Use of the Chinese Version Modified Checklist for Autism in Toddlers in a High-risk Sample in Taiwan

Yi Shan Wong, Kaohsiung Medical University
Ching-Chi Yang, Kaohsiung Medical University
Lydia Stewart, Johns Hopkins Bloomberg School of Public Health
Chung-Hsin Chiang, National Cheng Chi University
Chin-Chin Wu (corresponding author), Kaohsiung Medical University; Kaohsiung Medical University Hospital
Lai-Sang Iao, Nottingham Trent University

Abstract

This study examined the utility of the Chinese version Modified Checklist for Autism in Toddlers (C-M-CHAT) in a high-risk sample. This study recruited 236 children aged from 18 to 47 months old in rural area of Southern Taiwan, including 113 children with Autism Spectrum Disorder (ASD) and 123 children with Developmental Delay (DD). Using failing any four of the 23 C-M-CHAT items as cutoff, it showed a sensitivity of 77.0% and a specificity of 72.4%. The positive predictive value (PPV) was 71.9% while the negative predictive value (NPV) was 77.4%. Using failing any three of the 14 new critical items--referred to as the *"Brief 14"*--as cutoff, it yielded a sensitivity of 70.8% and a specificity of 82.1%. The PPV of the *"Brief 14"* was 78.4% while the NPV was 75.4%. The preliminary results of **the** C-M-CHAT and the *"Brief 14"* performance demonstrated an acceptable predictive validity and promising utility for use in high-risk, rural populations in Taiwan.

Keywords: autism spectrum disorder, screening, sensitivity, specificity, C-M-CHAT

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social interaction and communication, as well as repetitive interests and stereotyped behavior (American Psychiatric Association, APA, 2013). Global prevalence estimates of ASD vary by nations, with current estimates in the United States as high as 1 in 100 school aged children (Autism and Developmental Disabilities Monitoring Network, 2012, 2014; Baird et al., 2006; Christensen et al., 2016). In Taiwan, the prevalence of ASD increased from 9.1 per 10,000 to 16.1 per 10,000 between 2007 and 2010 (Lai, Tseng, Hou, & Guo, 2012). Some of the possible factors contributing to the increasing prevalence rate include an increase in awareness of ASD, as well as an improvement of early screening and service availability (Elsabbagh et al., 2012; Wallace et al., 2012).

The lower prevalence of ASD in Taiwan compared to Western countries may be due in part to challenges in implementing ASD early detection. Early diagnosis of ASD through using of Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) and Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore & Risi, 1999) by clinicians were difficult to implement due to time-constraints, limited clinical infrastructure to enable multidisciplinary assessment, and low health insurance reimbursements (Durkin et al., 2015). In order to aid in early diagnosis, some screening tools were developed and used in early-age population. **Screening tools for ASD can be divided into two levels (Robins, 2008). Level 1 screening tools are used for screening in community when a public health nurse first encounters a child and determines whether he/she should be further assessed. When a child is found to be developing atypically and is transferred for confirmation, Level 2 screening tools are used. The Taiwan version of Screening Tool for Autism in Two-Year-Olds (T-STAT; Chiang et al., 2013) is a Level 2 screening tool. Currently, the T-STAT is the only measure** had been validated **and replicated** in Taiwan for use in high-risk samples. Therefore, high-validity screening tools that can be implemented in a short period of time are needed to improve the process of early identification for children who are at high-risk of ASD in Taiwan so that they can start early interventions.

Among the many commonly used ASD screening tools, the Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001) constitutes one of the most widely-used and well-studied in early childhood. The M-CHAT consists of 23 yes/no questions designed to assess the development and behavior of children aged 16-30 months and can be completed by a parent or a caregiver within 5-10 minutes. Children who fail either 3 of the 23 items or 2 of six critical items are considered a “positive screen” and at high-risk for ASD or developmental delay (DD) (Chlebowski, Robins, Barton, & Fein, 2013; Kleinman et al., 2008; Robins et al., 2001). The six critical items include item 2 (interest in other children), 7 (pointing for interest), 9 (showing), 13 (imitation), 14 (response to name), and 15 (following pointing). The six critical items possess good standardized canonical discriminant function coefficients, and the failure rate is more than 50% in children with autism. **The M-CHAT showed high sensitivity and low specificity for detecting high-risk children with ASD aged 17 to 48 months in English-speaking populations (Eaves et al., 2006; Snow & Lecavalier, 2008).** Recently, the M-CHAT authors suggested a "Best7" score (failing 2 of Best7) as an alternative method (Robins et al., 2010). The Best7 items include item 2 (interest in other children), 5 (pretend play), 7 (pointing for interest), 9 (showing), 14 (response to name), 15 (following pointing), and 20 (suspected deaf). To reduce the number of the false-positive outcomes, the Modified Checklist for Autism in Toddlers-Revised with Follow-Up (M-CHAT-R/F; Robins et al., 2014) was further developed.

The use of **the** M-CHAT was well-studied in English-speaking populations and had been translated into different language versions (Kleinman et al., 2008; Robins, 2008). Some studies showed that the M-CHAT **had good utility in screening for ASD** in **children from** Saudi Arabia (Seif Eldin et al., 2008), Spain (Canal-Bedia et al., 2011), Japan (Inada, Koyama, Inokuchi, Kuroda, & Kamio, 2011; Kamio et al., 2014), **South Korea** (Seung et al., 2015), Iran (Samadi & McConkey, 2015), and Turkey (Kondolot et al., 2016). Inada et al. (2011) suggested the use of a different cutoff score to obtain better sensitivity and specificity for the Japanese version of the M-CHAT in Japanese children aged 18 to 24 months. A recent systematic review of ASD screening in low resource settings found that many studies using western-developed tools in non-western populations reported improved validity through using cutoff scores and critical items that differed from what was recommended (Stewart & Lee, 2017). Beyond simple translation, differences in cultural norms must be considered through rigorous validation and pilot testing of tools before they are considered for general use.

The Chinese-language version of the M-CHAT has previously been used in clinical and community-based samples. Wong et al. (2004) translated the M-CHAT (23 questions) into Chinese with graded scores, and combined it with five observation items from the original Checklist for Autism in Toddlers (CHAT; Baron-Cohen, Allen & Gillberg, 1992) to constitute a new checklist for aged 13 to 86 months, which is called the CHAT-23 in Hong Kong. **Seven key items were identified, included items** 2 (interest in other children), 5 (pretend play), 7 (pointing for interest), 9 (showing), 13 (imitation), 15 (following pointing), and 23 (social referencing), **named the '7 key items'**. High sensitivity and specificity value of 77-93% were reported for the CHAT-23. In Singapore, a South-East Asia multi-racial population consisting of 74% Chinese participants, Koh et al. (2014) examined clinical utility of the M-CHAT for detecting high-risk children with ASD aged 18 to 48 months. Their findings also showed good sensitivity and specificity value of 71-81% were reported for both the six critical items and Best7 score of **the** M-CHAT.

When considering the use of the M-CHAT in Taiwan, the findings of these case-control studies showed the importance of factors such as choice of scoring method, the age-range of the sample, selection of typically developing children along with children with DD to serve as control groups, dialectical differences between Mandarin and Cantonese, and potential variation in western-cultural influenced between Hong Kong, Singapore, and Taiwan. **Lung, Chiang, Lin, and Shu (2011) used the Taiwan Birth Cohort Study (TBCS) developmental instrument in the first stage and the M-CHAT in the second stage were utilized to screen children below 3 years old in a community-based setting. Their study examined whether the TBCS developmental instrument can be used for detecting ASD based on the M-CHAT category, and** found that 18% children were screened out as being high-risk for ASD. **However, they did not confirm the diagnosis of children for examining validity of the M-CHAT.** Recently, failing 13/14 of the 23 items was considered optimal for screening of children with ASD at 66 months in a community-based setting (Lung, Chiang, Lin, & Shu, 2017). No study has previously examined the validity and utility of the M-CHAT in a sample of children at high-risk for ASD in Taiwan.

The implementation of culturally appropriate and valid screening for ASD is a critical first step towards improving the pathways to care for children with ASD in Taiwan. Previous studies have found an urbanization disparity in the distribution of ASD diagnosis (Lai et al., 2012; Lung et al., 2017), it highlights the importance of early detecting of children with ASD, especially for rural areas. The M-CHAT is widely used globally in both community and clinical settings screening. In Taiwan, the M-CHAT was only used in community settings with no adaptation from the original form used in Lung et al. (2011). However, adaptations to instrument administration and cutoff scores may be necessary when using tools outside the settings in which they were developed (Stewart & Lee, 2017). In addition, using **the** M-CHAT in **a** high-risk sample, cutoff score is different from in a low-risk sample. The current study investigated the validity and utility of the Chinese version M-CHAT (C-M-CHAT) for screening a high-risk sample of children aged 18 to 47 months in Chia-Yi, a rural and agricultural area in Taiwan with a low-socioeconomic status (low-SES) and a large indigenous population. Through validating the C-M-CHAT in a high-risk, low-SES population, this study will inform future ASD screening efforts in Taiwan.

2. Methods

2.1 Participants

The study sample included 236 participants. The range of chronological age (CA) of the participants ranged from 18 to 47 months. The participants were recruited from the Chia-Yi area in Southwestern Taiwan. As part of recruitment, a social worker or physician explained the study to caregivers of children with suspected developmental problems during routine visits. If the caregivers or parents agreed to participate in this study, they were referred to the researcher team, which included clinical psychologists and children/adolescent psychiatrists. The researchers would then contact the caregivers and arrange the time of the assessment. Before starting the research, the researchers explained the purpose, content and process of the study, including filling in the questionnaires. After confirming that the parents clearly understood, a consent form was given to them. All of the parents had achieved at least an elementary education, and thus the researchers did not provide an explanation regarding the completion of the C-M-CHAT. 90% of the C-M-CHAT were completed by mothers of the participants. There were 8 families **in** which mothers were foreigners **who could not speak and read Chinese adequately** and fathers were **unavailable to complete the C-M-CHAT. These eight families were excluded from the study due to language barriers.**

Majority of the participants were diagnosed and classified into ASD or DD group based on DSM-5 criteria. Using the suggestion of lower criteria from Frazier et al. (2012), and Young and Rodi (2014), **children with ASD would meet criteria for an ASD diagnosis based on DSM-5 criteria: (1) three from social**

communication/interaction and one restricted/repetitive behaviors; (2) two from social communication/interaction and two restricted/repetitive behaviors. There were 108 children with ASD based on DSM-5 criteria. **In addition, there were 5 children meet criteria for an ASD diagnosis based on DSM-IV-TR criteria, but not based on DSM-5 criteria. The multidisciplinary team also reached consensus to include these 5 additional children in the ASD group because DSM-IV-TR was still widely used in Taiwan clinical setting.** If there were cases with diagnostic uncertainty, a team meeting was held within one month to review and discuss the specific developmental history and measured data, and final a diagnosis were made by the team.

All participants were assessed by the Mullen Scales of Early Learning (MSEL; Mullen, 1995) for obtaining four domains of development abilities. The children in the DD group were determined by failing to reach a **T** score of 40 in any one of the four domains, or that their **composite standard** score was lower than 85. Furthermore, we interviewed parents for more development milestones (i.e., language, motor) of their children and referred to Adaptive Behavior Assessment System-II (ABAS-II, Harrison & Oakland, 2003) which was filled in by one of the parents. There are 81 children with overall delayed development, 41 children with language delayed, and 1 child with fine motor delayed, which did not fit the diagnosis of ASD either in DSM-IV-TR or DSM-5. The study sample finally included 113 children with a diagnosis of ASD and 123 children with a diagnosis of DD to serve as controls.

According to the method of Rogers, Hepburn, Stackhouse and Wehner (2003), the raw scores of the four domains in MSEL were calculated, and the corresponding age equivalents were found. The age equivalents were summed together and divided by four, and this constituted the participants' overall mental age (MA). The numbers of years of education of the parents were coded by primary school (6 years), junior high school (9 years), senior high school (12 years), junior college (14 years), university (16 years), and master (18 years). Due to the divorce of the caregivers for one of the participants, the number of years of education for the father was not obtained. Using independent **t-tests and Chi-square test**, we compared the basic information and the developmental abilities of both groups of participants (see Table 1). **Most mothers of both ASD and DD group would be classified as the last two job categories, according to the system of 7 categories of occupations (Huang, 1998), meaning that their jobs do not require a specific skill, such as cleaners.** In addition, five mothers of the ASD group and three mothers of the DD group lacked the information of occupation.

Insert Table 1 about here

2.2 Measures and procedure

Mullen Scales of Early Learning (MSEL; Mullen, 1995): The MSEL is a standardized developmental test for assessing children from birth to 68 months. It consists of four cognitive scales, which include visual reception, fine motor, receptive language, and expressive language. The MSEL has demonstrated concurrent validity with other well-known development tests of language and cognitive development (Bayley Scales of Infant Development — Second Edition, BSID-II; Bayley, 1993). In addition, it has acceptable reliability in internal consistency and retesting.

Modified Checklist for Autism in Toddlers (M-CHAT; Robins et al., 2001): The M-CHAT consists of 23 yes/no questions, and it was designed to screen for early detection of autism for toddlers aged 16 to 30 months. The questions are scored as pass or fail, and are then divided into critical and non-critical items. According to the rating criteria, a child is screened as being at high-risk for ASD if he/she failed any 3 of the 23 items or he/she failed two of the critical items (Robins et al., 2001). The Chinese version M-CHAT (C-M-CHAT) was used in this study and obtained from the M-CHAT authors' website (mchatscreen.com). Robins et al. (2001) reported that sensitivity range from 87 to 97% while specificity from 95 to 99%, depending on the number of **the** M-CHAT items used and cutoff point chosen.

Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999): The ADOS is a semi-structured play-based and observational assessment. It is divided into four modules, and each module is selected according to **individuals' expressive language ability**. It takes 30 to 45 minutes to finish the assessment, and it provides a standard context for observing and scoring language and communication, social and stereotyped behavior, and restricted interests. **All participants were administered the module 1 based on their levels of expressive language ability. Classification was made on the basis of exceeding cutoffs in each of two domains: social behavior and communication, and exceeding a cutoff for a combined social-communication total. The cutoffs of social behavior, communication, and combined social-communication total as autism were 4, 7, and 12, respectively. While, the cutoffs of social behavior, communication, and combined social-communication total as ASD were 2, 4, and 7, respectively.**

None of the children in our sample had received a developmental diagnosis before. All of the children received a set of measures and then being diagnosed with reference to developmental history, current concerns from parents, the children's usual activity and performance in daily life, observations of the child, and the results of ADOS (Lord et al., 1999) by a multidisciplinary team, which included senior child psychologists with Ph.D. degrees and child psychiatrists. Two of the authors had previously received research training on ADOS in the U.S. (with

Dr. Catherine Lord's team at Michigan) and/or in Taiwan (with Dr. Catherine Rice's team at Pingtung county).

2.3 Data analysis

IBM SPSS 20 was used to conduct the statistical analysis of this study. The coefficient of internal consistency was used to examine reliability of **the** C-M-CHAT. In addition, both Chi-square test and discriminant analysis were used to identify the critical items of **the** C-M-CHAT. **To avoid alpha inflation, only *ps* < .002 (.05/23) were considered statistically significant**. Finally, Receiver Operating Characteristics (ROC) curve was used to decide the optimal range of the cut-off score.

3. Results

The coefficient of internal consistency (Kuder-Richardson, KR) for 23 items in the C-M-CHAT was 0.80, while for each item was between 0.78 and 0.82, thus showed high internal consistency. However, the correlations of the total score and items 1 (enjoy being swung, etc.), 3 (climbing), 11 (oversensitive to noise), 16 (walking), and 18 (unusual finger movements) were below .2.

Using the Chi-square tests for testing differences between two groups, two groups of participants showed significant differences in most of the 23 items, including items 2 (interest in other children), 5 (pretend play), 6 (pointing for requesting), 7 (pointing for interest), 9 (showing), 12 (smile in response to face or smile), 13 (imitation), 14 (response to name), 15 (following pointing), 17 (following gaze), 19 (attract attention), 20 (suspected deaf), 21 (language comprehension), 22 (wandering without purpose), and 23 (social referencing). The children with ASD exhibited more impairment in these items. There were no significant differences in the rest of the items (see Table 2). In addition, children with ASD failed more items than children with DD (6.7 versus 2.6).

Insert Table 2 about here

Using the coefficient of structure matrix (> .32) in discriminant analysis, we found that there were 14 items which showed high power of discriminant, including items 14 (response to name), 2 (interest in other children), 17 (following gaze), 19 (attract attention), 21 (language comprehension), 23 (social referencing), 15 (following pointing), 22 (wandering without purpose), 9 (showing), 7 (pointing for interest), 5 (pretend play), 6 (pointing for requesting), 20 (suspected deaf) and 13 (imitation), which were ascendingly ranked by their coefficient of structure matrix. We used these 14 items as screening criteria, and we referred to as the *"Brief 14"*, to determine whether the screening results were consistent with the diagnosis (see Table 3). Using discriminant analysis, 23 items predicted ASD 69.9% correctly, while predicting DD, 89.4% correctly. The *"Brief 14"* items classified ASD 64.6% correctly and DD 85.4% correctly.

Insert Table 3 about here

Using ROC, 3-5 was the optimal range of the cut-off score (see Table 4). Details of sensitivity and specificity for failing any three or four of the 23 items could be seen in Table 5. Using cutoff score of 3, the positive predictive value (PPV) was 63.1%, while the negative predictive value (NPV) was 82.3%. The area under the curve (AUC) was 0.82. Using cutoff score of 4, the PPV was 71.9%, while the NPV was 77.4%.

Insert Tables 4 and 5 about here

This study found several items, referred to as the *"Brief 14"*, that were highly correlated with the diagnosis by discriminant analysis. Using the ROC to determine the optimal cut-off score of the *"Brief 14"*, 2-4 was the best choice (see Table 6). Details of sensitivity and specificity for the *"Brief 14"* could be seen in Table 7. Using cutoff score of 3 in the *"Brief 14"*, the PPV of the "Brief 14" was 78.4%, while the NPV was 75.4%. The AUC was 0.83. In addition, details of specificity, sensitivity for Robins et al.'s (2001) and Wong et al.'s (2004) shorter items could be seen in Table 7. The results of both studies showed poor sensitivity and high specificity. The AUC were 0.77, 0.79, and 0.77, respectively.

Insert Tables 6 and 7 about here

---

4. Discussion

Few Level 2 screening tools for **young** children with ASD **under the age of four years** are available in Taiwan. The aim of this study **was** to examine the utility of the C-M-CHAT for detecting children with ASD before four-year-old in a high-risk sample in a rural and agricultural area of Southern Taiwan. The range of participants' CA **was** between 18 to 47 months, while the range of MA **was** between 6.5 to 44.25 months.

Five items were found to be inconsistent with the others including: items 1 (enjoy being swung etc.), 3 (climbing), 11 (oversensitive to noise), 16 (walking) and 18 (unusual finger movements). None of these 5 items **belongs to the *"Brief 14"* and the current results supported** eliminating some items in the C-M-CHAT may construct a more efficient and brief ASD screening tool.

In this study, items 1 (enjoy being swung, etc.), 3 (climbing), 4 (playing peek-a-boo), 11 (oversensitive to noise), 16 (walking), and 18 (unusual finger movements) did not show differences between ASD and DD groups. According to Wong et al.'s (2004) study, which included a sample of children with DD, items 1 (enjoy being swung, etc.), 3 (climbing), 4 (playing peek-a-boo) and 16 (walking) could not differentiate between ASD and DD well. Robins et al. (2014) suggested that 3 items should be dropped, including items 4 (playing peek-a-boo), 8 (play properly with toys), and 22 (wandering without purpose). Eaves et al. (2006) also suggested that 4 of the items should be dropped, whereas item 1 (enjoy being swung, etc.), 3 (climbing), 16 (walking) were not considered as ASD core impairments. In addition, the item 11 (oversensitive to noise) had poor internal consistency and could not differentiate between children with ASD from children with DD. The results of this study supported their findings and these 5 items could reasonably be dropped. However, it is worth noting that item 18, which belongs to repetitive interests/stereotyped behaviors, also could not distinguish children with ASD from those with DD. Consistent with the results of Robins et al. (2001), we also found that young children with ASD rarely exhibited stereotyped or repetitive motor movement. Thus, there was no significant difference between the two groups on item 18.

Consistent with previous studies **(Eaves et al., 2006; Snow & Lecavalier, 2008)**, failing any three of the 23 items as screening criteria had high sensitivity for screening ASD in a high-risk sample. **The current results also supported findings from those other studies in showing high sensitivity** (77.0-87.6%) and poor specificity (52.8-72.4%). **The number of the failure items** of control group in Robins et al.'s (2001) study was mostly 0, whereas in this study, there was only 13 children with DD which their **number of** failure item was 0. Besides, Eaves et al. (2006) suggested deleting 4 items, and yet their cutoff score suggested was 3, which supported our findings that cutoff score should be raised in a high-risk sample. The results of this study also indicated that researchers should consider using **higher the number of the failure items** as **cutoff** for detecting children with ASD in a high-risk sample, especially for communities similar to those studied in Taiwan.

This study further examined the validity of Robins et al.'s (2001) six critical items and Best7 items, and Wong et al.'s (2004) 7 key items. Inconsistent with previous studies (Eaves et al., 2006; Koh et al., 2014; Snow & Lecavalier, 2008; Wong et al., 2004), the results of this study produced poor sensitivity (46.0-56.6%) and high specificity (87.0-93.5%). In both Koh et al.'s (2014) and Wong et al.'s (2004) studies, typically developing children were included in the non-ASD group and there were more severe cases of children with ASD than in the current study, which could constitute a reason that this study could not replicate those findings. However, using discriminant analysis, the *"Brief 14"* was determined to have critical items for children with ASD, including items 2 (interest in other children), 5 (pretend play), 6 (pointing for requesting), 7 (pointing for interest), 9 (showing), 13 (imitation), 14 (response to name), 15 (following pointing), 17 (following gaze), 19 (attract attention), 20 (suspected deaf), 21 (language comprehension), 22 (wandering without purpose), and 23 (social referencing). Using failing any 3 of the *"Brief 14"* as criteria, the results of this study also demonstrated acceptable sensitivity and specificity for screening children with ASD in high-risk sample.

General standards for adequate levels of sensitivity and specificity have been suggested. Specifically, a sensitivity level of 70-80% and a specificity level of about 80% were recommended (Glascoe, 2005). In addition, AUC of 0.70-0.90 **was suggested to** indicate a moderate accuracy rate, **with an** AUC >.90 **considered** a high accuracy rate (Swets, 1988). **Using failing any 4 of the 23 items as cutoff, the findings of this study showed a bit higher specificity, PPV and NPV compared to previous studies (Eaves et al., 2006; Snow & Lecavalier, 2008). If using a criterion of failing any 3 of the *"Brief 14"* instead of using of failing four of 23 items as cutoff**, the sensitivity was **a bit lower, the specificity, AUC, PPV and NPV is a bit higher**. Both cutoffs of the 23 items and the *"Brief 14"* were used, and acceptable validity was shown for detecting children with ASD in high-risk sample. **However, the *"Brief 14"* showed a better accuracy in detecting young children with ASD in Taiwan.**

Comparing the critical items among different **the** M-CHAT studies, there were 3 items that overlapped, including items 7 (pointing for interest), 9 (showing), and 15 (following pointing). These 3 items belonged to the joint attention and supported the impairment of joint attention in children with ASD again. This study also replicated previous findings using behavior observation method, which children with ASD exhibited robust impairments in joint attention skills (Chiang, Soong, Lin, & Rogers, 2008; Wetherby, Watt, Morgan, & Shumway, 2007). This study showed that low-SES mothers of children with ASD would underestimate their children's symptoms overall, but would still report their impairments in social interaction. **Low-SES mothers of children with ASD could concern single symptom (i.e. interest in other) instead of noting the overall symptoms of ASD. It might lead to low-SES mothers do not fully know social and communicative impairments of their children. In addition, Taiwanese mothers of children with ASD mainly focus on language problems of their children, it might overlook the social deficits.** The result provided a reference for early screening.

Previous studies have shown some cultural differences in the utility of **the** M-CHAT and our results also support this trend (Inada et al., 2011; Samadi & McConkey, 2015; Wong et al., 2004; Stewart & Lee, 2017). However, compared to previous studies (Eaves et al., 2006; Koh et al., 2014; Snow & Lecavalier, 2008), the scores of the total 23 items were lower and the number of the critical items were higher in this study. There was no item with a failure rate of more than 50%, which is different from previous studies that had a failure rate of more than 50% in children with ASD (Robins et al., 2001; Wong et al., 2004). A possible explanation is that Taiwanese parents have more tolerance towards their children's behavior or they might think that it is not very serious when their children do not perform some important skills. It is also possible that parents experience social desirability bias and social stigma in Taiwanese culture (i.e., "saving face"), which can influence parents to underreport or deny the symptoms of their children with ASD in order to avoid stigmatization. In addition, compared to other areas, Chia-Yi is considered as a rural and agricultural area in Taiwan. Consequently, low-SES parents tend to be busy with their work, and may not be as deeply aware about their children's behaviors and might misinterpret them. The results suggested that early detection of ASD in rural areas of Southern Taiwan should provide parents with more information about ASD and milestones of development, a trend found in other low-SES populations globally (Kakooza-Mwesige et al, 2014). In order to reduce the denying behavior of parents, it is necessary to empathize with their stress and distress, and combine ASD screening with de-stigmatization and awareness outreach campaigns.

The results of this study showed that the current M-CHAT has not adequately reflected children's behaviors in a way that is easily recognizable for lower SES Taiwanese parents who may underreport their children's behavior. Therefore, a new cutoff score may be needed for this population. However, additional studies are needed for replication and verification of these findings. There are some limitations of this study. First, due to limitations of funding and the number of qualified researchers, the "gold standard" caregiver interview ADI-R was not included in this study. Although we are confident of the accuracy of our clinical diagnosis, including the ADI-R could allow for a better diagnostic assessment. Second, since we focused on parents of children with ASD from a rural and agricultural area of Southern Taiwan, it is not yet clear if the criteria and results would be the same in urban areas such as Taipei. Third, this study only collected data from one of the parents. Obtaining data from the other parent as well may be helpful for attaining a broader perspective of child behavior. Future studies should consider inclusion of all caregivers, potentially including grandparents and other non-parental caregivers, in completion of ASD screening tools.

5. Conclusions

Consistent with previous studies, caregivers from different cultures tended to explain the behavior of children with ASD differently. Some adaptation to screener content, structure or cutoff score may be needed when using tools outside of the settings in which they were first developed and tested. In this study, **the** C-M-CHAT produced acceptable sensitivity and specificity in screening ASD in rural area of Taiwan. Further research in this rural population should examine ways to increase parental awareness of child developmental milestones and ASD behaviors to improve early detection and promote health. Future screening efforts are needed to verify the validity and utility of the C-M-CHAT in this population and explore use in more urban areas in Taiwan as well.

Acknowledgements

We are grateful to the parents and children who generously donated their time to this project. Much thanks to Li-Ching Lee at the John Hopkins University for her invaluable comments and suggestions about the English editing. This study was supported by the Ministry of Science and Technology (formerly National Science Council) (NSC102-2410-H-037-002; MOST-103-2628-H-037-001-MY2).

References

Autism and Developmental Disabilities Monitoring Network (2012). Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *Morbidity and Mortality Weekly Reports. Surveillance Summaries,* **61,** 1-19.

Autism and Developmental Disabilities Monitoring Network (2014). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Reports. Surveillance Summaries,* **63,** 1-21.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washington, DC: American Psychiatric Association.

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

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry, 161*(6), 839-843.

Bayley, N. (1993). *Bayley scales of infant development: Manual*. Psychological

Corporation.

Canal-Bedia, R., García-Primo, P., Martín-Cilleros, M. V., Santos-Borbujo, J., Guisuraga-Fernández, Z., Herráez-García, L., et al. (2011). Modified checklist for autism in toddlers: Cross-cultural adaptation and validation in Spain. *Journal of Autism and Developmental Disorders, 41*(10), 1342-1351.

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.

Chiang, C. H., Wu, C. C., Hou, Y. M., Chu, C. L., Liu, J. H., & Soong, W. T. (2013). Development of T-STAT for early autism screening. *Journal of Autism and Developmental Disorders, 43*(5), 1028-1037.

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-e1127.

Christensen, D. L., Bilder, D. A., Zahorodny, W., Pettygrove, S., Durkin, M. S., Fitzgerald, R. T., ... & Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among 4-year-old children in the autism and developmental disabilities monitoring network. *Journal of Developmental & Behavioral Pediatrics*, *37*(1), 1-8.

Durkin, M. S., Elsabbagh, M., Barbaro, J., Gladstone, M., Happe, F., Hoekstra, R. A., ... & Tager- Flusberg, H. (2015). Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide. *Autism Research*, *8*(5), 473-476.

Eaves, L. C., Wingert, H., & Ho, H. H. (2006). Screening for autism agreement with diagnosis. *Autism*, *10*(3), 229-242.

Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcín, C., et al. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research, 5*(3), 160-179.

Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., et al. (2012). Validation of proposed DSM-5 criteria for autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(1), 28-40.

Glascoe, F. P. (2005). Screening for developmental and behavioral problems. *Developmental Disabilities Research Reviews, 11*(3), 173-179.

Harrison, P., & Oakland, T. (2003). *Adaptive behavior assessment system (ABAS-II).* San Antonio, TX: The Psychological Corporation.

Huang, Y. Z. (1998). Construction and evaluation of new occupational classification in Taiwan. *Survey Research, 5*, 5-32. (in Chinese)

Inada, N., Koyama, T., Inokuchi, E., Kuroda, M., & Kamio, Y. (2011). Reliability and validity of the Japanese version of the modified checklist for autism in toddlers (M-CHAT). *Research in Autism Spectrum Disorders, 5*(1), 330-336.

Kakooza-Mwesige, A., Ssebyala, K., Karamagi, C., Kiguli, S., Smith, K., Anderson, M. C., ... & Grether, J. K. (2014). Adaptation of the "ten questions" to screen for autism and other neurodevelopmental disorders in Uganda. *Autism, 18*(4), 447-457.

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.

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.

Koh, H. C., Lim, S. H., Chan, G. J., Lin, M. B., Lim, H. H., Choo, S. H. T., & Magiati, I. (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.

Kondolot, M., Özmert, E. N., Öztop, D. B., Mazıcıoğlu, M. M., Gümüş, H., & Elmalı, F. (2016). The modified checklist for autism in Turkish toddlers: A different cultural adaptation sample. *Research in Autism Spectrum Disorders, 21*, 121-127.

Lai, D. C., Tseng, Y. C., Hou, Y. M., & Guo, H. R. (2012). Gender and geographic differences in the prevalence of autism spectrum disorders in children: Analysis of data from the national disability registry of Taiwan. *Research in Developmental Disabilities, 33*(3), 909-915.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *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.

Lung, F. W., Chiang, T. L., Lin, S. J., & Shu, B. C. (2011). Autism-risk screening in the first 3 years of life in Taiwan birth cohort pilot study. *Research in Autism Spectrum Disorders, 5*(4), 1385-1389.

Lung, F. W., Chiang, T. L., Lin, S. J., & Shu, B. C. (2017). Urban and education disparity for autism spectrum disorders in Taiwan birth cohort study. *Journal* *of Autism and Developmental Disorders, 47*(3), 599-606.

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

Robins, D. L. (2008). Screening for autism spectrum disorders in primary care settings. *Autism, 12*(5), 537-556.

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M. A., 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.

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.

Robins, D. L., Pandey, J., Chlebowski, C., Carr, K., Zaj, J., Arroyo, M., et al. (2010). *M-CHAT Best7: A new scoring algorithm improves positive predictive power of the M-CHAT.* Paper presented at the International Meeting for Autism Research (IMFAR), Philadelphia, PA.

Rogers, S. J., Hepburn, S. L., Stackhouse, T., & Wehner, E. (2003). Imitation performance in toddlers with autism and those with other developmental disorders. *Journal of Child Psychology and Psychiatry, 44*(5), 763-781.

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.

Seif Eldin, A., Habib, D., Noufal, A., Farrag, S., Bazaid, K., Al-Sharbati, M., et al. (2008). Use of M-CHAT for a multinational screening of young children with autism in the Arab countries. *International Review of Psychiatry, 20*(3), 281-289.

Seung, H., Ji, J., Kim, S. J., Sung, I., Youn, Y. A., Hong, G., et al. (2015). Examination of the Korean modified checklist of autism in toddlers: Item response theory. *Journal of Autism and Developmental Disorders, 45*(9), 2744-2757.

Snow, A. V., & Lecavalier, L. (2008). Sensitivity and specificity of the modified checklist for autism in toddlers and the social communication questionnaire in preschoolers suspected of having pervasive developmental disorders. *Autism, 12*(6), 627-644.

Stewart, L. A., & Lee, L. C. (2017). Screening for autism spectrum disorder in low- and middle-income countries: A systematic review. *Autism, 21*(5), 527-539.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*(4857), 1285-1293.

Wallace, S., Fein, D., Rosanoff, M., Dawson, G., Hossain, S., Brennan, L., et al. (2012). A global public health strategy for autism spectrum disorders. *Autism Research, 5*(3), 211-217.

Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders, 37*(5), 960-975.

Wong, V., Hui, L. H. S., Lee, W. C., Leung, L. S. J., Ho, P. K. P., Lau, W. L. C., et al. (2004). A modified screening tool for autism (Checklist for autism in toddlers [CHAT-23]) for Chinese children. *Pediatrics, 114*(2), e166-e176.

Young, R. L., & Rodi, M. L. (2014). Redefining autism spectrum disorder using DSM-5: The implications of the proposed DSM-5 criteria for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(4), 758-765.

Table 1. Participants' demographics

	ASD group (n = 113)	DD group (n = 123)	*p*	***Effect size***
CA (month)				
Mean (SD)	31.20(7.60)	29.18(8.71)	.06	**0.25**
Range	18.00-47.00	18.00-47.00		
MA (month)				
Mean (SD)	19.48(8.00)	22.02(7.30)	.01	**0.33**
Range	6.50-37.75	10.50-44.25		
Number of years of education of the mother				
Mean (SD)	14.00(2.63)	13.60(2.50)	.24	**0.16**
Range	6.00-18.00	6.00-18.00		
Number of years of education of the father				
Mean (SD)	14.21(2.74)	13.31(2.50)	.01	**0.34**
Range	6.00-18.00	6.00-18.00		
Last 2 job categories[a]	**69(63.9%)**	**83(69.2%)**	**.40**	**0.06**
Gender				
Ratio (male : female)	8.42(101:12)	2.62(89:34)	.00	**0.22**
ADOS score				
Mean (SD)	17.49(2.91)	3.25(1.52)	.00	**6.13**
Range	11.00-22.00	0.00-6.00		

[a]Mothers' jobs

Table 2. Failure rate of each item between two groups

Item	ASD (n = 113) Failure rate (Yes : No)	DD (n = 123) Failure rate (Yes : No)	χ^2	*p*
1 enjoy being swung, etc.	0.16 (95:18)	0.12 (108:15)	.68	.46
2 interest in other children	0.34 (75:38)	0.06 (116:7)	29.79	.00*
3 climbing	0.12 (99:14)	0.06 (116:7)	3.26	.11
4 playing peek-a-boo	0.20 (90:23)	0.16 (103:20)	.66	.50
5 pretend play	0.34 (75:38)	0.11 (110:13)	18.49	.00*
6 pointing for requesting	0.26 (84:29)	0.07 (115:8)	16.35	.00*
7 pointing for interest	0.27 (82:31)	0.07 (115:8)	18.70	.00*
8 play properly with toys	0.28 (81:32)	0.18 (101:22)	3.63	.06
9 showing	0.19 (91:22)	0.02 (121:2)	18.62[b]	.00*
10 eye contact	0.19 (91:22)	0.08 (113:10)	6.46	.01
11 oversensitive to noise [a]	0.43 (64:49)	0.34 (81:42)	2.11	.18
12 smile in response to face or smile	0.15 (96:17)	0.02 (120:3)	10.49[b]	.00*
13 imitation	0.36 (72:41)	0.14 (106:17)	716.03	.00*
14 response to name	0.30 (79:34)	0.03 (119:4)	29.44[b]	.00*
15 following pointing	0.33 (76:37)	0.08 (113:10)	22.37	.00*

16 walking	0.02 (111:2)	0.02 (121:2)	.00[b]	1.00
17 following gaze	0.36 (72:41)	0.07 (114:9)	29.59	.00*
18 unusual finger movements [a]	0.27 (83:30)	0.23 (95:28)	.46	.55
19 attract attention	0.44 (63:50)	0.13 (107:16)	28.53	.00*
20 suspected deaf [a]	0.49 (58:55)	0.24 (94:29)	16.18	.00*
21 language comprehension	0.39 (69:44)	0.11 (109:14)	24.13	.00*
22 wandering without purpose [a]	0.48 (59:54)	0.20 (99:24)	21.28	.00*
23 social referencing	0.35 (73:40)	0.10 (111:12)	22.54	.00*

[a] reversed item, [b] Yates's correction.

*$p < .001$

Table 3. Coefficient of structure matrix of each item in **the** M-CHAT (descending order)

Item of questions	Coefficient of structure matrix
14 response to name	.51
2 interest in other children	.49
17 following gaze	.49
19 attract attention	.48
21 language comprehension	.44
23 social referencing	.42
15 following pointing	.42
22 wandering without purpose	.41
9 showing	.40
7 pointing for interest	.38
5 pretend play	.38
6 pointing for requesting	.35
20 suspected deaf	.35
13 imitation	.35
12 smile in response to face or smile	.30
10 eye contact	.22
8 play properly with toys	.16
3 climbing	.15
11 oversensitive to noise	.12
1 enjoy being swung, etc.	.07
4 playing peek-a-boo	.07
18 unusual finger movements	.06

16 walking	.01

Table 4. Sensitivity and specificity for different **the** C-M-CHAT cutoff

Cutoff	Sensitivity	Specificity
1	.97	.11
2	.94	.33
3	.88	.53
4	.77	.72
5	.63	.85
6	.51	.94
7	.43	.95

Table 5. Classification between **the** C-M-CHAT and clinical diagnosis

Cutoff score	True Positive	False Negative	False Positive	True Negative	Sensitivity	Specificity	PPV	NPV
3	99	14	58	65	87.6%	52.8%	63.1%	82.3%
4	87	26	34	89	77.0%	72.4%	71.9%	77.4%

PPV = positive predictive value, NPV = negative predictive value.

Table 6. Sensitivity and specificity for different ***"Brief 14"*** of **the** C-M-CHAT cutoff

Cutoff	Sensitivity	Specificity
1	.93	.33
2	.83	.58
3	.71	.82
4	.56	.92
5	.44	.97
6	.40	.99
7	.28	.99

Table 7. Classification between *"Brief 14"* and different critical items of **the** C-M-CHAT and clinical diagnosis

Different Critical Item	True Positive	False Negative	False Positive	True Negative	Sensitivity	Specificity	PPV	NPV
"Brief 14"[a]	94	19	52	71	83.2%	57.7%	64.4%	78.9%
"Brief 14"[b]	80	33	22	101	70.8%	82.1%	78.4%	75.4%
Robins' six critical items	52	61	8	115	46.0%	93.5%	86.7%	65.3%
Robins' Best7 items	60	53	16	107	53.1%	87.0%	78.9%	66.9%
Wong's 7 key items	64	49	14	109	56.6%	88.6%	82.1%	69.0%

[a] cutoff = 2, [b] cutoff = 3, PPV = positive predictive value, NPV= negative predictive value.