Research in Developmental Disabilities 85 (2019) 205–216

Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

Validation of the modified checklist for autism in toddlers, revised with follow-up in Taiwanese toddlers

Jung-Mei Tsai[a], Lu Lu[b], Suh-Fang Jeng[a], Pou-Leng Cheong[c], Susan Shur-Fen Gau[d], Yen-Hsun Huang[e], Yen-Tzu Wu[a,b,*]

[a] School and Graduate Institute of Physical Therapy, College of Medicine, National Taiwan University, Taipei, Taiwan
[b] Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei, Taiwan
[c] Department of Pediatrics, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan
[d] Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan
[e] Department of Child and Adolescent Psychiatry, Taipei City Psychiatric Center, Taipei City Hospital, Taipei, Taiwan

ARTICLE INFO

Number of reviews completed is 2

Keywords:
ASD
M－CHAT-R/F
Screening
Toddlers
Taiwan

ABSTRACT

Background: The Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M－CHAT-R/F) is a two-stage screening scale for determining the risk of autism spectrum disorder (ASD) in toddlers. However, the validity of the M－CHAT-R/F for Asian populations has not yet been established.

Aims: This study investigated the psychometric properties of the M－CHAT-R/F, Taiwan version (M－CHAT-R/F–T), among low- and high-risk Taiwanese toddlers aged 16–30 months. The associations among M－CHAT-R/F－T scores, developmental performance at 24 and 30 months, and ASD diagnosis prediction at 36 months were examined.

Methods and Procedures: A two-stage screening of the M－CHAT-R/F－T was applied to a study sample comprising 25 toddlers with ASD and 71 atypically developing (ATD) and 221 typically developing (TD) toddlers.

Outcomes and Results: The M－CHAT-R/F–T exhibited acceptable internal consistency and test–retest reliability. The M－CHAT-R/F–T scores were significantly correlated with several syndrome scores of the Child Behavior Checklist for Ages 1.5–5 and were significantly higher among toddlers with ASD than among ATD or TD toddlers. Furthermore, M－CHAT-R/F–T scores were negatively correlated with developmental scores in the Mullen Scales of Early Learning at 24 and 30 months. Moreover, the screening exhibited acceptable predictive validity (sensitivity = 0.86; specificity = 0.96) for ASD diagnosis at 36 months.

Conclusions and Implications: The findings indicate that the M－CHAT-R/F–T is a valid and reliable tool for the developmental screening of low- and high-risk Taiwanese toddlers in community and clinical settings.

What this paper adds?

The prevalence of ASD has increased significantly in Asia over the past decades, thus underscoring the need for reliable and valid screening tools for early detection of ASD in Asian populations. The two-stage ASD screening scale, M－CHAT-R/F, has been

* Corresponding author at: School and Graduate Institute of Physical Therapy, College of Medicine, National Taiwan University, Room 317, Floor 3, No.17, Xu-Zhou Road, Taipei 10055, Taiwan.
E-mail address: yenwu@ntu.edu.tw (Y.-T. Wu).

https://doi.org/10.1016/j.ridd.2018.11.011
Received 8 October 2017; Received in revised form 17 November 2018; Accepted 24 November 2018
Available online 19 December 2018
0891-4222/

validated in a large sample of U.S. toddlers. However, the psychometric properties of the M−CHAT-R/F for Asian populations have not been established. The results of the study reveal that the M−CHAT-R/F–T exhibited satisfactory reliability and validity when used with low- and high-risk Taiwanese toddlers. The findings confirmed that the two-stage screening scale, M−CHAT-R/F–T, is a promising instrument for assessing the risk of ASD in Taiwanese toddlers. The findings contribute to the literature by providing a useful cultural adaptation of the M−CHAT-R/F, which is applicable to non-Western populations.

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social interaction and communication and by the presence of repetitive and restricted behaviors or interests (American Psychiatric Association, 2013). In the United States, the most recent estimate of the prevalence of ASD among 8-year-old children was 14.7 per 1000 children (or 1 in 68 children) (Christensen et al., 2016). In Asia, the average prevalence of ASD in the general population was approximately 1.9/10,000 in 1980, and it increased to approximately 14.8/10,000 in 2008 (Sun & Allison, 2010). Studies have reported that the most recent estimation of prevalence of ASD in the ethnic Chinese populations in Taiwan, Hong Kong, and China is 2.8–29.5 per 10,000 (Chiang, Soong, Lin, & Rogers, 2008; Feng et al., 2013; Wong & Hui, 2008). The increasing prevalence of ASD in Asian societies suggests that an increasing number of Asian children require screening for early detection of ASD and for effective early interventions.

Identifying children with ASD during the early developmental period enables early interventions to be provided to enhance their developmental outcomes (Barton, Dumont-Mathieu, & Fein, 2012). Previous studies have revealed that the early signs of ASD manifested during the age range of 6–18 months (Bhat, Landa, & Galloway, 2011; Dumont-Mathieu & Fein, 2005). Developmental regression characterized by a loss of previously-acquired ability in social reciprocity and verbal repertoire, and deviation from normative language, social, and motor developments becomes increasingly apparent between 18 and 36 months (Al Backer, 2015; Bhat, Galloway, & Landa, 2012; Maestro et al., 2002; Osterling, Dawson, & Munson, 2002; Rogers, 2004; Werner & Dawson, 2005). Therefore, the developmental screening of children prior to 36 months of age is critical for detecting ASD and providing necessary optimal early interventions. The American Academy of Pediatrics recommends that ASD-specific standardized screening be conducted at 18- and 24-month well-child care (WCC) visits as preventive check-ups (Johnson & Myers, 2007). Although Taiwanese parents do have their toddlers immunized during WCC visits that each visit includes health examinations and professional consultations, no regular ASD-specific screenings are currently conducted at WCC visits. Hence, a combination of a short checklist of ASD screening and clinical interview during WCC visits could be helpful for ruling out ASD in Taiwanese toddlers.

Several ASD screening scales that were developed in Western countries have been classified as “level-one” screening scales. Level-one screening refers to the widespread screening of a population at a low risk of a disorder; such screening can be administered to all children in primary care settings (Barton et al., 2012; Dumont-Mathieu & Fein, 2005). These screening tests are mainly based on parental reports and include several items for rapid administration. A “screen-positive” patient is defined as a toddler whose risk scores equals or exceed the cutoff points of the scale; the case is subsequently referred for further assessment (e.g., a structured interview) to confirm the possibility of risk. Among the existing level-one screening scales for ASD, the Modified Checklist for Autism in Toddlers (M−CHAT) (Robins, Fein, Barton, & Green, 2001) is among the most widely used instruments worldwide. Furthermore, the M−CHAT with Follow-Up (M−CHAT/F) was subsequently developed to add a second-stage follow-up screening to reduce the number of toddlers who initially screen positive according to the initial M−CHAT checklist (Chlebowski, Robins, Barton, & Fein, 2013; Dumont-Mathieu & Fein, 2005; Kleinman et al., 2008).

Several studies have translated the M−CHAT and M−CHAT/F from English into many languages so that they can be used for ASD screening in Asian populations. For instance, a recent study validated a Thai version of the M−CHAT. It reported a two-stage screening test that combined a short checklist and a follow-up interview, which showed a high sensitivity of 0.91 and a high specificity of 0.99 (Srisinghasongkram, Pruksananonda, & Chonchaiya, 2016) for low- and high-risk samples of Thai children. Furthermore, acceptable validity results have also been found for Japanese (Kamio et al., 2014) and Iranian (Samadi & McConkey, 2015) versions of the M−CHAT in large population-based samples of children. However, Perera, Wijewardena, and Aluthwelage, (2009) used a Sri Lankan version of the M−CHAT to screen 18- to 24-month-old toddlers in a suburban area; the results showed a low sensitivity of 0.25. These findings suggest a need for social and cultural adaptation of the M−CHAT in the application of translated versions of the screening scale in Eastern societies.

Recently, the M−CHAT was revised and renamed M−CHAT, Revised with Follow-Up (M−CHAT-R/F) (Robins et al., 2014) and comprises simplified items that use simpler language than do those in the original versions of the M−CHAT and M−CHAT/F. The new version is a two-stage screening test comprising an initial checklist and a follow-up scoring sheet. A recent validation study on a large sample of U.S. toddlers indicated that the two-stage screening of M−CHAT-R/F exhibited an improved ability (compared with the original version) for detecting ASD, with an estimated sensitivity of 0.91 and a specificity of 0.96, and effectively reduced the age of diagnosis in low-risk toddlers (Robins et al., 2014). However, the M−CHAT-R/F has not yet been validated in countries other than the United States, with the exception of a study in Serbia, which revealed that the M−CHAT-R/F, Serbian version, exhibited acceptable reliability for screening Serbian toddlers (Carakovac et al., 2016).

Therefore, this study validated the use of the M−CHAT-R/F, Taiwan version (M−CHAT-R/F–T), in a population of low- and high-risk Taiwanese toddlers consisting of typically developing (TD) toddlers, atypically developing (ATD) toddlers with developmental disabilities other than ASD, and toddlers with ASD to investigate the psychometric properties of the M−CHAT-R/F–T. Furthermore, the associations between the two-stage M−CHAT-R/F–T scores and developmental performance at the ages of 24 and 30 months as well as the predictive value of the M−CHAT-R/F–T scores for the diagnosis of ASD at 36 months old were examined in this study.

2. Material and methods

2.1. Participants and procedures

In this study, the sample population consisted of low- and high-risk Taiwanese toddlers; the toddlers were recruited from community and clinical settings during the years 2015–2017. Toddlers and their parents were enrolled by distributing flyers at a day-care center for infants and young children, a research laboratory for the developmental surveillance of preterm infants, well child care visits at outpatient clinics, and the child development clinics at National Taiwan University Children's Hospital and Taipei City Hospital. Toddlers and their parents were included in the study if the toddlers were 16–30 months old, one of the parents was the primary caregiver, the parents' ages were ≥20 years at enrollment, and the parent had at least 9 years of education and could read Chinese. Children with severe sensory or motor deficits, including uncorrected vision or hearing loss, were excluded from the study. High-risk toddlers with developmental disorders were clinically referred by the physicians in the child psychiatry, pediatric neurology, or physical medicine and rehabilitation departments at child developmental clinics. The study was approved by the institutional review boards of the participating hospitals. Written informed consent was obtained from the parents after they had received a complete description of the study.

The primary caregivers were mailed the initial M−CHAT-R/F–T checklist at enrollment, and 369 responses from the primary caregivers were returned (Fig. 1). After excluding responses that were collected for children aged > 32 months, the final sample included responses for 317 toddlers (167 boys [52.7%] and 150 girls [47.3%]). Among the included toddlers, 150 (47.3%) were screened at 16–24 months old at enrollment, and 167 (52.7%) were screened at 25–32 months old. Most respondents were mothers (97.2%). The mean age of all respondents was 41 years (standard deviation [SD] = 13.2), and three-quarters of the respondents (72%) had > 12 years of education. Of the respondents, 284 (89.9%) simultaneously completed the Child Behavior Checklist for Ages 1.5–5 (CBCL/1.5–5) (Achenbach & Rescorla, 2000). The initial M−CHAT-R/F checklist was re-sent to 317 caregivers 2 weeks after the initial assessment, and 103 (32.5%) responses were returned. Based on the scores of the initial M−CHAT-R/F checklist, 57 toddlers had scores indicating medium or high risk (i.e., total scores of 3–20). The caregivers of 52 toddlers were successfully contacted by telephone for the second-stage follow-up interview (FUI), and 28 of these caregivers completed a second interview by telephone after a 2-week interval. A total of 95 toddlers who were > 24 months of age at enrollment underwent developmental assessment based on the Mullen Scales of Early Learning (MSEL) (Mullen, 1995) at 24 months old; and 179 toddlers received MSEL assessments at 30 months old.

In this study, all toddlers were divided into three groups based on the developmental diagnosis received when they approached 36 months of age: (1) toddlers with ASD (n = 25), that is, toddlers with formal ASD diagnoses; (2) ATD toddlers (n = 71), that is, toddlers who were diagnosed with any developmental disorders other than ASD (e.g., global developmental delay, intellectual disability, communication disorder, or motor delay); and (3) TD toddlers (n = 221), that is, toddlers whose primary caregivers or pediatric physicians have never been concerned about any developmental problems, and the toddlers had not received any diagnosis of developmental disorder at 36 months of age. These groupings were determined on the basis of criteria listed in the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) (American Psychiatric Association, 2013) by a multidisciplinary team of experts, including medical doctors in child psychiatry, physical medicine and rehabilitation and pediatric neurology. Diagnostic interviews and evaluations were conducted for each child with all clinicians. If necessary, the child further received developmental and behavioral assessments by clinical psychologists, developmental pediatricians, neurologists, physical therapists, occupational therapists, and speech and language pathologists using standardized developmental and behavioral assessment tools. The clinicians had reviewed and discussed all available measures to integrate the results of developmental and behavioral assessments that the final results of group classification reflected the impressions of clinicians. Among 25 toddlers with ASD, seven toddlers exhibited co-occurring developmental diagnoses, including global developmental delay (n = 4), communication disorders (n = 2), and motor delay (n = 1).

2.2. ASD screening tool

The M−CHAT-R/F–T was translated from the original version of the M−CHAT-R/F. The translation was permitted by the author, Dr. Robins. The initial checklist and follow-up scoring sheet were translated into Mandarin Chinese by two bilingual experts with doctoral degrees, who are experienced in assessing child development and who are fluent in Chinese and English. The vocabulary and sentence structure of all items in the initial checklist and follow-up scoring sheet were modified to accommodate the terms used in Taiwan. A pilot study of 10 Taiwanese mothers indicated that the descriptions of items 1, 5, 9 and 16 were confusing. Therefore, we reworded the descriptions to enhance understanding and prevent misinterpretation of the children's behavior. Moreover, we added examples of activities in the follow-up scoring sheet that were more suitable for Taiwanese culture. For example, Taiwanese parents do not say the words "peek-a-boo" when playing this game. We therefore added the examples that involves hiding behind something and suddenly reappearing of face or eyes such as hide-and-seek or making funny faces. The resulting version was subsequently back-translated into English and compared with the original version to allow minor modifications to be made, if required. The suitability of sentences and wordings in the M−CHAT-R/F–T was finally reviewed by a second pilot sample of 25 Taiwanese mothers to determine the final version of M−CHAT-R/F–T (Appendix M−CHAT-R/F–T).

The two-stage screening of the M−CHAT-R/F–T is consistent with the original version of M−CHAT-R/F, which contains an initial checklist and a follow-up scoring sheet (Robins et al., 2014). The caregiver is required to answer 20 yes/no questions during the initial stage, and the total number of risk items is subsequently used to classify a child as having a low- (total score: 0–2), medium-

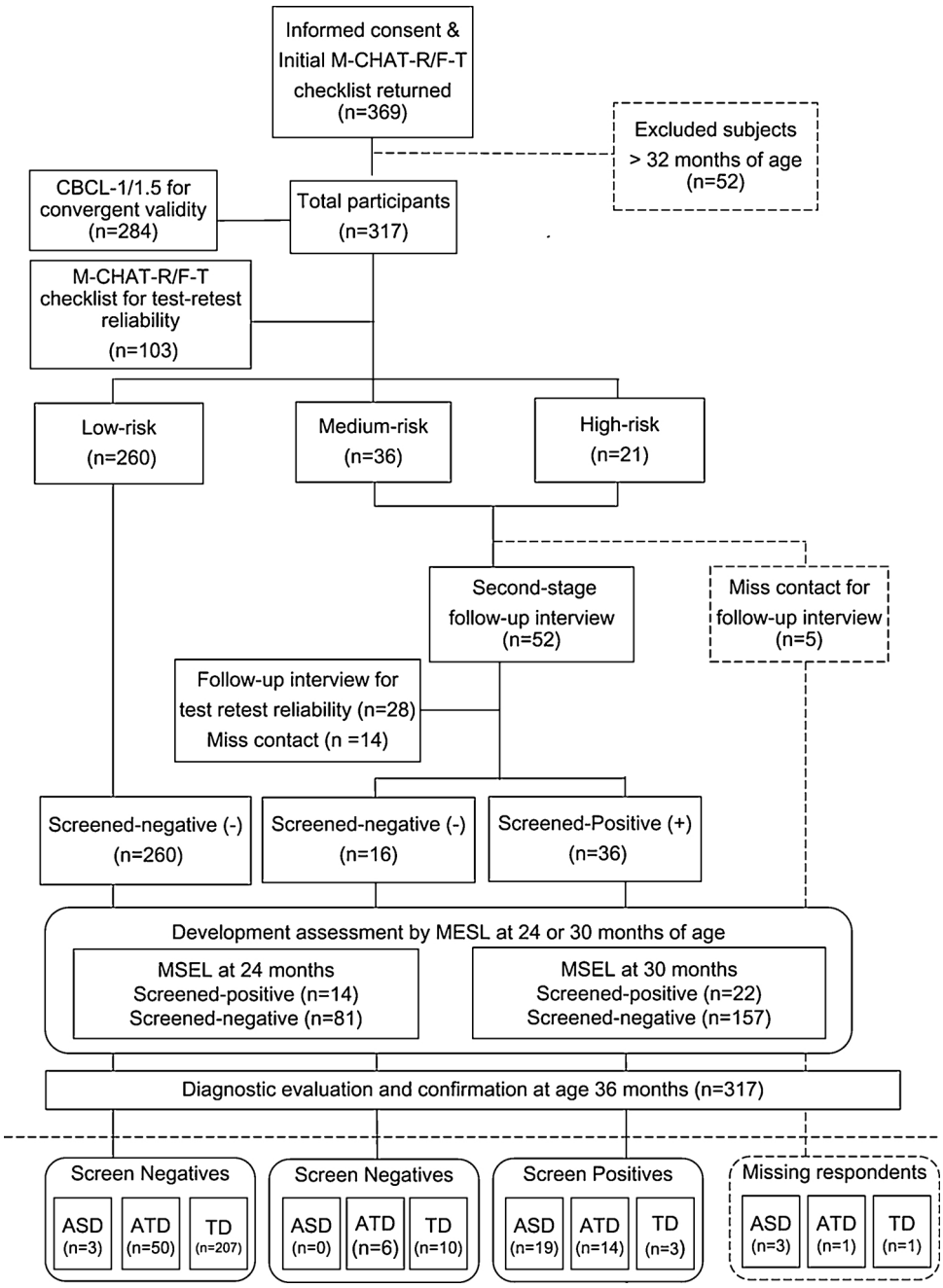

Fig. 1. Flowchart of study design and participant enrollment.

(total score: 3–7), or high-level (total score: 8–20) of risk based on the original classification in the M − CHAT-R/F. For children with a medium- or high-risk (cutoff score: ≧3), the caregivers were subsequently asked structured questions during the second-stage FUI. The caregiver's responses were used for the follow-up scoring sheet. Each item had a structured flowchart format, and a trained research assistant conducted the FUI by following an algorithm for each of the risk behaviors until a pass or fail score was obtained. The interview was considered to yield a positive screening if the child failed on any two or more items on the follow-up scoring sheet.

2.3. Assessment of behavioral, emotional, and social functioning

The CBCL/1.5–5 (Achenbach & Rescorla, 2000) is a report by caregivers regarding the children's behavioral, emotional, and social functioning for children aged 1.5–5 years; it consists of 100 items. Each item is rated on a 0–2 point scale. The sum of the 100 items is counted as the total problems score, with 67 of the items being scored specifically with regard to seven narrowband behavioral

syndromes: emotionally reactive, anxious/depressed, somatic complaints, withdrawn, sleep problems, attention problems, and aggressive behavior syndromes. Furthermore, scores for two broadband behavioral syndromes were derived, with the first four listed syndromes constituting the internalizing syndrome and the last two listed syndromes constituting the externalizing syndrome. Moreover, the pervasive developmental problems (PDD) scale is one of five DSM oriented scales in the CBCL/1.5-5 that are used to assess ASD related behavioral problems. The CBCL/1.5–5 was translated into Chinese, and the translated version was reported to have an acceptable reliability and validity when used to assess Taiwanese preschoolers (Wu et al., 2012). The subscales used in this study for assessing the convergent validity of the M−CHAT-R/F−T were the CBCL/1.5–5 PDD, emotionally reactive syndrome, withdrawn syndrome, internalizing syndrome, and total problems scales. These subscales were used because they were previously reported to be correlated with the total scores of the original version of M−CHAT (Albores-Gallo et al., 2012).

2.4. Assessment of cognitive and motor development

The MSEL is a comprehensive norm-referenced developmental test for children aged 0–68 months (Mullen, 1995). The test contains the following five subscales: the visual reception, expressive language, receptive language, gross motor, and fine motor scales. The scores for each test item range from 0 to 2 or from 0–5. The raw scores of each of the subscales can be converted to standard T scores and developmental-equivalent ages for comparison with the data in a normative sample of 1849 U.S. children between 2 days and 69 months old. The early learning composite score is representative of cognitive function and comprises the combined T scores of four subscales (specifically, the expressive language, receptive language, visual reception, and fine motor subscales). The MSEL was reported to have acceptable reliability, and the MSEL scores exhibit moderate to high correlations with the Bayley mental development index (r = 0.53–0.59) and the Peabody fine motor scale (r = 0.65–0.82) scores (Mullen, 1995). In addition, studies have found that the MSEL has acceptable discriminative validity because children with ASD or children with developmental delay obtained lower MSEL scores than did the TD children (Akshoomoff, 2006; Burns, King, & Spencer, 2013; Hellendoorn et al., 2015). The examination was conducted by experienced psychologist or pediatric physical therapist that the examiners had > 6 months of experience of using the MSEL in clinical practices prior to the study.

3. Statistical analysis

Demographic variables of the toddlers and caregivers were compared among the three groups; the Kruskal–Wallis test was used for analyzing continuous variables, and the chi-squared test was used for analyzing categorical variables.

The internal consistency of the M−CHAT-R/F–T was confirmed by calculating Cronbach's α coefficient. A cutoff α value of ≥0.8 represented excellent internal consistency and an α value of 0.6–0.8 indicated acceptable internal consistency (Cohen, 1988). A subsample of 103 respondents who completed the second assessment of the checklist and 28 caregivers who completed the second assessment of a telephone interview, respectively, were assessed for the test–retest reliability coefficients by calculating the item-by-item Cohen's κ coefficient values and the intrarater correlation coefficient (ICC) for the total scores. The κ values 0.21–0.4, 0.41–0.6, 0.61–0.8, and ≥0.81 represented reasonable, moderate, substantial, and perfect agreement, respectively (Landis & Koch, 1977). The ICC values < 0.4, 0.4–0.75, and > 0.75 indicated poor, acceptable, and high reliability, respectively.

The convergent validity was evaluated by determining the Pearson correlation coefficients between the total and subscales scores of the CBCL/1.5–5 and the total M−CHAT-R/F−T scores by using the CBCL/1.5–5 subscales as comparative measures. The values of the Pearson correlation coefficients 0.8–1, 0.6–0.79, 0.4–0.59, and 0.2–0.39 indicate very strong, strong, moderate, and weak correlations, respectively (Evans, 1996). To determine discriminative validity, the total scores of the two-stage M−CHAT-R/F–T screening test were compared among the ASD, ATD, and TD groups by using the Kruskal–Wallis test. To determine predictive validity, linear regression analysis was used to analyze the associations among the MSEL subscale scores and the total M−CHAT-R/F–T scores. Furthermore, the predictions of an ASD diagnosis at 36 months were examined by estimating predictive values, such as sensitivity, specificity, positive predictive value (PPV), negative predictive value, and likelihood ratios. The receiver operating characteristic (ROC) curve, area under the curve (AUC), and optimal cutoff scores for both the initial checklist and the second-stage FUI were estimated. All statistical analyses were performed using IBM SPSS Statistics version 20.

4. Results

4.1. Sample characteristics

The demographic characteristics of the participants and overall scores of the M−CHAT-R/F–T, CBCL/1–1.5, and MSEL scales are presented in Table 1. Each variable is shown separately in the ASD, ATD, and TD groups. The percentage of toddlers referred from clinical settings was significantly higher in the ASD group than in the ATD or TD groups ($p < 0.001$). Furthermore, the results revealed a higher ratio of boys to girls in the ASD group than in the ATD and TD groups. Moreover, the toddlers in the ASD group were older at enrollment than those in the ATD and TD groups (both $p < 0.001$). However, caregivers' sex, age, and years of education did not differ significantly among the groups (all $p > 0.05$). A comparison of the percentages of toddlers participating in two-stage M−CHAT-R/F–T screening among the groups revealed that the percentage of toddlers who missed the second-stage FUI was higher in the ASD group (16%) than in the other groups ($p < 0.001$). Furthermore, the interval between the initial checklist and second-stage FUI was significantly longer for the toddlers in the ASD group than in the other groups ($p = 0.03$). The three groups were compared for participation in each reliability or validity assessment. The completion of test-retest reliability assessment of the

Table 1
Demographic and evaluation characteristics of participants.

Characteristics	Total Sample (N = 317)	Classification at 36 Months of Age			P value [a]
		TD (N = 221)	ATD (N = 71)	ASD (N = 25)	
Recruitment of toddlers, n (%)					< 0.001
Community settings	191 (60.3)	154 (69.4)	37 (52.1)	0 (0)	
Clinical settings	126 (39.7)	67 (30.6)	34 (47.9)	25 (100)	
Gender, n (%)					< 0.001
Boy	167 (52.7)	100 (45.2)	45 (63.4)	22 (88)	
Girl	150 (47.3)	121 (54.8)	26 (36.6)	3 (12)	
Toddler's age at enrollment (month), M (SD, range)	24.3 (4.4, 16-32)	23.7 (4.1, 16-32)	24.5 (4.4, 16-32)	28.4 (3.9, 18-32)	< 0.001 (ASD > ATD; ASD > TD)
Primary caregivers, n (%)					0.45
Mother	308 (97.2)	218 (98.6)	67 (94.4)	23 (92)	
Father	9 (2.8)	3 (1.4)	4 (5.6)	2 (8)	
Age of primary caregivers (year), M (SD, range)	40.2 (10.8, 24-72)	40.7 (11.3, 24-72)	41.4 (12.8, 28-67)	38.5 (8.6, 24-61)	0.79
Educational years of primary caregivers, n (%)					0.16
9-12 years	84 (26.4)	58 (26.2)	21 (29.6)	5 (20)	
> 12 years or above	228 (72)	161 (72.9)	47 (66.2)	20 (80)	
Missing	5 (1.6)	2 (0.9)	3 (4.2)	0 (0)	
M-CHAT-R/F-T initial checklist, M (SD, range)	1.7 (2.8, 0-14)	0.8 (1.1, 0-7)	2.4 (3.2, 0-13)	7.6 (3.5, 1-14)	< 0.001 (ASD > ATD > TD)
Two-stage screening of M-CHAT-R/F-T, n (%)					< 0.001
Completion of two-stage screening	312 (98.4)	220 (99.5)	71(100)	21 (84)	
Missing at second-stage FUI	5 (1.6)	1 (0.5)	0 (0)	4 (16)	
Two-stage M-CHAT-R/F-T, M (SD, range)	1.1 (1.7, 0-12)	0.6 (0.7, 0-3)	1.6 (1.98, 0 -12)	4.7 (2.9, 1-11)	< 0.001 (ASD > ATD > TD)
Interval between initial checklist and second-stage FUI (day), M (SD, range)	47 (32.3, 0-144)	36.7 (27.9, 0-81)	61.2 (36.9, 0-144)	40.4 (32.3, 1-144)	0.03 (ASD > ATD; ASD > TD)
Test-retest reliability for initial checklist, n (%)					0.80
Completion of second assessment	103 (32.5)	73 (33)	21 (29.6)	9 (36)	
No second assessment	214 (67.5)	148 (67)	50 (70.4)	16 (64)	
Test-retest reliability for second-stage FUI, n (%)					< 0.001
Completion of second assessment	28 (8.8)	8 (3.6)	14 (19.7)	6 (24)	
No second assessment	289 (91.2)	213 (96.4)	57 (80.3)	19 (76)	
CBCL/1.5-5 assessment, n (%)					0.003
Completion of assessment	284 (89.9)	196 (88.7)	69 (98.6)	19 (76)	
No assessment	32 (10.1)	25 (11.3)	1 (1.4)	6 (24)	
CBCL/1.5-5 total scores, M (SD, range)	38.2 (22.8, 0-118)	34.3 (19, 0-111)	42.7 (26.9, 9-118)	62.1 (25.6, 25-101)	< 0.001 (ASD > ATD; ASD > TD)
MESL assessment at 24 months, n (%)					
Completion of assessment	95 (30)	59 (22.2)	29 (40.8)	7 (28)	0.75
No assessment	222 (70)	162 (73.3)	42 (59.2)	18 (72)	
MESL ELC scores at 24 months, M (SD, range)	199.9 (65.9, 0-296)	207.3 (69, 0-296)	194 (61.2, 68-268)	161.1 (46.2, 101-204)	0.01 (ASD > TD)
MESL assessment at 30 months, n (%)					
Completion of assessment	179 (56.5)	126 (57)	36 (50.7)	17 (68)	0.31
No assessment	138 (43.5)	95 (43)	35 (49.3)	8 (32)	
MESL ELC scores at 30 months, M (SD, range)	209.9 (51.3, 54-313)	223.6 (37.5, 135-313)	205.7 (51.3, 93-260)	117.6 (44.1, 54-175)	< 0.001 (ASD > ATD; ASD > TD)

Categorical variables are presented as n (%); Continuous variables are presented as M (SD, range); TD: typically developing; ATD: atypically developing; ASD: Autism Spectrum Disorder; FUI: Follow-up interview; M−CHAT-R/F–T: M−CHAT-R/F Taiwan version; CBCL/1-1.5: Child Behavior Checklist for Ages 1.5–5; MESL: Mullen Early Learning Scales; ELC: early learning composite.

[a] Kruskal-Wallis test for analyzing continuous variables and the Chi-square test for analyzing categorical variables.

second-stage FUI was higher among parents in the ASD group than the parents in the other groups. Furthermore, the completion of CBCL/1–1.5 assessment was higher among toddlers in the ATD group than the toddlers in the other groups (both $p < 0.001$).

4.2. Reliability

The M−CHAT-R/F–T exhibited excellent internal consistency; the value of Cronbach's α was 0.84 ($p < 0.05$). The test–retest

Table 2
Test-retest reliability of initial checklist and second-stage follow-up interview of the M−CHAT-R/F–T.

M-CHAT-R/F-T item		Initial checklist (N = 103) Kappa (κ)	Follow-Up Interview (N = 28) Kappa (κ)
1.	Looks where you point	0.63*	0.91**
2.	Acts like he/she is deaf	0.64*	1**
3.	Plays pretend games	0.56*	1**
4.	Likes climbing on things	0.74*	1**
5.	Unusual finger movements near eyes	0.67*	1**
6.	Points with one finger to request	0.64*	0.71**
7.	Points with one finger to show something	0.66*	1**
8.	Interested in other children	0.40*	1**
9.	Shows something by bringing it or holding it up	0.70*	0.84**
10	Responds when you call his/her name	0.64*	0.78**
11.	Smiles back at you	0.56*	1**
12.	Gets upset by everyday noises	0.37*	0.47**
13	Does your child walk?	0.88*	0.84**
14.	Looks you in the eye when interacting	0.33*	0.63**
15.	Tries to copy what you do	1*	1**
16.	Looks around to see what you are looking at	0.70*	0.86**
17.	Tries to get you to watch him/her	0.47*	0.65**
18.	Understands when you tell him/her to do something	0.70*	0.71**
19	Looks at your face to see how you feel about something	0.48*	0.85**
20.	Likes movement activities	0.58*	1**

M−CHAT-R/F–T: M−CHAT-R/F Taiwan version; *p value < 0.05; **p value < 0.001.

reliability results are shown in Table 2. The initial checklist of the M−CHAT-R/F–T revealed moderate to perfect agreement in terms of test–retest reliability for 18 of 20 items (κ ranging from 0.4 to 0.88, all $p < 0.05$), whereas only fair agreement was achieved for the remaining items (κ = 0.33 and 0.37, both $p < 0.05$). However, all items at the second-stage FUI exhibited moderate to perfect agreement in terms of test–retest reliability (κ ranging from 0.47 to 1, all $p < 0.05$). In addition, the initial checklist's total scores [ICC (95% CI) = 0.94 (0.91–0.96)] and the second-stage FUI [ICC (95% CI) = 0.96 (0.9–0.98)] revealed high ICCs on test–retest reliability.

4.3. Convergent validity

A significantly strong correlation was observed between the total scores of the two-stage M−CHAT-R/F−T and the CBCL/1.5–5 withdrawn syndrome scores (r = 0.63, $p < 0.001$). Furthermore, the total scores of two-stage M−CHAT-R/F−T were moderately correlated with the PDD (r = 0.58, $p < 0.001$), emotionally reactive syndrome (r = 0.37, $p < 0.001$), internalizing syndrome (r = 0.45, $p < 0.001$), and total problems (r = 0.37, $p < 0.001$) scores.

4.4. Discriminant validity

The toddlers with ASD (M = 7.6, SD = 3.5, R = 1–14) showed significantly higher scores on the initial checklist of the M−CHAT-R/F−T than did the ATD (M = 2.4, SD = 3.2, R = 0–13) and TD groups (M = 0.8, SD = 1.1, R = 0–7) (both $p < 0.001$). Similarly, the scores in the ASD group were significantly higher (M = 4.7, SD = 3, R = 1–11) than those in the ATD (M = 1.6, SD = 2, R = 0–12) and TD (M = 0.6, SD = 0.7, R = 0–3) groups after completing the second-stage FUI (both $p < 0.001$). Furthermore, the total scores of the initial checklist and two-stage screening were significantly higher in the ATD group than in the TD group (all $p < 0.001$).

4.5. Predictive validity

Linear regression analysis revealed that the total M−CHAT-R/F–T scores were negatively associated with each of the MESL

Table 3
The association of the total scores of M−CHAT-R/F–T with the MSEL subscales scores derived at 24 and 30 months of age using the linear regression analysis.

Age	N	Gross motor		Visual reception		Fine motor		Receptive language		Expressive language	
		β	p	β	p	β	p	β	p	β	p
24 months	95	−0.40	< 0.001	−0.48	< 0.001	−0.46	< 0.001	−0.64	< 0.001	−0.35	< 0.001
30 months	179	−0.36	< 0.001	−0.53	< 0.001	−0.39	< 0.001	−0.59	< 0.001	−0.43	< 0.001

MSEL: Mullen Scales of Early Learning.

Table 4
Predictive validity properties of initial and two-stage M−CHAT-R/F–T screening.

	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	LR+ (95%CI)	LR- (95%CI)
Initial checklist scoring (n = 317)						
Total scores ≥ 3 (cutoff score = 3)	0.88 (0.69–0.97)	0.89 (0.84–0.92)	0.39 (0.31–0.47)	0.99 (0.97–0.99)	7.34 (5.2–10.3)	0.14 (0.05–0.39)
Total scores ≥ 4 (cutoff score = 4)	0.88 (0.68–0.97)	0.92 (0.88–0.95)	0.48 (0.38–0.58)	0.99 (0.97–1.00)	10.7 (7.11–16.13)	0.13 (0.05–0.38)
Two-stage M-CHAT-R/F-T (n = 312)						
Initial checklist ≥ 3 and FUI ≥ 2	0.86 (0.65–0.97)	0.94 (0.91–0.97)	0.53 (0.41–0.65)	0.99 (0.97–1.00)	14.73 (9.0–24.1)	0.14 (0.05–0.41)
Initial checklist ≥ 4 and FUI ≥ 2	0.86 (0.65–0.97)	0.96 (0.93–0.98)	0.59 (0.46–0.72)	0.99 (0.7–1.00)	19.3 (11.0–33.6)	0.14 (0.05–0.41)

M−CHAT-R/F–T: M−CHAT-R/F Taiwan version; LR+, likelihood ratio of positive screen; LR-, likelihood ratio of negative screen; NPV, negative predictive value; 95%CI: 95% confidence interval.

subscale scores at 24 and 30 months old (all $p < 0.001$) (Table 3), indicating that high total M−CHAT-R/F–T scores were associated with low developmental scores. The results of ROC analysis indicated an optimal cutoff score of 4 and 2 points for the initial checklist and the second-stage FUI, respectively. The specificity of the initial checklist increased from 0.89 to 0.92, but the sensitivity remained 0.88 with the original cutoff scores of ≥3 points and with cutoff scores of ≥4 points (Table 4). In addition, when the optimal cutoff scores (initial checklist ≥4 and FUI ≥ 2) were used for the positive-screen criterion of the two-stage M−CHAT-R/F–T screening, the predictive values revealed a sensitivity of 0.86, a specificity of 0.96, and a PPV of 0.59. The AUC of the initial checklist and two-stage M−CHAT-R/F–T screening based on each cutoff score criterion was 0.93–0.94 (Fig. 2).

5. Discussion

In the present study, the psychometric properties of the Taiwan version of the M−CHAT-R/F were investigated by applying it to a sample of low- and high-risk Taiwanese toddlers aged 16–30 months, including TD toddlers, ATD toddlers, and toddlers with ASD. The results indicate that the psychometric properties of the M−CHAT-R/F–T are satisfactory for screening Taiwanese toddlers in terms of various indices of reliability and validity. The findings confirmed that two-stage M−CHAT-R/F−T screening is a promising tool for screening for ASD in Taiwanese toddlers. In accordance with the literature, which considers the M−CHAT-R/F a useful tool for screening U.S. and Serbian toddlers, this study contributes to non-Western Asian populations. The translation and cultural modification of the M−CHAT-R/F–T provides an evidence-based screening method for other regions of Taiwan or other Asian countries with a similar cultural background.

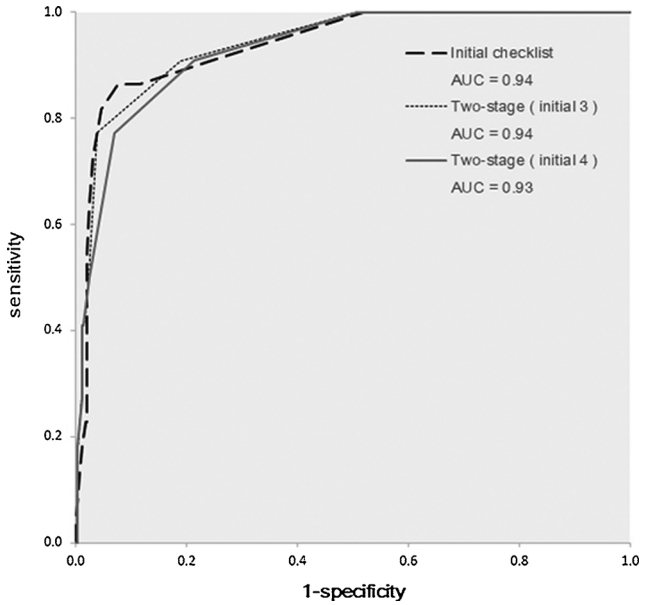

Fig. 2. Receiver operating characteristic (ROC) curve and area under the curve (AUC) for the two-stage M−CHAT-R/F–T screening test.

5.1. Reliability and convergent validity

The high Cronbach's α value indicated sufficient internal consistency for the M−CHAT-R/F–T. This statistic was similar to the internal consistency of 0.91 of the Serbian version of the M−CHAT-R/F (Carakovac et al., 2016). In addition, the high ICC values of the M−CHAT-R/F−T scores indicated that the overall screening results were highly consistent. Moreover, although the κ values of two items in the initial checklist exhibited only fair agreement over a 2-week interval, all items of the second-stage FUI exhibited perfect test–retest reliability. This discrepancy in the test–retest reliability results for the two stages of screening might be because caregivers misunderstood certain individual items or because the phrasing of some items was shortened and unclear. Parents may choose between one of two answers for each item depending on their own interpretation at the time. Moreover, a structured algorithm and specific criterion at the second-stage FUI could elicit more reliable and consistent results from caregivers compared with at the first screening. For example, some parents exhibited confusion or uncertainty for the item "Looks you in the eye when interacting" and the item "Does your child show you things by bringing them to you, not to get help, but just to share." During the second-stage FUI, we noticed that some parents were confused about whether their children were looking them in the eye or at other facial features; some parents misunderstood the meaning of "showing something to people to share (to elicit an identical level of interest or share ownership of the object)" as following a request to give an object to others. The interview during the second-stage screening helped parents clarify their children's behavior. We also noticed that some Taiwanese parents may not have clearly observed their children's eye contact and may have failed to recognize the relevance of eye contact, or the lack thereof, to ASD.

Moderate to high correlations were observed between M−CHAT-R/F−T total scores and the CBCL/1.5–5 PDD, withdrawn syndrome, emotionally reactive syndrome, internalizing syndrome, and total problems scores in our sample of Taiwanese toddlers. These results were similar to those of Albores-Gallo et al. (2012), who determined that the total scores of a Mexican version of the M−CHAT were correlated with the CBCL/1.5–5 PDD and withdrawn syndrome scores (r = 0.65–0.66). Furthermore, Muratori et al. (2011) and Narzisi et al. (2013) have reported that the CBCL/1.5–5 PDD and withdrawn syndrome scores could be used to distinguish between TD toddlers and those later diagnosed with ASD. Because children with ASD have previously been found to exhibit high levels of many behavioral problems, including inattention and hyperactivity (Mayes, Calhoun, Mayes, & Molitoris, 2012), mood disturbances and temper tantrums (Dominick, Davis, Lainhart, Tager-Flusberg, & Folstein, 2007), sleep problems (Mazurek & Sohl, 2016; Souders et al., 2009), and withdrawn behaviors (Albores-Gallo et al., 2012; Hartley, Sikora, & McCoy, 2008; Limberg, Gruber, & Noterdaeme, 2016). Our findings suggest that high scores derived from M−CHAT-R/F–T are associated with high scores of ASD-related problems and behaviors.

5.2. Discriminate and predictive validity

The toddlers in the ASD group in the present study had significantly higher total M−CHAT-R/F–T scores than those in the TD and ATD toddler groups. The results were consistent with those of Albores-Gallo et al. (2012), which revealed that Mexican TD toddlers had significantly lower M−CHAT scores than did Mexican toddlers with ASD. However, notably, no previous studies have examined the discriminative validity of M−CHAT or M−CHAT-R/F among toddlers with TD, ATD, or ASD. Our findings indicate that the M−CHAT-R/F–T scores may identify those who will be diagnosed with ASD among TD toddlers or those with developmental problems other than ASD at 36 months old.

The results of the present study demonstrated that the M−CHAT-R/F–T was generally effective at predicting developmental outcomes for children 24 and 30 months old; thus, high M−CHAT-R/F–T scores are likely predictive of poor developmental performance in Taiwanese toddlers. Numerous studies have indicated that the problematic aspects of development are usually observed in children with ASD; they include the core symptoms of social interaction, communication deficits, and poor cognitive performance (Dawson et al., 2010; Oono, Honey, & McConachie, 2013; Warren et al., 2011). Therefore, these findings suggest the need for standardized screening tools that can be widely used for the earliest possible ASD detection in toddlers and providing optimal and necessary early interventions.

Both the initial checklist and two stages of the M−CHAT-R/F–T screening revealed sufficient sensitivity and specificity according to the present study. The high predictive values of sensitivity, specificity, and AUC of the M−CHAT-R/F–T were consistent with previous reports of the two-stage M−CHAT-R/F in U.S. toddlers (Robins et al., 2014), the M−CHAT checklist only for toddlers in Singapore (Koh et al., 2014), and the two-stage M−CHAT screening for toddlers in Japan (Kamio et al., 2014), Thailand (Srisinghasongkram et al., 2016), and Iran (Samadi & McConkey, 2015). The predictive values indicate that the initial checklist reached the optimal screening accuracy when the threshold was altered to ≥4. Similarly, both the specificity and PPV increased and met the optimal prediction in the two-stage M−CHAT-R/F–T screening when the cutoff score for the initial checklist was ≥4 and FUI ≥ 2. Our findings suggest that referring Taiwanese toddlers to further checks at second-stage FUI if they score ≥4 in the initial checklist, rather than the original cutoff score (≥3), is appropriate. The discrepancy of cutoff points between M−CHAT-R/F–T and the original version might be attributed to the methodological differences of sampling. A high proportion of clinically-referred and high-risk toddlers in our study may potentially elevate the overall risk scores compared to the study by Robins et al. (2014) that the original version was validated in a large and low-risk pediatric sample. In addition, the specificity and PPV exhibited improvements after the second-stage FUI was conducted, indicating that the second-stage FUI was effective in excluding toddlers without ASD. Our findings suggest that the M−CHAT-R/F–T has acceptable validity in terms of its capacity to distinguish between toddlers with and without ASD.

In the present study, the results of the initial checklist and two-stage M−CHAT-R/F–T screening had PPV values of 0.39 and 0.53, respectively. With specific respect to the 36 toddlers who were screened as positive using the M−CHAT-R/F−T in the present study,

19 (52.7%) subsequently received a diagnosis of ASD at 36 months old, whereas 14 (38%) were found to have other developmental disorders. Thus, although the M−CHAT-R/F–T had a high false-positive rate (0.47) for ASD, it demonstrated a high PPV (0.91) for detecting developmental delay and ASD. Studies have revealed inconsistent findings for the predictive values of PPV in two-stage M−CHAT or M−CHAT-R/F. For example, Robins et al. (2014) reported a PPV of 0.47 for an M−CHAT-R/F screening of ASD in a large low-risk sample of U.S. toddlers. By contrast, a higher PPV of 0.96 was reported by Srisinghasongkram et al. (2016) in a high- and low-risk sample of Thai toddlers. This discrepancy might be due to differences in methods, including differences in the study population (e.g., only low-risk samples, low- and high-risk samples, or only clinically referred high-risk samples), resulting in a different prevalence, and differences in screening method (e.g., only parent completion of the M−CHAT versus two-stage screening of the M−CHAT or M−CHAT-R/F). In addition, Srisinghasongkram et al. (2016) demonstrated that the conducting FUI with clinicians with expertise in screening for ASD may reduce false-positive rates and contribute to improving the PPV.

5.3. Limitation

The present study had several limitations. First, the M−CHAT-R/F–T, had high predictive value for the study sample in terms of sensitivity, specificity, and AUC, and approximately 7.8% of the toddlers in the sample were ultimately diagnosed with ASD. However, the generalizability of the results cannot be determined because the rate of 7.8% is considerably higher than the prevalence in the general population. Furthermore, standardized diagnostic instruments were not used to provide sufficiently reliable diagnostic assessments of ASD was a shortcoming of this study that could have influenced the classification of toddlers and the results of the predictive validity indices. Second, our sample included clinically referred high-risk toddlers who had already been assessed at child developmental clinics centers. The sampling bias in which toddlers with ASD were more likely to be included in clinical settings than in community settings may potentially influence the subject characteristics among groups. In addition, the caregivers of high-risk toddlers may have answered differently from caregivers who are unaware of their child's atypical development in the general population. Because the M−CHAT-R/F is designed as a level-one screening tool, further research involving a large, unselected community sample is required to achieve an accurate characterization of the instrument's validity and reliability for the screening of Taiwanese toddlers in general.

Third, the interval between the assessment of initial checklist and FUI in the toddlers with ASD was longer than those in the other groups. The FUI was not conducted for some of the toddlers because of the difficulty involved in contacting parents or caregivers by telephone. The results of the validity analyses may have been influenced by the time interval between the two stages of screening and missing data. Finally, because the toddlers in the study were enrolled at various ages ranging from 16 to 30 months, we determined that the toddlers with ASD were older than those in the other two groups. The age differences in the sample and the possibility that some of the older toddlers may have begun treatment through early interventions during the study period, as well as the effects of any such interventions, may have affected the children's development or their parents' responses during the screening.

6. Conclusion

This study was the first study to investigate the reliability and validity of the translated Taiwan version of the M−CHAT-R/F for screening Taiwanese toddlers. The preliminary findings suggest that the M−CHAT-R/F–T has acceptable reliability and validity for the developmental screening of ASD in low- and high-risk Taiwanese toddlers in community or clinical settings. However, additional research is required to validate the results of this study in a large population-based sample of Taiwanese toddlers.

Financial disclosure

The authors have no financial relationships relevant to this article to disclose.

Potential conflicts of interest

The sponsor for the data collection is the Ministry of Science and Technology in Taiwan. The sponsor had no involvement in (1) study design; (2) analysis, and interpretation of data; (3) the writing of the report; or (4) the decision to submit the paper for publication. The authors do not have any potential conflict of interest, real or perceived.

Acknowledgments

This study was supported by a grant from the Ministry of Science and Technology for the Screening and Follow-Up for Autism Spectrum Disorder in Taiwanese Toddlers (MOST 104-2314-B-002-003). We thank the toddlers and their parents for participating in this study. This manuscript was edited by Wallace Academic Editing.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ridd.2018.11.011.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA school-age forms & profiles : An integrated system of multi-informant assessment /Thomas M. Achenbach & Leslie A. Rescorla.* Burlington: VT ASEBA.

Akshoomoff, N. (2006). Use of the mullen scales of early learning for the assessment of young children with autism Spectrum disorders. *Child Neuropsychology, 12*(4-5), 269–277. https://doi.org/10.1080/09297040500473714.

Al Backer, N. B. (2015). Developmental regression in autism spectrum disorder. *Sudanese Journal of Paediatrics, 15*(1), 21–26.

Albores-Gallo, L., Roldán-Ceballos, O., Villarreal-Valdes, G., Betanzos-Cruz, B. X., Santos-Sánchez, C., Martínez-Jaime, M. M., et al. (2012). M-CHAT mexican version validity and reliability and some cultural considerations. *ISRN Neurology, 2012*, 7. https://doi.org/10.5402/2012/408694.

American Psychiatric Association (2013). *Autism spectrum disorderDiagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing50–59.

Barton, M. L., Dumont-Mathieu, T., & Fein, D. (2012). Screening young children for autism spectrum disorders in primary practice. *Journal of Autism and Developmental Disorders, 42*(6), 1165–1174.

Bhat, A. N., Galloway, J. C., & Landa, R. J. (2012). Relation between early motor delay and later communication delay in infants at risk for autism. *Infant Behavior & Development, 35*(4), 838–846. https://doi.org/10.1016/j.infbeh.2012.07.019.

Bhat, A. N., Landa, R. J., & Galloway, J. C. (2011). Current perspectives on motor functioning in infants, children, and adults with autism spectrum disorders. *Physical Therapy, 91*(7), 1116–1129. https://doi.org/10.2522/ptj.20100294.

Burns, T. G., King, T. Z., & Spencer, K. S. (2013). Mullen scales of early learning: The utility in assessing children diagnosed with autism spectrum disorders, cerebral palsy, and epilepsy. *Applied Neuropsychology Child, 2*(1), 33–42. https://doi.org/10.1080/21622965.2012.682852.

Carakovac, M., Jovanovic, J., Kalanj, M., Rudic, N., Aleksic-Hil, O., Aleksic, B., et al. (2016). Serbian language version of the modified checklist for autism in toddlers, revised, with follow-up: Cross-cultural adaptation and assessment of reliability. *Scientific Reports, 6*, 38222. https://doi.org/10.1038/srep38222.

Chiang, C. H., Soong, W. T., Lin, T. L., & Rogers, S. J. (2008). Nonverbal communication skills in young children with autism. *Journal of Autism and Developmental Disorders, 38*(10), 1898–1906. https://doi.org/10.1007/s10803-008-0586-2.

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*(4), e1121–1127.

Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., et al. (2016). Prevalence and characteristics of autism spectrum disorder among 4-year-old children in the autism and developmental disabilities monitoring network. *MMWR Surveillance Summary, 65*(3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., et al. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics, 125*(1), e17–23. https://doi.org/10.1542/peds.2009-0958.

Dominick, K. C., Davis, N. O., Lainhart, J., Tager-Flusberg, H., & Folstein, S. (2007). Atypical behaviors in children with autism and children with a history of language impairment. *Research in Developmental Disabilities, 28*(2), 145–162. https://doi.org/10.1016/j.ridd.2006.02.003.

Dumont-Mathieu, T., & Fein, D. (2005). Screening for autism in young children: The modified Checklist for Autism in Toddlers (M-CHAT) and other measures. *Developmental Disabilities Research Reviews, 11*(3), 253–262. https://doi.org/10.1002/mrdd.20072.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences.* Pacific Grove, CA: Brooks/Cole Publishing Company.

Feng, L., Li, C., Chiu, H., Lee, T. S., Spencer, M. D., & Wong, J. C. (2013). Autism spectrum disorder in Chinese populations: A brief review. *Asia-Pacific Psychiatry, 5*(2), 54–60. https://doi.org/10.1111/appy.12079.

Hartley, S. L., Sikora, D. M., & McCoy, R. (2008). Prevalence and risk factors of maladaptive behaviour in young children with Autistic Disorder. *Journal of Intellectual Disability Research, 52*(10), 819–829. https://doi.org/10.1111/j.1365-2788.2008.01065.x.

Hellendoorn, A., Wijnroks, L., van Daalen, E., Dietz, C., Buitelaar, J. K., & Leseman, P. (2015). Motor functioning, exploration, visuospatial cognition and language development in preschool children with autism. *Research in Developmental Disabilities, 39*, 32–42. https://doi.org/10.1016/j.ridd.2014.12.033.

Johnson, C. P., & Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*(5), 1183–1215. https://doi.org/10.1542/peds.2007-2361.

Kamio, Y., Inada, N., Koyama, T., Inokuchi, E., Tsuchiya, K., & Kuroda, M. (2014). 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. *Journal of Autism and Developmental Disorders, 44*(1), 194–203. https://doi.org/10.1007/s10803-013-1864-1.

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*(5), 827–839. https://doi.org/10.1007/s10803-007-0450-9.

Koh, H. C., Lim, S. H., Chan, G. J., Lin, M. B., Lim, H. H., Choo, S. H., et al. (2014). The clinical utility of the modified checklist for autism in toddlers with high risk 18-48 month old children in Singapore. *Journal of Autism and Developmental Disorders, 44*(2), 405–416. https://doi.org/10.1007/s10803-013-1880-1.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 159–174.

Limberg, K., Gruber, K., & Noterdaeme, M. (2016). The German version of the Child Behavior Checklist 1.5-5 to identify children with a risk of autism spectrum disorder. *Autism, 21*(3), 368–374. https://doi.org/10.1177/1362361316645932.

Maestro, S., Muratori, F., Cavallaro, M. C., Pei, F., Stern, D., Golse, B., et al. (2002). Attentional skills during the first 6 months of age in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(10), 1239–1245. https://doi.org/10.1097/00004583-200210000-00014.

Mayes, S. D., Calhoun, S. L., Mayes, R. D., & Molitoris, S. (2012). Autism and ADHD: Overlapping and discriminating symptoms. *Research in Autism Spectrum Disorders, 6*(1), 277–285. https://doi.org/10.1016/j.rasd.2011.05.009.

Mazurek, M. O., & Sohl, K. (2016). Sleep and behavioral problems in children with autism Spectrum disorder. *Journal of Autism and Developmental Disorders, 46*(6), 1906–1915. https://doi.org/10.1007/s10803-016-2723-7.

Mullen, E. M. (1995). *Mullen scales of early learning: AGS edition.* Circle Pines, MN: American Guidance Services.

Muratori, F., Narzisi, A., Tancredi, R., Cosenza, A., Calugi, S., Saviozzi, I., ... Calderoni, S. (2011). The CBCL 1.5-5 and the identification of preschoolers with autism in Italy. *Epidemiology and Psychiatric Sciences, 20*(4), 329–338.

Narzisi, A., Calderoni, S., Maestro, S., Calugi, S., Mottes, E., & Muratori, F. (2013). Child Behavior check List 1(1/2)-5 as a tool to identify toddlers with autism spectrum disorders: A case-control study. *Research in Developmental Disabilities, 34*(4), 1179–1189. https://doi.org/10.1016/j.ridd.2012.12.020.

Oono, I. P., Honey, E. J., & McConachie, H. (2013). Parent-mediated early intervention for young children with autism spectrum disorders (ASD). *Cochrane Database System Review,* 4. https://doi.org/10.1002/14651858.CD009774.pub2 Cd009774.

Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology, 14*(2), 239–251.

Perera, H., Wijewardena, K., & Aluthwelage, R. (2009). Screening of 18-24-month-old children for autism in a semi-urban community in Sri Lanka. *Journal of Tropical Pediatrics, 55*(6), 402–405. https://doi.org/10.1093/tropej/fmp031.

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for Autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics, 133*(1), 37–45. https://doi.org/10.1542/peds.2013-1813.

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*(2), 131–144.

Rogers, S. J. (2004). Developmental regression in autism spectrum disorders. *Developmental Disabilities Research Reviews, 10*(2), 139–143. https://doi.org/10.1002/mrdd.20027.

Samadi, S. A., & McConkey, R. (2015). Screening for autism in Iranian preschoolers: Contrasting M-CHAT and a scale developed in Iran. *Journal of Autism and Developmental Disorders, 45*(9), 2908–2916. https://doi.org/10.1007/s10803-015-2454-1.

Souders, M. C., Mason, T. B., Valladares, O., Bucan, M., Levy, S. E., Mandell, D. S., et al. (2009). Sleep behaviors and sleep quality in children with autism spectrum disorders. *Sleep, 32*(12), 1566–1578.

Srisinghasongkram, P., Pruksananonda, C., & Chonchaiya, W. (2016). Two-step screening of the modified checklist for autism in toddlers in Thai children with language delay and typically developing children. *Journal of Autism and Developmental Disorders, 46*(10), 3317–3329. https://doi.org/10.1007/s10803-016-2876-4.

Sun, X., & Allison, C. (2010). A review of the prevalence of autism Spectrum disorder in Asia. *Research in Autism Spectrum Disorders, 4*, 156–167.

Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-Vanderweele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics, 127*(5), e1303–1311. https://doi.org/10.1542/peds.2011-0426.

Werner, E., & Dawson, G. (2005). Validation of the phenomenon of autistic regression using home videotapes. *Archives of General Psychiatry, 62*(8), 889–895. https://doi.org/10.1001/archpsyc.62.8.889.

Wong, V. C., & Hui, S. L. (2008). Epidemiological study of autism spectrum disorder in China. *Journal of Child Neurology, 23*(1), 67–72. https://doi.org/10.1177/0883073807308702.

Wu, Y. T., Chen, W. J., Hsieh, W. S., Chen, P. C., Liao, H. F., Su, Y. N., et al. (2012). Maternal-reported behavioral and emotional problems in Taiwanese preschool children. *Research in Developmental Disabilities, 33*(3), 866–873. https://doi.org/10.1016/j.ridd.2011.11.018.