**K-SADS-PL-2009/Autism spectrum disorders (ASD) inter-rater reliability**

Patricia Zavaleta-Ramírez, Omar Náfate-López, Gabriela Villarreal-Valdés, Rosa Elena Ulloa-Flores, Lilia Albores-Gallo

**SUMMARY**

Autism spectrum disorders (ASD) have demonstrated an increase in prevalence and are no longer considered a rare condition. The Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule-Generic (ADOS-G) have been gold standard instruments to diagnose ASD, but neither of these tools identifies associated comorbidity. The new Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Present and Life time version K-SADS-PL-2009 version added a supplement for assessing ASD.

**Objective**

The purpose of this study was to translate, adapt and analyze the inter-rater reliability of the Kiddie SADS-PL-2009/ASD.

**Method**

The sample consisted of 40 children and adolescents, both male and female, with an age range between four and 17 years old, and presumptive ASD diagnosis. The original ASD screen and supplement interview was translated into Spanish by two of the authors and retranslated by a certified translator blind to the study.

**Results**

Inter-rater reliability: Intra-class correlation coefficients varied from moderate to excellent for the following diagnosis both in the present and the past: Autism 0.79, 0.74; Asperger’s 0.85, 1.0; PDDNOS 0.72, 0.41. Kappa coefficients for expert evaluations in the present and the past were as follows: Autism 0.89, 0.87; Asperger’s 0.77, 1.00; PDDNOS 0.69, 0.64. Diagnostic concordance between the initial clinical diagnosis and the following diagnosis through the interview was: 37.5%, with a 67.5% error margin.

**Conclusions**

In this study, the K-SADS-PL-2009/ASD showed moderate inter-rater reliability indicating that it can be used for clinical or research purposes in Mexican children and adolescents.

**Key words:** Validity, reliability, K-SADS, autism, autism spectrum disorders.

**RESUMEN**

El incremento en la prevalencia de los trastornos del espectro autista (TEA) provocó que dejaran de considerarse como poco frecuentes. La Entrevista Diagnóstica de Autismo Revisada (ADI-R; Autism Diagnostic Interview-Revised) y la Cédula de Observación General para el Diagnóstico del Autismo (ADOS-G, Autism Diagnostic Observation Schedule-Generic) se consideran estándares de oro para diagnosticar TEA. Sin embargo, ninguno de estos instrumentos evalúa la comorbilidad asociada. La nueva versión de la Entrevista de Diagnóstico Psiquiátrico para Niños y Adolescentes, K-SADS-PL-2009, agregó un apartado para el diagnóstico de los TEA.

**Objetivo**

El propósito de este estudio fue traducir, adaptar y analizar la confiabilidad interevaluadora del K-SADS-PL-2009/TEA.

**Método**

La muestra se conformó por 40 niños y adolescentes, de ambos sexos, con un rango de edad de cuatro a 17 años, con un diagnóstico presuntivo de trastorno del espectro autista. La versión original del K-SADS-PL-2009/TEA fue traducida al español por dos de los autores y retraducida al inglés por un traductor certificado ciego al estudio.

**Resultados**

Confiable interevaluador. Los coeficientes de correlación intraclase fueron de buenos a excelentes para los siguientes diagnósticos en el presente y pasado: autismo 0.79 y 0.74; trastorno de Asperger 0.85 y 1.0; trastorno generalizado del desarrollo no especificado (TGDNE) 0.72 y 0.41. Los coeficientes kappa para las evaluaciones realizadas por los expertos fueron de buenos a excelentes para los siguientes diagnósticos en el presente y pasado: autismo 0.89 y 0.87; Asperger 0.77 y 1.00; TGDNE 0.69 y 0.64. La concordancia entre el diagnóstico realizado en el servicio de urgencias y el corroborado posteriormente por medio del instrumento diagnóstico fue de 37.5%, con un índice de error diagnóstico de 67.5%.
**INTRODUCTION**

Correct and early diagnosis of autism is important given that it is known that timely intervention with applied behavioral analysis techniques improves children’s prognosis. Although in the last two decades there has been a boost in the development of instruments to assess autism spectrum disorders (ASD) oriented towards screening, diagnosis, and detection of changes (Albores et al., 2008), few of these tools are available for the Mexican population. The validation of the M-CHAT made by Albores et al. in 2012 and of the ABC by Varela and Albores in 2012 show that the transcultural equivalence of instruments is difficult to achieve for reasons such as cultural bias, as recognized by several authors. Because of this, it is not enough to simply translate the instruments; they also need to be culturally adapted in order to be used efficiently.

Although there are interviews to make a sure diagnosis of ASD, such as the Autism Diagnostic Observation Schedule-Generic (ADOS-G) and the Autism Diagnostic Interview-Revised (ADI-R), these tools have the inconvenience of a long application, they require training, and their use is restricted to staff with much experience in this area. Furthermore, they do not assess psychiatric comorbidity, which is very elevated in ASD. Equally, psychiatric interviews, such as Angold and Costello’s Child and Adolescent Psychiatric Assessment (CAPA) and the most recent version of the Preschool Age Psychiatric Assessment (PAPA), Egger et al., 2006, do not include questions to assess developmental problems. The consequence is that autism is not studied in psychiatry surveys and that there are few studies on psychiatric comorbidity and longitudinal follow-up in minors with autism.

Despite not containing a specific section to assess symptoms of autism, the K-SADS-PL was used by Ivarzon in 2001 and Sinzing in 2008 to study the association of psychiatric disorders in children with ASD. Lee also used it in 2006 to assess patients with Attention Deficit Hyperactivity Disorder and its relation to pervasive development disorders not otherwise specified (PDDNOS).

In 2006 Leyfer made a modification to the whole K-SADS-PL interview, calling it the Autism Comorbidity Interview-Present and Lifetime Version (ACI-PL), in order to study disability associated with autism. These studies demonstrated the need for the K-SADS-PL, an instrument which assesses psychopathology and function throughout a lifetime, to have a section to assess children and adolescents with ASD.

In Mexico, Ulloa et al. assessed the K-SADS-PL in 2006, showing excellent inter-rater reliability with greater kappa indexes for major depressive disorder, 0.76; dysthyemic disorder, 0.77; attention deficit hyperactivity disorder, 0.91; and oppositional defiant disorder, 0.71. However, it was not until 2009 that a section to assess ASD was added to the K-SADS-PL. Unlike the ADI-R and the ADOS-G, this section contains algorithms for diagnosis of autism, Asperger’s, and PDDNOS, and does not require any training. Furthermore, its shorter application time allows for observation of the minor with ASD while the parent or tutor is being interviewed.

The aim of this study was to assess the inter-rater reliability of the K-SADS-PL-2009/ASD diagnostic interview (screening and supplement) in Mexican children and adolescents.

**MATERIAL AND METHOD**

**Participants**

Some 40 children and adolescents of both sexes were included, between four and 17 years of age. They were recruited in the outpatients of a child psychiatric hospital, with a presumptive diagnosis of ASD. Once they were made aware of the proposal of the study, the parents signed the informed consent form and the adolescents gave their consent. Participants with any psychotic disorders were excluded, and participants who did not correctly complete the questionnaires were eliminated.

**Measurement instruments**

A semi-structured interview was used to assess the ASD based on DSM-IV criteria in the present and past. This interview has been used before as a gold standard and has shown its effectiveness.

**Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS-PL)**. This is a semi-structured interview that assesses psychopathology in a categorical manner, in accordance with the diagnostic criteria of the DSM-IV. It enables the starting and/or remission age of the symptoms in the present and throughout a lifetime to be established. It was designed to assess children and adolescents between six and 17 years old, although in 2009, Birmaher demonstrated its use and validity in children under six years old.
The instrument consists of three sections: 1. Introductory interview that collects data on medical, educational, and family history, as well as global function (C-GAS). 2. The screening section consists of two or three questions which assess key symptoms. The responses are coded as follows: 0=no information, 1=absent, 2=present, and 3=threshold. 3. The supplementary diagnostic section is applied when any symptom in the screening section is coded as threshold=3 in the clinician’s summary.

The informants’ (parent and child/adolescent) responses are independently recorded. The interviewer prepares a summary which is the best clinical estimate based on the reports of the parents, the minor, and from direct observation of the patient.

The section of screening counts with five questions that assess repetitive motor mannerisms and stereotypies, inflexible adherence to routines or rituals, persistent preoccupation with restricted/stereotypied interests, deterioration in non-verbal behavior, such as eye contact, expressions, and/or gestures, and delay in language. The latter is coded as “threshold” when single words or the use of sentences are expressed three months later than expected (<24 months, and sentences at 33 months). As well as this question, another was drafted to assess delay in the development of language when the child does not express 50 single words at 24 months, or uses sentences at 27 months or later. If the child lacks language, their responses would be coded as “no information”.

Furthermore, the supplement assesses the lack of motivation to share enjoyments, interests, or achievements with others, and inability to develop appropriate relationships with peers of the same age. It also explores social or emotional reciprocity and persistent preoccupation with parts of objects. The last section examines social and language deficiencies, such as pragmatic impairment, prosody, echolalia, verbosity, social isolation, and self-help skills in Asperger’s patients. It also assesses functional deterioration. The final diagnosis is considered definitive if all of the DSM-IV criteria are met, probable when 75% of the diagnostic criteria of a disorder are met, and absent when the diagnostic criteria are not met.

Autism Diagnostic Interview-Revised (ADI-R) (Lord et al., 1994). This instrument is considered as the gold standard to diagnose autism. It is a semi-structured interview administered by clinicians with broad experience in autism, in order to evaluate symptoms of the disorder throughout development. The results are organized into three areas: a) lack of social reciprocity, b) communication problems, and c) restrictive and stereotyped behavior and interests. The ADI-R is based on the frequency of symptoms throughout a lifetime with special emphasis on the age of four or five. This instrument distinguishes children with autism from others with psychiatric disorders or with typical development. It is more precise with children over 40 months of age, as demonstrated by Cox in 1999, and it has greater validity for identifying autism than Asperger syndrome (Gilchrist et al., 2001).

Cicchetti et al. (2008) studied the reliability of the ADI-R and found raised coefficients of correlation (94-96%) in all questions, with kappas (K) between .80 and .88, and .40 and .47.

PROCEDURE The original version of the screening and supplement of the K-SADS-PL-2009/ASD was translated into Spanish by two of the authors, and later retranslated into English by a certified translator who was blind to the study. The face validity and content of the version in Spanish was assessed through the opinion of three mental health experts. Before formally using the instrument, 12 pilot interviews took place to evaluate aspects of the interview such as its flow and understanding on the part of the parents, children, and adolescents.

Participants were recruited in the Juan N. Navarro Child Psychiatric Hospital from May 2010 through January 2011. Doctors in outpatients of the hospital referred children with a presumptive diagnosis of autism, Asperger syndrome, and PDDNOS. After being explained the aim of the study, the parents signed an informed consent letter, and the adolescents and children over eight years of age gave their consent to participate. The application of the K-SADS-PL/ASD (screening and supplement) interviews was carried out by two experts: a psychologist and a child psychiatrist, and two residents in child and adolescent psychiatry who received formal training for three months. A group of residents independent to the study applied the semi-structured interview which assessed the subcategories of autistic disorder, Asperger disorder, and PDDNOS in the present and past.

**Ethical implications**

The present study was approved by the Ethics Committee of the participating hospital. Informed consent was obtained from the parents of each of the children and/or adolescents participating.

**Statistical analysis**

Analysis of demographic variables. Measures of central tendency were used in order to describe the demographic variables. In this way, frequencies were reported for the categorical variables, medians, and standard deviations for the continual variables. To assess the reliability of the interview (screening and supplement), the intra-class coefficient of correlation was used. Furthermore, Cohen’s kappa coefficient was calculated in order to assess reliability between the two experts. There was considered to be excellent reliability when the value of the kappa coefficient was greater than 0.75; good if it was between 0.59 and 0.74; moderate from 0.40 to 0.58, and poor if it was less than 0.40. Finally, the percentage of concordance was calculated with the diagnosis made by the emergency service and that obtained by means of the interview.
**RESULTS**

A total of 45 participants were recruited, and from these, exclusions were made of those who did not fill out the form of sociodemographic data and/or those whom an assessor considered provided unreliable information during the diagnostic interview of DSM-IV and/or ADI-R criteria. The group considered for analysis (N=40) had an average age of 8.96 years (interval from 3.0 to 17.1 years); 80% (n=32) were male.

**Inter-rater reliability**

The intra-class coefficients of correlation were from good to excellent for the majority of the diagnoses (Table 1). The kappa coefficients between the assessments made by the experts were also from good to excellent (Table 2). Finally, the coincidence between the diagnosis made by the emergency service and that corroborated by means of the diagnostic instrument was 37.5%, with a misdiagnosis rate of 67.5%.

**DISCUSSION**

This study evaluated the inter-rater reliability of the K-SADS-PL-2009/ASD diagnostic interview for ASD. Unlike the ADI-R and the ADOS-G, the ASD section of the K-SADS allows a differentiated diagnosis of autism, Asperger disorder, and PDDNOS, according to DSM-IV criteria. In the recent edition of the DSM-5,27 these categories of the DSM-IV were integrated into a single category of ASD, with the aim of increasing the reliability and predictive value of the prognosis, compared to the category determined in the classification of the DSM-IV.28,29

Our results show that the ASD section had good inter-rater reliability for the subcategories. The kappa coefficients for autism were higher among the expert evaluators, where the importance of clinical experience in assessing ASD stood out, as indicated by various researchers.30

The raised concordance for the diagnosis of Asperger’s “in the past” corresponded with the use of quantitative criteria to identify a delay in language development (not expressing words <24 months of age and sentences at 33 months, and/or expressing ≤50 words at 24 months or using sentences at 27 months), whose threshold value of three in the screening distinguishes autism from Asperger’s. The number of words that the child says at 24 months is a very important variable for the diagnosis of ASD; as demonstrated by Jeans in 2014, who established that a cutoff point of ≤25 words at two years of age correctly differentiated autism from Asperger’s.31 This criteria is lower than our own proposal of ≤50 words at two years of age, and this should be taken into account in future studies.

In spite of the importance of the presence/absence of delayed language development as a useful criteria to diagnose ASD and their subcategories, this was excluded from the DSM-5, as more emphasis was given to the pragmatic-semantic problems of language. This fact likely reduces the early identification of autism, given that language delay is the primary cause of parents seeking medical care for their children, as shown by various studies.32,33

Although all of the interview questions should be addressed to the child and/or adolescent, in our study, 90% of these informants were not reliable. This result is consistent with a study by Mazefsky (2011), who demonstrated a low concordance between the responses of adolescents with ASD above symptoms of depression and anxiety and parental responses by means of the ACI-PL interview.34

In spite of this, the ASD section in the K-SADS-PL interview demonstrated that the information provided by the parent, direct observation of the child or adolescent, and the clinician’s experience were sufficient to obtain a better estimate, whose diagnosis had an inter-rater reliability of moderate to good.

Future studies should investigate the construct validity of this interview in order to diagnose ASD in subjects of different ages and with different levels of functioning.

The K-SADS-PL/ASD does bring with it certain limitations, among which is the fact that it does not contain a section to assess the loss of language or motor skills like the ADI-R, and which allows for a classification of autism in its regressive variety. Neither does it assess the special skills of individuals with ASD which are characteristic of so-called savant individuals. Due to the interview being designed before the publication of the DSM-5, its compatibility with this

---

**Table 1.** Inter-rater reliability for the different diagnoses of autistic spectrum disorders in the K-SADS-PL 2009 diagnostic interview

<table>
<thead>
<tr>
<th>K-SADS-PL diagnosis</th>
<th>Intra-class correlation</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present autism</td>
<td>.79</td>
<td>.70 -.87</td>
</tr>
<tr>
<td>Past autism</td>
<td>.74</td>
<td>.62 -.84</td>
</tr>
<tr>
<td>Present Asperger’s</td>
<td>.85</td>
<td>.74 -.93</td>
</tr>
<tr>
<td>Past Asperger’s</td>
<td>1.00</td>
<td>1.0 - 1.0</td>
</tr>
<tr>
<td>Present PDDNOS</td>
<td>.72</td>
<td>.60 -.82</td>
</tr>
<tr>
<td>Present PDDNOS</td>
<td>.41</td>
<td>.25 -.58</td>
</tr>
</tbody>
</table>

CI = Confidence Interval.

**Table 2.** Kappa coefficients between experts of diagnoses obtained in the K-SADS-PL 2009 ASD diagnostic interview

<table>
<thead>
<tr>
<th>K-SADS-PL diagnosis</th>
<th>Kappa correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present autism</td>
<td>.89</td>
</tr>
<tr>
<td>Past autism</td>
<td>.87</td>
</tr>
<tr>
<td>Present Asperger’s</td>
<td>.77</td>
</tr>
<tr>
<td>Past Asperger’s</td>
<td>1.00</td>
</tr>
<tr>
<td>Present PDDNOS</td>
<td>.69</td>
</tr>
<tr>
<td>Past PDDNOS</td>
<td>.64</td>
</tr>
</tbody>
</table>
classification is partial, given that it does not explore sensorial hypersensitivity, which is a criteria added to the DSM-5.

This study has important limitations, among which we should mention the small sample size and the lack of a measure of IQ to prove the effect of this on questions that measure language delay, which is very common in intellectual disability.\textsuperscript{35} In spite of these limitations, the ASD section of the K-SADS-PL is a useful instrument that integrates the report of the minor’s development by means of parental interview and observation of the child or adolescent. This information, gathered in a single instrument, is an advantage if compared with the cost and time of administrating the ADOS-G (observational) and ADI-R (parental interview) instruments, whose combination obtains the best results for sensitivity and specificity, as suggested by several authors.\textsuperscript{36,37}

Furthermore, the K-SADS-PL/ASD is brief and simple to apply, which makes it ideal for use in busy clinical environments.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the cooperation of all the families involved with this study.

REFERENCES

4. Living-Varela K. Confiability y validez del inventario de comportamiento autista (ABC) en niños con trastornos del espectro autista. 2012. Available at: https://sites.google.com/site/hipcomicisioninvestigacion/psiquiatria-infantil-y-de-la-adolescencia.