

# Diagnostic Validity of Autism Diagnostic Observation Schedule, Second Edition (K-ADOS-2) in the Korean Population

**So Yoon Kim**

Duksung Women's University

**Miae Oh** (✉ [miae612@naver.com](mailto:miae612@naver.com))

Kyung Hee University <https://orcid.org/0000-0002-1927-3637>

**Guiyoung Bong**

Seoul National University Bundang Hospital

**Da-Yea Song**

Seoul National University Bundang Hospital

**Nan-He Yoon**

Wonkwang University

**Joo Hyun Kim**

Seoul National University Bundang Hospital

**Hee Jeong Yoo**

Seoul National University College of Medicine <https://orcid.org/0000-0003-0521-2718>

---

## Research Article

**Keywords:** autism, diagnosis, validity, ADOS-2

**Posted Date:** February 3rd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1313371/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

## Background

Although the Korean version of the Autism Diagnostic Observation Schedule-2 (K-ADOS-2) is widely being used to diagnose autism in South Korea, no previous study has examined the validity and reliability of all modules of K-ADOS-2 across a wide age range, particularly older children, adolescents, and adults.

## Method

Data from 2,177 participants were included (mean age = 79.2 months; 73.5% male); 1480 participants with autism and 697 participants without autism (Module-T, n = 296; Module-1, n = 654; Module-2 n = 574; Module-3 n = 411; Module-4, n = 233). Participants completed a battery of tests, including the K-ADOS or K-ADOS-2 and other existing diagnostic instruments. Sensitivity, specificity, area under the receiver operating characteristic (ROC) curve, positive predictive value (PPV), negative predictive value (NPV), Cohen's kappa ( $k$ ), and agreement with existing diagnostic instruments were computed. Cronbach's  $\alpha$  values were also calculated.

## Results

All five modules of the K-ADOS-2 showed sufficient ranges of sensitivity 98.2–100.0%; specificity, 83.1–93.5%; area under the ROC curve, .91-.97; PPV, 88.1–97.6%; NPV, 87.3–100.0%; and  $k$  values, .83–.92. The kappa agreement with existing diagnostic instruments ranged from .20 to .90. Cronbach's  $\alpha$  values ranged from .83 to .90 across all modules.

## Limitation

: The best-estimate clinical diagnoses made in this study were not independent of the K-ADOS-2 scores. Some modules did not include balanced numbers of participants in terms of gender and diagnostic status.

## Conclusion

The K-ADOS-2 is a valid and reliable instrument in diagnosing autism in South Korea. Future studies exploring the effectiveness of the K-ADOS-2 in capturing restricted, repetitive behaviors and differentiating autism from other developmental disabilities are needed.

## Background

Autism is a neurodevelopmental disorder characterized by social communication difficulties and the presence of repetitive and stereotyped behaviors, interests, and activities [1]. Due to the heterogeneity in symptom presentation of autism, the clinical diagnosis is most valid and reliable when made on comprehensive, gold-standard diagnostic instruments [2, 3] such as *Autism Diagnostic Observation Schedule (ADOS; [4])* or *ADOS-2 [5]* and *Autism Diagnostic Interview-Revised (ADI-R; [6]) [7, 8]*.

The ADOS-2 is a semi-structured, standardized observational instrument designed to assess and diagnose autism across all ages. Initially developed in 2000 [9], the ADOS has been updated into the ADOS-2 to improve the accuracy and versatility of the assessment. It revised diagnostic algorithms, amended protocols of administration, included the additional module for toddlers between 12-30 months, and created new criteria for comparison scores, which allow the examination of autism symptom severity across different modules [10]. The ADOS-2 classification of autism requires an individual's score to meet or exceed the algorithm threshold for the two domains: social affect and restricted, repetitive behavior (RRB), with higher scores reflecting greater symptom severity. The schedule consists of five developmentally sequenced modules, each of which has a different combination of activities based on developmental age and expressive language skills. The modules range from the Toddler Module, for children aged younger than 30 months, to Module 4, for verbally fluent older adolescents and adults. The diagnostic algorithm for the Toddler Module, Modules 1, and 2 are further subdivided into developmental cells based on age/language. The algorithm for the Toddler Module is divided into two developmental cells: 12–20 months/nonverbal 21–30 months toddlers (12-20/NV21-30) and 21–30 months toddlers with some words (21-30SW). The algorithm of the Module 1 is divided into two developmental cells based on expressive language level: no words (NW) and some words (SW). The algorithm of the Module 2 is divided into two developmental cells based on age groups: < 5 years and  $\geq$  5 years.

The ADOS-2 has become more internationally accessible, driven by increased autism awareness as well as the efforts to administer the ADOS-2 in different countries [11]. Currently, the ADOS-2 has been translated into more than 20 languages [12], and the clinical validity of the ADOS-2 has been well-established in various international samples [13]. Previous studies have underscored that autism diagnostic instruments developed in Western countries can be properly translated and adapted in non-Western countries [11, 14-16]. Adapting the diagnostic tools that were originally developed for different cultures requires a re-examination of reliability and validity [7, 17]. Since culture influences the language, play materials, and social norms concerning developmentally appropriate behaviors, it can consequently affect how people in specific cultural contexts evaluate the appropriateness and severity of autistic symptoms [18, 19]. However, the majority of studies investigating the validity of translated versions of the ADOS-2 have been conducted in Western, English-speaking

countries, such as the United States [20], Canada [21], and the United Kingdom [22]. Only a handful of studies have examined the validity of the ADOS-2 in non-Western populations (e.g., in Chinese [14], Indian [15], and South Korean [11, 16]).

The Korean versions of the ADOS/ADOS-2 (i.e., K-ADOS/K-ADOS-2) have been used in South Korea for more than a decade [23]. To date, only two studies have partially validated the K-ADOS/K-ADOS-2 in South Korea. Kim et al. [16] conducted a study including 292 school students (aged 7-14 years old) to show that Module 3 of the K-ADOS had sufficient specificity and sensitivity. More recently, after the ADOS-2 was translated into Korean [24], Lee et al. [11] evaluated the validity of the Toddler Module and Modules 1 and 2 of the K-ADOS-2 on 143 South Korean toddlers and preschoolers. They found that the modules had adequate sensitivity, specificity, and internal consistency with respect to age. However, these previous studies were limited by their small sample size and relatively narrow age range of participants; therefore, research on the use and applicability of the K-ADOS-2, particularly on older children, adolescents, and adults, is still limited.

Further, researchers have emphasized the importance of establishing diagnostic utility of diagnostic instruments in differentiating autism from other disabilities because autism is often accompanied by and shows behavioral overlap with many neurodevelopmental and behavioral disorders (e.g., intellectual disabilities, anxiety disorders, and attention-deficit/hyperactivity disorders) [25-28], complicating the diagnostic process. Lee et al. [11] showed that the sensitivity and specificity of the K-ADOS-2 in distinguishing children and toddlers with autism from those without autism but have other developmental delays or language delays ranged from 94%–100% and 82%–100%, respectively. Yet, no other studies have examined the clinical validity of the K-ADOS-2 in distinguishing autism from other developmental disabilities (OD), notably in Modules 3 and 4. It is particularly important to examine the diagnostic accuracy of the K-ADOS-2 in differentiating autism from OD in adolescent and adult populations because the diagnostic process is considered more complicated due to increased comorbidities [29, 30]. Indeed, Langmann et al. [27] reported that the diagnostic validity of the ADOS Module 4 in distinguishing autism from other clinical samples (e.g. personality disorders, behavioral and emotional disorders, anxiety and/or compulsive disorders) was low for older adults and individuals with high IQ and verbal ability, suggesting the need for further research.

Therefore, the purpose of this study was to expand on previous findings [11, 16], examine the psychometric properties and establish the diagnostic validity of the K-ADOS-2 across all modules (i.e., Toddler Module and Module 1–4) with a larger number of participants. Specifically, we aimed to investigate 1) the diagnostic validity of all modules of the K-ADOS-2 algorithms, 2) its agreement with existing autism diagnostic instruments, and 3) the reliability of all modules of the K-ADOS-2 to examine whether it can be applied to the South Korean population across all ages. Additionally, we preliminarily explored if the K-ADOS-2 could be used to differentiate autism from OD.

## Methods

### Participants

This study is a secondary analysis of pooled data from projects aimed at identifying an autism biomarker, training social skills in adolescents with autism and developing an early autism screening instrument based on clinical diagnostic evaluations conducted at (Redacted for Review) from 2008 to 2017. Participants were recruited via patient referrals from psychiatric clinics and the Pediatrics and Child Rehabilitation clinic at (Redacted for Review), local clinics and daycare centers, recruitment posters on online/offline bulletin boards of public institutions, and online parenting communities. Participants in each project consented to the retrospective use of their collected data. A total of 2,177 participants (mean age [standard deviation] = 79.2 [64.0] months; age range = 9–393 months; 1600 males; 1480 participants with autism and 697 participants without autism) were included in this study (Module-T, n = 296; Module-1, n = 654; Module-2 n = 574; Module-3 n = 411; Module-4, n = 233). Further, participants who were diagnosed as not having autism based on clinical best-estimate diagnosis and obtained scores lower than 80 in either the full-scale intelligence quotients (FSIQ) or Korean Vineland Adaptive Behavior Scales, Second Edition (K-VABS; [31]) were categorized as OD. Diagnostic procedures are presented in the Procedures section below. Detailed characteristics of the total participants and participants by module are included in Tables 1 and 2. Information on participant characteristics for each developmental cell of the Toddler Module, Module 1, and Module 2 are available in Supplementary Table S1. Detailed characteristics of the OD participants are available in Supplementary Table S2.

[INSERT TABLE 1 and 2 HERE]

### Procedures

Participants and their parents completed a battery of tests during their one-time visit, including the K-ADOS or K-ADOS-2, ADI-R, the Korean version of Childhood Autism Rating Scale (K-CARS), K-SMS, and cognitive tests measuring FSIQs. Questionnaires, such as the Social Responsiveness Scale-2 (SRS-2), Social Communication Questionnaire (SCQ), the Korean version of Childhood Autism Rating Scale (K-CARS) and K-VABS, were mailed and filled out prior to the visit. The K-ADOS or K-ADOS-2 and ADI-R were administered by research-reliable professionals or research assistants who worked alongside them in the same lab on a daily basis and were trained prior to the actual administration. The scales were administered only after an adequate level of inter-reliability with the research-reliable professionals (>80%) was reached. All administrations of the K-ADOS or K-ADOS-2 and ADI-R were video-taped and double-checked by these professionals to confirm the quality and reliability.

Subsequently, two board-certified psychiatrists made the best-estimate clinical diagnostic criteria for autism (including pervasive developmental disorder and Asperger's syndrome) and non-autism based on DSM-IV™ [32] and DSM-V [1]. The clinical best-estimate diagnosis was made according to the information gathered collectively from all tests administered, including the K-ADOS/K-ADOS-2, ADI-R, SCQ, SRS-2, K-CARS, SMS, VABS, and IQ assessments, as well as observed clinical impressions. The study was approved by the Institutional Review Board (IRB) of (Redacted for Review) (IRB no. B-2110-716-102).

### Measures

*Autism Diagnostic Observation Schedule and Autism Diagnostic Observation Schedule-2 (ADOS and ADOS-2; [4, 5])*

This study used the Korean translated versions of the ADOS/ADOS-2, approved by its publisher Western Psychological Services. Data collected prior to July 2017, when the ADOS-2 was published in Korea, was administered using the original K-ADOS. The results from the K-ADOS were rescored based on the K-ADOS-2 algorithm for this study. All modules provide two cut-off points in the diagnostic algorithms. For Modules 1 through 4, there is a higher cut-off in the diagnostic algorithms for stringent classification (i.e., *autism*) and a lower cut-off in the diagnostic algorithms for more inclusive classification (that is, autism spectrum disorder; *ASD*). For Module 4, we applied the revised algorithm and cut-off points from Hus and Lord [33].

Similarly, the Toddler Module has a higher cut-off in the diagnostic algorithms for stringent classification (*moderate-severe concern*) and a lower cut-off in the diagnostic algorithms for more inclusive classification (*mild-moderate concern*). Meeting the *ASD cut-off* in Modules 1 through 4 and *mild-moderate concern* in the Toddler Module were categorized as autism in the K-ADOS-2 diagnosis.

#### ***Autism Diagnostic Interview-Revised (ADI-R; [6])***

The ADI-R is a semi-structured caregivers' interview used to diagnose or evaluate the core symptoms of autism. Each item is scored and converted on a scale of 0, 1, and 2, with higher scores indicating more severe autism-related symptoms. ADI-R includes 93 items describing four diagnostic domains: social interaction, communication, RRBs, and abnormality of development evident at or before 36 months. Each domain has a diagnostic criterion, but individuals must exceed all four cut-off scores to be classified as autism. While the majority of the algorithm score consists of parents' descriptions of a child's behaviors between the ages of 4–5 years, some items ask whether the behavior has ever been present during the child's lifetime. For children under 4 years of age, ratings on current behaviors are used. The Korean translation of the ADI-R [23], approved by its publisher Western Psychological Services, was used in this study.

#### ***Social Communication Questionnaire [34]***

The SCQ is a caregiver-report screening instrument for autism designed to evaluate an individual's behavior in three domains: social interaction, language and communication, and RRB. SCQ includes 40 items to be rated as either "yes" or "no." It consists of two forms: the Lifetime Form, which focuses on an individual's developmental history, and the Current Form, which inspects an individual's behaviors over the past three months. The total score in the Lifetime Form is used to determine if an individual is likely to have autism, and whether a more extended diagnostic evaluation needs to be undertaken. In this study, we used a cutoff score of 10, for children under 47 months of age, and 12, for children over 48 months, based on a standardization study conducted in Korea [35].

#### ***Social Responsiveness Scale-2 (SRS-2; [36])***

The SRS-2 is a 65-item parent report questionnaire that assesses the severity of autism-related symptoms on a 4-point scale, with higher total scores reflecting more severe autism symptomatology. It consists of five subscales: social awareness, social cognition, social communication, social motivation, and autistic mannerisms. The SRS-2 has been used extensively in the autism literature as a diagnostic measure [37] and is reported to have good internal consistency and concurrent, discriminant validity [38]. Chun et al. [39] demonstrated adequate levels of sensitivity and specificity of the Korean translated version of the SRS-2. A cut-off T-score of 65 was applied regardless of gender in the preschool form of the SRS-2, and cut-off T-scores of 70 and 63 were used for female and male participants, respectively, for the school-age and adults SRS-2 versions because these values are widely used across clinical settings in South Korea.

#### ***Korean version of the Childhood Autism Rating Scale (K-CARS; [40])***

The CARS [41] is a clinician-rated scale developed to screen for autism. Consisting of 15 items rating the presence and severity of symptoms associated with autism, the CARS is scored from 1 (no impairment observed or reported) to 4 (severe impairment). There is no consensus on the cut-off score of the K-CARS; Shin and Kim [42] suggested a cut-off score of 28, while others recommend 24 [43]. Therefore, we utilized both cut-off scores in this study.

#### ***Full Scale Intelligence Quotients (FSIQ)***

The following instruments were used to calculate FSIQ in this study: the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) [44] for children aged 2 years and 6 months to 6 years, Wechsler Intelligence Scale for Children (WISC) [45] for children aged 6–16 years, and Wechsler Adult Intelligence Scale (WAIS) [46] for individuals over 16 years of age. These instruments utilize chronological age standardization with a mean of 100 and a standard deviation of 15.

#### ***Korean version of the Vineland Adaptive Behavior Scale, second edition (K-VABS; [31, 47])***

The VABS is a parent or other caregiver's ratings of a person's adaptive functioning and social self-sufficiency from birth to adulthood. The VABS consists of five domains: communication, daily living skills, socialization, motor skills, and maladaptive behavior. It is scored on a 0–2 rating scale, with a higher score representing skills used more frequently. The five domains together yield a total adaptive behavior composite score. The normative mean of the composite score is 100, with standard deviation of 15.

#### ***Korean Vineland Social Maturity Scale (K-SMS; [48])***

The K-SMS is a clinician-rated instrument that assesses social and adaptive maturity. Originally developed using the Doll's Vineland Social Maturity Scale [49], the K-SMS includes 89 items grouped by behavioral milestones that are expected at each age. It consists of eight subdomains (communication, general self-help, locomotion, occupation, self-direction, self-help eating, self-help dressing, and socialization skills) and provides a global social age and social quotient.

## **Statistical Analyses**

Initially, the scores from the K-ADOS-2, ADI-R, K-CARS, SCQ, SRS-2, mean age, and FSIQ of participants with autism and those without autism were compared using a set of independent sample t-tests. Calibrated severity scores were used to compare the K-ADOS-2 scores.

To address the first aim, the sensitivity, specificity, PPV, NPV, and Cohen's kappa ( $k$ ) were calculated to check for consistency between the best-estimate clinical diagnosis and diagnosis based on *ASD cut-off* in K-ADOS-2 Modules 1–4 and *mild-moderate concern* in Toddler Module. This analysis was conducted on all modules combined, each module (including Module-T, 1, 2, 3, and 4) individually, and each developmental cell (12-20/NV21-30 and 21-30SW in Toddler Module, NW and SW in Module 1, and under and over 5 years of age in Module 2). We also computed the area under the receiver operating characteristic (ROC) curve of all items by developmental cell to explore if all items included in the algorithm have sufficient diagnostic accuracy according to the area under the curve (AUC).

To investigate the second aim, we computed Pearson's  $r$  correlation coefficients between the total scores of K-ADOS-2 and those of existing autism diagnostic instruments (i.e., ADI-R, K-CARS, SCQ, and SRS-2) for all modules combined, each module individually, and each developmental cell. Additionally,  $k$  values were calculated between the diagnosis based on the K-ADOS-2 ASD cutoff and the diagnosis based on the existing autism diagnostic instruments. The  $k$  values were interpreted based on McHugh's [50] criteria (0–.2, none; .21–.39, minimal; .40–.59, weak; .60–.79, moderate; .8–.9, strong; above .9, almost perfect). For the third aim, Cronbach's  $\alpha$  values for the algorithm items and values after an item was removed were computed to examine the internal consistency of each developmental cell and module.

Finally, we calculated the sensitivity, specificity, PPV, NPV, and  $k$  values to examine how accurately the K-ADOS-2 ASD cut-off can distinguish autism from OD for all modules combined, each module individually, and each developmental cell. We did not compare the diagnostic validity between OD and the remaining participants without autism (i.e., participants who were not diagnosed as autism and did not have FSIQ or VABS scores lower than 80) because this sample included a few participants for whom we did not have all FSIQ and VABS scores and therefore would have been categorized as OD if all relevant information was available.

All analyses except for the calculation of Cronbach's  $\alpha$  values were repeated using the *Autism cut-off* in Modules 1–4 and *moderate-severe concern* in the Toddler Module, but we primarily relied on the results from the *ASD cut-off* to make the decisions regarding validity. All statistical analyses were performed using Excel and SPSS Statistics (version 23.0; IBM Corp., Armonk, NY, USA).

## Results

There were statistically significant inter-group differences (i.e., autism vs. non-autism) in all algorithm scores of the K-ADOS-2, ADI-R, K-CARS, SCQ, and SRS-2 ( $p < 0.001$ ) in the composite K-ADOS-2 and across all modules and developmental cells, except for in the ADI-R RRB domain in the 21-30W developmental cell group (Table 1, Table 2, and Table S1).

All five modules had high sensitivity 98.2–100.0%; specificity, 83.1–93.5%; AUC, .91–.97; PPV, 88.1–97.6%; and NPV, 87.3–100.0%.<sup>[1]</sup> The agreement between all modules of K-ADOS-2 and clinical diagnosis was also at strong to almost perfect levels (Toddler Module,  $k = .90$ ; Module-1,  $k = .83$ ; Module-2,  $k = .92$ ; Module-3,  $k = .88$ ; Module-4,  $k = .90$ ). Detailed results of the sensitivity, specificity, AUC, PPV, NPV, and  $k$  values by module and developmental cell are presented in Table 3. The AUC values of the majority of algorithm items in each developmental cell exceeded .7 (range = .70~.93). The list of algorithm items with AUC values lower than .7 is presented in the Supplementary Table S3 by developmental cell. Across all developmental cells, the AUCs of *Hand Finger and Other Complex Mechanism item* was consistently lower than .7, and all items with AUC lower than .7 were from the RRB algorithm.

[INSERT TABLES 3 AND 4 HERE]

The total scores of the K-ADOS-2 were significantly and positively correlated with those of ADI-R, SCQ, SRS-2, and K-CARS scores across all modules and developmental cells. Pearson's  $r$  correlations ranged between .60 - .75 for Toddler Module, .54 - .80 for Module 1, .64 - .87 for Module 2, .52 - .82 for Module 3, and .47 - .84 for Module 4. The kappa agreements between all K-ADOS-2 modules and existing diagnostic instruments ranged between .47 - .89 for Toddler Module, .41 - .68 for Module 1, .51 - .77 for Module 2, .33 - .61 for Module 3, and .25 - .77 for Module 4, suggesting weak to strong agreement. Detailed results of Pearson's correlations and kappa values with existing diagnostic instruments by module and developmental cell are presented in Table 4.

All modules and developmental cells had high internal consistencies, with  $\alpha$  values ranging from .83 to .90. Removing an item inflicted no to minimal changes (that is, a change of less than .02 change in  $\alpha$  values). The complete results of the reliability analysis are presented in Table 5.

[INSERT TABLE 5 HERE]

When using the *ASD cut-off* to distinguish OD from autism, all modules and developmental cells of the K-ADOS-2 also had sufficient sensitivity, specificity, AUC, PPV, and NPV except for NPV in Module 1. Sensitivity across the modules ranged from 98.2–100.0%; specificity, 71.8–92.3%; AUC, .83–.96; PPV, 97.6–99.4%; and NPV, 46.7–100% (Supplementary Table S4). The  $k$  values ranged between .70–.85, suggesting moderate to almost perfect agreement based on McHugh's [50] criteria.

The sensitivity, specificity, AUC, PPV, NPV, and  $k$  values, as well as the agreement with existing diagnostic instruments based on the *Autism cut-off* score are available in Supplementary Tables S4 and S5, respectively. Supplementary Table S7 presents the specificity, sensitivity, PPV, NPV, and  $k$  values between autism and OD calculated based on the *Autism cut-off*.

**Footnote:**

<sup>[1]</sup> We conducted a set of sensitivity tests excluding the participants categorized as OD, and the changes in values of sensitivity, specificity, AUC, PPV, NPV, kappa, and  $\alpha$  were minimal (i.e., all changes of sensitivity, specificity, PPV, and NPV were in the tenth digits, and all changes of AUC,  $k$ , and  $\alpha$  were in hundredths).

## Discussion

This study showed that the K-ADOS-2 has excellent diagnostic validity in distinguishing individuals with autism from those without autism with sufficiently high sensitivity, specificity, AUC, PPV, and NPV across a wide age group (24 months to 34 years). Moreover, all modules and developmental cells of the K-ADOS-2 demonstrated sufficient reliability. These findings provide additional evidence that the K-ADOS-2 can be adapted for various cultural settings [7, 11, 14, 51]. This suggests that although there can be cultural differences in the interpretation of severity and appropriateness of autistic behaviors (18, 19), the behavior patterns that need to be considered when diagnosing autism may not differ across cultures.

Further, compared to previous adaptation studies conducted in different countries such as the Netherlands [51] and Poland [7], the K-ADOS-2 exhibited higher sensitivity and specificity values. As Lee et al. [11] also highlight the importance of highly trained, research-reliable clinicians in establishing strong validity and specificity for a measure, we postulate that this positive result may be due to the strict, reliable administration and coding process in which research professionals double-checked all K-ADOS-2 administration.

The examination of the AUC under each item indicated that all algorithm items in the social affect domain have an acceptable ability to distinguish autism from non-autism. Meanwhile, the *Hand Finger and Other Complex Mechanism item* showed consistently low AUC across all developmental cells, and items with AUC lower than .7 were all from the RRB algorithm. Similarly, previous studies have also suggested that social communicational items tend to distinguish the individuals with autism from those without autism more accurately than the RRB items [52–54]. Given the brevity of the time allotted for the observation during the K-ADOS-2 and the variability of frequency and types of RRBs depending on the observational contexts (i.e., clinic vs. home) [55], it is possible that clinicians are not offered a sufficient opportunity to observe these types of RRBs during the K-ADOS-2. We therefore suggest the importance of complementing the results of K-ADOS-2 with other diagnostic instruments such as the ADI-R that rely on a more long-term observation by parents or teachers, particularly when assessing the RRBs.

The K-ADOS-2 scores were correlated with the ADI-R, SRS-2, SCQ, and K-CARS scores across all developmental cells and modules, suggesting sufficient concurrent validity. Interestingly, the Pearson's  $r$  coefficients between K-ADOS-2 and ADI-R and K-CARS tend to be greater than those between K-ADOS-2 and SCQ and SRS-2. This pattern could possibly be explained by the inherent shortcomings of parent-report questionnaires (i.e., the SRS-2 and SCQ). Caregivers may respond to the questions based on their interpretations without an accurate understanding of the concepts captured in each question [56]. Specifically, some parents may have difficulties reporting their child's social interaction abilities because young children's social dynamics tend to be passive or exclusively with adults [57]. Caregivers' beliefs, characteristics, acceptance, and awareness of autism may influence how they interpret their child's behaviors [58, 59].

It is noteworthy that kappa agreement between diagnoses made by the K-ADOS-2 and SCQ, SRS-2, and K-CARS were weak in some modules and developmental cells. In particular, the kappa agreements between K-ADOS-2 and K-CARS were minimal in Modules 3 and 4. However, considering that the Pearson's correlations between them were strong and significant, we postulate that this discrepancy may signal the need for more studies adjusting cutoff scores on the K-CARS in the Korean population, especially for the verbally fluent children, adolescents, and adults with autism. Indeed, due to the lack of consistency in the K-CARS cut-off score used in Korea, we applied two cutoff scores (i.e., [19, 34]) [42, 43] to calculate agreement with K-ADOS-2 scores. Further, we recommend that the K-ADOS-2 scores should be interpreted in conjunction with the results from other diagnostic instruments to make an accurate diagnosis.

Similar to the findings from the previous K-ADOS-2 adaptation study of the Toddler Module and Modules 1, and 2 [11], applying an autism (i.e., higher) cutoff lowered the sensitivity and specificity compared to using an ASD (i.e., lower) cutoff. However, previous validation studies of the ADOS-2 conducted in Western countries such as the US [60] and Sweden [61] have reported more balanced specificities and sensitivities when applying *an autism cutoff*, suggesting that sample variability may impact the diagnostic validity of the ADOS-2 [11].

In our preliminary examination of the K-ADOS-2's validity in differentiating autism from OD, we found promising results that sensitivity, specificity, AUC, PPV, and Cohen's kappa were satisfactory for all developmental cells. However, most modules and developmental cells, particularly the NW algorithm of Module 1, yielded low NPV values. This suggests that children with developmental difficulties, especially those who do not use words to communicate, should be examined with additional diagnostic instruments even if the K-ADOS-2 identifies them as non-autistic. Notably, however, it is unclear if some of the participants categorized as having OD in this study had a formal developmental disability diagnosis. Different patterns could have emerged if we had included individuals with a confirmed diagnosis of non-autistic developmental disabilities (e.g., intellectual disabilities) as a separate clinical control group, and future studies should investigate this possibility.

## Limitations

This study has several limitations, which suggest promising avenues for future studies. First, while the best-estimate clinical diagnosis is based on the combination of direct observation, caregiver report and other psychological assessments, the final diagnosis was not independent of the K-ADOS-2 scores. To establish its validity more accurately, we suggest that separate institutions independently implement the standard diagnostic procedures (which may or may not also include K-ADOS-2). Second, the ratios of autism-to-non-autism and male-to-female participants were unbalanced for some modules and developmental cells in our sample. For instance, in Module 4, 93.2% of the 192 participants with autism were male, while 38.0% of the 50 participants without autism were male. We recommend that future iterations of the study recruit a balanced number of participants in terms of diagnostic status and gender. Moreover, as more studies are reporting sex differences in symptom presentation (e.g., fewer RRBs in female individuals), which may be contributing to sex

biases in diagnostic tools and practices [62], future studies should examine if there are sex differences in the validity of and symptom presentations captured by the K-ADOS-2.

Third, the discriminant validity of the instrument (autism vs. OD) should be interpreted with caution because of the small number of participants with OD in each developmental cell and module. Fourth, due to missing FSIQ and VABS data, some participants without autism who may have been categorized as the OD were not categorized as such. Finally, we did not conduct additional or follow-up assessments to verify whether participants categorized as having OD actually have some other developmental disabilities. Future studies should utilize a larger and more balanced sample including participants with a confirmed diagnosis in each diagnostic group to confirm the validity of the K-ADOS-2 in differentiating autism from OD.

## Conclusions

This study demonstrates evidence of the K-ADOS-2's validity and reliability in diagnosing autism by virtue of its high sensitivity, specificity, AUC, PPV, and NPV, k value, Cronbach's alpha as well as its moderate agreement with existing autism diagnostic instruments. To our knowledge, this study has the largest sample size among the studies conducted in Korea and is the first to examine the validity and reliability of all modules and developmental cells of the K-ADOS-2. We recommend that future studies should compare K-ADOS-2 scores with best-estimate clinical diagnoses made using independent administration of standard diagnostic procedures, as well as include balanced numbers of participants in terms of gender and diagnostic status. Further, we suggest the need for studies recruiting larger total samples and participants with formal diagnoses of developmental disabilities.

## Declarations

### Ethics Approval and consent to participate

The study procedure including informed consent, recruitment, and participation procedures was approved by the Institutional Review Board (IRB) of Seoul National University Bundang Hospital (IRB no. B-2110-716-102).

### Consent for publication

We did not include any individual person's data in any form.

### Data Accessibility Statement

The datasets generated and/or analyzed during the current study are not publicly available because we do not have the permission from the IRB to share or make the unidentified participant information available online and did not receive the consent from the participants but are available from the corresponding author on reasonable request.

### Competing Interests

HJY receives royalties from Hakjisa for sales of the Korean versions of the Autism Diagnostic Observation Scale-2, Autism Diagnostic Interview-Revised, and Social Communication Questionnaire.

### Funding

This work was supported by the Original Technology Research Program for Brain Science of the NRF, funded by the Korean government, MSIT (NRF-2017M3C7A1027467) and MIST (NRF-2021M3E5D9021878).

### Authors' contribution

Conceptualization: HJY

Data curation: GB, NHY, JHK, DS

Formal analysis: SYK, MO

Funding acquisition: HJY

Investigation: SYK, MO

Methodology: SYK, MO, HJY

Project administration: GB, JHK

Supervision: HJY

Writing: SYK, MO, DS

**All authors** approved the final version of the submitted manuscript.

### Acknowledgments

## References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). Washington, DC: Author; 2013.
2. Guthrie W, Swineford LB, Nottke C, Wetherby AM. Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *J Child Psychol Psychiatry*. 2013;54(5):582–90.
3. Kim SH, Lord C. Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *J Child Psychol Psychiatry*. 2012;53(2):143–51.
4. Lord C, Rutter M, DiLavore PC, Risi S. Autism Diagnostic Observation Schedule (ADOS). Los Angeles: Western Psychological Services; 1999.
5. Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL. Autism diagnostic observation schedule, (ADOS-2) modules 1-4. Los Angeles: Western Psychological Services; 2012.
6. Rutter M, Le Couteur A, Lord C. Autism diagnostic interview-revised (ADI-R). Los Angeles: Western Psychological Services; 2003b.
7. Chojnicka I, Pisula E. Adaptation and Validation of the ADOS-2, Polish Version. *Front Psychol*. 2017;8:1916.
8. Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *Eur Child Adolesc Psychiatry*. 2013;22(6):329–40.
9. Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*. 2000;30(3):205–23.
10. Dorlack TP, Myers OB, Kodituwakku PW. A Comparative Analysis of the ADOS-G and ADOS-2 Algorithms: Preliminary Findings. *J Autism Dev Disord*. 2018;48(6):2078–89.
11. Lee KS, Chung SJ, Thomas HR, Park J, Kim SH. Exploring diagnostic validity of the autism diagnostic observation schedule-2 in South Korean toddlers and preschoolers. *Autism Res*. 2019;12(9):1356–66.
12. Western Psychological Services. Published translations 2018 [Available from: <http://www.wpspublish.com/app/OtherServices/PublishedTranslations.aspx>].
13. Hong JS, Singh V, Kalb L, Ashkar A, Landa R. Replication study of ADOS-2 Toddler Module cut-off scores for autism spectrum disorder classification. *Autism Res*. 2021;14(6):1284–95.
14. Sun X, Allison C, Auyeung B, Zhang Z, Matthews FE, Baron-Cohen S, et al. Validation of existing diagnosis of autism in mainland China using standardised diagnostic instruments. *Autism: the international journal of research practice*. 2015;19(8):1010–7.
15. Rudra A, Banerjee S, Singhal N, Barua M, Mukerji S, Chakrabarti B. Translation and usability of autism screening and diagnostic tools for autism spectrum conditions in India. *Autism Res*. 2014;7(5):598–607.
16. Kim SH, Kim YS, Koh Y-J, Lim E-C, Kim S-J, Leventhal BL. Often Asked but Rarely Answered: Can Asians Meet DSM-5/ICD-10 Autism Spectrum Disorder Criteria? *J Child Adolesc Psychopharmacol*. 2016;26(9):835–42.
17. Hambleton RK, Merenda PF, Spielberger CD. Adapting educational and psychological tests for cross-cultural assessments. Mahwah: Lawrence Erlbaum Associates, Inc; 2009.
18. Matson JL, Matheis M, Burns CO, Esposito G, Venuti P, Pisula E, et al. Examining cross-cultural differences in autism spectrum disorder: A multinational comparison from Greece, Italy, Japan, Poland, and the United States. *Eur Psychiatry*. 2017;42:70–6.
19. Pacifico MC, de Paula CS, Namur VS, Lowenthal R, Bosa CA, Teixeira M. Preliminary evidence of the validity process of the Autism Diagnostic Observation Schedule (ADOS): translation, cross-cultural adaptation and semantic equivalence of the Brazilian Portuguese version. *Trends Psychiatry Psychother*. 2019;41(3):218–26.
20. Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *J Autism Dev Disord*. 2007;37(4):613–27.
21. Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. 2006;45(9):1094–103.
22. Le Couteur A, Haden G, Hammal D, McConachie H. Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: the ADI-R and the ADOS. *J Autism Dev Disord*. 2008;38(2):362–72.
23. Yoo HJ, Kwak Y. Korean version of autism diagnostic observation schedule (ADOS). Seoul: Hakjisa; 2007.
24. Yoo HJ, Bong GY, Kwak YS, Lee MS, Cho SH, Kim BN. Korean Autism Diagnostic Observation Schedule-2 (K-ADOS-2). Seoul: Hakjisa; 2018.
25. Collin L, Bindra J, Raju M, Gillberg C, Minnis H. Facial emotion recognition in child psychiatry: a systematic review. *Res Dev Disabil*. 2013;34(5):1505–20.
26. Gjevik E, Eldevik S, Fjæran-Granum T, Sponheim E. Kiddie-SADS reveals high rates of DSM-IV disorders in children and adolescents with autism spectrum disorders. *J Autism Dev Disord*. 2011;41(6):761–9.
27. Langmann A, Becker J, Poustka L, Becker K, Kamp-Becker I. Diagnostic utility of the autism diagnostic observation schedule in a clinical sample of adolescents and adults. *Research in Autism Spectrum Disorders*. 2017;34:34–43.
28. Sappok T, Diefenbacher A, Budczies J, Schade C, Grubich C, Bergmann T, et al. Diagnosing autism in a clinical sample of adults with intellectual disabilities: how useful are the ADOS and the ADI-R? *Res Dev Disabil*. 2013;34(5):1642–55.
29. Lombardo MV, Baron-Cohen S. The role of the self in mindblindness in autism. *Conscious Cogn*. 2011;20(1):130–40.

30. Lombardo MV, Barnes JL, Wheelwright SJ, Baron-Cohen S. Self-referential cognition and empathy in autism. *PLoS One*. 2007;2(9):e883.
31. Volkmar F. Vineland adaptive behavior scales, second edition. New York, NY: Springer; 2013.
32. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (4th ed., Text Revision). Washington, DC: Author; 2000.
33. Hus V, Lord C. The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. *J Autism Dev Disord*. 2014;44(8):1996–2012.
34. Rutter M, Bailey A, Lord C. The social communication questionnaire: Manual. Los Angeles: Western Psychological Services; 2003a.
35. Kim JH, Sunwoo HJ, Park SB, Noh DH, Jung Y, Cho SC, et al. A validation study of the Korean version of social communication questionnaire. *Journal of the Korean Academy of Child Adolescent Psychiatry*. 2015;26:197–208.
36. Constantino JN, Gruber CP. Social Responsiveness Scale, Second Edition (SRS-2). Torrance: Western Psychological Services; 2012.
37. Bölte S, Poustka F, Constantino JN. Assessing autistic traits: cross-cultural validation of the social responsiveness scale (SRS). *Autism Res*. 2008;1(6):354–63.
38. Constantino JN, Gruber CP. Social Responsiveness Scale (SRS). Los Angeles: Western Psychological Services; 2005.
39. Chun J, Bong G, Han JH, Oh M, Yoo HJ. Validation of Social Responsiveness Scale for Korean Preschool Children With Autism. *Psychiatry Investig*. 2021;18(9):831–40.
40. Kim T, Park R. Korean version of childhood autism rating scale. Seoul: Special Education; 1995.
41. Schopler E, Reichler RJ, Renner BR. The childhood autism rating scale (CARS). Los Angeles: Western Psychological Services; 1988.
42. Shin MS, Kim YH. Standardization study for the Korean version of the Childhood Autism Rating Scale: Reliability, validity and cut-off score. *Korean Journal of Clinical Psychology*. 1998;17:1–15.
43. Kwon HJ, Yoo HJ, Kim JH, Noh DH, Sunwoo HJ, Jeon YS, et al. Re-adjusting the cut-off score of the Korean version of the Childhood Autism Rating Scale for high-functioning individuals with autism spectrum disorder. *Psychiatry Clin Neurosci*. 2017;71(10):725–32.
44. Wechsler D. Wechsler preschool and primary scale of intelligence—third edition. San Antonio: Psychological Corporation; 2002.
45. Wechsler D. Wechsler intelligence scale for children-Fourth Edition. San Antonio: Psychological Corporation; 2003.
46. Wechsler D. Wechsler adult intelligence scale-revised (WAIS-R). New York: Psychological Corporation; 1981.
47. Hwang S, Kim JH, Hong S. Korean Vineland Adaptive Behavior Scales-II. Daegu: Korea Psychology Co; 2014.
48. Kim S, Kim O. Korean Vineland Social Maturity Scale. Seoul: Chung-Ang Jeokseong Press; 2002.
49. Doll EA. The measurement of social competence: A manual for the Vineland Social Maturity Scale. Educational Test Bureau Educational Publishers; 1953.
50. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–82.
51. Oosterling I, Roos S, de Bildt A, Rommelse N, de Jonge M, Visser J, et al. Improved diagnostic validity of the ADOS revised algorithms: a replication study in an independent sample. *J Autism Dev Disord*. 2010;40(6):689–703.
52. Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: diagnostic validity. *Br J Psychiatry*. 1999;175:444–51.
53. Dow D, Day TN, Kutta TJ, Nottke C, Wetherby AM. Screening for autism spectrum disorder in a naturalistic home setting using the systematic observation of red flags (SORF) at 18-24 months. *Autism Res*. 2020;13(1):122–33.
54. Nilsson Jobs E, Bölte S, Falck-Ytter T. Preschool Staff Spot Social Communication Difficulties, But Not Restricted and Repetitive Behaviors in Young Autistic Children. *J Autism Dev Disord*. 2019;49(5):1928–36.
55. Stronach S, Wetherby AM. Examining restricted and repetitive behaviors in young children with autism spectrum disorder during two observational contexts. *Autism*. 2014;18(2):127–36.
56. Volkmar FR, Lord C, Bailey A, Schultz RT, Klin A. Autism and pervasive developmental disorders. *J Child Psychol Psychiatry*. 2004;45(1):135–70.
57. Gustafson GE, Green JA, West MJ. The infant's changing role in mother-infant games: The growth of social skills. *Infant Behavior Development*. 1979;2:301–8.
58. Kim SY, Kim YA, Song D-Y, Bong G, Kim J-M, Kim JH, et al. State and Trait Anxiety of Adolescents with Autism Spectrum Disorders. *Psychiatry Investig*. 2021;18(3):257–65.
59. Oh M, Song DY, Bong G, Yoon NH, Kim SY, Kim JH, et al. Validating the Autism Diagnostic Interview-Revised in the Korean Population. *Psychiatry Investig*. 2021;18(3):196–204.
60. Molloy CA, Murray DS, Akers R, Mitchell T, Manning-Courtney P. Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism*. 2011;15(2):143–62.
61. Zander E, Sturm H, Bölte S. The added value of the combined use of the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule: diagnostic validity in a clinical Swedish sample of toddlers and young preschoolers. *Autism*. 2015;19(2):187–99.
62. Kreiser NL, White SW. ASD in females: are we overstating the gender difference in diagnosis? *Clin Child Fam Psychol Rev*. 2014;17(1):67–84.

## Tables

Table 1 *Participant Characteristics*

	Total		<i>t</i>	
	Autism Mean (SD)	Non-Autism Mean (SD)		
Total n	1480	697		
% Male	82.3%	54.8%		
Age (months)	86.6 (65.9)	64.0 (56.8)	-7.7**	
FSIQ	84.5 (23.6)	101.6 (19.2)	9.9**	
ADOS-2 <sup>a</sup>	SA	6.5 (1.7)	1.9 (1.3)	-49.8**
	RRB	5.8 (2.1)	1.7 (1.6)	-34.8**
	Total	7.4 (1.7)	1.8 (1.1)	-61.7**
ADI-R	SI	18.9 (6.3)	3.5 (3.5)	-60.2**
	C	14.0 (5.1)	2.0 (2.5)	-49.7**
	NC	9.4 (3.5)	1.7 (2.2)	-48.4**
	RRB	5.2 (2.6)	.8 (1.2)	-42.8**
K-CARS	33.1 (5.4)	17.7 (3.3)	-58.9**	
SCQ	14.4 (7.1)	4.5 (5.2)	-25.9**	
SRS	89.6 (29.8)	36.9 (21.4)	-36.7	

*Note.* \* $p < .05$ , \*\* $p < .01$ . FSIQ, full-scale intelligence quotient; ADOS-2, Autism Diagnostic Observation Schedule-2; SA, Social Affect; RRB, Restricted Repetitive Behavior; ADI-R, Autism Diagnostic Interview-Revised; SI, Social Interaction; C, Communication Total; NC, Nonverbal communication; RRB, Restricted Repetitive; K-CARS, Korean Child Autism Rating Scale; SCQ, Social Communication Questionnaire; SRS, Social Responsiveness Scale. <sup>a</sup> = reported in calibrated severity scores.

Table 2 *Participant Characteristics by Module*

	Module T			Module 1			Module 2			Module 3			Module 4		
	Autism Mean (SD)	Non- Autism Mean (SD)	<i>t</i>	Autism Mean (SD)	Non- Autism Mean (SD)	<i>t</i>	Autism Mean (SD)	Non- Autism Mean (SD)	<i>t</i>	Autism Mean (SD)	Non- Autism Mean (SD)	<i>t</i>	Autism Mean (SD)	Non- Autism Mean (SD)	
n	96	200		571	83		183	75		233	178		192	50	
% Male (%)	70.8	56.5		80.2	63.9		82.0	60.0		86.3	47.2		93.2	38	
Age (months)	24.9 (4.4)	21.2 (5.5)	-5.8**	52.1 (22.0)	37.3 (19.0)	-5.8**	66.3 (30.2)	43.8 (11.8)	-9.8**	116.5 (32.9)	105.8 (33.8)	-3.2**	223.0 (53.8)	20	
FSIQ	NA	NA	NA	60.4 (17.5)	97 (28.6)	3.4**	72.1 (21.4)	93.4 (18.8)	5.6**	89.2 (21.3)	103.8 (18.0)	6.8**	97.5 (18.4)	10	
ADOS-2 <sup>a</sup>	SA	7.8 (2.0)	2.4 (1.2)	-18.0**	6.1 (1.4)	1.9 (1.5)	17.4**	6.3 (1.3)	1.6 (.9)	28.9**	6.7 (1.9)	1.48 (1.0)	29.0**	8.9 (1.1)	2.1 (1)
	RRB	5.1 (2.6)	2.9 (2.3)	-4.6**	5.9 (2.1)	2.0 (1.9)	11.5**	5.9 (2.0)	2.0 (1.8)	14.3**	5.7 (2.6)	1.2 (.9)	20.1**	5.7 (1.8)	1.1 (1)
	Total	7.5 (2.0)	2.1 (.9)	-20.9**	7.1 (1.7)	2.1 (1.51)	17.7**	7.2 (1.5)	1.7 (1.1)	28.5**	7.5 (1.7)	1.5 (1.1)	35.6**	1.8 (1.2)	8.1 (1)
ADI-R <sup>a</sup>	SI	15.1 (5.0)	4.1 (3.2)	-22.7**	19.4 (6.3)	5.5 (4.3)	-19.6**	18.2 (6.0)	2.7 (2.7)	-33.6**	19.1 (6.5)	2.9 (3.6)	-29.6**	20.1 (6.5)	1.2 (2)
	C	6.2 (7.9)	1.3 (1.7)	-4.0**	12.3 (5.1)	3.7 (3.1)	-11.2**	14.8 (4.4)	1.8 (1.9)	-37.8**	14.4 (5.2)	2.0 (2.8)	-28.3**	15.2 (5.1)	1.2 (2)
	NC	9.5 (2.8)	2.3 (2.4)	-22.3**	10.5 (3.1)	2.6 (2.5)	-20.8**	8.8 (3.5)	1.0 (1.4)	-25.7**	8.0 (3.7)	1.3 (2.1)	-18.6**	9.0 (3.7)	1.2 (2)
	RRB	3.5 (2.2)	.9 (1.1)	-13.7**	4.8 (2.2)	1.3 (1.5)	-14.2**	5.8 (2.8)	.9 (1.3)	-22.7**	5.8 (2.7)	.5 (.9)	-24.7**	5.3 (2.6)	.6 (1)
K-CARS	31.6 (5.0)	17.4 (3.4)	-28.3**	34.4 (6.5)	18.9 (5.3)	-21.5**	32.7 (5.0)	17.3 (2.8)	-33.3**	31.7 (5.0)	18.2 (3.4)	-23.6**	32.1 (4.0)	1.2 (2)	
SCQ	15.2 (4.5)	8.0 (5.3)	-7.9**	17.0 (6.5)	6.0 (5.3)	-10.1**	13.7 (7.0)	3.1 (3.6)	-14.8**	11.5 (7.3)	3.4 (5.3)	-11.6**	12.9 (7.2)	3.4 (4)	
SRS	64.5 (16.3)	48.3 (8.1)	-10.4**	89.3 (30.4)	46.0 (19.5)	-11.0**	82.8 (28.8)	35.3 (19.4)	-19.6**	89.1 (30.3)	24.8 (22.4)	-22.6**	93.7 (28.9)	3.4 (3)	

Note. \* $p < .05$ , \*\* $p < .01$ . FSIQ, full-scale intelligence quotient; ADOS-2, Autism Diagnostic Observation Schedule-2; SA, Social Affect; RRB, Restricted Repetitive Behavior; ADI-R, Autism Diagnostic Interview-Revised; SI, Social Interaction; C, Communication Total; NC, Nonverbal Communication; RRB, Restricted Repetitive; K-CARS, Korean Child Autism Rating Scale; SCQ, Social Communication Questionnaire; SRS, Social Responsiveness Scale. <sup>a</sup> = reported in calibrated severity scores

Table 3 Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Cohen's Kappa value Based on ASD Cut-off Criteria

	n	Sensitivity (%)	Specificity (%)	AUC	PPV (%)	NPV (%)	Kappa (p-value)
Total Modules	2177	98.9	89.7	.94	95.3	97.5	.91 ( <i>p</i> <.001)
Toddler Module	296	100.0	93.5	.97	88.1	100.0	.90 ( <i>p</i> <.001)
Toddler Module (12-20/NV21-30)	213	100.0	93.1	.97	90.1	100.0	.91 ( <i>p</i> <.001)
Toddler Module (21-30 SW)	83	100.0	94.2	.97	77.8	100	.85 ( <i>p</i> <.001)
Module 1	654	98.2	83.1	.91	97.6	87.3	.83 ( <i>p</i> <.001)
Module 1 (NW)	320	98.0	92.6	.95	99.3	80.6	.92 ( <i>p</i> <.001)
Module 1 (SW)	334	98.6	78.6	.89	95.8	91.7	.87 ( <i>p</i> <.001)
Module 2	574	99.2	91.4	.95	96.0	98.3	.92 ( <i>p</i> <.001)
Module 2 (< 5 yo)	350	98.9	92.7	.96	93.9	98.7	.92 ( <i>p</i> <.001)
Module 2 (≥ 5 yo)	224	99.5	81.8	.91	98.0	94.7	.87 ( <i>p</i> <.001)
Module 3	411	99.6	87.1	.93	91.0	99.4	.88 ( <i>p</i> <.001)
Module 4	233	99.0	88.0	.94	96.9	95.7	.90 ( <i>p</i> <.001)

*Note.* AUC, area under curve; PPV, Positive Predictive Value; NPV, Negative Predictive Value; 12-20/NV21-30, 12-20 months toddlers/nonverbal 21-30 months toddlers; 21-30 SW, 21-30 months toddlers with some words; NW, no words; SW, some words; yo, years old.

Table 4 Agreement with Existing Instrument Based on ASD Cut-off Criteria

	Pearson's r (Kappa values)				
	ADI-R	SCQ	SRS	K-CARS (cut-off 24)	K-CARS (cut-off 28)
Total Modules	.77** (.73**)	.65** (.51**)	.62** (.60**)	.86** (.69**)	. (.58**)
Toddler Module	.75** (.67**)	.62** (.47**)	.60** (.50**)	.90** (.89**)	. (.79**)
Toddler Module (12-20/NV21-30)	.89** (.68**)	.62** (.43**)	.58** (.48**)	.89** (.88**)	. (.77**)
Toddler Module (21-30SW)	.85** (.60**)	.57** (.47**)	.59** (.89**)	.90** (.90**)	. (.81**)
Module 1	.70** (.53**)	.54** (.41**)	.53** (.43**)	.80** (.68**)	. (.51**)
Module 1 (NW)	.52** (.45**)	.44** (.42**)	.47** (.34**)	.80** (.80**)	. (.63**)
Module 1 (SW)	.78** (.58**)	.65** (.39**)	.55** (.50**)	.77** (.60**)	. (.43**)
Module 2	.82** (.77**)	.66** (.51**)	.64** (.58**)	.87** (.69**)	. (.58**)
Module 2 (< 5 yo)	.85** (.78**)	.66** (.54**)	.66** (.61**)	.88** (.69**)	. (.59**)
Module 2 (≥ 5 yo)	.66** (.49**)	.55** (.28**)	.58** (.31**)	.68** (.37**)	. (.20*)
Module 3	.78** (.61**)	.64** (.50**)	.52** (.61**)	.82** (.44**)	. (.33**)
Module 4	.74** (.77**)	.59** (.44**)	.47** (.55**)	.84** (.28**)	. (.25**)

Note. \*\*p-value <1.001. \*p-value < .01 12-20/NV21-30, 12-20 months toddlers/nonverbal 21-30 months toddlers; 21-30 SW, 21-30 months toddlers with some words; NW, no words; SW, some words; ADI-R, Autism Diagnostic Interview-Revised; SCQ, Social Communication Questionnaire; SRS, Social Responsiveness Scale; K-CARS, Korean Child Autism Rating Scale; NA, not available or can't be calculated due to small sample size.

Table 5 Results of Reliability Analysis

	Toddler (12-20/NV21-30)	Toddler (21-30SW)	Module 1 (No Words)	Module 1 (Words)	Module 2	Module 3	Module 4
Cronbach's alpha	.90	.85	.83	.87	.89	.88	.89
Item discrimination (range)	.89-.90	.83-.85	.81-.83	.85-.87	.88-.89	.87-.88	.87-.89

Note. 12-20/NV21-30, 12-20 months toddlers/nonverbal 21-30 months toddlers; 21-30 SW, 21-30 months toddlers with some words; NW, no words; SW, some words; yo, years old.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ADOSValiditySupplementaryTable.docx](#)