Research in Developmental Disabilities 35 (2014) 1599–1608

Contents lists available at ScienceDirect

Research in Developmental Disabilities

Comparison between a Mandarin Chinese version of the Childhood Autism Spectrum Test and the Clancy Autism Behaviour Scale in mainland China

Xiang Sun [a,b,c,*], Carrie Allison [b], Bonnie Auyeung [b,d], Fiona E. Matthews [e], Zhixiang Zhang [f], Simon Baron-Cohen [b], Carol Brayne [a]

[a] Cambridge Institute of Public Health, University of Cambridge School of Clinical Medicine, Forvie Site, Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK
[b] Autism Research Centre, Department of Psychiatry, University of Cambridge, Douglas House, 18b Trumpington Road, CB2 2AH, UK
[c] The Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, NT, Hong Kong
[d] Department of Psychology, University of Edinburgh, EH8 9JZ, UK
[e] MRC Biostatistics Unit, Cambridge Institute of Public Health, Forvie Site, Robinson Way, CB2 0SR, UK
[f] Paediatrics Department, Peking University First Hospital, No. 8, Xishiku Avenue, Xicheng District, Beijing 100034, China

ARTICLE INFO

Article history:
Received 10 October 2013
Received in revised form 9 February 2014
Accepted 14 February 2014
Available online 25 April 2014

Keywords:
Autism
Screening
Diagnosis
Epidemiology
CAST
CABS
China

ABSTRACT

A Mandarin Chinese version of the Childhood Autism Spectrum Test (CAST) and Clancy Autism Behaviour Scale (CABS) were applied to 150 children aged 4–11 years old from clinical settings and mainstream schools in Beijing. All the children were further assessed using the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R). The validity of two instruments on screening of ASC was examined and compared using receiver operating characteristic (ROC) curve analysis. The validity of CAST (sensitivity: 89%, specificity: 80%, PPV: 70%) was better than the CABS (sensitivity: 58%, specificity: 84%, PPV: 65%). The area under the curve (AUC) of the CAST (AUC = 0.90) was significantly higher than the CABS (AUC = 0.79, $p = 0.0002$). The Mandarin CAST demonstrated a better validity in distinguishing children with ASC from children without ASC. It is an acceptable candidate as a screening instrument for ASC in large epidemiological study in Chinese population.

1. Introduction

Autism Spectrum Conditions (ASC) are neurodevelopmental disorders characterised by impairments in social interaction and communication, alongside unusually repetitive and stereotyped behaviours, and unusually narrow interests and activities (American Psychiatric Association, 2000). In the West, recent epidemiological studies of ASC have adopted a two-phase method for case identification. The first phase is population-based screening using appropriate screening instruments.

* Corresponding author at: Cambridge Institute of Public Health, University of Cambridge School of Clinical Medicine, Forvie Site, Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK. Tel.: +44 01223 763833; fax: +44 01223 330300.
E-mail address: xs227@medschl.cam.ac.uk (X. Sun).

http://dx.doi.org/10.1016/j.ridd.2014.02.005
0891-4222/

The second phase comprises diagnostic assessments in a smaller sample of children considered to be at risk of having ASC according to screening results (Fombonne, 2009). Studies using this method have estimated that ASC occurs in approximately 1% of the general population (Baird et al., 2006; Baron-Cohen et al., 2009).

Many explanations for the apparent rise in prevalence have been proposed by researchers (Blaxill, 2004; Fombonne, 2009). One contribution could be the adoption of the spectrum definition (Rutter, 2005). The changes in definition led to the revision of diagnostic criteria for ASC which shifted the boundary to include people on the borderline of the spectrum (King & Bearman, 2009). Autism was first described by Leo Kanner in 1943 (Kanner & Eisenberg, 1957) based on the case histories and observations of 11 children who showed a similar pattern of behaviour including social remoteness, stereotypy and echolalia (Croen, Grether, Hoogstrate, & Selvin, 2002). At that time, the term autism was used to describe early infantile autism or infantile autism (Blaxill, 2004). In 1944, Hans Asperger independently described a syndrome now known as Asperger syndrome (Asperger, 1991; Williams, 2003). With the accumulation of research and clinical experience, more behavioural symptoms have been described and categorised as autistic traits (Croen et al., 2002). The term "autism spectrum" was proposed by Wing and Gould in 1979 in order to capture a wider presentation of autistic features (Wing & Gould, 1979). In 1993, *the International Classification of Disease, 10th revision* (ICD-10), suggested that the following categories should be grouped under the autism spectrum: childhood autism, atypical autism, pervasive developmental disorder-not otherwise specified (PDD-NOS) and Asperger's Syndrome (World Health Organisation, 1993). In 2013, the DSM-V revised the diagnostic criteria of ASC by combining the impairments in social interaction and communication into a single subgroup. The three domains of core impairments becomes two: (1) social/communication deficits; (2) fixated interests and repetitive behaviours. A single diagnosis of ASC replaces separate diagnostic subtypes in the *Diagnostic and Statistical Manual of Mental Disorders, 5th edition* (DSM-V), so the diagnosis such as Asperger Syndrome and PDD-NOS are no longer used (American Psychiatric Association, 2012).

Another contribution could be the development in screening and diagnostic instruments for the identification of ASC (Williams, Higgins, & Brayne, 2006). The changes in diagnostic criteria have led to the development of screening and diagnostic instruments. Within the autism spectrum, children with AS may have different autistic behaviours from children with childhood autism since the former do not have delays in language or cognitive development (Baron-Cohen, Wheelwright, Robinson, & Woodbury-Smith, 2005). The impairments due to ASC in children with AS may not be as obvious as those seen in children with childhood autism (Tantam & Girgis, 2009). The borderline diagnostic criteria require instruments to be sensitive enough to capture more subtle and milder autistic traits. The purpose of capturing subtle traits is to make sure instruments can be used for case detection across the whole spectrum. As a developmental condition, autistic features at different developmental stages would not be expected to be identical. Thus, screening instruments in the West have been designed to be age specific. The *Checklist for Autism in Toddlers* (CHAT) targets children as young as 18 months old (Baron-Cohen et al., 2000), and the *Autism Spectrum Quotient* (AQ) has child (Auyeung, Baron-Cohen, Wheelwright, & Allison, 2008), adolescent (Baron-Cohen, Hoekstra, Knickmeyer, & Wheelwright, 2006) and adult versions (Ketelaars et al., 2008).

In mainland China, epidemiological studies have mainly focused on childhood autism (Wu et al., 2010; Zhang et al., 2011). The prevalence of ASC in mainland China has not been well described (Sun & Allison, 2009; Sun, Allison, Auyeung et al., 2013). Earlier studies in mainland China have adopted varied research methodologies for case definition in terms of screening and diagnosis (Sun, Allison, Matthews, et al., 2013). However, the results from previous studies cannot be compared with Western studies directly for many reasons. One reason is that most screening instruments only target the most severe subtype, childhood autism, but not the ASC. The most frequently used screening instrument in Chinese studies is the *Clancy Autism Behavioural Scale* (CABS), the use of which was first reported in 1969 in the West (Clancy, Dugdale, & Rendle-Short, 1969; Sun, Allison, Auyeung, et al., 2013). There has been almost no research using CABS among the Western populations since the 1970s. However, the CABS has been widely used in epidemiological studies of ASC in mainland China. Another reason is the lack of standard diagnostic instruments to assess and further confirm the screening results (Sun, Allison, Auyeung, et al., 2013). In developed countries, the combination of the *Autism Diagnostic Observation Schedule* (ADOS) and the *Autism Diagnostic Interview-Revised* (ADI-R) has been frequently used in epidemiological studies for case confirmation (Lord, Rutter, DiLavore, & Risi, 2001; Rutter, LeCouteur, & Lord, 2003). By adopting these two instruments, the validity of the screening instruments can be tested. In China, these diagnostic instruments have not been well adopted yet. Thus, with the limited usage of CABS in Western populations, as well as the limited usage of ADOS and ADI-R in Chinese populations, it is difficult to compare prevalence estimates from previous studies in mainland China with more recent results from developed countries directly. In order to investigate the current situation of ASC in China, an important question would be whether we need to adopt more recent developed instruments for prevalence studies instead of using those which are already there.

The *Childhood Autism Spectrum Test* (CAST) was developed and validated among a general population from the United Kingdom and demonstrated good validity and reliability (Williams, Allison, et al., 2006; Williams et al., 2005). Sensitivity was 100% and specificity was 97% in primary school age children (Williams et al., 2005). The test-retest reliability of the CAST was good with a kappa statistic of 0.7 (Williams, Allison, et al., 2006). The CAST was used as a screening instrument in an ASC prevalence study in the UK in 2009 (Baron-Cohen et al., 2009). Evidence has demonstrated that the CAST is a relatively robust screening instrument in epidemiological studies on ASC in general population. However, the performance of the CAST in different cultures has not been thoroughly investigated. This study adopted Mandarin versions of both the CAST and the CABS for use in a Chinese sample to compare the utility of these two instruments as screening tools for ASC in Chinese population.

2. Method

2.1. Participants

The participants in this study were drawn from two samples. The first sample comprised children in clinical settings in mainland China who already had an ASC diagnosis made by Chinese clinicians before this study ($n = 50$). The cases were recruited from the Beijing China Disabled Persons' Federation (BCDPF) and a rehabilitation centre for autism in Qingdao. The second sample was an assessment sample including children from two ordinary primary schools in Beijing. The second sample was recruited for the purposes of detecting new autistic cases in the ordinary school population ($n = 103$). Only children who completed both Mandarin CAST and the CABS, as well as standardised assessments, were included in this study. Thus, 3 children in the case sample were excluded as they did not complete the CABS. In total, the data from 150 children were available for analysis (47 in the first sample, 103 in the second sample).

2.2. Procedure

Ethical approval for this research was sought and obtained from the Peking University First Hospital (PUFH) Ethics Committee and the University of Cambridge Psychology Ethics Committee. Written informed consent was obtained from all the participants. Those participants with an existing diagnosis of ASC were invited for an assessment. The measurements included the Mandarin CAST, the CABS, the ADOS and the ADI-R. When they came in for assessment, their parents were asked to complete the Mandarin CAST and the CABS while their child was being assessed. After completion, both questionnaires were collected, along with the assessment results. In the ordinary school sample, the CAST was distributed through schools to parents to complete at home. About 2–4 months after the completion of the CAST, the participants were invited to a diagnostic assessment using the ADOS and ADI-R, at which point they were also asked to complete the CABS during assessment. After assessment, if the child scored above the cut-off point for autism or ASC on both the ADOS and the ADI-R, he or she was given a research diagnosis of ASC. Children who scored above the cut-off point for ASC only in the ADOS or the ADI-R were referred to the clinical child psychiatrists in PUFH using a clinical judgement based on the international criteria of the ICD-10 and the DSM-IV. A consensus diagnosis was made by the clinical child psychiatrists and the first author together. The final diagnosis was the result of the consensus diagnosis.

2.3. Measures

2.3.1. The Mandarin CAST

The UK CAST consists of 37 parent-completed items, of which 31 are scorable. The scores range from 0 to 31. An ASC-positive response receives a score of 1 and an ASC-negative response receives a score of 0. To avoid response bias, some items are reverse scored (Scott, Baron-Cohen, Bolton, & Brayne, 2002). It is designed specifically for a primary school age population between 4 and 11 years old (Scott, Baron-Cohen, Bolton, & Brayne, 2002). It is developed according to a variety of behavioural descriptions of the core features of the autism spectrum as described in the ICD-10 (World Health Organisation, 1993), and the DSM-IV (American Psychiatric Association, 1994). The recommended cut-off point for the CAST is 15 and above for ASC screening (Scott, Baron-Cohen, Bolton, & Brayne, 2002). The CAST was translated into Mandarin Chinese and then back-translated into English with culture-relevant expressions which are described elsewhere (Sun et al., 2012). The final version was approved by the authors of the original CAST. The scoring of the Mandarin CAST is the same as the UK CAST.

2.3.2. Clancy Autism Behaviour Scale

Since the first published study using the *Clancy Autism Behaviour Scale* (CABS) in the US (Clancy et al., 1969), no additional published articles have been made available on the validity and reliability of the CABS in either PubMed or the Web of Knowledge database. Thus, very limited literature is available for examining the effectiveness of the CABS in a Western population. The Mandarin version of the CABS is designed to be completed by parents. It contains 14 items with each item rating three frequency levels, including "Never" (score of 0), "Occasionally" (score of 1) and "Often" (score of 2). Scores on the CABS are usually divided into three groups according to cut-off points of 14 and 21 (Wang, Wang, & Wang, 2003; Zhang, 2006). If a child scores equal to or higher than 14 and has fewer than 3 items scored as "Never" and more than 6 items as "Often", the child is considered to have a potential case of autism. Using this cut-off, the agreement between the CABS and the ICD-10 is reported to be 95–96% (Wang, Wang, & Wang, 2003; Wang, Wang, & Shen, 2003). The agreement between the CABS and the DSM-IV was reported to be 87% in a sample of 28 children with autism and 34 children with typical development (Sun, Allison, Auyeung, et al., 2013). Another study examined the performance of the CABS compared with clinical judgement and reported a sensitivity of 88% and specificity of 82% (Zhang, 2006).

2.3.3. Diagnostic instruments

Diagnostic assessments have been conducted using standardised instruments: ADOS (Li, Zhong, Cai, Chen, & Zhou, 2005) and the ADI-R (Le-Couteur et al., 1989; Lord, Rutter, & Le Couteur, 1994; Rutter et al., 2003). The combined use of the ADOS and the ADI-R has been widely adopted in both research and clinical settings (Papanikolaou et al., 2009), although its

limitations are acknowledged (e.g., it is better at detecting classic autism than Asperger Syndrome, especially in adulthood). The Chinese versions of both instruments were provided by the publisher (World Psychological Service, WPS).

The ADOS is a diagnostic assessment involving interactions and observations (Lord et al., 2001). It has 4 comparable models for administration which are designed to be applicable to individuals with different chronological ages and different levels of expressive language (Berument et al., 2005). Each ADOS assessment takes approximately 45 min. In this study, module 3 was usually used for the primary school children. For children who already had an ASC diagnosis, module 2 was generally chosen. For children who did not have language, module 1 was chosen. The ADI-R is a face-to-face parent-report diagnostic instrument (Lord et al., 1994). It uses semi-structured interviews with parents or caregivers. Each Chinese ADI-R assessment takes about 2–3 h to complete.

2.3.4. Supplementary tests

Raven's Progressive Matrices (RPM) are used as an IQ test for primary school children. The RPM was developed in 1938 and is a commonly used test in clinical neuropsychology to assess general intellectual abilities (Raven, 1938). The Chinese version of the RPM was validated in 1989 and can be applied to individuals from 5 to 75 years old (Li, 1989). The RPM contains 60 single choice questions, listed in order of difficulty. In each test item, the subject is asked to identify the missing element that completes a pattern which requires cognitive capacity to encode and analyse information. The cut-off point for low IQ is 80, median IQ is 81–109, high IQ is 110–129 and extraordinary IQ is 130 or higher (Li, 1989). After the ADOS, each child in this study was given the RPM to complete on his/her own. Each child was given, at most, 1 hour to complete the test.

Those participants with an existing diagnosis of ASC were given the RPM first. If the children were not able to complete the Chinese RPM, another test called the *Adaptive Scale of Infant and Children* (ASIC) was given to examine their psychological age and degree of social adaptation (Zhang et al., 1995). The ASIC was developed in Japan. After validation in China, it was used as one of the survey instruments in the National Survey of Mental Retardation in children aged 0–14 years old in mainland China (Zhang et al., 1995). It contains six domains: Self-help, Locomotion, Daily Activity, Communication, Socialisation and Self-Management.

2.4. Statistical analysis

Statistical methods were used to examine the differences in the characteristics of children and parents between clinical sample and school sample. Unpaired *t*-tests and one-way ANOVA were used to compare means, and Chi-square test was used to examine differences in proportions. Whenever the numbers were small, a Fisher's exact test was used.

Each missing item on the CAST and the CABS was given a value of 0 (ASC-negative value) to generate an observed score for further analysis. The score distributions of the Mandarin CAST and the CABS were described in terms of the mean, median and inter-quartile ranges.

The validities of the CAST and the CABS were first examined by calculating the following. First, the agreement between the screening instrument and standardised diagnostic assessment was examined using Cohen's kappa statistics (Cohen, 1968). Sensitivity was calculated using the percentage of children with ASC who test positively on the instrument. Specificity was calculated using the percentage of children without ASC who test negatively on the instrument. A positive predictive value (PPV) was calculated based on the percentage of all children who test positively who really have ASC. A negative predictive value (NPV) was determined by the percentage of all children who test negatively who really do not have ASC (Webb, 2005).

Second, group differences in the scores between children with ASC and children without ASC were investigated using independent sample *t*-tests when the distribution was robustly normal. The Mann–Whitney test was used when the distribution was skewed. The normality of the distribution was examined by using the Skewness–Kurtosis test. The null hypothesis of the Skewness–Kurtosis test is that the distribution is normal. Thus, a "not significant" test result suggests that the distribution is normal ($p > 0.05$).

Third, the discriminant power of the two instruments in differentiating children with ASC from those without was investigated using a receiver operating characteristic (ROC) area-under-curve (AUC) analysis (Charman et al., 2007; Corsello et al., 2007). All the analysis was conducted in STATA 10.0.

3. Results

3.1. Participant characteristics

The characteristics of two samples were shown in Table 1. After statistical comparison, no significant differences were observed between the children from clinical sample and children from mainstream schools. In total, 150 children took part in this study, the consensus diagnosis of ASC were give to 46 (out of 47) children who had an existing ASC diagnosis before this study. Another 6 children in ordinary school were given a consensus diagnosis of ASC after the assessment. The mean age of participants was 7.7 years old (SD = 1.6). The mean age of autistic cases was 8.4 years old (SD = 1.2), while the mean age of school students was 6.5 years old (SD = 1.6). There were 98 (66%) boys and 51 (34%) girls in this sample. The sex ratio was 1.9:1 (male:female). Among the 52 children with ASC, the sex ratio was 6.3:1 (male:female).

Table 1
Characteristics of the sample.

Characteristics	Category	Mainstream group		Clinical group	
Age	Mean	8.4 (SD: 1.2)		6.3 (SD: 1.6)	
		Number	(%)	Number	(%)
Sex	Boys	58	(56.3)	41	(87.2)
	Girls	45	(43.7)	6	(12.8)
	Missing	0	(0.0)	0	(0.0)
Siblings	Only child	78	(75.7)	37	(78.7)
	Having sibling	18	(17.5)	7	(14.9)
	Missing	7	(6.8)	3	(6.4)
Father's occupation	Worker or farmer	23	(22.3)	7	(14.9)
	Clerk	25	(24.3)	8	(17.0)
	Technical staff	21	(20.4)	24	(51.1)
	Manager	3	(2.9)	0	(0.0)
	Own-business	19	(18.5)	5	(10.7)
	Missing	12	(11.7)	3	(6.4)
Mother's occupation	Worker or farmer	29	(28.2)	10	(21.3)
	Clerk	20	(19.4)	8	(17.0)
	Technical staff	22	(21.4)	21	(44.7)
	Manager	1	(1.0)	0	(0.0)
	Own-business	22	(21.4)	4	(8.5)
	Missing	9	(8.7)	4	(8.5)
Father's education	Junior high school	19	(18.5)	2	(4.3)
	High school	25	(24.3)	12	(25.5)
	College	45	(43.7)	18	(38.3)
	Master or higher	5	(4.9)	12	(25.5)
	Missing	9	(8.7)	3	(6.4)
Mother's education	Junior high school	18	(17.5)	1	(2.1)
	High school	35	(34.0)	13	(27.7)
	College	38	(36.9)	21	(44.7)
	Master or higher	5	(4.9)	7	(14.9)
	Missing	7	(6.8)	5	(10.7)

3.2. Overall distribution of the Mandarin CAST and CABS

A number of 125 Mandarin CAST questionnaires (83.3%) were completed. Twenty questionnaires had one missing item, three had two missing items and two questionnaires had either three or four items missing. Using the cut-off point of 15, 66 children (44%) scored higher or equal to 15 on the Mandarin CAST. The mean score on the CAST was 14.7 (SD = 5.7) and the median was 14 (IQR: 12, 18; range: 2–19). All 150 CABS questionnaires (100%) were completed. Using the cut-off point of 14, 46 children (31%) scored higher or equal to the cut-off point on the CABS. The mean score on the CABS was 10.7 (SD = 5.4) and median score was 10 (IQR: 7, 15; range: 0–25).

3.3. Test accuracy

Using the cut-off point of 15, the overall agreement between the Mandarin CAST and the consensus diagnosis was 83% (95%CI: 76%, 88%). The kappa statistic demonstrated good agreement ($kappa = 0.64$, $p < 0.001$). The sensitivity of the Mandarin CAST was 89% (95%CI: 77%, 96%). The specificity was 80% (95%CI: 70%, 87%). The PPV was 70% (95%CI: 57%, 80%). The NPV was 93% (95%CI: 85%, 97%).

Using the cut-off point of 14, the overall agreement between the CABS and the consensus diagnosis was 75% (95%CI: 67%, 81%). The kappa statistic demonstrated moderate agreement ($kappa = 0.43$, $p < 0.001$). The sensitivity of the CABS was 58% (95%CI: 43%, 71%). The specificity was 84% (95%CI: 75%, 90%). The PPV was 65% (95%CI: 50%, 79%). The NPV was 79% (95%CI: 70%, 86%).

3.4. Difference in score distributions between ASC and non-ASC

The Skewness–Kurtosis test showed that the distribution of the CAST score was normal ($p = 0.42$) and the distribution of the CABS score was also normal ($p = 0.06$). As the score distribution of two screening instruments was normal, an independent *t*-test was performed to compare mean scores in children with ASC and children without. The mean CAST score

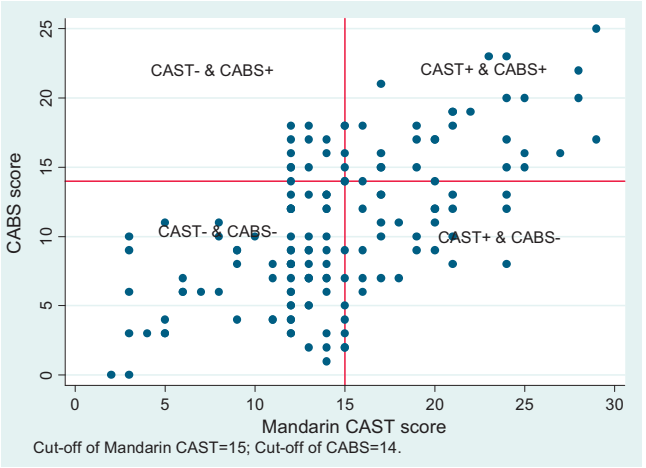

Fig. 1. Scores on the CAST and CABS scores for the same individual. (Note: + indicates score positive; — indicates score negative).

in children with ASC was significantly higher than that of children without ASC ($p < 0.001$). The same association was also found in the CABS scores in children with and without ASC ($p < 0.001$). The correlation between the CAST and CABS scores was moderate (*Spearman's rho* = 0.57, $p < 0.001$). Fig. 1 shows the CAST and CABS scores for the same individual. The mean CAST and CABS scores for children with and without ASC are shown in Table 2.

3.5. Correlations between the scores on the two instruments and the scores on the

3.5.1. ADOS and ADI-R algorithms

Module 1 was administered to 18 participants (12%), module 2–21 (14%) and module 3–111 participants (74%). The mean ADOS score was 7.54 (SD = 10.0). The median ADOS score for children with ASC was 20 (IQR: 15, 25; range: 3–34), while the median for children without ASC was 0 (IQR: 0, 1; range: 0–13). The median ADI-R score for children with ASC was 54.5 (IQR: 42.5, 65; range: 5–77), while the median for children without ASC was 3 (IQR: 1, 7; range: 0–59).

The Skewness–Kurtosis test showed that the distributions of scores for both the ADOS and the ADI-R were highly skewed. Thus, the Mann–Whitney test was used which showed that the ADOS score for children with ASC was significantly higher than that of children without ASC ($p = 0.0061$). The Mann–Whitney test showed that the ADI-R score for children with ASC was significantly higher than that of children without ASC ($p = 0.0084$).

The correlation between the Mandarin CAST scores and the ADOS and ADI-R algorithm scores was examined. The CAST score was significantly associated with the ADOS (*Spearman's rho* = 0.64) and ADI-R (*Spearman's rho* = 0.59) scores on the algorithm. The CABS score was also significantly associated the ADOS (*Spearman's rho* = 0.44) and ADI-R (*Spearman's rho* = 0.49) scores.

3.5.2. Comparison of the CAST and the CABS

The performance of the CAST and CABS were compared through a receiver-operating characteristic (ROC) area-under-curve (AUC) analysis. Fig. 2 shows the CAST had a higher AUC (0.90) than the CABS (0.79). The Mandarin CAST performed significantly better than the CABS in distinguishing children with ASC from those without ($p = 0.0002$). The performance of the two instruments was examined in a subsample of children with normal IQ (IQ $\geq$ 80). The CAST had an AUC at 0.88 which is significantly higher than that of the CABS (0.79, $p = 0.008$). These findings indicated that the Mandarin CAST performed significantly better than the CABS in distinguishing children with ASC from those without. The performance of the two instruments was also examined in separate subsamples of boys and girls. For boys, the CAST had a significantly higher AUC (0.92) than the CABS (AUC = 0.80, $p = 0.0025$). For girls, the CAST had a significantly higher AUC (0.92) than the CABS (AUC = 0.67, $p = 0.0106$). The performance of the two screening instruments overall and in different subgroups, including a normal IQ group (IQ $\geq$ 80), boys and girls, are shown in Table 3.

Table 2
Scores on the two screening instruments by diagnostic groups.

	ASC ($n = 52$)	Non-ASC ($n = 98$)
	Mean (SD)	Mean (SD)
CAST	11.9 (15.8)	19.9 (4.7)
CABS	14.3 (5.1)	8.7 (20.7)

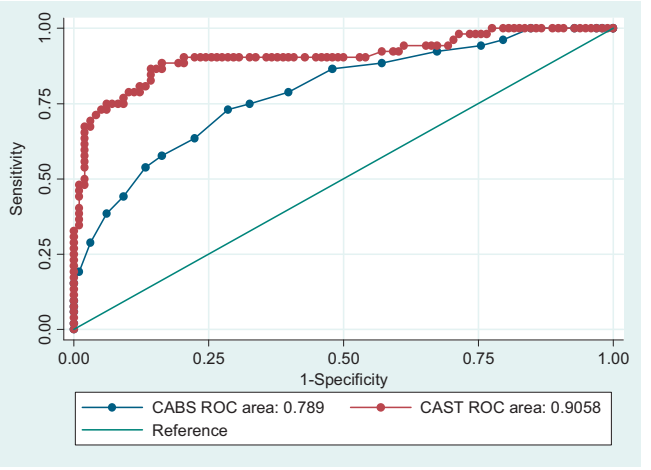

Fig. 2. Receiver operating characteristics (ROC) curves for the CAST and the CABS.

Table 3
Performance of Mandarin CAST and CABS.

	CAST cut-off ≥ 15	CABS cut-off ≥ 14
Whole sample (n = 150)		
AUC (95%CI)	0.90 (0.84–0.96)	0.79 (0.71–0.87)
Sensitivity (95%CI)	0.89 (0.77–0.96)	0.58 (0.43–0.71)
Specificity (95%CI)	0.80 (0.70–0.87)	0.84 (0.75–0.90)
PPV (95%CI)	0.70 (0.57–0.80)	0.65 (0.50–0.79)
NPV (95%CI)	0.93 (0.85–0.97)	0.79 (0.70–0.86)
Normal IQ (n = 127)		
AUC (95%CI)	0.88 (0.80–0.96)	0.79 (0.69–0.88)
Sensitivity (95%CI)	0.88 (0.71–0.97)	0.56 (0.38–0.74)
Specificity (95%CI)	0.79 (0.70–0.87)	0.84 (0.75–0.90)
PPV (95%CI)	0.58 (0.43–0.72)	0.53 (0.35–0.70)
NPV (95%CI)	0.95 (0.87–0.99)	0.85 (0.77–0.92)
Boys (n = 99)		
AUC (95%CI)	0.92 (0.85–0.98)	0.79 (0.71–0.89)
Sensitivity (95%CI)	0.89 (0.76–0.96)	0.60 (0.44–0.74)
Specificity (95%CI)	0.80 (0.67–0.89)	0.83 (0.71–0.92)
PPV (95%CI)	0.78 (0.65–0.89)	0.75 (0.58–0.88)
NPV (95%CI)	0.90 (0.77–0.97)	0.71 (0.59–0.82)
Girls (n = 51)		
AUC (95%CI)	0.92 (0.82–1.00)	0.67 (0.42–0.91)
Sensitivity (95%CI)	0.86 (0.42–1.00)	0.43 (0.10–0.82)
Specificity (95%CI)	0.80 (0.65–0.90)	0.84 (0.70–0.93)
PPV (95%CI)	0.40 (0.16–0.68)	0.30 (0.07–0.65)
NPV (95%CI)	0.97 (0.86–1.00)	0.90 (0.77–0.97)

4. Discussion

4.1. Main findings on test utility

This study was the first to apply standardised diagnostic instruments to examine and compare the utility of the up-to-date Western screening instrument with the currently used screening instrument in China. Both the Mandarin CAST and the CABS can distinguish children with ASC from children without these conditions. The Mandarin CAST was found to have better validity than the CABS and performed significantly better than the CABS according to ROC-AUC analysis. The Mandarin CAST score was more closely associated with the ADOS and ADI-R algorithm scores. Overall, the Mandarin CAST demonstrated higher validity in distinguishing children with ASC from children without ASC.

4.2. Limitations

The limitations of this study include the fact that the children with an existing diagnosis of ASC were younger than the children from primary schools. However, the children were within the designated age range for the CAST (4–11 years old). The CABS has been considered to apply to all ages in previous autism research in China (Chen, Chen, & Hu, 2007). The difference in ages is unlikely to influence the performance of two instruments, since they were applied to the same sample. A relatively modest sample size, especially the subsample of girls, was another limitation that led to relatively wide confidence intervals in the test accuracy. Further research should use larger sample sizes.

One limitation is that the ADOS and ADI-R examiner was aware of the previous diagnosis of the case sample before assessment. Thus, ASC diagnosis based on ADI-R and ADOS was not completely independent from the assessment by CAST and CABS. However, the examiner was blind to the CAST status of the ordinary school sample until completion of the assessment. Thus, it is unlikely this could have influenced the performance of the two screening instruments.

The test validity of the two screening instruments might not be representative of the population nationwide since 30% of the sample comprised children with a previous diagnosis whose parents could be more familiar with or have more knowledge of the autism spectrum. They might have been much clearer about what the items on the questionnaire were looking for than parents in the ordinary schools, leading to better responses to items on the Mandarin CAST and the CABS. Thus, there should be some cautions about applying these figures to the whole population.

4.3. Validity of the Mandarin CAST and the CABS

The results of the Mandarin CAST showed that this measure has better sensitivity than the CABS. The sensitivity of the Mandarin CAST was the same in the subsamples of both boys and girls, while the CABS performed relatively better in boys than in girls. The low sensitivity of the CABS is likely to generate more false negatives. As a result, the CABS could potentially miss more children with ASC by incorrectly indicating that some children are not at risk of having ASC. This could delay referrals to further diagnosis and appropriate intervention if the condition is not accurately identified early on (South et al., 2002). The specificity of the Mandarin CAST and the CABS were both good, with the latter showing better results than the former. The PPV of the Mandarin CAST was better than the CABS. However, both of them were quite low. This result might be due to the fact that the PPV value depends on the prevalence value of the disease. In terms of prevalence, the prevalence of ASC in girls is generally much lower than that of boys (Auyeung et al., 2009; Williams et al., 2008), which could partly explain the much lower PPV within samples of girls.

4.4. Possible explanations for the better performance of the Mandarin CAST

The ROC-AUC results demonstrated that, as a screening instrument for ASC, the Mandarin CAST performed significantly better than the CABS. One explanation might be that the CABS was developed in 1969 for the purpose of identifying children with infantile autism whose diagnosis was much narrower than the diagnosis of ASC (Clancy et al., 1969). The CAST was developed in 2002 with the aim of identifying children along the whole autism spectrum in the school population. Thus, it is possible that children with less obvious autistic features that could be picked up by the Mandarin CAST were missed by the CABS (see Fig. 1).

Another possible explanation could be that the Mandarin version of the CABS was introduced to China during 1980s and since then it has been used without any updating of the items to reflect ongoing diagnostic changes. In addition, the test accuracy and validity of the CABS was not thoroughly validated or investigated before being put to use. The Mandarin CAST has been piloted and validated twice (Williams et al., 2005), and it has been adopted for large population epidemiological research in the UK (Baron-Cohen et al., 2009). The usage of the CABS is also not age-specific in China. Although it was designed for infantile autism, it has been used for all age groups in the Chinese population in previous studies (Sun & Allison, 2009; Zhang & Ji, 2005). By contrast, the Mandarin CAST was designed to be used among primary school-aged children from the very beginning.

Finally, this study adopted standardised diagnostic assessments using a combination of the ADOS and the ADI-R. This study also used a consensus diagnosis as the reference standard that is more comparable to methodologies implemented in Western studies. The previous Chinese studies on the CABS did not use the ADOS or the ADI-R for diagnosis but often adopted clinical judgement as the gold standard (Li, Chen, Song, Du, & Zheng, 2011). However, the validity of the CAST has been established through previous studies using the ADOS and the ADI-R (Allison et al., 2007; Scott, Baron-Cohen, Bolton, & Brayne, 2002; Williams et al., 2005).

4.5. Conclusions

The Mandarin CAST showed good validity in identifying children with ASC in this sample drawn from both clinical settings and the general population. It is a potential new screening instrument for ASC in primary school-aged children in China. It performed better at distinguishing children with ASC from children without ASC than the CABS, a currently used screening instrument for ASC in mainland China. The Mandarin CAST can be considered as a feasible and reasonable choice for a screening instrument for ASC in large epidemiological studies in the Chinese population.

Acknowledgements

We are grateful to the parents and children who participated in this study. We are grateful to the Waterloo Foundation, the Peking University First Hospital, and the Cambridge Institute of Public Health and Autism Research Centre in the University of Cambridge for their generosity in funding this study. We are grateful to Qingdao Elim Autism Research and Intervention Centre. SBC, CA, CB and BA were funded by the MRC UK, the Wellcome Trust, and by the NIHR CLAHRC for Cambridgeshire and Peterborough NHS Foundation Trust during the period of this work. XS was partly funded by the Cambridge Commonwealth Trust, Clare Hall of the University of Cambridge, and the Great Britain-China Educational Trust during the data collection and then funded by the International Development Fund - Cambridge-CUHK Collaboration on Autism Research in Hong Kong and China during the writing up of this paper.

References

Allison, C., Williams, J., Scott, F., Stott, C., Bolton, P., & Baron-Cohen, S. (2007). The Childhood Asperger Syndrome Test (CAST): Test-retest reliability in a high scoring sample. *Autism, 11*(2), 177–190.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington DC: APA: American Psychiatric Association.

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

American Psychiatric Association. (2012). Proposed revision of DSM-V. In *Autism spectrum disorder* from http://www.dsm5.org/ProposedRevision/Pages/proposedrevision.aspx?rid=94.

Asperger, H. (1991). Autistic psychopathy in childhood (1944)-translated and annotated by U. Frith. In U. Frith (Ed.), *Autism and Asperger Syndrome* (pp. 37–92).

Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The Autism Spectrum Quotient: Children's Version (AQ-Child). *Journal of Autism and Developmental Disorders., 38*(7), 1230–1240.

Auyeung, B., Wheelwright, S., Allison, C., Atkinson, M., Samarawickrema, N., & Baron-Cohen, S. (2009). The children's Empathy Quotient and Systemizing Quotient: Sex differences in typical development and in autism spectrum conditions. *Journal of Autism and Developmental Disorders, 39*(11), 1509–1521.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., & Meldrum, D. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: The Special Needs and Autism Project (SNAP). *Lancet, 368*(9531), 210–215.

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The Autism-Spectrum Quotient (AQ) – Adolescent version. *Journal of Autism and Developmental Disorders, 36*(3), 343–350.

Baron-Cohen, S., Scott, F. J., Allison, C., Williams, J., Bolton, P., & Matthews, F. E. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. *The British Journal of Psychiatry, 194*(6), 500–509.

Baron-Cohen, S., Wheelwright, S., Cox, A., Baird, G., Charman, T., & Swettenham, J. (2000). Early identification of autism by the Checklist for Autism in Toddlers (CHAT). *Journal of Royal Society of Medicine, 93*(10), 521–525.

Baron-Cohen, S., Wheelwright, S., Robinson, J., & Woodbury-Smith, M. (2005). The Adult Asperger Assessment (AAA): A diagnostic method. *Journal of Autism and Developmental Disorders, 35*(6), 807–819.

Berument, S. K., Starr, E., Pickles, A., Tomlins, M., Papanikolauou, K., & Lord, C. (2005). Pre-linguistic Autism Diagnostic Observation Schedule adapted for older individuals with severe to profound mental retardation: A pilot study. *Journal of Autism and Developmental Disorders., 35*(6), 821–829.

Blaxill, M. F. (2004). What's going on? The question of time trends in autism. *Public Health Reports, 119*(6), 536–551.

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

Chen, Y., Chen, Z., & Hu, R. (2007). Clinical application of Clancy Autism Behavior Scale. *Guangdong Medical Journal, 28*(3), 376–377.

Clancy, H., Dugdale, A., & Rendle-Short, J. (1969). The diagnosis of infantile autism. *Developmental Medicine and Child Neurology, 11*(4), 432–442.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213–220.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Jr., & Leventhal, B. L. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, 48*(9), 932–940.

Croen, L. A., Grether, J. K., Hoogstrate, J., & Selvin, S. (2002). The changing prevalence of autism in California. *Journal of Autism and Developmental Disorders, 32*(3), 207–215.

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatrics in Review, 65*(6), 591–598.

Kanner, L., & Eisenberg, L. (1957). Early infantile autism, 1943–1955. *Psychiatric Research Reports American Psychiatric Association,* (7), 55–65.

Ketelaars, C., Horwitz, E., Sytema, S., Bos, J., Wiersma, D., & Minderaa, R. (2008). Brief report: adults with mild autism spectrum disorders (ASD): Scores on the autism spectrum quotient (AQ) and comorbid psychopathology. *Journal of Autism and Developmental Disorders., 38*(1), 176–180.

King, M., & Bearman, P. (2009). Diagnostic change and the increased prevalence of autism. *International Journal of Epidemiology, 38*(5), 1224–1234.

Le-Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., & Holdgrafer, M. (1989). Autism diagnostic interview: A standardized investigator-based instrument. *Journal of Autism and Developmental Disorders., 19*(3), 363–387.

Li, D. (1989). *Chinese Raven Test*. Shanghai: East China Normal University Press.

Li, J., Zhong, J., Cai, L., Chen, Y., & Zhou, M. (2005). Comparison of clinical application of three autism rating scale. *China Journal of Contemporary Paediatrics, 7*(1), 59–62.

Li, N., Chen, G., Song, X., Du, W., & Zheng, X. (2011). Prevalence of autism-caused disability among Chinese children: A national population-based survey. *Epilepsy & Behavior, 22*(4), 786–789.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2001). *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles, CA: Western Psychological Services.

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*(5), 659–685.

Papanikolaou, K., Paliokosta, E., Houliaras, G., Vgenopoulou, S., Giouroukou, E., & Pehlivanidis, A. (2009). Using the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule-Generic for the diagnosis of autism spectrum disorders in a Greek sample with a wide range of intellectual abilities. *Journal of Autism and Developmental Disorders, 39*(3), 414–420.

Raven, J. C. (1938). *Standard progressive matrices: Sets A, B, C, D, and E*. London: Lewis HK.

Rutter, M. (2005). Incidence of autism spectrum disorders: Changes over time and their meaning. *Acta Paediatrica, 94*(1), 2–15.

Rutter, M., LeCouteur, A., & Lord, C. (2003). *Autism diagnostic interview-revised manual*. Los Angeles, CA: Western Psychological Services.

Scott, F. J., Baron-Cohen, S., Bolton, P., & Brayne, C. (2002a). Brief report: Prevalence of autism spectrum conditions in children aged 5–11 years in Cambridgeshire, UK. *Autism, 6*(3), 231–237.

Scott, F. J., Baron-Cohen, S., Bolton, P., & Brayne, C. (2002b). The CAST (Childhood Asperger Syndrome Test): Preliminary development of a UK screen for mainstream primary-school-age children. *Autism, 6*(1), 9–31.

South, M., Williams, B. J., McMahon, W. M., Owley, T., Filipek, P. A., & Shernoff, E. (2002). Utility of the Gilliam Autism Rating Scale in research and clinical populations. *Journal of Autism and Developmental Disorders., 32*(6), 593–599.

Sun, X., & Allison, C. (2009). A review of the prevalence of Autism Spectrum Disorder in Asia. *Research in Autism Spectrum Disorders, 4*(2), 156–167.

Sun, X., Allison, C., Auyeung, B., Matthews, F., Baron-Cohen, S., & Brayne, C. (2013). What is available for case identification in autism research in mainland China? *Research in Autism Spectrum Disorders., 7*(5), 579–590.
Sun, X., Allison, C., Matthews, F., Zhang, Z., Auyeung, B., & Baron-Cohen, S. (2014). *An exploration of the underdiagnosis of autism in mainland China using screening and diagnostic instruments*. Cambridge, (submitted for publication).
Sun, X., Allison, C., Matthews, F. E., Sharp, S. J., Auyeung, B., & Baron-Cohen, S. (2013). Prevalence of autism in mainland China, Hong Kong and Taiwan: A systematic review and meta-analysis. *Molecular Autism, 4*(1), 7.
Tantam, D., & Girgis, S. (2009). Recognition and treatment of Asperger syndrome in the community. *British Medical Bulletin, 89*, 41–62.
Wang, Y., Wang, G., & Wang, Y. (2003). Analysis of childhood autism by using Clancy Autism Behavior Scale and Autism Behavior Checklist. *Journal of Shandong University (Health Science), 41*(2), 213–214.
Wang, Y., Wang, Y., & Shen, Y. (2003). Investigation of intelligence and behavior impairment in autistic children. *Chinese Journal of Child Health Care, 11*(2), 133–134.
Webb, P. (2005). *Essential Epidemiology: An introduction for students and health professionals*. Cambridge: Cambridge University Press.
Williams, J. (2003). *Screening for autism spectrum disorders*. Cambridge: University of Cambridge.
Williams, J., Allison, C., Scott, F., Stott, C., Bolton, P., & Baron-Cohen, S. (2006). The Childhood Asperger Syndrome Test (CAST): Test–retest reliability. *Autism, 10*(4), 415–427.
Williams, J., Scott, F., Stott, C., Allison, C., Bolton, P., & Baron-Cohen, S. (2005). The CAST (Childhood Asperger Syndrome Test): Test accuracy. *Autism, 9*(1), 45–68.
Williams, J. G., Allison, C., Scott, F. J., Bolton, P. F., Baron-Cohen, S., & Matthews, F. E. (2008). The Childhood Autism Spectrum Test (CAST): Sex differences. *Journal of Autism and Developmental Disorders., 38*(9), 1731–1739.
Williams, J. G., Higgins, J. P., & Brayne, C. E. (2006). Systematic review of prevalence studies of autism spectrum disorders. *Archives of Diseases in Childhood, 91*(1), 8–15.
Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders, 9*(1), 11–29.
World Health Organisation. (1993). *International statistical classification of diseases and related health problems 10th edition (ICD-10)*. Geneva: World Health Organization.
Wu, X., Lu, Y., Wang, Y., Zheng, Q., Wang, T., & Lin, J. (2010). Investigation of childhood autism status in Lianyungang city. *Journal of Modern Medicine and Hygiene, 26*(24), 3724–3726.
Zhang, F., Liu, Y., Xie, Y., Wei, Y., Zhang, L., & Yang, J. (2011). An investigation of awareness of childhood autism in Wuxi city. *Jiangsu Medical Journal, 37*(14), 1704–1705.
Zhang, Q. (2006). The evaluation of diagnosis of Childhood Autism by DSM-IV and Clancy Autism Behaviour Scale. *Journal of Medical Theory and Practice, 19*(5), 586–587.
Zhang, X., & Ji, C. Y. (2005). Autism and mental retardation of young children in China. *Biomedical and Environmental Sciences, 18*(5), 334–340.
Zhang, Z., Zuo, Q., Lei, Z., Chen, R., Huang, L., & He, G. (1995). The restandardization of the Adaptive Scale of Infant and Children. *Journal of China Clinical Psychology, 3*(1), 12–15.