the five-factor structure of the SDQ; for the parent and teacher ratings, respectively, the comparative fit index was 0.83 and 0.86, the goodness of fit index was 0.93 and 0.89, the adjusted goodness of fit index was 0.91 and 0.86, and the root mean square error of approximation was 0.06 and 0.07. In addition, the 3 items (7, 10, 14) mentioned above were found to load onto the predicted factor with factor loadings >0.40 (0.43-0.75).

Cross-scale correlations

Table 5 presents cross-scale correlations among five subscales by rater and gender. Correlations between externalizing-externalizing scales, that is, between conduct problems and hyperactivity/inattention, were strong (parent $\rho = 0.48$, teacher $\rho = 0.53$). By contrast, those between internalizing-externalizing scales were small (between emotional symptoms and conduct problems: parent $\rho = 0.28$, teacher $\rho = 0.25$; between emotional symptoms and hyperactivity/inattention: parent $\rho = 0.28$, teacher $\rho = 0.32$). Prosocial behavior was

Table 3 Normative banding of total difficulties score for parent- and teacher-rated SDQs for Japanese children

SDQ		7-9 years				10-12 years				13-15 years			
		Boys		Girls		Boys		Girls		Boys		Girls	
		Raw score	(%)	Raw score	(%)	Raw score	(%)	Raw score	(%)	Raw score	(%)	Raw score	(%)
Parent rating	Normal	0-13	82.0%	0-11	81.0%	0-11	79.8%	0-10	82.0%	0-10	79.7%	0-10	81.5%
	Borderline	14-16	9.0%	12-14	9.7%	12-14	9.9%	11-13	8.2%	11-14	11.3%	11-13	8.9%
	Clinical	17-40	9.0%	15-40	9.3%	15-40	10.3%	14-40	9.8%	15-40	9.0%	14-40	9.6%
Teacher rating	Normal	0-11	78.9%	0-7	80.5%	0-10	78.1%	0-6	81.4%	0-9	81.3%	0-6	82.5%
	Borderline	12-16	11.6%	8-11	10.2%	11-14	10.8%	7-9	9.6%	10-12	8.9%	7-9	7.8%
	Clinical	17-40	9.5%	12-40	9.3%	15-40	11.1%	10-40	9.0%	13-40	9.8%	10-40	9.7%

Note. SDQ, strengths and difficulties questionnaire. There were no significant differences in proportion by age band between parent and teacher ratings for either boys or girls.

Table 4 Results of exploratory factor analysis (Varimax Rotation) of parent- and teacher-rated SDQs for Japanese children

SDQ items	Parent ratings (n = 24,519)						Teacher ratings (n = 7,977)					
	Factor I	Factor II	Factor III	Factor IV	Factor V	Communality	Factor I	Factor II	Factor III	Factor IV	Factor V	Communality
	Pro	Hyper	Emotion	Conduct	Peer		Pro	Hyper	Emotion	Conduct	Peer	
Initial eigenvalue	4.88	2.60	1.70	1.21	1.12	11.52	7.07	2.60	1.82	1.24	1.08	13.80
% of variance	9.06	16.82	23.68	28.39	33.03		16.68	28.53	38.89	47.25	55.22	
Prosocial behavior												
1 considerate	-.63				.45		-.75					.65
4 shares	-.49				.26		-.64					.44
9 caring	-.66				.45		-.81					.69
17 kind to kids	-.53				.29		-.74					.57
20 helps out	-.58				.37		-.78					.63
Hyperactivity/inattention												
2 restless		.56				.46		.80				.74
10 fidgety		.27	.34			.27		.61				.56
15 distractive		.69				.63		.82				.77
21 reflective (*)		.56				.44		.57				.63
25 persistent (*)		.64				.50		.59				.59
Emotional symptoms												
3 somatic complaints			.31			.14			.54			.37
8 worries			.55			.37			.75			.59
13 unhappy			.44			.30			.65			.48
16 clingy			.62			.43			.68			.56
24 fears			.51			.29			.68			.51
Conduct problems												
5 temper				.45		.33				.57		.49
7 obedient (*)	.44			.28		.30	.54			.38		.44
12 fights				.46		.25				.67		.60
18 lies, cheats				.41		.31				.62		.54
22 steals				.23		.07				.58		.34
Peer problems												
6 solitary				.41	.21						.70	.55
11 good friend (*)				.38	.18						.61	.48
14 popular (*)				.42	.34	.55					.42	.59
19 picked on, bullied				.44	.33						.52	.44
23 best with adults				.50	.31						.68	.54

Note. SDQ, strengths and difficulties questionnaire. *indicates a reverse item and inverted scores were analyzed.

negatively correlated with externalizing behaviors (conduct problems, hyperactivity/inattention: parent $\rho = 0.32$, 0.31; teacher $\rho = 0.50$, 56, respectively) but showed little correlation with internalizing behaviors (emotional symptoms: parent $\rho = -0.03$, teacher $\rho = -0.17$). These findings were in line with the theoretical predictions, and common in boys and girls. All correlations were statistically significant at $p < 0.01$.

Table 5 Cross-scale correlations for parent- and teacher-rated SDQs of Japanese children aged 7–15 years (Spearman's rho)

SDQ subscale		Parent rating (n = 24,519)				Teacher rating (n = 7,977)			
		Conduct problems	Hyperactivity/ inattention	Peer problems	Prosocial behavior	Conduct problems	Hyperactivity/ inattention	Peer problems	Prosocial behavior
Emotional symptoms	Boys	.29*	.31*	.33*	-.05*	.27*	.34*	.37*	-.18*
	Girls	.28*	.28*	.31*	-.04*	.23*	.33*	.37*	-.16*
	Total	.28*	.28*	.32*	-.03*	.25*	.32*	.37*	-.17*
Conduct problems	Boys		.50*	.24*	-.30*		.57*	.41*	-.50*
	Girls		.45*	.25*	-.33*		.45*	.41*	-.46*
	Total		.48*	.25*	-.32*		.53*	.42*	-.50*
Hyperactivity/inattention	Boys			.31*	-.28*			.41*	-.53*
	Girls			.28*	-.30*			.44*	-.52*
	Total			.30*	-.31*			.43*	-.56*
Peer problems	Boys				-.24*				-.46*
	Girls				-.25*				-.47*
	Total				-.24*				-.47*

Note. *SDQ*, strengths and difficulties questionnaire. Parent ratings: boys (n = 12,472), girls (n = 12,047). Teacher ratings: boys (n = 4,010), girls (n = 3,967).
*p < 0.01.

Internal consistency
Table 6 shows that internal consistencies were generally good, with those of teacher ratings tending to be stronger than those of parent ratings. The relatively weak internal consistencies of conduct problems and peer problems might be explained by the cross-loadings of items 7 and 11 mentioned above. Cronbach's α coefficients were very similar for boys and girls.

Inter-rater reliability
In a smaller subsample, parent-teacher correlations were found to be moderate for total difficulties scores (n = 63, 44 boys, 19 girls, mean age 9.0 ± 1.3 years, 42 with clinical diagnoses, 21 with no diagnoses; $\rho = 0.40$). Spearman's rank correlation coefficients varied by subscale: emotional symptoms $\rho = 0.49$, conduct problems $\rho = 0.33$, hyperactivity/inattention $\rho = 0.34$, peer problems $\rho = 0.50$, and prosocial behavior $\rho = 0.28$. All were statistically significant ($p < 0.01$ for all scales except for prosocial behavior, $p < 0.05$ for prosocial behavior).

Test-retest reliability
Thirty-four parents of a subsample (17 boys, 17 girls, mean age 10.4 ± 2.7 years, 19 with clinical diagnoses, 15 with no diagnoses) and 18 classroom teachers of children from community schools (12 boys, 6 girls, mean age 10.3 ± 2.8 years, 4 with clinical diagnoses, 14 with no diagnoses) completed the SDQ on two occasions (intervals: mean 54 ± 43 days, [14–137 days], mean 25 ± 25 days [10–107 days] for parents and teachers, respectively). Test-retest correlations of both parent and teacher ratings were excellent for total difficulties and all subscales (total difficulties $\rho = 0.79$, 0.95; emotional symptoms $\rho = 0.80$, 0.76; conduct problems $\rho = 0.76$, 0.88; hyperactivity/inattention $\rho = 0.70$, 0.84; peer problems $\rho = 0.74$, 0.79; prosocial behavior $\rho = 0.87$, 0.72; parent and teacher, respectively; all $p < 0.01$). Both parent and teacher ratings on two occasions did not significantly differ for any of the subscales except teacher-rated peer problems ($Z = -2.14$, $p < 0.05$, two-tailed test), indicating overall temporal stability.

Table 6 Cronbach's alpha coefficients for SDQ scores of Japanese children aged 7–15 years

SDQ	Parent rating (n = 24,519)			Teacher rating (n = 7,977)		
	Boys	Girls	Total	Boys	Girls	Total
Total difficulties score	.82	.79	.81	.86	.84	.86
Emotional symptoms	.64	.65	.64	.72	.72	.72
Conduct problems	.56	.50	.54	.69	.62	.67
Hyperactivity/inattention	.78	.73	.76	.85	.75	.84
Peer problems	.62	.54	.59	.70	.64	.68
Prosocial behavior	.72	.71	.73	.84	.82	.84

Note. *SDQ*, strengths and difficulties questionnaire.

Convergent and divergent validity
Table 7 shows the correlations between parent-rated SDQ and CBCL/4-18 scores for 46 clinical patients (36 boys, 10 girls, mean age 8.0 ± 0.8 years) and those between teacher-rated SDQ and TRF scores for 29 clinical patients (23 boys, 6 girls, mean age 7.9 ± 0.7 years). SDQ total difficulties scores were strongly correlated with CBCL total scores for ratings by both parents and teachers (parent $\rho = 0.56$, teacher $\rho = 0.77$). Correlations between corresponding subscales of the SDQ and the CBCL were also moderate to strong: those between SDQ conduct problems scores and externalizing scores of the CBCL4-18/TRF (externalizing, delinquent behavior, aggressive behavior subscales) were strong (parent $\rho = 0.50$-0.66, teacher $\rho = 0.66$-0.80), whereas those between SDQ emotional symptoms scores and internalizing scores of the CBCL4-18/TRF (internalizing, withdrawal problems, somatic complaints, anxiety/depressed subscales) were moderate to strong (parent $\rho = 0.40$-0.52, teacher $\rho = 0.50$-0.57). All correlations were statistically significant ($p < 0.01$). By contrast, there were no significant correlations among subscales measuring conceptually different behaviors, as shown in Table 7.

Similarly, Table 8 shows that SDQ hyperactivity/inattention subscale scores were strongly correlated with the ADHD-RS total scores as well as the inattention and hyperactivity/compulsion subscale scores for parent ratings (n = 41 from local schools, 25 boys, mean age 8.1 ± 1.5 years) and teacher ratings (n = 43 from local schools, 27 boys, mean age 8.1 ± 1.5 years). Strong correlations were also found between SDQ conduct problems subscale scores and ADHD-RS total and two subscales scores. By contrast, no significant correlation existed between the teacher-rated emotional symptoms subscale score and ADHD-RS score, although the correlation was moderate for the parent ratings.

Discussion
Our results provided normative data of parent and teacher SDQs for Japanese schoolchildren aged 7 to 15 years, and confirmed its reliability and validity.

Gender and age effects in the general population
As for gender effects, both parents and teachers reported higher levels of difficulties for boys than for girls, except for emotional symptoms. Such gender differences in SDQ scores are well in line with previous SDQ studies across ages and countries [13,15-19,21-24] and in the original U.K. study [35]. In our study, observed gender differences were more pronounced in teacher ratings than parent ratings, a tendency that has also been reported in previous studies using SDQ [13,16,23,35,36]. A possible explanation for this tendency is that girls might be more able to adjust their behaviors to social situations

than boys. Thus, we should exercise caution when interpreting information from parents and teachers when assessing clinical severity. Our finding of gender differences emphasizes the need to establish a culturally calibrated gender-specific norm for each SDQ rater version.

As for age effects, both parents and teachers reported the highest levels of difficulties for the youngest children, aged 7–9 years, although we found no systematic differences for either peer problems or prosocial behaviors. In our study, we found a robust line of descending tendency with age only for parent ratings; the effect size for teacher ratings was negligible. Many studies have reported a similar descending tendency of parent ratings with age [13,18,23,24,36], although no such age effect was found in community samples in Holland [19] or Hong Kong [16] or in an epidemiological sample in the United Kingdom [37]. By contrast, except for a study from Shanghai, China [13], almost all studies, including ours, found no systematic age difference for teacher ratings [16,23,36,38]. A Dutch study that examined parent, teacher, and self-ratings of the SDQ reported no age effect except in parent ratings [23]. Although ADHD prevalence decreases with development [39], a recent prospective and longitudinal study revealed that childhood-onset psychiatric disorders are relatively stable, and homotypic or heterotypic continuity is found for each disorder, especially behavioral disorders such as ADHD [37]. In other words, the descending tendency of parent ratings might reflect a phenotypic transition in their child rather than a true change in severity. Instead, as children get older, they might begin to conceal worries and problems from their parents. Therefore, researchers and clinicians might want to consider the clinical significance of gender and age differences when applying normative bandings to specific child populations [12].

Mean and cut-off scores of the Japanese version of the SDQ were lower than those for Europe, the United States, and China, although they were similar to those for Israel and Holland. These studies cannot be easily compared because the age ranges studied in their samples were not identical. However, the tendency for Japanese parents or teachers to give lower scores to children's behaviors appears consistent among questionnaires such as the CBCL [29], ADHD-RS [33,34], and Social Responsiveness Scale [40,41]. One partial explanation for the relatively lower scores of Japanese children on behavioral measures such as the SDQ is that Japanese informants tend to respond to Likert-type ratings by choosing the scale's midpoint, whereas U.S. informants tend to choose the scale's extreme values [42]. In fact, if the original U.K. cut-off were applied to Japanese children, some Japanese children in the "clinical" range instead would be labeled "borderline", and some labeled "borderline" would fall into

Table 7 Correlations between the SDQ and CBCL for each rater (Spearman's rho)

CBCL SDQ	Withdrawal problems	Somatic complaints	Anxiety/ dep	Social problems	Thought problems	Attention problems	Delinquent behaviors	Aggressive behavior	Internalizing	Externalizing	Total
Parent rating (n = 46)											
Total difficulties score	.32*	.44**	.25	.48**	.23	.62**	.54**	.43**	.36*	.51**	.56**
Emotional symptoms	.40**	.48**	.44**	.23	.20	.19	-.03	.07	.52**	.05	.34*
Conduct problems	.19	.21	.16	.06	.00	.37*	.66**	.50**	.21	.59**	.39**
Hyperactivity/inattention	.06	.27	.00	.35*	.12	.58**	.49**	.39**	.09	.44**	.39**
Peer problems	.20	.09	.05	.50**	.13	.32*	.11	.00	.10	.04	.18
Prosocial behavior	-.26	-.15	-.03	-.07	-.06	-.27	-.30*	-.34*	-.16	-.37*	-.21
Teacher rating (n = 29)											
Total difficulties score	.44**	.29	.49**	.75**	.48**	.82**	.55**	.68**	.48**	.68**	.77**
Emotional symptoms	.23	.57**	.56**	.37*	.21	.33	.11	.18	.50**	.18	.36
Conduct problems	.24	.05	.18	.60**	.33	.71**	.66**	.79**	.22	.80**	.66**
Hyperactivity/inattention	.30	.22	.33	.52**	.36	.74**	.40*	.62**	.31	.59**	.66**
Peer problems	.46*	-.05	.31	.75**	.55**	.66**	.51**	.53**	.33	.54**	.64**
Prosocial behavior	-.34	-.03	-.15	-.43*	-.28	-.44*	-.14	-.46*	-.23	-.40*	-.40*

Note. SDQ, strengths and difficulties questionnaire. CBCL, child behavioral checklist. The subsample from which parent ratings were obtained (n = 46) consisted of clinical patients (36 boys, mean age 8.0 ± 0.8). The subsample from which teacher ratings were obtained (n = 29) consisted of clinical patients (23 boys, mean age 7.9 ± 0.7). *p < 0.05, **p < 0.01.

Table 8 Correlations between the SDQ and ADHD-RS for each rater (Spearman's rho)

ADHD-RS SDQ	Parent rating (n = 41)			Teacher rating (n = 43)		
	Inattention	Hyperactivity/impulsivity	Total	Inattention	Hyperactivity/impulsivity	Total
Total difficulties score	.76**	.67**	.77**	.73**	.65**	.74**
Emotional symptoms	.34*	.33*	.34*	.28	.17	.26
Conduct problems	.70**	.70**	.75**	.53**	.57**	.60**
Hyperactivity/inattention	.73**	.63**	.73**	.83**	.81**	.85**
Peer problems	.58**	.51**	.59**	.36*	.22	.33*
Prosocial behavior	-.29	-.26	-.31*	-.42**	-.86**	-.48**

Note. SDQ, strengths and difficulties questionnaire. ADHD-RS: ADHD-Rating Scale-IV. The subsample from which parent ratings were obtained (n = 41) consisted of primary schoolchildren (25 boys, mean age 8.1 ± 1.5). The subsample from which teacher ratings were obtained (n = 43) consisted of primary schoolchildren (27 boys, mean age 8.1 ± 1.5). *p < 0.05, **p < 0.01.

the "normal" range. Thus, for both culturally appropriate use and cross-cultural research, we must establish national norms based on population distribution.

Factor analysis

We confirmed the proposed five-factor structure for the Japanese version of the parent and teacher SDQs using EFA and CFA.

Reliability and validity

Internal consistency, inter-rater reliability, and test-retest reliability of the Japanese version of the parent and teacher SDQs were generally satisfactory and comparable to the original version [14], and on the whole fell well within previously reported ranges [43]. On all subscales of internal consistency, teacher ratings were more reliable, a tendency that is in line with those of previous studies [43]. The test-retest interval of 10 days to 5 months in our study was wider than that in conventional measurement, but the test-retest reliability from our sample is comparable to that of samples with shorter intervals of 2 weeks to 2 months [13,16,19]. Therefore, the true test-retest reliability with a shorter interval might be even higher than the finding in the present study [14,15].

Regarding convergent validity, strong correlations between the SDQ and CBCL support that, overall, the Japanese SDQ measures the same construct that the Japanese CBCL measures, as shown in many studies [43]. Again, the correlation was higher for teacher ratings than for parent ratings. At the subscale level, correlations between SDQ behavioral difficulties subscales (e.g., conduct problems and hyperactivity/inattention subscales) and corresponding CBCL subscales were higher than the correlation between the SDQ emotional symptoms subscale and the corresponding CBCL subscale for both parent and teacher ratings. In addition, the SDQ hyperactivity/inattention subscale was highly correlated with the ADHD-RS measures for both parent and teacher ratings. This parent-teacher discrepancy or externalizing-internalizing discrepancy appears to be consistent with the studies reviewed by Stone [43].

Limitations

This study has a number of limitations. First, despite a sufficiently large-sized normative sample, the validation sample was small and the clinical information was based on experts' clinical judgment obtained without a validated structured interview in some cases. Thus, we could establish neither discriminant validity nor calculated sensitivity or specificity against psychiatric diagnoses. Second, the parent SDQ response rate was low (29.4%), although that of the teacher SDQ was acceptable (78.8%). Van Widenfelt et al. [23] pointed out that children of non-responding parents but not non-responding schools are likely to show higher scores. Also, we did not obtain demographic information (e.g., parental education level, income, and age; one- or two-parent family; number of siblings; teachers' age and gender) that might be related to SDQ scores [12]. Therefore, the representativeness of our normative sample for parent ratings is unclear, although the normative sample rated by teachers was representative. Also, the influence of demographic factors on parents' or teachers' ratings is unclear. Third, because the age range of participants in the present study was restricted to school age (7–15 years), the applicability of the Japanese version of the SDQ for preschoolers is unknown. Fourth, we did not study the self-report version for adolescents aged approximately 11 to 16 years, who are an important target for community mental health service planning. Thus, a future study examining its usefulness as a screening tool must include detailed clinical data from a larger clinical sample and investigate its ability to discriminate between community and clinical samples and receiver operating characteristic curves. In addition, Japanese norms and psychometric properties of parent and teacher ratings for preschoolers and self-report for adolescents should be examined.

Conclusions

This study provides gender- and age-specific norms by rater for Japanese schoolchildren and further evidence that the psychometric properties of the Japanese version of the parent and teacher SDQs are satisfactory. The findings indicate that the SDQ will serve as an efficient assessment tool of broad mental health problems in Japanese schoolchildren for research and clinical purposes, and that it is comparable to the original version and many other language versions. Our findings also emphasize the importance of establishing culturally calibrated norms and boundaries for each instrument's use.

References
1. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, Rohde LA, Srinath S, Ulkuer N, Rahman A: Child and adolescent mental health worldwide: evidence for action. *Lancet* 2011, 378:1515–1525.
2. Caspi A, Moffitt TE, Newman DL, Silva PA: Behavioral observations at age 3 years predict adult psychiatric disorders: longitudinal evidence from a birth cohort. *Arch Gen Psychiatry* 1996, 53:1033–1039.
3. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE: Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National comorbidity survey replication. *Arch Gen Psychiatry* 2005, 62:593–602.
4. Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Benjet C, Georgiades K, Swendsen J: Lifetime prevalence of mental disorders in U.S. adolescents: results from the National comorbidity survey replication—adolescent supplement (NCA-A). *J Am Acad Child Adolesc Psychiatry* 2010, 49:980–989.
5. Denda K, Kako Y, Kitagawa N, Koyama T: Assessment of depressive symptoms in Japanese school children and adolescents using the Birleson depression self-rating scale. *Int J Psychiatry Med* 2006, 36:231–234.
6. Kondo N, Sakai M, Kuroda Y, Kiyota Y, Kitabata Y, Kurosawa M: General condition of hikikomori (prolonged social withdrawal) in Japan: psychiatric diagnosis and outcome in mental health welfare centers. *Int J Soc Psychiatry* 2013, 59:79–86.
7. Nishida A, Tanii H, Nishimura Y, Kajiki N, Inoue K, Okada M, Sasaki T, Okazaki Y: Associations between psychotic-like experiences and mental health status and other psychopathologies among Japanese early teens. *Schizophr Res* 2008, 99:125–133.
8. Achenbach TM: *Manual for the Child Behavior Checklist and 1991 Profile.* Burlington, VT: University of VT, Department of Psychiatry; 1991.
9. Goodman R: A modified version of the Rutter parent questionnaire including extra items on children's strengths. *J Child Psychol Psychiatry* 1994, 35:1483–1494.
10. Goodman R: The strength and difficulties questionnaire: a research note. *J Child Psychol Psychiatry* 1997, 38:581–586.
11. SDQ: Information for researchers and professionals about the Strengths and Difficulties Questionnaire. http://www.sdqinfo.com/.
12. Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz D: The strengths and difficulties questionnaire: U.S. normative data and psychometric properties. *J Am Acad Child Adolesc Psychiatry* 2005, 44:557–564.
13. Du Y, Kou J, Coghill D: The validity, reliability and normative scores of the parent, teacher and self report versions of the strengths and difficulties questionnaire in China. *Child Adolesc Psychiatry Ment Health* 2008, 2(8). http://dx.doi.org/10.1186/1753-2000-2-8.
14. Goodman R: Psychometric properties of the strength and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry* 2001, 40:1337–1345.
15. Hawes DJ, Dadds MR: Australian data and psychometric properties of the strengths and difficulties questionnaire. *Aust N Z J Psychiatry* 2004, 38:644–651.
16. Lai KYC, Luk ESL, Leung PWL, Wong ASY, Law L, Ho K: Validation of the Chinese version of the strengths and difficulties questionnaire in Hong Kong. *Soc Psychiatry Psychiatr Epidemiol* 2010, 45:1179–1186.
17. Mansbach-Kleinfeld I, Apter A, Farbstein I, Levine SZ, Ponizovsky AM: A population-based psychometric validation study of the strengths and difficulties questionnaire-Hebrew version. *Front Psychiatry* 2010, 1:151. http://dx.doi.org/10.3389/fpsyt.2010.00151.
18. Matsuishi T, Nagano M, Araki Y, Tanaka Y, Iwasaki M, Yamashita Y, Nagamitsu S, Iizuka C, Ohya T, Shibuya K, Hara M, Matsuda A, Tsuda A, Kakuma T: Scale properties of the Japanese version of the strengths and difficulties questionnaire (SDQ): A study of infant and school children in community samples. *Brain Dev* 2008, 30:410–415.
19. Muris P, Meesters C, van den Berg F: The strength and difficulties questionnaire (SDQ): further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *Eur Child Adolesc Psychiatry* 2003, 12:1–8.
20. Obel C, Heiervang E, Rodriguez A, Heyerdahl S, Smedje H, Sourander A, Guðmundsson OÓ, Clench-Aas J, Christensen E, Heian F, Mathiesen KS, Magnússon P, Njarðvík U, Koskelainen M, Rønning JA, Stormark KM, Olsen J: The strengths and difficulties questionnaire in the Nordic countries. *Eur Child Adolesc Psychiatry* 2004, 13(Suppl 2):II32–II39.
21. Shojaei T, Wazana A, Pitrou I, Kovess V: The strength and difficulties questionnaire: Vvalidation study in French school-aged children and cross-cultural comparisons. *Soc Psychiatry Psychiatr Epidemiol* 2009, 44:740–747.
22. Syed EU, Hussein SA, Mahmud S: Screening for emotional and behavioral problems amongst 5-11-year-old school children in Karachi, Pakistan. *Soc Psychiatry Psychiatr Epidemiol* 2007, 42:421–427.
23. Van Widenfelt BM, Goedhart AW, Treffers PDA, Goodman R: Dutch version of the strengths and difficulties questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 2003, 12:2891–2891.
24. Woerner W, Becker A, Rothenberger A: Normative data and scale properties of the German parent SDQ. *Eur Child Adolesc Psychiatry* 2004, 13(Suppl 2):II3–II10.
25. Woerner W, Fleitlich-Bilyk B, Martinussen R, Fletcher J, Cucchiaro G, Dalgalarrondo P, Lui M: The strengths and difficulties questionnaire overseas: evaluations and applications of the SDQ beyond Europe. *Eur Child Adolesc Psychiatry* 2004, 13(2):47–54. http://dx.doi.org/10.1007/s00787-004-2008-0.
26. Achenbach TM, McConaughy SH, Howell C: Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychol Bull* 1987, 101(2):213–232. doi:10.1037/0033-2909.101.2.213.
27. Goodman R, Ford T, Richards H, Gatward R, Meltzer H: The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000, 41:645–655.
28. Youthinmind SDQ: Japanese. http://www.sdqinfo.com/py/sdqinfo/b3.py?language=Japanese.
29. Itani T, Kanbayashi Y, Nakata Y, Kita M, Fujii H, Kuramoto H, Negishi T, Tezuka M, Okada A, Natori H: Development of child behavior checklist/4-18 Japanese version. *Seishin Shinkeigaku Zasshi* 2001, 41(4):243–252.
30. Kawauchi M, Kihara N, Setoya Y, Makino H, Kita M, Kanbayashi Y: Standardization of child behavior checklist for ages 6–18. *Seishin Shinkeigaku Zasshi* 2011, 51(2):143–155.
31. Ivanova MY, Achenbach TM, Dumenci L, Harder VS, Ang RP, Bilenberg N, Bjarnadottir G, Capron C, De Pauw SSW, Dias P, Dobrean A, Doepfner M,

Duyme M, Eapen V, Erol N, Esmaeili EM, Ezpeleta L, Frigerio A, Gonçalves MM, Gudmundsson HS, Jeng S-F, Jetishi P, Jusiene R, Kim Y-A, Kristensen S, Lecannelier F, Leung PWL, Liu J, Montirosso R, Oh KJ, *et al*: Testing the 8-syndrome structure of the child behavior checklist in 30 societies. *J Clin Child Adolesc Psychol* 2007, 36:405–417. http://dx.doi.org/10.1080/15374410701444363.
32. DuPaul GJ, Power TJ, Anastropoulos AD, Reid R: *ADHD Rating Scale IV: Checklists, Norms, and Clinical Interpretation.* New York, NY: Guilford Press; 1998.
33. Ohnishi M, Okada R, Tani I, Nakajima S, Tsujii M: Japanese version of school form of the ADHD-RS: an evaluation of its reliability and validity. *Res Dev Disabil* 2010, 31(6):1305–1312. http://dx.doi.org/10.1016/j.ridd.2010.07.011.
34. Tani I, Okada R, Ohnishi M, Nakajima S, Tsujii M: Japanese version of home form of the ADHD-RS: an evaluation of its reliability and validity. *Res Dev Disabil* 2010, 31(6):1426–1433. http://dx.doi.org/10.1016/j.ridd.2010.06.016.
35. Youthinmind SDQ: British means and standard deviations for the 5–15 year old sample split by gender. http://www.sdqinfo.com/norms/UKNorm2.pdf.
36. Youthinmind SDQ: Australian means and standard deviations for the sample split by gender and age. http://www.sdqinfo.com/norms/AusNorm2.pdf.
37. Copeland WE, Adair CE, Smetanin P, Stiff D, Briante C, Colman I, Fergusson D, Horwood J, Poulgon R, Jane Costello E, Angold A: Diagnostic transitions from childhood to adolescence to early adulthood. *J Child Psychol Psychiatry* 2013, 54:791–799.
38. Youthinmind SDQ: British means and standard deviations for the sample split by age band. http://www.sdqinfo.com/norms/UKNorm3.pdf.
39. Faraone SV, Biederman J, Mick E: The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychol Med* 2006, 36:159–165.
40. Kamio Y, Inada N, Moriwaki A, Kuroda M, Koyama T, Tsujii H, Kawakubo Y, Kuwabara H, Tsuchiya KJ, Uno Y, Constantino JN: Quantitative autistic traits ascertained in a national survey of 22,529 Japanese schoolchildren. *Acta Psychiatr Scand* 2013, 128:45–53.
41. Kamio Y, Moriwaki A, Inada N: Utility of teacher-report assessments of autistic severity in Japanese school children. *Autism Res* 2013. http://dx.doi.org/10.1155/2013/373240.
42. Chen C, Lee S, Stevenson HW: Response style and crosscultural comparisons of rating scales among East Asian and North American students. *Psychol Sci* 1995, 6:170–175.
43. Stone LL, Otten R, Engels RCME, Vermlst AA, Janssens JMAM: Psychometric properties of the parent and teacher version of the strengths and difficulties questionnaire for 4- to 12- year-olds: a review. *Clin Child Fam Psychol Rev* 2010, 13:254–274. http://dx.doi.org/10.1007/s10567-010-0071-2.

発達障害の診察室で考えていること

中井昭夫　福井大学 子どものこころの発達研究センター
医学部附属病院子どものこころ診療部
特命准教授

発達障害バブル

世の中、ちょっとした？ いや、かなりの発達障害ブームである。「発達障害バブル」という言葉まででてきている。学会でも発達障害関連の演題は多く、少し前の「脳科学」ブームを引き続いで大型研究も盛んに行われている。書店に行けば一般向けの解説本から新書、専門書までたくさんの書籍が並んでいる。医師・保健師、心理士、保育士・教師向けはもちろん、一般市民向けの講演会、研修会も各地で多く開催されている。また、このようなことを背景にしてか、保育所や学校で、ちょっと変わっている、ちょっとうまく行かない、ちょっとお勉強についていけないと、すぐに園や学校から保護者が呼び出され「病院に行ってきてお薬をもらってきて下さい」「発達障害だと思うので診断書をもらってきて下さい」と言われ、納得のいかないまま、あるいは怒りを抱えながら受診されるケースも多い。また、医療の側も、身体疾患の鑑別のための診察や検査もきちんと行わず、いわゆるチェックリストのみで発達障害と診断し、「お子様は発達障害です」「このお薬が必要です」…ということも多く耳にする。昨日まで、優しく、まじめで、字は少し汚いがスポーツもでき、成績優秀で、歴史に詳しくクラスの尊敬を一手に集め、教師からも信頼の厚かった子が、クラスメイトからの心ない誹謗中傷によりキレて暴れたのをパニックとして大人３人がかりで引きずられて医療機関に連れて来られ、チェックリストでアスペルガー障害と診断された途端、癇癪・パニックを抑えるためにその日から薬物療法が開始され、教師や友人からも障害者扱いとなり、特別支援学校への進学を進められてしまうという現実。確かに本人の特性からくる「困り感」への早期の気づきと適切な予防的対応・支援は重要であり、自分もその中で診療、研究、教育、地域・社会貢献を行ってはいるのだが、このような流れの中でいつも何かしら違和感のようなものを抱えているのが実際のところである。

発達障害とは…
まず「発達障害者支援法」を正しく紐解く

超党派の議員立法として平成16年末に成立、翌年4月施行の「発達障害者支援法」によれば、「発達障害」とは「自閉症、アスペルガー症候群その他の広汎性発達障害、学習障害、注意欠陥多動性障害、その他これに類する脳機能の障害であってその症状が通常低年齢において発現するものとして政令で定めるものをいう。」(第2条第1項)と定義され、この広汎性発達障害(PDD)、学習障害(LD)、注意欠陥多動性障害(ADHD)の3つについては社会的にも認知が広がってきていることは事実である。

その一方で、この3つの他にも「その他」の脳機能の障害と記されていることに留意すべきである。すなわち、この「政令」には「言語の障害、協調運動の障害、その他厚生労働省令で定める障害」とされており、言語障害などのコミュニケーション障害や、筆者が複数の国際・国内共同研究を進めている発達性協調運動障害(DCD)が「脳機能の障害」である「発達障害」であることは、保育・教育現場はもちろん、医療・療育現場でもあまり認知されていない。更に、「厚生労働省令」で定める障害として「WHO(世界保健機関)のICD-10(疾病及び関連保健問題の国際統計分類)における「心理的発達の障害(F80－F89)」及び「小児<児童>期及び青年期に通常発症する行動及び情緒の障害(F90-F98)」に含まれる障害」とされ、この中には例えば、2013年発表されたDSM-5では神経発達障害 Neurodevelopmental disorders として位置づけられたトゥレット障害を含むチック障害の他、愛着障害、不安障害など含め、中には生物学的・脳科学的には様々な議論も想定されるものも含まれるが、少なくとも法律上の「発達障害」に該当し、特別支援教育や福祉行政的な様々な支援を受けることができる対象であるという認識は低い。加えて、社会的な認知が進みつつある、いわゆる成人での「高次脳機能障害」に相当する「てんかんなどの中枢神経系の疾患、脳外傷や脳血管障害の後遺症が、上記の障害を伴うものである

場合においても、法の対象とするものである。(法第2条関係)」ということもほとんど普及していないように感じている。

「障害」って？

「発達障害」ということになると、しばしば「うちの子は障害者ですか？」「病気ですか？」「治りますか？」というような質問があったりする。また、「障」「害」「碍」という漢字のイメージから「障害」を「障碍」「障がい」「しょうがい」と表記すべきであるという、当事者自身・当事者団体自体からも「単なる『言葉遊び』だ」「表記変更で解決できる問題ではない」と批判・揶揄されてしまうような議論に陥っている側面もある。

一方、英語では Disorder 障害、Defect 欠損・欠陥、Disease / Trouble 疾病・疾患、Impediment 言語障害、Impairment 機能障害、Disability 能力障害、Difficulty 困難、Barrier/Hurdle/Obstacle 障壁・障害物、Handicapped 社会的不利などというように、様々なニュアンスで、これらの語彙がきちんと使い分けられている。WHOの新しい障害の捉え方である国際生活機能分類(International Classification of Functioning, Disability and Health；ICF)モデルでも、「健康状態」として何らかの遺伝的な素因による脳機能の発達のアンバランスがあり、結果「心身機能」として認知の偏りや行動パタンの特徴があったとしても、年齢や性別、ライフスタイルなどの「個人因子」との関係の中で、「環境因子」として周囲の理解や社会的サービス、環境整備などの「促進因子」を強化すれば、「生活機能」としての「活動」や社会「参加」が可能となり「社会的不利」が生じないようにすることが可能ということが示されている。

自閉症スペクトラム障害、学習障害、注意欠陥・多動性障害、発達性協調運動障害などにおける「障害」とは、英語では Disorder であって、あくまでも「バセドウ病」「クローン病」などの Disease(疾患／疾病)ではない。Disorder の語源は、秩序・整然としている状態を表す Order に、否定を表す接頭語である Dis がついたもので、本来は「秩序が乱れている状態、不調、何らかの支障・困難が発生している状態」という意味である。実際に、国際的診断基準である DSM や ICD においても、ほとんどの発達障害の診断基準に「社会的、職業的、または他の重要な領域における機能の臨床的に著しい障害を引き起こしている」、「学業成績あるいは日常生活の活動に明らかな支障をきたしてい

ること」等という記載がある。すなわち、どんなにその特性があったとしても、生活するのに支障がなければ発達障害と診断してはいけないのである。逆に言えば、現在は診断基準を満たしてしまうような「生き辛さ」「生活困難」を、子育て、保育・教育、医療・療育、福祉などによるリエゾン支援で、いわゆる「発達障害」とよばれる特性のある方を理解し支援することで、日常生活などにおける困難や支障がなくなり、その優れた特性を活かして社会参加が可能になれば、かつて「発達障害」と診断された方も、もはや「障害」ではなく「支援の必要な強い脳の個性」となるのである。これらの事を山梨県の本田秀夫先生は「非障害自閉症スペクトラム」と呼ばれている。筆者は「社会参加を目指してADHDやASDの「D」を失くしていく支援を行い、むしろADH、ASを持っているんだ、と誇れるようにしていきましょう」「LDは Learning Disorder(学習障害)、Learning Difficulties、Learning Disabilities(学習困難)のLDではなく、Learning Differences(学び方の違う子)のLDです」とお話している。

発達障害の身体性について
～当事者研究から見えてきた発達障害の身体のこと～

筆者はこれまで発達障害当事者・保護者と設立したNPO法人活動なども含め、多くの発達障害の支援を行なってきたが、その中で、支援者の「思い」と、当事者・保護者の実際の「困り感」のニーズにはしばしば大きなギャップがあるということに気づかされた。すなわち、現在、支援者は主に各発達障害のそれぞれの特性、例えば、ASDに対するソーシャルスキル・トレーニングなど療育プログラム、LDに対する特別支援教育や合理的配慮、ADHDに対する社会心理学的アプローチや薬物療法などを行なっている。もちろん、これらは非常に重要かつ必要な支援であるが、一方で、発達障害当事者・保護者の様々な生活場面での一番の「困り感」はいわゆる感覚過敏・鈍麻などと呼ばれる「感覚」の問題や、身体の使い方、すなわち「協調 Coordination」など、「身体機能の調整障害」からくる「生活障害」なのである。様々な当事者研究からも「自閉は身体障害」(ニキ・リンコ)、「当事者にとっての問題の大半は、対人関係以前の、知覚・運動のレベルにある」(綾屋、熊谷)など同様の観点が報告されている。しかし、これら日常生活の中での感覚や協調の問題による困り感は、保護者・支援者はもとより、当事者自身も気づいていないことも多く、結果、不安・疲労・焦燥などのストレスから不適切な

対応やネガティブな養育スタイルに繋がり、高い虐待のリスクとなったり、自分の指導力のせい、子どものやる気の問題、怠慢、練習不足などと誤解され、不適切な対応が続けられることで、子どものセルフエスティームの低下を引き起こし、問題を悪化させることが報告されている。

発達性協調運動障害（DCD）はいわゆる「不器用」と呼ばれる状態で、様々な感覚入力を「まとめあげ」、運動制御として出力する統合脳機能のひとつである「協調」の発達の問題である。「協調」はバランスや姿勢制御、手と目の協応を必要とする運動・スポーツに限らず、会話、食事、衣類の着脱、描画・書字、楽器操作、道具の使用、姿勢保持など子ども達の様々な日常・学校生活に必要な重要な「脳機能」である。また、50～70％と高い頻度で成人になっても残存し、書字や細かい手作業、料理、メーキャップ・髭剃りなど日常生活や職業上の大きな困難となり、うつ病・不安障害や肥満・糖尿病・高血圧など生活習慣病、心筋梗塞や脳卒中など心血管障害につながることも問題となっている。DCDの頻度は約6～10％と非常に高いが、我が国では保育・教育現場や職場はもとより、医療・療育においても「不器用」が「脳機能」である「協調」の「発達障害」であるという理解や認知は非常に低い。そこで、筆者は国際発達性協調運動障害研究学会（http://psych.brookes.ac.uk/isrdcd/）日本代表committeeとして複数の国際・国内共同研究を推進しているところである。

また、ASDでは、感覚の過敏や鈍麻と表現される独特の感覚の問題も多い。これまでASDでの感覚の問題は単なる併存状態と理解されていたが、最新の当事者研究からは自閉症の本質は、実は身体感覚や視聴覚等の情報統合の困難であり、空腹感や疲れの感覚、目の前の人の顔や表情の認知も含めて、「大量の身体内外の情報を絞り込み、意味や行動にまとめあげるまでがゆっくりな状態で、しかも一度できた意味や行動のまとめあげパタンも容易にほどけやすい」という「情報のまとめあげ困難説」が提唱されている。「社会性」も「協調」も自己を基準に他者や周囲の環境を認識するというプロセスが必要で、両者に共通の「身体化による認知 Embodied Cognition」という神経基盤が存在する可能性が示唆されている。DSM-5では、これら感覚の過敏・鈍麻、感覚刺激への強い関心などがASDの診断基準に再び盛り込まれた。このことは、すなわち、今後ASDの診断にはこれら感覚の問題をきちんと評価する必要があるということである。

また、発達障害に（を）伴う睡眠障害も重要な課題である。睡眠・覚醒リズムとはすなわち、様々な時計遺伝子から様々な神経伝達物質、ホルモンなどの生体リズムである。臨床的には後に発達障害と診断されたケースで乳幼児期からの睡眠の問題はよく経験されることである。最近、発達障害は生体リズム障害であるというアプローチの研究も多く、更に、近年、胎児期からの生体リズムを観察することも可能となっている。今後、これらの研究からエビデンスに基づく、妊娠中から胎児への、また乳幼児期からの生体リズムへの介入ということが期待される。

脳の多様性

発達障害の頻度を考えた時、例えば、ADHDの頻度は約3～5％、ASDは約1％、LDは約6％、発達性協調運動障害（DCD）は約10％と、非常に高い。ADHDやLDの約半数にDCDを伴うこと、更にDSM-IV-TR(2000)までは認められていなかった、ASDとADHDの併存、ASDとDCDの併存もDSM-5(2013)で認められるなど、一口に発達障害といってもひとりひとりは全て異なるのである。このように、人口の約10％以上と頻度の非常に高い、いわゆる発達障害とされている状態は人類に必要な「遺伝子プール」、「脳の多様性 Neurodiversity」と捉えるべきであるとう考え方がある。「脳の多様性（Neurodiversity）」という言葉は1990年代にアスペルガー障害の保護者であるジュディー・シンガーが考案した言葉と言われている（彼女はアスペルガー障害を「アスピー」と呼んでいる）。「生物多様性」「文化的多様性」のように、「脳」にも多様性が存在し、1つとして同じ脳はないのだという考え方である。喩え話によく使われるABO血液型では、A型、B型の両方の転移酵素を持たないO型は劣性遺伝であり、インディオではO型がほとんどであるが、日本人ではA型が多く、AB型は10％である。だからといって血液型で正常・健常である、障害者であるなどという人はいない。（日本人で約10％しかいないAB型である筆者は血液型占いなどで肩身の狭い思いをしてきたのは事実だが…）

トーマス・アームストロングによれば、19世紀フランスの統計学者ケトレーが身長、体重などの変数に関するデータを集め、「平均的な人」（平均人）の平均値を割り出したという。この「平均値」という概念から、特定の個人の集まりという領域から飛び出て、純粋に数学的な値に位置することになったという。身長、体重、BMIなども連続的な数字（スペクトラム）であり、例えば、ある年の日本人男性の身長の平均値と

いうものも算出できる。医学的な高身長・低身長の定義は標準偏差の2倍を超えるものと人為的に定義されただけである。知能指数もまた連続するスペクトラムで、IQ70以下を知的障害とするというのもただ便宜的に定義されただけである。更に、発達検査で、その時の調子により、ひとつ、2つ出来た、出来なかっただけの違いで、知的障害と判定されてしまったり、逆に知能は正常と判定され、支援の対象から外されてしまうことは臨床現場でよく耳にすることである。本来、DSM-IV-TRの「精神遅滞」の診断基準にも「現在の適応機能、すなわち、その文化圏でその年齢に対して期待される基準に適合する有能性の欠如または不全が以下のうち、2つ以上の領域で存在すること」として、「意志伝達、自己管理、家庭生活、社会的・対人的技能、地域社会資源の利用、自律性、発揮される学習能力、仕事、余暇、健康、安全」が挙げられている。その診断には、IQの数字だけでなく、様々な生活や社会参加の領域での「生き辛さ」を評価する必要があるのである。全ての人のIQを測定することはないが、もしIQが69であったとしても、これらの領域で支援を受けながらもいきいきと生活し、社会参加ができていれば「精神遅滞」と診断する必要もないのではと考えられる。DSM-5では「精神遅滞」は「知的発達障害」に変更となったが、その診断にはIQテストは必須ではなくなり、また社会適応や支援の必要度で重症度を判定することとされた。

「やわらかな遺伝子」の著者であるマット・リドレーはヒトゲノム・プロジェクトには大きな誤解があり、「これぞヒトゲノム」というものは実は存在しないという。このように、脳にも多様性があり、むしろ「正常な」「平均的な」脳というものはあるのかという疑問さえ生ずる。また、平均（統計）をとることで、個々の重要な特徴を見失うことにもなるのである。

発達障害研究でよく使われる言葉に、科学論文で使用される「正常コントロール」の代わりに用いられる「定型発達 Typical development:TD、Neurotypical:NT」がある。元々、定型発達とは、英国の自閉症コミュニティから出てきた語彙であるとされている。日本でも発達障害当事者の方々が、多数派を指す時に「定型さん」という呼び方を好んで使われる事も増えている。また、定型発達症候群 Neurotypical Syndrome や 定型発達スペクトラム障害 Neurotypical Spectrum Disorder（NSD）という概念を提唱するグループもある。彼らの「定型発達症候群」の診断基準 DSN（The Diagnostic and Statistical

Manual of 'Normal' Disorders）では、例えば、「社会的相互関係における非依存性の質的障害」の例として「苦悩時に、極端に、あるいは異常なまでに慰めを求める」、「言葉あるいは言葉によらないコミュニケーションと想像遊びにおける質的障害」の例として「露骨に過度なコミュニケーションの全ての方法の使用、コミュニケーションのための喃語、顔の表情、ジェスチャー、模倣、話し言葉など。過度に空想的で無意義な活動、大人役、ファンタジーキャラクター、動物などのごっこ遊びなど。コンピューターや他の論理的に遂行する遊びへの興味の欠如」と皮肉たっぷり？に定義されている。

2006年国連で採択された「障害者の権利に関する条約」にも、「障害」とは「発展する概念」であり、「障害者と障害者に対する態度及び環境による障壁との間の相互作用」であって、「障害者が他の者と平等に社会に完全かつ効果的に参加することを妨げるものによって生ずる」ものとしている。このように「障害」とはむしろ、周囲の無知・無理解や Stigma（汚辱・烙印）という双方の間の Barrier/Hurdle/Obstacle（障壁・障害物）であることを全ての人が認識し、これらを取り除いていく必要がある。

実はこの世界は発達障害が創ってきた

世界の偉人と呼ばれる人々の中にも、発達障害ではと言われている人が多く存在していることはよく知られている。レオナルド・ダ・ヴィンチ、ヴォルフガング・アマデウス・モーツァルト、トーマス・エジソン、アルバート・アインシュタイン、坂本龍馬など、枚挙に暇がなく、いずれも素晴らしい芸術や発明や革新的技術、歴史的改革を生み出した天才達であり、もはや、実は世界は発達障害が創ってきたと言っても過言ではない。いかに伝記になるような人物でも、様々な周囲の無理解や生き辛さがあったことは間違いはないだろう。一方で「弱み」も含めた「強い個性」としての周囲の理解や受容、素晴らしい才能や創造力に対する支援があったからこそ様々な偉業が達成できたと考えられる例も多く存在する。加えて、このように伝記になるような天才でなくても、数多くの起業家や技術者、芸術家、政治家達が、我が国のみならず、歴史・文化・社会に貢献してきたはずである。

我が国での「個性」とは大前提として、平均した能力があり、更にそれに加えられた優れた能力を求められる。また我が国での「ギフテッド」とは、英才児、優秀児、天才児などと訳され、IQ>130のいわゆる天

才を指すことが多いが、海外では知能を通常の発達検査で算出される従来のIQだけでなく、「多重知能・多重知性 Multiple Intelligences：MI」という概念で捉えている。例えば、ハーバード大学のガードナーがいうMI理論では言語的知能、論理数学的知能、音楽的知能、身体運動的知能、空間的知能、対人的知能、内省的知能の７つの知能に、博物的知能、霊的知能、実存的知能の３つを追加した10の知能を想定している。実際にこれらMI理論を取り入れた「プロジェクト・スペクトラム」を実践・研究する学校もある。他にも、高い能力を持つ一方、発達障害などを抱える Twice exceptional（2E）と呼ばれる人への大学教育を含めたギフテッド・タレンテッド教育や支援が行われており、今後我が国でもこのような視点での子どもの理解や教育の推進が望まれる。

発達障害の遺伝子研究への無用な危惧

これまで述べてきたことと関連して、無用な危惧であることを強く望んでいるのが、発達障害の遺伝子研究についてである。もちろん、さまざまな疾病の機序に関する分子生物学的な理解を進めることは、将来の客観的な早期の診断や創薬を含めた治療方法の開発のためには重要であることは間違いない。

ただ、発達障害の場合は果たしてどうなのだろう？という疑問が生ずる。近年の様々な医学、脳科学研究から「発達障害」とは、遺伝的素因と環境との相互作用による、高次脳機能の発達のアンバランス・偏り（発達不均等：Developmental Imbalance）と理解されている。逆に言えば、ある特定の遺伝子の異常や多型のみでは説明できず、生後の生育環境によるものだけでもなく、更に言えば、ある脳の領域の機能のみでも説明できず、これら複雑な相互作用の連続的変化の、ある時点での側面を見ているだけではないかと考えられている。

実際に、ごく一部のものを除けば、単一の責任遺伝子は見つかっておらず、更に、例えば、注意欠陥・多動性障害（ADHD）の遺伝子多型研究でも、ほとんどの研究からその危険度（オッズ比）は平均1.3程度と決定的なものはない。自閉症スペクトラム障害に関連するとされるいくつかの遺伝子もエピジェネティクスの影響を受ける場合。

しかし、現在のような発達障害の理解の中では、遺伝子研究が進み、何らかの関連遺伝子が発見された場合、遺伝子操作や出生前診断によりこれらの「個性的な脳」がこの世界から取り除かれる可能性はないのだろうか？近年メディアなどでも取り上げられるいわゆる「新型出生前診断」ではダウン症もその対象となっており議論となっている。実際、十分な遺伝カウンセリングもないまま、ややもすれば商業ベースで診断結果のみが依頼者に伝わり中絶に繋がることもあるという。また、Googleの共同創業者セルゲイ・ブリン氏の妻であるアン・ウォジスキ氏が共同創設者を務める 23andMe の遺伝子検査キット「Personal Genome Service（PGS）」に対し、アメリカ食品医薬品局（FDA）が販売停止命令を下したことが昨年話題となった。PGSは利用者が唾液を採取してキットを返送すると検査完了後にメールで通知があり、専用ウェブサイトにログインして結果を確認、難病や薬物応答性、体質など254項目を200米ドルで検査できるというものである。もし、ASDの多数の遺伝子多型や関連遺伝子が発見され、誰でも簡便に人知れず網羅的に検査できるようになったその時、人はどのような行動をとるのだろうか？実際に、前述の多重知能理論を提唱するガードナーは進むディスレキシアの遺伝子研究について、「遺伝子の時代の到来とともに危険は増大する…（中略）…人間にとって大切な能力、例えば、空間能力、あるいは図形認識能力のどれが危険にさらされるのか、考えておいたほうがいいだろう」とすでに警告を発している。

逆転の発想

ADHDの著名な研究者で、自らもADHDの当事者である、トム・ハートマンは、ADHDはファーマー（農耕民族）が多数派を占める社会に生きるハンター（狩猟民族）である「ハンター・ファーマー説」を唱えている。ADHDの特性はある時代・文化・社会では非常に好ましい特徴であるはずだが、ファーマーが多数派を占めるようになった複雑な現代社会では生き辛さを感じる事が多いとされる。しかし、チェックリストでマイナスポイントがつくような「多動」「衝動性」も異なる視点から見れば「活動的」「創造力」「自発性」「実行力」などと捉えられ、「不注意」に関しても好きなこと得意なことに関しては「過集中」と呼ばれるほどずば抜けた精神力を発揮する。米国では子どもがアスペルガー障害とわかると「まぁ、一体どんな個性を発揮してくれるのでしょう？ 楽しみですね！」と周囲から祝福されるという。

大正末期から昭和初期にかけて活躍した童謡詩人である金子みすゞの有名な「わたしと小鳥とすずと」の中に「みんなちがって、みんないい」というフレーズがある。これは福井県「ふくいっ子　みんなちがってみんないい応援プロジェクト」のスローガンにもなっており、その他にも各地で広まっている。今後、我が国でも、発達障害を正しく理解することで、発達に凸凹のある彼らの「凹」「弱み」を認めつつ、「逆転の発想」を含めた「凸」「強み」として活かす支援が進み、我が国でも「多様性の受容」や、優しく、強い「共生社会」が実現されることを強く願っている。

おわりに

このように述べてきた、これまでの自分の発達障害の医療・支援への姿勢、発達障害に関する哲学めいたもの、そして研究の方向性などは、独断や偏った意見ではないか、異端なのではないかと不安になることもあったが、本論文を執筆するにあたり、今回、多く引用させていただいた以外にも沢山の文献や書籍に改めて触れ、世界的に観ても同じような考え方、捉え方をもつ方が増えてきていることを確認し、あながち間違いではなかったことを再認識することができた。もちろん、このような考え方や姿勢には多くのご批判があるのはもとより覚悟の上であるが、いずれにせよ本論文が、発達障害に対する正しい理解と支援のあり方への議論の契機になれば幸いである。

最後に、自分自身がADHDだったのではと言われている偉大な発明家であるトーマス・エジソンの言葉を紹介して筆を置くことにする。我々の社会はエジソンが描いていた「未来」と呼べるほどまだ成熟していない。

「未来の医者は薬をださず、患者に身体の手入れや正しい食生活、病気の原因や予防に関心を持たせる」

〈引用文献〉
1）中井昭夫：厚生労働省平成24年度障害者総合福祉推進事業 指定課題21「医療や福祉分野の発達障害支援者の人材育成体制の調査」：非特定営利活動法人 AOZORA福井『発達障害児・者のニーズやライフステージに応じたトランジション・リエゾン支援のための医療・福祉分野等の人材育成に関する調査」報告書 http://ci.nii.ac.jp/ncid/BB12927130
2）ニキリンコ、藤家寛子：『自閉っ子、こういう風にできてます！』花風社（2004）
3）綾屋紗月、熊谷晋一郎：『発達障害当事者研究―ゆっくりていねいにつながりたい（シリーズ ケアをひらく）』医学書院（2008）
4）本田秀夫：『自閉症スペクトラム -10人に1人が抱える「生きづらさ」の正体』ソフトバンククリエイティブ（2013）
5）中井昭夫：「アセスメントツールの活用の仕方：発達性協調運動障害（Developmental Coordination Disorder: DCD）辻井正次（監修）、明翫光宜（編）：『発達障害児者支援とアセスメントのガイドライン』金子書房（2014）
6）トーマス・アームストロング、中尾ゆかり（訳）：『脳の個性を才能にかえる―子どもの発達障害との向き合い方』NHK出版（2013）
7）中井昭夫：巻頭言『療育とは』再考　環境の中で身体が脳を創り、運動がこころを創る　脳と発達 43:432.（2011）
8）中井昭夫：「発達障害の子どもの不器用さのアセスメント・診断と治療の実際」特集「不器用さのある発達障害の子どもたちへの支援」『アスペ☆ハート』33:26-33,2013
9）中井昭夫：「発達障害者雇用は戦略である ～まずは発達障害への正しい理解と合理的配慮から～」特集　ものづくりを行うための人材育成　教育・医療・福祉との連携、協働に向けて」日本設備管理学会誌 25:2-8.（2013）
10）中井昭夫「発達障害は身体障害？～協調運動からの発達障害へのアプローチ―」小児の精神と神経（印刷中）
11）トム・ハートマン、片山奈緒美（訳）：『ADD/ADHDという才能』ヴォイス（2003）

〈筆者プロフィール〉
中井 昭夫（なかい あきお）
福井大学 子どものこころの発達研究センター／医学部附属病院子どものこころ診療部／連合小児発達学研究科・特命准教授。医学博士、小児科専門医、臨床発達心理士、日本小児精神神経学会認定医、日本臨床薬理学会特別指導医。昭和61年 福井医科大学卒業（一期生）。平成3年 同大学院博士課程修了。福井医科大学救急部助手、小児科助手、McGill大学モントリオール神経研究所ブレインイメージングセンター留学、福井県こども療育センター主任医長等を経て、平成23年より現職。平成26年4月より兵庫県立リハビリテーション中央病院「子どもの睡眠と発達医療センター」・副センター長、兼 診療部・神経小児科部長、小児科部長、ならびに、福井大学医学部附属病院子どものこころ診療部・客員教授に就任予定。
日本発達神経科学学会理事、日本赤ちゃん学会理事・学会誌「ベビーサイエンス」編集委員長、日本子ども学会理事、日本小児神経学会評議員・学会誌「脳と発達」編集委員、国際発達性協調運動障害研究学会日本代表committee、アジア・オセアニア小児神経学会終身会員。また発達障害当事者らとNPO法人 AOZORA福井を設立しその理事、JDDネット福井顧問を務める。主な著書に「子どものPTSD：診断と治療」（診断と治療社、印刷中）、『発達障害児者支援とアセスメントに関するガイドライン』（金子書房、2014）、『ADHD: Cognitive Symptoms, Genetics and Treatment Outcomes』（Nova Science Publishers, Inc.、2012）、『小児科学レクチャー：プライマリケアで使える子どもの発達と心の問題への対応 Q&A』（総合医学社、2012）、『0歳からやっておきたい教育―06歳の能力を伸ばす知育トレーニング』（日本経済新聞出版社、2011）、『日本における子供の認知・発達に影響を与える要因の解明：すくすくコホート実践記録集』（独立行政法人 科学技術振興機構 社会技術研究開発センター、2011）、『臨床医とコメディカルのための最新クリニカルPET』（先端医療技術研究所、2010）等。