

# Validación del Examen del Estado Mental en Autismo en español: sensibilidad y especificidad en una muestra clínica argentina

*Validation of the Autism Mental Status Exam in Spanish: sensitivity and specificity in an Argentine clinical sample*

Sebastián H. Cukier<sup>1</sup>, Ana Maglio<sup>2</sup>, Jonathan Berman<sup>3</sup>, Marco Arduini<sup>4</sup>, Natalia Barrios<sup>5</sup>, Maranada Ngue<sup>6</sup>, Cecilia Montiel<sup>7</sup>, Mónica Elsa Ferrea<sup>8</sup>, Karina Gutson<sup>9</sup>, Emilio Zieba<sup>10</sup>, Marcela Menassé<sup>11</sup>, David Grodberg<sup>12</sup>

<https://doi.org/10.53680/vertex.v36i170.941>

## Resumen

**Introducción:** El examen del estado mental en autismo (AMSE, por sus siglas en inglés) es un instrumento breve completado por el clínico que estructura la observación y documentación de los signos y características socio-comunicativas y conductuales del autismo. En su versión original en inglés demostró alta precisión en la identificación del trastorno del espectro autista (TEA). El presente estudio explora la sensibilidad y especificidad de una versión de AMSE en lengua española en una muestra argentina de 313 sujetos, frente al diagnóstico de consenso clínico utilizando los criterios del DSM-5. **Materiales y métodos:** Se calcularon los valores de corte mediante el análisis de la curva ROC (Característica Operativa del Receptor por sus siglas en inglés: Receiver Operating Characteristic), identificando la sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN) correspondientes. La homogeneidad interna de los ítems fue evaluada a través del coeficiente alfa de Cronbach. Asimismo, se calculó el coeficiente Kappa de Cohen para estimar la confiabilidad interevaluador. **Resultados:** Los hallazgos indican una sensibilidad optimizada del 90,71 % y una especificidad del 92,17 % para esta muestra de 313 pacientes. AMSE mostró una consistencia interna aceptable de .75 y una alta confiabilidad entre evaluadores ( $K = .97$ ). **Discusión y conclusiones:** AMSE se perfila como una herramienta prometedora para la evaluación diagnóstica de TEA en niños, adolescentes y adultos en riesgo, destacándose por su alta utilidad clínica. Su aplicación es especialmente relevante en regiones como América Latina, donde

RECIBIDO 7/6/2025 - ACEPTADO 8/9/2025

<sup>1</sup>Médico especialista en Psiquiatría infantojuvenil, Jefe de Salud Mental y Psicopatología, Hospital Pediátrico "Pedro de Elizalde", CABA, Argentina, Co-director médico de Fundación Ados. <https://orcid.org/0000-0002-1497-4709>

<sup>2</sup>Directora Fundación Equipo de Terapia Cognitiva Infantojuvenil, Profesora Titular Psicología Clínica de niños y adolescentes, Universidad Favaloro.

<sup>3</sup>MBA, Former Fulbright Researcher, US Fulbright Program, Atlanta, EE. UU.

<https://orcid.org/0009-0005-3287-4583>

<sup>4</sup>Médico especialista en Psiquiatría. Hospital infantojuvenil "Dra. Carolina Tobar García". <https://orcid.org/0009-0005-4556-8796>

<sup>5</sup>Médica especialista en Psiquiatría. Perito oficial del Ministerio Público Fiscal de la Nación. Argentina. <https://orcid.org/0000-0003-0687-4510>

<sup>6</sup>Harvard University, Doctor of Dental Medicine Degree Candidate, University of Pennsylvania School of Dental Medicine.

<sup>7</sup>PhD, Profesor Titular, University of Texas Rio Grande Valley (UTRGV), EE. UU. <https://orcid.org/0000-0002-4293-5416>

<sup>8</sup>Médica neuróloga infantil. Coordinadora del Centro de rehabilitación infantil, Fleni, sede Escobar, Argentina. <https://orcid.org/0000-0002-9673-9170>

<sup>9</sup>Pediatra especialista en Desarrollo Infantil (UBA). Médica de planta, Centro de Rehabilitación, Hospital de Día, Fleni, sede Escobar, Argentina. <https://orcid.org/0000-0001-5083-7798>

<sup>10</sup>Psiquiatra infantojuvenil. Fundación Ados, CABA, Argentina. <https://orcid.org/0009-0008-0222-3554>

<sup>11</sup>Licenciada en Psicología. Centro de Desarrollo Infantil, Municipalidad de San Carlos de Bariloche, Río Negro, Argentina.

<sup>12</sup>M. D. Assistant Clinical Professor, Yale School of Medicine, New Haven, Connecticut, EE. UU. <https://orcid.org/0000-0003-1961-9252>

## Autor correspondiente:

Sebastián H. Cukier

[sebastiancukier@gmail.com](mailto:sebastiancukier@gmail.com)

**Institución en la que se realizó la investigación:** Fundación Lucha contra la Enfermedad Neurológica Infantil (Fleni), Fundación Ados (Ciudad de Buenos Aires), Centro de Desarrollo Infantil (Municipalidad de San Carlos de Bariloche), Facultad de Ciencias Médicas de la Universidad Nacional de Rosario (CPENI -Rosario), Programa Argentino para Niños, Adolescentes y Adultos con Condiciones del Espectro Autista (Martínez) y Hospital Nacional en Red "Lic. Laura Bonaparte" (Ciudad de Buenos Aires).



el acceso a la capacitación y uso de las escalas que son estándar de oro es limitado. Estos resultados respaldan el uso de AMSE como una herramienta diagnóstica breve, confiable y culturalmente adaptada para apoyar el diagnóstico clínico de autismo en contextos de habla hispana con recursos limitados.

**Palabras clave:** Examen del estado mental del autismo, trastorno del espectro autista, evaluación diagnóstica del autismo, Argentina, Latinoamérica

## Abstract

*Introduction: The Autism Mental State Exam (AMSE) is a brief clinician-completed instrument that structures the observation and documentation of the social, communicative, and behavioral signs and characteristics of autism. In its original English version, it demonstrated high accuracy in identifying autism spectrum disorder (ASD). This study explores the sensitivity and specificity of a Spanish version of AMSE in an Argentine sample of 313 subjects, against DSM-5 clinical consensus diagnosis as the primary outcome. Materials and methods: The cutoff values were calculated using the receiver operating characteristic (ROC) curve, identifying the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The internal homogeneity of the items was determined using Cronbach's alpha, and the Cohen Kappa coefficient was calculated for inter-rater reliability. Results: Findings indicate an optimized sensitivity of 90.71% and a specificity of 92.17% for this sample of 313 patients. AMSE exhibited fair internal consistency (alpha coefficient of .75) and high inter-rater reliability (.97) in this sample of subjects evaluated for ASD in Argentina. Discussion & conclusions: AMSE is emerging as a promising tool for diagnostic assessment of ASD in children, adolescents and adults at risk, standing out for its high clinical utility. Its application is particularly relevant in regions like Latin America, where access to the Autism Diagnostic Observation Schedule training, materials and use is limited. These results support the use of AMSE as a brief, reliable, and culturally adapted diagnostic tool to support the clinical diagnosis of autism in Spanish-speaking contexts with limited resources.*

**Keywords:** Autism mental status exam, autism spectrum disorder, autism diagnostic assessment, Argentina, Latin America

## Introducción

El trastorno del espectro autista (TEA) es una condición del neurodesarrollo caracterizada por desafíos en el funcionamiento social y comunicativo, comportamientos repetitivos e intereses restringidos, hiper o hiposensibilidad a diferentes estímulos sensoriales, y por tener una manifestación temprana (durante las primeras etapas del desarrollo) (APA, 2022). Una vez considerado raro, con una prevalencia de 4-5/10,000, el autismo hoy en día tiene una prevalencia en el rango de 1.09/10,000 a 436/10,000, con una media de 100/10,000 (Zeidan et al., 2022). A nivel mundial, el análisis longitudinal de datos a lo largo de los años dentro de las mismas áreas geográficas confirma el aumento de las estimaciones de prevalencia que ha llamado repetidamente la atención de los científicos en los últimos veinte años (Chiarotti et al., 2020).

Los niños, adolescentes y adultos con TEA tienen una presentación muy heterogénea y requieren un enfoque diagnóstico estandarizado que sea factible en todos los entornos clínicos. La evaluación del TEA ha sido investigada durante las últimas décadas y se

han generado varios instrumentos complementarios para estandarizar y facilitar la evaluación diagnóstica. La Guía de Observación Diagnóstica del Autismo (ADOS) es hoy en día un instrumento principal en los esquemas de evaluación para el TEA en todo el mundo (Lord et al. 2001; Lord et al., 2012). ADOS requiere una formación profesional específica extensa antes de su uso, materiales específicos y un entorno clínico adecuado en el que se pueda realizar. También es un proceso que insume un tiempo prolongado para implementar y puntuar (frecuentemente más de 1 hora) y tiene un costo elevado. Por estas razones, aunque relativamente preciso y confiable, ADOS es inaccesible para muchas personas en Argentina y otros países de la región. Numerosas familias deben viajar grandes distancias (a veces más de 100 kilómetros) para una evaluación diagnóstica porque no hay herramientas de evaluación estandarizadas para apoyar la impresión clínica en su área de residencia (Paula et al., 2020).

Con el propósito de validar una herramienta de diagnóstico accesible en entornos clínicos y de investigación para individuos remitidos para evaluación neuropsiquiátrica, incluyendo tanto ítems de observación

como información obtenida de los padres o cuidadores, Grodberg et al. (2012) desarrollaron el Examen del Estado Mental del Autismo (AMSE). AMSE estructura la evaluación de signos y síntomas autistas complementando el criterio de los clínicos e investigadores que evalúan a individuos con riesgo de TEA, abordando la falta de evaluación observacional estandarizada en poblaciones clínicas desatendidas y con pocos recursos (Grodberg et al., 2012).

AMSE es un instrumento estandarizado breve, fácil de usar y gratuito, basado en la observación clínica directa y el informe de los padres, que tiene como objetivo apoyar el juicio clínico en el proceso diagnóstico del TEA. Se compone de ocho categorías que se puntúan individualmente con 0, 1 o 2 y se suman para obtener una puntuación de hasta 14 (Grodberg et al., 2012). AMSE puntúa los siguientes ocho ítems: contacto visual, interés en los demás, habilidades de señalización, nivel de lenguaje, pragmática del lenguaje, comportamientos repetitivos y estereotipados, preocupaciones inusuales o persistentes, y sensibilidades inusuales. Todos estos rasgos pueden observarse durante un chequeo regular del paciente, lo que facilita al trabajador de la salud y a la familia llegar a un diagnóstico estandarizado más preciso. AMSE no añade trabajo extra a un encuentro con el profesional evaluador, no requiere materiales o actividades específicas, sino que proporciona un método claro y cuantitativo de documentar comportamientos observados y reportados durante la entrevista clínica (Grodberg et al., 2016).

La versión original de AMSE (en inglés) ha sido validada en relación con el diagnóstico de consenso del DSM-5, con el ADI-R, así como con ADOS, y se ha demostrado que tiene buena confiabilidad, con una tasa de sensibilidad y una tasa de especificidad que superan el 80 % en comparación con estos instrumentos (según las normas de validación de Cicchetti et al. (1995)). Un estudio que aplicó AMSE a una población de alto riesgo de 45 niños pequeños de 18 meses a 5 años, mostró que una puntuación de corte de 6 produjo una sensibilidad del 94 % y una especificidad del 100 % en la identificación de niños con TEA según los criterios del Manual Diagnóstico y Estadístico de los Trastornos Mentales, quinta edición (APA, 2013) (Grodberg et al., 2016). También un estudio con adultos (18 a 45 años) con lenguaje verbal fluido y sospecha de TEA mostró una sensibilidad de 91 % y una especificidad de 93 % al compararlo con el diagnóstico final de consenso según criterios del DSM-5 (Grodberg et al., 2014).

Se han realizado estudios de validación de AMSE en diferentes países. Un estudio en Suecia validó AMSE

utilizando los criterios diagnósticos de la Clasificación Internacional de Enfermedades, décima edición (CIE-10) basado en una entrevista con el Cuestionario Diagnóstico para los Trastornos Sociales y de Comunicación (DISCO-11) con padres/cuidadores en un grupo de 124 niños derivados para evaluación. La capacidad intelectual o índice de desarrollo, y la capacidad de lenguaje expresivo de los participantes se compararon con los resultados de AMSE y del cuestionario Vineland-II (completado por los padres). Según este estudio, existe buena correlación entre AMSE y los criterios de la CIE-10 para TEA. Para los niños que no cumplían los criterios para TEA, sus puntuaciones medias de AMSE fueron significativamente más bajas en comparación con los niños con TEA. Además, puntuaciones medias más altas de AMSE se correlacionaron con puntuaciones estándar más bajas de funcionamiento adaptativo global en Vineland-II. En este estudio con población sueca, el nivel de corte óptimo para un probable TEA se encontró en una puntuación de 7 puntos en el AMSE, con una sensibilidad del 75 % y una especificidad del 78 % (Cederlund, 2019).

Øien et al. (2018) exploraron las posibles diferencias sexuales en el puntaje de la escala AMSE y las observaciones de 123 niños noruegos derivados para evaluación del autismo. Los resultados indicaron que la puntuación de AMSE fue similar en la identificación de TEA y no-TEA en mujeres en comparación con hombres (Øien et al., 2018). Otro protocolo de Noruega investigó el rendimiento de AMSE en la detección de TEA en una muestra de 45 niños con síntomas de trastorno por déficit de atención e hiperactividad (TDAH) y TEA. El análisis de la curva de características operativas del receptor (ROC) indicó una fuerte sensibilidad y especificidad en esta población, concluyendo que AMSE tiene potencial como herramienta de evaluación de TEA en niños con síntomas de TDAH y TEA (Øien et al., 2020).

Betz et al. (2019) exploraron la viabilidad de AMSE tanto en muestras de pediatría del desarrollo como en atención primaria en Estados Unidos. Cincuenta y tres niños pequeños con TEA y otras condiciones del desarrollo fueron evaluados usando AMSE y comparados con 55 niños pequeños con desarrollo típico. Las puntuaciones de AMSE diferían significativamente entre los grupos TEA, discapacidad del desarrollo no TEA y neurotípicos. Un punto de corte en AMSE de  $\geq 5$  para TEA maximizó la sensibilidad (81.2 %) y la especificidad (90.5 %). Las diferencias de puntuación entre los grupos sugirieron que AMSE también puede ser útil en un entorno clínico para ayudar a identificar a los niños pequeños con posible TEA (Betz et al., 2019).

AMSE también se investigó en una muestra brasileña de 260 niños de 3 a 18 años, de los cuales 130 cumplían criterios de autismo según la escala de puntuación del autismo infantil (CARS-BR). Los participantes se sometieron tanto a AMSE como a CARS-BR, que se utilizó para aproximar la severidad del autismo y el punto de corte en AMSE. AMSE mostró buena consistencia interna (.74), sensibilidad (.91) y especificidad (.98); y alta correlación con el CARS-BR ( $r = .91, p < .01$ ) (Galdino et al. 2020). En Chile se realizó un estudio con 64 niños de entre 15 meses y 17 años (52 de ellos con TEA) que consultaron por desafíos en la comunicación y el lenguaje, en el que compararon los resultados de AMSE con los de ADOS-2, encontrando una sensibilidad de 79 % y una especificidad de 92 % (Irrarrázaval et al., 2023).

Un estudio chino de 2023 evaluó AMSE en una muestra de 66 niños de entre 2 y 11 años con riesgo de TEA, comparándolo con los criterios del DSM-5. El análisis de la curva ROC arrojó un área bajo la curva de 0.98. El valor de corte óptimo de la versión china se estimó en 6, produciendo la mayor sensibilidad del 98 % y una especificidad del 87 % (Yang et al., 2023).

Los hallazgos preliminares de estos estudios sugieren que AMSE tiene propiedades psicométricas prometedoras como herramienta de evaluación para identificar síntomas de TEA y apoyar la toma de decisiones diagnósticas en la población de diferentes países y culturas.

El objetivo de este estudio fue examinar la validez de la versión en español de AMSE en una muestra de niños, adolescentes y adultos residentes en Argentina, comparando su rendimiento con el diagnóstico clínico de consenso y, secundariamente, con los resultados de ADOS.

El objetivo específico primario fue determinar una puntuación de corte de AMSE que proporcione una sensibilidad y especificidad óptimas al diferenciar entre personas con TEA y controles sin TEA. Con este objetivo la muestra fue estratificada por sexo, edad, nivel de lenguaje y capacidad cognitiva, para analizar diferentes puntos de corte según estas características. El segundo objetivo del estudio fue comparar los resultados de AMSE con los resultados de la Guía de Observación Diagnóstica del Autismo (ADOS) en la diferenciación entre TEA y no-TEA en un subgrupo de los sujetos participantes.

Tabla 1. Estudios previos de validación del AMSE en diferentes países

1er autor / año	Población	Nro. casos	Comparación con	Resultados principales
Grodberg 2014	Población de riesgo de 18 a 45 años	50	Criterios DSM-5	Sensibilidad 91 % Especificidad 93 % Punto de corte $\geq 5$
Grodberg 2016	Población de alto riesgo de 18 meses a 5 años	45	Criterios DSM-5	Sensibilidad 94 % Especificidad 100 % Punto de corte $\geq 6$
Cederlund 2019	Niños 2-5 años derivados para evaluación	124	Criterios CIE-10 basado en DISCO-11 a padres y Vineland II	Sensibilidad 75 % Especificidad 78 % Punto de corte $\geq 7$
Betz 2019	Niños 18-68 meses: con TEA vs con CND no-TEA vs desarrollo típico	108	Dx clínico usando la Checklist for Autism Spectrum Disorder (CASD)	Sensibilidad 81.2 % Especificidad 90.5 % Punto de corte $\geq 5$
Øien 2020	Niños 3-12 años con síntomas de TDAH y TEA	45	Criterios DSM-5 utilizando ADOS y ADI-R	Sensibilidad 93 % Especificidad 77 % Punto de corte $\geq 5$
Galdino 2020	Niños de 3 a 18 años con TEA (130) u otros diagnósticos psiquiátricos (130)	260	Criterios DSM-5 y CARS-BR	Sensibilidad 91 % Especificidad 98 % Punto de corte $\geq 4$
Irrarrázaval 2023	Niños de entre 1 y 17 años con desafíos en la comunicación o riesgo de TEA	64	ADOS-2	Sensibilidad de 79 % Especificidad de 92 % Punto de corte $\geq 6$
Yang 2023	Niños 2-11 años con riesgo de TEA	66	Criterios DSM-5	Sensibilidad 98 % Especificidad 87 % Punto de corte $\geq 6$



## Materiales y métodos

### Participantes

La muestra de este estudio estuvo conformada por niños y jóvenes que consultaron por preocupaciones sobre su desarrollo socio-comunicacional, y que fueron evaluados en la Fundación Lucha contra la Enfermedad Neurológica Infantil (Fleni – Sede Escobar y Sede Belgrano), la Fundación Ados (Ciudad de Buenos Aires), el Centro de Desarrollo Infantil (Municipalidad de San Carlos de Bariloche), el Programa Argentino para Niños, Adolescentes y Adultos con Condiciones del Espectro Autista (Martínez), la Facultad de Ciencias Médicas de la Universidad Nacional de Rosario (CPE-NI - Rosario) y el Hospital Nacional en Red “Lic. Laura Bonaparte” (Ciudad de Buenos Aires), en Argentina entre abril de 2017 y marzo de 2020. La muestra del estudio, incluyó a todas las personas (o bien sus padres o cuidadores) que consintieron participar en el protocolo.

### Procedimiento

En los seis centros participantes se aplicó un protocolo de evaluación diagnóstica con un enfoque estandarizado que incluyó múltiples componentes. Primero, se realizó una evaluación diagnóstica clínica, que fue llevada a cabo por un psiquiatra, pediatra del desarrollo o psicólogo clínico con extensa experiencia en el diagnóstico de TEA y otras condiciones del desarrollo. Esta evaluación incluyó una entrevista clínica semiestructurada con los padres o cuidadores y una observación basada en el juego/conversación con el sujeto. La entrevista semiestructurada relevó el motivo principal de consulta, la historia de la enfermedad actual, el funcionamiento ecológico actual y los antecedentes psiquiátricos, médicos y de desarrollo del sujeto. Los criterios del DSM-5 fueron utilizados para guiar la formulación diagnóstica clínica de TEA versus no-TEA.

AMSE se administró en el contexto de la primera evaluación clínica o inmediatamente después. La puntuación fue cargada en un formulario Google y no fue utilizada otra vez por el examinador hasta el procesamiento de los datos del estudio. Cuando fue factible, un segundo evaluador se ubicó en la sala con el evaluador principal durante las entrevistas clínicas y utilizó AMSE para co-evaluar al paciente de manera independiente con el fin de medir la confiabilidad entre evaluadores. Las puntuaciones no se compartieron en ningún momento entre los dos evaluadores.

Todos los clínicos habían recibido la capacitación para implementar AMSE utilizando el curso de formación en línea (original en inglés), que incluía ins-

trucciones sobre la puntuación y cuatro casos de práctica (<http://autismmentalstatusexam.com>). Además, realizaron una sesión de puntuación de fidelidad, a distancia utilizando videoconferencia, con el primer autor de este estudio, que incluía tres casos más. Para el estudio actual, todos los clínicos que administraron AMSE habían establecido previamente una confiabilidad inter-evaluador del 95 %. Los clínicos entrenados completaron AMSE basándose en el manual de puntuación adaptado y traducido al español por el primer autor, sin conocer el diagnóstico final del niño ni los resultados de otras evaluaciones estandarizadas. Los clínicos que realizaron la puntuación de AMSE no habían recibido ninguna capacitación específica sobre la herramienta antes de la formación para el estudio. Sin embargo, AMSE había sido utilizado por el primer autor en entornos clínicos durante aproximadamente tres años antes del estudio, por lo que el autor estaba familiarizado con el instrumento antes de su implementación en la investigación.

Cuando la evaluación clínica no era suficiente para llegar a un diagnóstico claro según los criterios del DSM-5, la persona regresaba a la clínica para que un neuropsicólogo independiente, sin conocimiento de la puntuación de AMSE ni de las impresiones diagnósticas del admisor, le administrara ADOS. Todos los datos recibidos de la evaluación ADOS fueron transportados al sistema de base de datos utilizando solo el diagnóstico categórico (TEA o no TEA) para simplificar el análisis, dado que la muestra incluía sujetos con amplios rangos de edad y niveles de lenguaje. Todas las evaluaciones con ADOS fueron realizadas por evaluadores con amplia experiencia en el campo neuropsicológico.

Este estudio se llevó a cabo de manera ciega, ya que diferentes examinadores independientes puntuaron AMSE y ADOS. Los evaluadores clínicos que completaron AMSE no conocían las puntuaciones de corte para el TEA, ya que estas se determinaron después del análisis estadístico. El consenso clínico final del diagnóstico según los criterios del DSM-5 se alcanzó considerando la información recopilada de la entrevista clínica y la historia del desarrollo, con los aportes de un neuropsicólogo y un psicólogo especializados, y considerando la información de ADOS cuando estuviera disponible, pero sin considerar la información de AMSE, solo conocida por el clínico admisor.

También se utilizaron evaluaciones de funcionamiento adaptativo, de coeficiente intelectual y de comunicación y lenguaje, según estuvieran indicados para guiar las decisiones de diagnóstico o tratamiento y las recomendaciones, así como para estratificar la

muestra para este estudio. El nivel de funcionamiento se determinó utilizando la segunda edición de la Escala de Comportamiento Adaptativo de Vineland (Vineland-II Cuestionario de Evaluación para Padres/Cuidadores -VABs-2) en la mayoría de los centros. El cuestionario de evaluación Vineland-II se distribuyó a los padres/cuidadores por correo electrónico para ser devuelto por el mismo medio al equipo de evaluación. Las puntuaciones del cuestionario de evaluación para padres/cuidadores de VABs-2 fueron procesadas en el programa de datos Vineland-II por el investigador para obtener las puntuaciones estándar para el Funcionamiento Adaptativo Global. Cuando no se disponía de VABs-2, el nivel de funcionamiento se basó en la puntuación de escalas de evaluación cognitiva o de desarrollo estandarizadas; específicamente, una de las escalas de Wechsler, es decir, Escala de Inteligencia de Wechsler para Preescolar y Primaria, Cuarta Edición (WPPSI-IV), Escala de Inteligencia de Wechsler para Niños, Cuarta Edición (WISC-IV), Escala de Inteligencia para Adultos de Wechsler (WAIS-IV), Escala Mullen de Aprendizaje Temprano, o Escala Leiter de inteligencia no-verbal, tercera edición (Leiter-3). Para los participantes para quienes no se realizó una prueba formal de nivel de inteligencia o de funcionamiento adaptativo, el nivel de funcionamiento fue estimado clínicamente en la visita a la clínica utilizando información sobre habilidades de desarrollo adquiridas de evaluaciones y reportes previos (por ejemplo, reportes escolares). A partir de esto los participantes se estratificaron según el nivel de funcionamiento, la capacidad intelectual o el nivel de desarrollo en los siguientes niveles: (1) Discapacidad intelectual (DI) profunda o severa o nivel de funcionamiento adaptativo muy descendido (CI 0-34 o equivalente), (2) DI moderada a leve o nivel de funcionamiento adaptativo bajo (CI 35-69 o equivalente), (3) Inteligencia normal baja, normal, o por encima del promedio o funcionamiento adaptativo promedio-bajo a muy superior (CI  $\geq 70$  o equivalente). La clasificación del nivel de lenguaje del niño se basó en los niveles de lenguaje expresivo presentados en la escala AMSE (es decir, "No verbal - Palabras sueltas" para una puntuación de 2, "Frasas - Oraciones no desarrolladas" para una puntuación de 1, y "Puede hablar de otro tiempo y lugar - lenguaje fluido" para una puntuación de 0).

### **Aprobación ética y consentimiento informado**

El estudio fue aprobado por la Junta de Revisión de la Fundación Lucha contra la Enfermedad Neurológica Infantil (Fleni). Durante la visita inicial a las clínicas, se proporcionó a los participantes y a los padres o cuidadores un resumen de la investigación y formularios

de consentimiento para llevar al hogar y revisar. Si la persona o el padre/cuidador decidía participar, firmaban el formulario de consentimiento y lo devolvían al examinador, quien lo enviaba al investigador principal. Los participantes adultos firmaron el consentimiento informado previamente a ingresar al estudio, los menores de 18 años o aquellos mayores con dificultades para comprender los términos del consentimiento, fueron consentidos por un padre o tutor, y proporcionaron asentimiento verbal cuando fue apropiado. Los participantes y sus familiares recibieron retroalimentación junto con un informe detallado de la evaluación diagnóstica.

### **Instrumentos de medición**

#### **Examen del Estado Mental del Autismo - AMSE**

AMSE es una herramienta de observación diagnóstica breve y gratuita que sirve para diferenciar el TEA de otras condiciones del desarrollo. Fue desarrollada originalmente por un equipo dirigido por el Dr. David Grodberg, incluyendo psiquiatras infanto-juveniles, pediatras del desarrollo y neurólogos pediátricos en el Centro Seaver de Investigación y Tratamiento del Autismo, en la Escuela de Medicina del Monte Sinaí. AMSE estructura las observaciones directas por parte del examinador para los ocho ítems incluidos, y proporciona la oportunidad de registrar la información clínica recibida del cuidador durante la evaluación. Los primeros 3 ítems son puntuados únicamente por el investigador y se relacionan con observaciones directas del niño, y los ítems 4-8 se puntúan basándose tanto en la presentación del niño en la consulta como en la información proporcionada por el cuidador. Las puntuaciones más altas reflejan una mayor severidad de los síntomas. Los síntomas o comportamientos observados para los ítems 5, 7 y 8 se puntúan más alto (2p) en comparación con la información sólo proporcionada por el cuidador (1p). Cada ítem se puntúa con 0, 1 o 2 puntos dependiendo de la severidad de los síntomas como se especifica a continuación. Los ocho ítems incluyen (1) Contacto visual (observado) ( $\geq 3$  segundos: 0 puntos, Fluctuante: 1p, Ninguno: 2p); (2) Interés en los demás (observado) (Inicia interacción con el examinador 0 puntos, Solo responde pasivamente 1p, Sin interés: 2p); (3) Habilidades de señalamiento (observado) (Puede señalar o gesticular hacia un objeto: 0 puntos, Solo sigue el señalamiento: 1p, Ninguno: 2p); (4) Lenguaje (reportado u observado) (Puede hablar sobre otro momento o lugar: 0p, Palabras sueltas/frases ( $\leq 3$  palabras)/oraciones no complejas: 1p, Sin lenguaje verbal: 2p); (5) Pragmática

del lenguaje (reportado u observado) (No deteriorado o No aplica: 0p, No puede manejar turnos o mantener reciprocidad en los tópicos, o Entonación monótona o extraña, reportados: 1p, observados: 2p); (6) Comportamientos repetitivos, estereotipias (reportados u observados) Ninguno: 0 p, Comportamientos compulsivos o insistencia en rutinas: 1p, Movimientos repetitivos, Ecolalia o Habla estereotipada: 2p); (7) Preocupaciones inusuales o persistentes (reportado u observado) (ninguna: 0 puntos, Presente reportado: 1p, observado: 2p); y (8) Sensibilidades inusuales (reportado u observado) (Ninguna: 0p, Sensibilidad aumentada o Alta tolerancia al dolor reportada: 1p, observada: 2p) (Grodberg et al. 2012).

Con la autorización del Dr. David Grodberg, AMSE fue traducido del inglés al español rioplatense por el primer autor de este estudio, quien pulió la traducción para que fuera más culturalmente apropiada y fácilmente comprensible, lo que se reflejó en el borrador final de la versión en español de AMSE. El trabajo de retro-traducción fue completado por un traductor profesional que desconocía la versión original en inglés y recibió la aprobación del Dr. Grodberg. La versión en español de AMSE se estableció en 2016 después de este ciclo de traducción y retro-traducción y posteriormente se utilizó para este estudio. Para asegurar el uso adecuado de la versión en español de AMSE, se realizó una capacitación específica para todos los investigadores y clínicos que la deseen utilizar. Detalles adicionales sobre el uso y la puntuación de AMSE en español estarán disponibles en el sitio web <https://ilina.com.ar> en el que estará subido el curso en línea gratuito. Este incluye casos simulados en video y viñetas en texto, adaptadas del curso original en inglés (ya disponible en <http://autismmentalstatusexam.com>). AMSE, al igual que cualquier otra herramienta de evaluación, no puede ser utilizado de manera independiente para diagnosticar TEA, sino que debe ser utilizado como instrumento estandarizado de observación directa para apoyo al juicio clínico del experto en el proceso diagnóstico de autismo y condiciones del desarrollo.

### **Guía de Observación para el Diagnóstico de Autismo – ADOS y ADOS 2**

ADOS es un instrumento observacional estandarizado y semiestructurado diseñado para ayudar a diagnosticar a individuos con trastorno del espectro autista (Lord et al. 2000). La evaluación dura entre 45 y 60 minutos y se divide en cuatro módulos según el nivel de lenguaje del individuo. ADOS combina juego no estructurado con actividades estructuradas y pre-

guntas de entrevista para indagar en las características sociales, comunicativas y conductuales asociadas con el TEA. ADOS-2 (Lord et al., 2012) es la segunda edición de la guía ADOS. ADOS-2 comprende cinco módulos basados en la edad y la capacidad lingüística (Módulo T para Niños Pequeños y Módulos 1-4). Existen dos puntuaciones límite: autismo y espectro autista. En este estudio, se utilizaron los criterios de corte del espectro autista en comunicación, socialización y puntuaciones totales de comunicación y socialización para determinar la clasificación diagnóstica en ADOS.

### **Escala de Comportamiento Adaptativo Vineland, Segunda Edición (Vineland-II) Cuestionario de Evaluación para Padres/Cuidadores**

Se trata de un cuestionario de evaluación para padres/cuidadores que ofrece una evaluación integral del comportamiento adaptativo en tres áreas principales: (1) Comunicación (receptiva, expresiva, escrita) (2) Habilidades de la Vida Diaria (HVD) (personal, doméstica, comunitaria) y (3) Socialización (relaciones interpersonales, juego y tiempo libre, habilidades de afrontamiento). Además, se incluye un dominio de Habilidades Motoras. A partir de las puntuaciones de estas escalas, se puede derivar una puntuación estándar de Funcionamiento Adaptativo General (GAF) (Sparrow et al. 2016).

### **Pruebas de inteligencia y desarrollo**

*Escala de Inteligencia de Wechsler para Preescolar y Primaria, Cuarta Edición (WPPSI-IV):* Es una prueba de inteligencia diseñada para niños de 2 años y 6 meses a 7 años y 7 meses, que proporciona subescalas y puntuaciones compuestas que representan el funcionamiento intelectual en comprensión verbal, capacidad visual-espacial, memoria de trabajo, razonamiento fluido y velocidad de procesamiento, así como una puntuación compuesta que representa la capacidad intelectual general del niño (Wechsler 2012).

*Escala de Inteligencia de Wechsler para Niños, Cuarta Edición (WISC-IV):* Prueba de inteligencia diseñada para niños de 6 años a 16 años y 11 meses, que proporciona subescalas y puntuaciones compuestas que representan el funcionamiento intelectual en comprensión verbal, razonamiento perceptivo, memoria de trabajo y velocidad de procesamiento, así como una puntuación compuesta que representa la capacidad intelectual general del niño (Wechsler, 2003).

*Escala de Inteligencia para Adultos de Wechsler, Cuarta Edición (WAIS-IV):* es una prueba ampliamente utilizada diseñada para medir la capacidad cognitiva y la inteligencia de adultos y adolescentes mayores entre



los 16 y los 90 años. Consta de 10 subescalas básicas y 5 complementarias, que en conjunto proporcionan una evaluación completa del funcionamiento intelectual de un individuo. La prueba arroja varias puntuaciones, incluido el coeficiente intelectual, escala completa (FSIQ), que se deriva de las subescalas principales (Wechsler, 2008).

*La Escala de Mullen de Aprendizaje Temprano:* es una herramienta de evaluación que mide el desarrollo temprano de los niños desde el nacimiento hasta los 5 años y 8 meses. Este sistema evalúa varias áreas del desarrollo, incluyendo habilidades motoras gruesas, recepción visual, habilidades motoras finas, lenguaje expresivo y receptivo (Mullen, 1995).

*Leiter-3:* La Escala de Desempeño Internacional Leiter es una prueba de inteligencia no verbal diseñada para evaluar las habilidades cognitivas sin depender de la comunicación verbal. Esto la hace especialmente útil para personas con discapacidades del habla o de la audición, hablantes no nativos y aquellos en el espectro autista. La versión más reciente, Leiter-3, cubre un rango de edad de 3 años a 75+ años e incluye tareas que miden diversas funciones cognitivas como la memoria, la atención y el razonamiento no verbal (Gordon & Miller, 2013).

### **Métodos estadísticos y análisis de datos**

Para validar esta versión en español de AMSE, se evaluaron la fiabilidad inter-evaluadores, la consistencia interna y el rendimiento diagnóstico del instrumento tomando como referencia el diagnóstico de consenso según los criterios del DSM-5. Además, se analizaron variables como edad, sexo, nivel de funcionamiento adaptativo (o en su defecto la capacidad intelectual / índice de desarrollo) y nivel de lenguaje expresivo de los participantes en comparación con las puntuaciones obtenidas en AMSE. El propósito de este análisis fue la estratificación de la muestra para evaluar si se podría mejorar la sensibilidad y especificidad variando los puntos de corte en función de la edad, el sexo, el lenguaje o el nivel de funcionamiento.

El conjunto de datos original recopilado fue ingresado en un formulario de Google para cada participante y exportado a una hoja de cálculo de Excel y de allí al paquete estadístico para las ciencias sociales SPSS en su versión 27. El análisis de datos incluyó el cálculo de medidas de tendencia central, variabilidad y distribución de frecuencias; las estadísticas descriptivas se presentan como media, desviación estándar, mediana, mínimo y máximo para las variables continuas y como números y porcentajes para las varia-

bles categóricas. El sexo del individuo (masculino, femenino u otro) se trató como variable categórica; el índice de funcionamiento y el nivel de lenguaje se trataron como ordinales; y la edad como una variable continua. Las comparaciones de grupos se realizaron utilizando pruebas de chi-cuadrado para variables dicotómicas, chi-cuadrado de Mantel-Haenszel para variables categóricas ordinales y pruebas t de muestras independientes para variables continuas. El coeficiente de correlación de Spearman se utilizó para analizar la relación entre las variables. Finalmente, se calcularon la sensibilidad, especificidad y valores predictivos positivos y negativos para diferentes cortes de AMSE, primero con la muestra completa y luego estratificando por sexo, edad, nivel adaptativo y de lenguaje. Se utilizó un corte del 5 % para la significación estadística. Se estableció un punto de corte para AMSE que maximiza la precisión diagnóstica general (basada en el diagnóstico clínico de consenso) y minimiza la diferencia entre sensibilidad y especificidad. Con este punto de corte, se calcularon la sensibilidad (porcentaje de sujetos correctamente identificados con TEA por la puntuación de AMSE sobre el total de niños con TEA) y la especificidad (porcentaje de niños correctamente identificados sin TEA sobre el total de niños sin TEA). Se compararon las distribuciones de las puntuaciones de AMSE entre las dos muestras (TEA y No-TEA) utilizando la prueba de U de Mann-Whitney.

Luego se realizó un análisis de curva de características operativas del receptor (ROC) para medir la discriminación dentro de la población de referencia. Se calcularon la sensibilidad y especificidad de AMSE utilizando el diagnóstico clínico en todos los umbrales enteros. Se trazaron curvas ROC generales y se calcularon las áreas bajo la curva utilizando el método trapezoidal. La curva ROC es un método que ilustra la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos de una prueba diagnóstica. El análisis de curva ROC muestra simultáneamente la línea que representa cada punto de corte y su tasa de verdaderos positivos (sensibilidad) y tasa de falsos positivos (1 - especificidad) asociada. El punto en la línea que está más distante de la diagonal verdadera indica el punto de corte con la mayor sensibilidad y especificidad. El área bajo la curva (AUC, por sus siglas en inglés "Area Under the Curve") se utilizó para reflejar la precisión de la prueba, indicando qué tan bien la prueba separa a los individuos con TEA de aquellos sin TEA. Un AUC de 1.0 indica una prueba con capacidad discriminativa perfecta, mientras que un AUC de 0.5 refleja una prueba sin capacidad discriminativa,



similar a un modelo que clasifica al azar. Se realizó un análisis post hoc de subgrupos para examinar los efectos del sexo, la edad, el nivel de funcionamiento y el nivel de lenguaje en las puntuaciones de AMSE dentro de cada muestra. Además, se evaluó si la variación del punto de corte según estas variables podría optimizar la sensibilidad y/o especificidad de AMSE. En última instancia, el presente estudio también llevó a cabo un análisis de la curva ROC, comparando las puntuaciones totales de AMSE con el resultado de ADOS (positivo o negativo para TEA) en un subgrupo de participantes ( $n = 249$ ) (objetivo secundario).

La confiabilidad inter-evaluador se midió mediante el coeficiente Kappa de Cohen. Kappa mide el grado de concordancia de las evaluaciones nominales u ordinales realizadas por diferentes evaluadores cuando se evalúan las mismas muestras. El valor de Kappa varía entre -1 y 1. Un valor de 1 indica un acuerdo perfecto, 0 indica un acuerdo equivalente al azar y -1 indica un desacuerdo completo. La consistencia interna se determinó utilizando el alfa de Cronbach, con el fin de estimar la confiabilidad de la prueba. Valores de alfa de Cronbach entre .70 y .90 indican una buena consistencia interna.

## Resultados

### Participantes

Como se muestra en la *Tabla 2*, participaron 313 sujetos consecutivos; se incluyeron 237 hombres (75.6 %) y 76 mujeres (24.4 %), con una proporción hombre:mujer de aproximadamente 3,12:1. La totalidad de los participantes recibieron un diagnóstico final de consenso (basado en los criterios del DSM 5), resultando dos subgrupos, un grupo con TEA ( $n=165$ , 52.72 %) y un grupo sin TEA ( $n=148$ , 47.28 %) (ver *Tabla 2*).

Con edades comprendidas entre 2 y 31 años ( $M=6.11$ ;  $DE=4.75$ ) al momento de la admisión. La mayoría (96 %) tenía menos de 18 años de edad, y el 68 % menos de 6 años de edad (ver *Tablas 3 y 4*).

De los participantes que no cumplieron con los criterios de TEA según DSM-5, 83 (56,08 %) recibieron otros diagnósticos del DSM-5 (sin contar DI) siendo el más frecuente el trastorno específico del lenguaje. De los que cumplieron criterios para TEA, 21 (12.7%) recibieron otros diagnósticos asociados (sin contar DI). La *Tabla 5* muestra los diagnósticos para todos los participantes. En términos de desarrollo del lenguaje, el 71.2 % no tenía lenguaje fluido (el 44.4 % era mínimamente verbal y el 26.6 % de la muestra total tenía lenguaje en frases); la diferencia entre los grupos puede verse en la *Tabla 6*.

El nivel de funcionamiento se determinó utilizando VABs-2 en la mayoría de los casos (91 %). En los centros en los que no se utilizaba VABs-2, el nivel de funcionamiento se estimó utilizando una de las medidas de coeficiente intelectual o de índice de desarrollo (ver "instrumentos de medición" en el apartado de Métodos). Para 19 participantes no se realizó una prueba formal de inteligencia o funcionamiento y su nivel adaptativo fue estimado clínicamente en la visita a la clínica utilizando información sobre habilidades del desarrollo adquiridas de evaluaciones y reportes previos (por ejemplo, reportes escolares recientes, antecedentes de repitencia o de adaptaciones de contenido, concurrencia a escolaridad especial o índices de desarrollo previos). El nivel de funcionamiento adaptativo identificó que el 47.6 % ( $n=149$ ) no tenía DI (CI por encima de 70 o funcionamiento adaptativo en el

**Tabla 2.** Datos demográficos de la muestra: sexo

Sexo / Dx	TEA		No-TEA		Total	
Varón	139	84.2 %	98	65.5 %	237	75.6 %
Mujer	26	15.8 %	50	34.5 %	76	24.4 %
Otro	-	-	-	-	-	-
Total	165	52,7 %	148	47,3 %	313	100 %

**Tabla 3.** Datos demográficos de la muestra: edad

Sexo / Dx	TEA (n: 165)	No-TEA (n: 148)	Total (n: 313)
Mínima	2	2	2
Máxima	23	31	31
Media	5.8	6.5	6.1
Desvío estándar	4	5.4	4.8

**Tabla 4.** Datos demográficos de la muestra: frecuencias por años de edad

Edad en años	Frecuencia	%
2	43	13.7
3	82	26.3
4	45	14.6
5	28	8.9
6	15	4.8
7	15	4.8
8	12	3.8
9	15	4.8
10	6	1.9
11	8	2.5
12	10	3.2
13	4	1.3
14	10	3.2
15	4	1.3
16	3	1.0
17	3	1.0
18	2	.6
19	1	.3
20	2	.6
23	2	.6
24	1	.3
26	1	.3
31	1	.3
Total	313	100.0

**Tabla 5.** Datos demográficos de la muestra: diagnósticos asociados (sin considerar discapacidad intelectual)

Diagnósticos asociados	TEA		No-TEA		Total	
<i>Sin otro diagnóstico psiquiátr.</i>	144	87.3 %	65	43.9 %	209	66.8 %
<b>Tr. lenguaje</b>	<b>13</b>	<b>7.9 %</b>	<b>32</b>	<b>21.6 %</b>	<b>45</b>	<b>14.5</b>
<b>Tr. conducta</b>	<b>1</b>	<b>.6 %</b>	<b>12</b>	<b>8.1 %</b>	<b>13</b>	<b>4.1</b>
<b>TDAH</b>	<b>3</b>	<b>1.8 %</b>	<b>13</b>	<b>8.8 %</b>	<b>16</b>	<b>5.2</b>
Trastorno de ansiedad	1	.6 %	11	7.4 %	12	3.8
TEPT	-	0 %	3	2.0 %	3	1.0
Trastornos afectivos	1	.6 %	6	4.1 %	7	2.2
Trastorno de aprendizaje	1	.6 %	1	.7 %	2	.6
Esquizofrenia	-	0 %	2	1.4 %	2	.6
Trastorno pragmático	-	0 %	1	.7 %	1	.3
Trastorno de personalidad	-	0 %	1	.7 %	1	.3
Trastorno de la coordinación	-	0 %	1	.7 %	1	.3
TOC	1	.6 %	-	0 %	1	.3
Total	165	100 %	148	100 %	313	100 %

**Tabla 6.** Datos demográficos de la muestra: nivel de lenguaje

Nivel de lenguaje	TEA		No-TEA		Total	
No verbal–Palabras sueltas (1)	99	60.4 %	40	27 %	139	44.6 %
Frases breves (2)	35	21.3 %	48	32.4 %	83	26.6 %
Fluido (habla de otro lugar y tiempo) (3)	31	18.3 %	60	40.6 %	91	28.8 %

**Tabla 7.** Datos demográficos de la muestra: nivel de funcionamiento

Nivel de funcionamiento	TEA		No-TEA		Total	
Normal o superior (ej. CI >70) (3)	54	32.7 %	96	64.9 %	150	47.6 %
Bajo (ej.: CI 69-40) (2)	75	45.5 %	44	29.7 %	119	38.1 %
Muy bajo (ej. CI <39) (1)	36	21.8 %	8	5.4 %	44	14.3 %

rango normal o superior), el 38.1 % (n=120) tenía DI leve a moderada, mientras que el 14.3 % (n=44) tenía una DI severa a profunda o un funcionamiento adaptativo muy bajo (ver nivel de funcionamiento según grupo TEA o no-TEA en *Tabla 7*).

La confiabilidad entre evaluadores se evaluó utilizando el coeficiente kappa de Cohen con 32 participantes (23 hombres, 9 mujeres) de edades comprendidas entre 2 y 16 años ( $M = 4,06$ ;  $DE = 2.76$ ) para diferenciar TEA de no-TEA entre 2 evaluadores independientes, arrojando un valor de kappa de .97

### Consistencia interna

La homogeneidad de los ítems de AMSE y su consistencia interna se determinaron utilizando el alfa de Cronbach. Los ocho ítems de AMSE obtuvieron un coeficiente alfa de .75, lo que sugiere una consistencia interna moderada (Cicchetti, 1994). La eliminación del ítem 5 (lenguaje pragmático) aumenta el coeficiente alfa a .80. Esto podría deberse al hecho de que la puntuación del ítem 5 está condicionada por el puntaje obtenido en el ítem 4 (nivel de lenguaje). Cada sujeto que obtuvo un nivel de lenguaje (ítem 4) peor que 0 (lenguaje fluido) automáticamente obtuvo un 0 en el ítem 5 (tal como indica el manual de puntuación de AMSE), lo que llevó a un mayor porcentaje de puntuación de 0 en el ítem 5 en la muestra (ver *Tabla 8*).

Las distribuciones de las puntuaciones de AMSE de los dos grupos, TEA y No-TEA se resumen en la *Tabla 9*. Las puntuaciones totales de AMSE para los participantes que cumplen con los criterios del DSM-5 de TEA oscilaron entre 1 y 14 puntos ( $M = 7.18$ ;  $DE = 2.44$ ); mientras que aquellos que no cumplieron con los criterios oscilaron entre 0 y 9 ( $M = 1.83$ ,  $DE = 1.80$ ). Al utilizar el diagnóstico de consenso de TEA o No-TEA, las puntuaciones de AMSE presentaron diferencias estadísticamente significativas ( $t(313) =$

22.1,  $p < .001$ ). Asimismo se observaron diferencias según el nivel de funcionamiento [ $F(2, 313) = 51.35$ ,  $p < .001$ ], nivel de lenguaje [ $F(2, 313) = 53.98$ ,  $p < .001$ ] y el sexo ( $t(313) = 4.52$ ,  $p < .001$ ). Se observaron puntuaciones más altas de AMSE en los participantes diagnosticados con TEA, aquellos con DI, los menos fluidos en el lenguaje y los varones.

### Precisión de AMSE basado en el diagnóstico final de consenso

La precisión diagnóstica del AMSE se evaluó mediante la medida no paramétrica del área bajo la curva de la característica operativa del receptor (ROC). El análisis de la curva ROC se utilizó para determinar el punto de corte óptimo para AMSE en comparación con un diagnóstico final de TEA utilizando el consenso clínico basado en los criterios del DSM-5. Del total de la muestra de 313 participantes, 165 (52 %) cumplieron con los criterios de TEA utilizando el consenso clínico. El área bajo la curva ROC (AUC) fue de 0,96 [intervalo de confianza (IC) del 95 % 0,94–0,98] (*Figura 1*). El puntaje de corte de AMSE que maximizó el acuerdo con el diagnóstico clínico y minimizó la diferencia entre sensibilidad y especificidad en la muestra completa fue mayor o igual a 5 ( $\geq 5$ ) (ver *Tabla 10*, *Figura 1*). Con este puntaje de corte, el AMSE arrojó una sensibilidad del 90,71 % y una especificidad del 92,17 %.

### Precisión de AMSE en comparación con el resultado de ADOS

ADOS fue implementado para 249 individuos (79 % de la muestra total) como parte del proceso de diagnóstico. La diferencia de tiempo entre la toma de AMSE y ADOS nunca fue superior a 30 días. En 64 individuos incluidos en el estudio (20 %), no se realizó una entrevista ADOS ya que no era clínicamente necesario (por ejemplo, la evaluación clínica fue su-

Tabla 8. Consistencia Interna AMSE

	Media	DS	Correlación elemento-total corregida	Alfa de Cronbach (si se elimina el elemento)
AMSE_1. Contacto visual	.3270	.54516	.570	.709
AMSE_2. Interés en los demás	.5460	.64364	.558	.705
AMSE_3. Habilidad de señalamiento	.4508	.74864	.531	.707
AMSE_4. Lenguaje	.7556	.74479	.425	.728
AMSE_5. Pragmática del lenguaje	.2635	.66565	-.028	.800
AMSE_6. Comp. repetitivos/estereotipias	1.0413	.90375	.578	.696
AMSE_7. Preocupaciones	.5810	.75013	.493	.715
AMSE_8. Sensibilidades inusuales	.6825	.63910	.520	.712

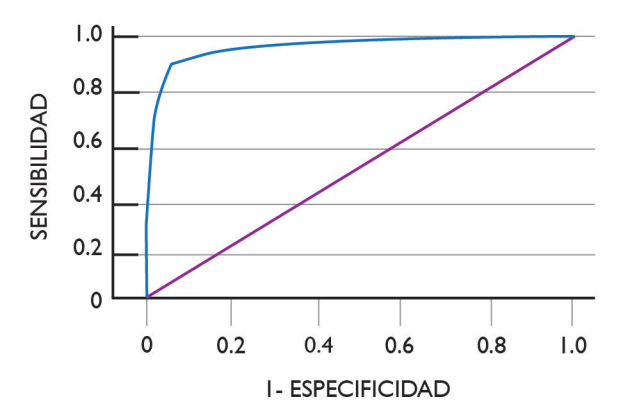
Tabla 9. Distribución de puntuaciones de AMSE

Ítem	TEA		No-TEA	
	x	SD	x	SD
1. Contacto visual	.57	.63	.05	.23
2. Interés en los demás	.85	.66	.20	.42
3. Habilidad de señalamiento	.77	.86	0.9	.32
4. Lenguaje	1.03	.77	.45	.59
5. Pragmática del lenguaje	.40	.79	.12	.44
6. Comportamientos repetitivos/estereotipias	1.61	.65	.42	.71
7. Preocupaciones	.97	.78	.16	.40
8. Sensibilidades inusuales	.99	.56	.35	.55

Tabla 10. Sensibilidad y especificidad de los puntajes de corte del AMSE para el diagnóstico clínico de mejor estimación de TEA basado en los criterios DSM-5.

Punto de corte	Sensibilidad	Especificidad
.5	1.0	.69
1.5	.99	.47
2.5	.98	.34
3.5	.96	.20
4.5	.90	.07
5.5	.73	.03
6.5	.55	.01
7.5	.42	.01
8.5	.29	.01
9.5	.16	0
10.5	.09	0
11.5	.05	0
12.5	.03	0
13.5	.01	0
15	0	0

Figura 1. Curva ROC para AMSE basado en el diagnóstico de consenso DSM-5



Área bajo la curva: variables de resultados de contraste AMSE total

Área	Error tip.	Sig. asintótica	Intervalo de confianza asintótico al 95%	
			Límite superior	Límite inferior
.962	.010	.000	.943	.982



ficiente para llegar a un claro diagnóstico según los criterios del DSM-5) (ver Tabla 11). El 64 % (n=160) cumplió con los criterios para el TEA según ADOS. La relación entre las puntuaciones totales de AMSE y la clasificación de ADOS también se examinó utilizando un análisis de curva ROC. El área bajo la curva ROC fue de .93 (IC 95 % [.90-.96]) (ver Tabla 12 y Figura 2) y una puntuación de  $\geq 5$  produjo un punto de corte con una sensibilidad del 88 % y una especificidad del 85 % (ver Tabla 8). Cuando se utilizó solo el ADOS para el diagnóstico, identificó el 95 % de verdaderos positivos y el 90.8 % de verdaderos negativos en comparación con el diagnóstico clínico final de consenso [ $\chi^2(1, 247) = 181.85, p < 0.001$ ].

**Precisión de AMSE basado en el diagnóstico final de consenso estratificando la muestra**

Se analizó si la variación del punto de corte según sexo, diferentes edades, niveles de funcionamiento adaptativo o niveles de lenguaje podría mejorar la sensibilidad y/o especificidad de AMSE. El cambio de punto de corte según el sexo no produjo diferencias significativas en la sensibilidad o especificidad de AMSE en base al diagnóstico clínico final de consenso. Tampoco se logró optimizar la confiabilidad de la escala variando el punto de corte mediante la división de la muestra según el nivel de lenguaje o de funcionamiento adaptativo. En el caso de la estratificación por edad no se vieron mejoras significativas al cambiar el punto de corte. Sin embargo, al combinar algunas de estas variables, sí se pudieron hallar subgrupos con potencialidad para mejorar la sensibilidad o la especificidad, bajando o elevando el punto de corte (respectivamente). Por ejemplo, si para el grupo de sujetos mayores de 13 años con nivel de funcionamiento promedio a superior se utiliza un puntaje de corte  $\geq 4$ , arroja un 95 % de sensibilidad de AMSE sin disminuir la especificidad, lo cual mejoraría la precisión al compararlo con el uso del punto de corte de 5 para toda la población.

**Discusión y conclusiones**

Este estudio tuvo como objetivo primario investigar la capacidad de la versión en español de AMSE para diferenciar correctamente a los pacientes con TEA de los no-TEA al compararlo con el diagnóstico clínico experto basado en los criterios del DSM-5, en una muestra de niños, adolescentes y adultos (n:313) que consultaron en 5 servicios de salud mental en cuatro ciudades de Argentina. Los resultados determinaron que una puntuación total de  $\geq 5$  es el corte óptimo de AMSE para individuos con características de TEA con una muy buena sensibilidad y especificidad. Aumentar el puntaje de corte a  $\geq 6$  mejoró la especificidad a 97 % pero redujo la sensibilidad de AMSE al 72 %. La versión en español del AMSE identificó el 90,3 % de los verdaderos positivos (TEA) (valor predictivo positivo) y el 92,6 % de los verdaderos negativos (No TEA) (valor predictivo negativo) en comparación con el diagnóstico clínico final de consenso [ $\chi^2(1, 313) = 214,42, p < 0,001$ ]. Un punto de corte de  $\geq 5$  también se determinó como el óptimo al comparar las puntuaciones totales de AMSE con los resultados de ADOS en términos de diferenciar TEA de no-TEA en una submuestra (objetivo secundario).

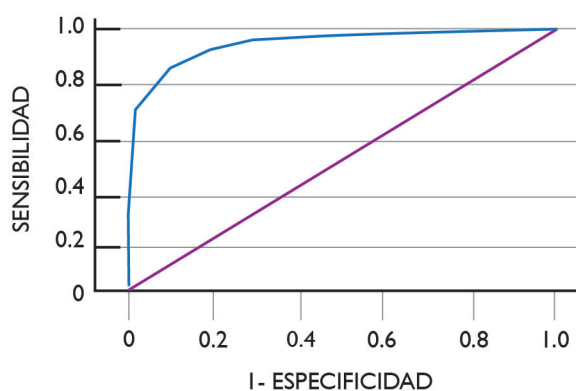
Se utilizó una muestra estratificada a partir de cuatro características, edad, sexo, nivel de lenguaje y nivel de funcionamiento adaptativo. Se evaluó si la variación del punto de corte según estas variables podría optimizar la sensibilidad y/o especificidad de AMSE. Para esto, se dividió la muestra en 54 grupos, combinando: cuatro categorías etarias (0-30 meses, 31 meses – 5 años, 6-12 años y  $\geq 13$  años), dos categorías sexuales (varón y mujer, ya que no hubo en esta muestra otras variantes), tres niveles de funcionamiento adaptativo (muy bajo, bajo y promedio a superior) y tres niveles de lenguaje expresivo (sin lenguaje a palabras sueltas, frases no complejas y lenguaje fluido). Se ana-

**Tabla 11. Diagnóstico según ADOS.**

Diagnóstico (TEA / no-TEA)	Frecuencia	%	% sobre los que tienen ADOS TOMADA (n:249)
ADOS No puntúa TEA	89	28.3	35.7
ADOS Puntúa TEA	160	50.8	64.3
Total	249	79.0	100.0
SIN ADOS TOMADA	64	20.0	
Total	313	100.0	

**Tabla 12.** Sensibilidad, especificidad y puntaje de corte del AMSE en comparación con el diagnóstico según ADOS.

Punto de corte	Sensibilidad	Especificidad
.5	1.0	.81
1.5	.99	.61
2.5	.98	.47
3.5	.94	.31
4.5	.88	.16
5.5	.73	.03
6.5	.55	.02
7.5	.42	.01
8.5	.28	.01
9.5	.14	.01
10.5	.08	.0
11.5	.05	.0
12.5	.04	.0
13.5	.01	.0
15.	.0	.0

**Figura 2.** Curva ROC para AMSE basada en el diagnóstico de TEA según ADOS

Los segmentos diagonales con productos por los empates

lizó si alguna de estas 54 combinaciones (por ejemplo “varones – preescolares – con nivel de funcionamiento muy bajo – y nivel de lenguaje hasta palabras sueltas”) podría tener un punto de corte diferente (por ejemplo 4) que mejore la sensibilidad o la especificidad (como se describe en Suren -2019- para el Cuestionario de Comunicación Social (SCQ), donde se mostró que para los niños más pequeños, descendiendo el punto de corte de SCQ mejoraba la sensibilidad de la escala). Al analizar para cada subgrupo los resultados de elevar a 6 o disminuir a 4 el punto de corte, casi no se hallaron subgrupos que aumentaron la especificidad de la escala sin empeorar la sensibilidad, ni que me-

joraron la sensibilidad sin perjudicar la especificidad. La excepción fue el subgrupo de  $\geq 13$  años con nivel de funcionamiento normal a superior (independientemente del nivel de lenguaje o sexo), en el cual un punto de corte de 6 podría mejorar la especificidad de la escala disminuyendo el número de falsos positivos; arrojando un 95 % de especificidad de AMSE sin disminuir la sensibilidad, lo cual mejoraría la precisión al compararlo con el uso del punto de corte de 5 para toda la población. Sin embargo, este subgrupo de sujetos de 13 años o más sin DI en esta muestra es demasiado pequeño (15 casos) como para basarse en esta submuestra para sugerir cambiar el punto de corte para los adolescentes y adultos sin DI en la utilización futura de la escala. Deberá investigarse esta posibilidad con un número mayor de adolescentes y adultos en un estudio futuro.

La muy buena sensibilidad y especificidad encontradas respalda tanto la utilidad clínica de AMSE como su capacidad para apoyar los diagnósticos de TEA en el contexto de protocolos de investigación. Es importante destacar que los sujetos participantes en este estudio fueron derivados a los centros de salud en base a preocupaciones en relación a la comunicación social y/o la flexibilidad, lo cual enfatiza la relevancia clínica del instrumento, destinado a individuos para quienes hay razones para creer que el resultado de la prueba puede ser positivo para TEA. La confiabilidad inter-evaluador resultó excelente ( $\kappa = .97$ ) en coincidencia con estudios anteriores (Grodberg et al. 2012). Se trató del primer estudio de validación de la versión de AMSE en español; el proyecto de este estudio se presentó en el congreso de la Sociedad Internacional de Investigación en Autismo (INSAR) de 2016 en Salt Lake City, los resultados preliminares en el congreso regional de INSAR en Puerto Varas, Chile, en 2019 y los resultados principales fueron presentados en el congreso de INSAR de 2022 en Austin, Texas.

A pesar de los resultados alentadores en cuanto a sensibilidad, especificidad y confiabilidad inter-evaluador, es fundamental subrayar que AMSE no debe considerarse un sustituto de una evaluación diagnóstica completa. Su valor radica en su utilidad como herramienta de apoyo al juicio clínico, especialmente en contextos donde el acceso a instrumentos más complejos como ADOS es limitado. Por ello, se recomienda su uso por profesionales con experiencia en el diagnóstico de TEA, como parte de un abordaje diagnóstico integral que contemple múltiples fuentes de información.

Por último, al revisar los estudios anteriores de validación de AMSE, se observó que los respectivos análisis estadísticos arrojaron diferentes puntos de corte óptimos (ver Tabla 1). Estas diferencias podrían deberse a diferencias en la composición de las muestras (diferentes distribuciones de edad, de nivel de funcionamiento y lenguaje, condiciones asociadas, pacientes con o sin sospecha de TEA, entre otros); a diferencias culturales entre las regiones; a diferentes nosografías y/o escalas utilizadas como estándar de comparación; a posibles diferencias en las traducciones y adaptaciones de AMSE o en los entrenamientos recibidos para el uso adecuado de la escala; entre otros factores. Sería importante realizar nuevos estudios con muestras estratificadas con tamaños sub muestrales adecuados, con población de diferentes regiones, teniendo en cuenta todos estos factores.

## Limitaciones

Aunque este estudio demuestra la alta precisión de la clasificación de AMSE, se deben abordar algunas limitaciones. En primer lugar, si bien los promedios de edad son similares entre los grupos de TEA y no-TEA en la muestra utilizada, se encontraron diferencias significativas entre los grupos, tanto en la distribución entre sexos (84 % de varones en el grupo TEA versus 65 % en el grupo no-TEA), como en el nivel de lenguaje (60,4 % sin lenguaje verbal o palabras sueltas en el grupo TEA versus 27 % en el grupo no-TEA) y en el nivel de funcionamiento (32,7 % del grupo TEA tienen CI de 70 o mayor; versus 64 % en el grupo no-TEA). Estas diferencias pudieron haber sesgado los resultados. A su vez, la representación de adolescentes y adultos (de 13 años o más) en la muestra es relativamente baja (n:34, 11,7 %) en comparación con la de niños ( $\leq 12$  años), lo que afecta comparativamente la confiabilidad de los resultados para aquella franja etaria (aunque en el estudio previo de Grodberg et al. (2014) con adultos con lenguaje fluido, AMSE había mostrado muy buena confiabilidad utilizando un punto de corte de 5). Este punto deberá ser abordado en investigaciones a futuro considerando la escasez de herramientas confiables para el apoyo al diagnóstico en adultos y la falta de consenso para evaluar el autismo en esta población (ver por ejemplo Conner et al., 2019), teniendo en cuenta que AMSE tiene potencial como medida estandarizada rápida para guiar el juicio clínico en mayores de 17 años. Otra limitación es que, si bien la muestra consta de 313 casos, la generalización de los resultados sigue siendo limitada a partir de

que la cantidad de casos en cada subgrupo continúa siendo reducida. Estudios futuros deben examinar el uso del AMSE en muestras más grandes con estratificación por sexo, edad, diagnósticos asociados, nivel de lenguaje y nivel de funcionamiento cognitivo y adaptativo, que permitan un adecuado tamaño muestral para cada uno de los subgrupos resultantes de la combinación de esas categorías.

Por otro lado, es importante mencionar que a pesar de la buena confiabilidad que muestran los resultados de este estudio, se encontró un porcentaje de falsos positivos y falsos negativos utilizando AMSE, por lo cual sigue siendo crítico que en los casos en los que un clínico no tenga experiencia en el diagnóstico de TEA o cuando surge una presentación clínica complicada, AMSE no debe sustituir una evaluación diagnóstica completa. En tales casos, se deben realizar las derivaciones necesarias para una evaluación adicional. También es importante destacar que el nivel de sensibilidad y especificidad encontrado en este estudio depende del uso confiable del AMSE. Se necesita más investigación para determinar las propiedades psicométricas de AMSE entre los clínicos con diferentes niveles de experiencia y pericia. Se recomienda que AMSE no sea utilizado por clínicos sin experiencia diagnóstica específica en TEA. Aunque los clínicos con menos experiencia pueden tener la capacidad de aprender rápidamente cómo administrar el AMSE, la puntuación y la interpretación sólo pueden ser realizadas por clínicos que tengan experiencia en realizar diagnósticos de TEA en primer lugar.

Los resultados de este estudio respaldan el uso de AMSE como una herramienta diagnóstica breve, confiable y culturalmente adaptada para apoyar el diagnóstico clínico de TEA en contextos clínicos de habla hispana. Su alta sensibilidad y especificidad, junto con su facilidad de administración y bajo costo, lo posicionan como una alternativa especialmente valiosa en regiones donde el acceso a instrumentos como ADOS es limitado por barreras económicas, geográficas o de capacitación profesional. En este sentido, se recomienda la incorporación de AMSE como parte de protocolos diagnósticos escalonados, en los que pueda utilizarse como herramienta de apoyo al juicio clínico en la primera línea de atención. Su implementación puede facilitar la identificación temprana de casos sospechosos de TEA, optimizando la derivación a evaluaciones más exhaustivas cuando sea necesario. Su implementación, puede contribuir a reducir la brecha diagnóstica en autismo, especialmente en poblaciones vulnerables o desatendidas.

**Conflicto de intereses:** los autores declaran no tener conflictos de intereses. Este estudio no contó con financiación de ningún tipo.

**Agradecimientos:** a todos los participantes y a sus familias, así como a los psiquiatras, pediatras, psicólogos y demás profesionales de la salud y coordinadores de investigación que ayudaron con la recopilación de datos, con la corrección del manuscrito final y con otras tareas y apoyo logístico durante el desarrollo del estudio. En particular queremos agradecer a Luciana Arancibia, Soledad Burrone, Claudia Cabrera, Arianna Gervaso, Ivana Gizzi, Ignacio Gómez de la Fuente, Brenda Grosskopf, Florencia Frizzera, Lorena Giselle Vetere, Pierina Landolfi, Mariela López, Gabriela Lupo, Anahí Luque, María Gabriela Morell, Ramiro Pérez Martín, Analía Prieto, Florencia Salaberry, Silvina Schwemm, Magdalena Urruspuru, Damián Ursino, Bernardo Vilariño, Antonella Zampini, Julio Cukier y Fundación Ados.

## Referencias bibliográficas

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. American Psychiatric Publishing.
- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders (5th ed., text rev.)*. <https://doi.org/10.5209/psic.84045>
- Betz, E., Hackman, N. M., Mayes, S., Chin, E. M., Ivy, J. W., & Tierney, C. (2019). Validity of the Autism Mental Status Exam in developmental pediatrics and primary care settings. *Global Pediatric Health*, 6, 2333794X19847905. <https://doi.org/10.1177/2333794X19847905>
- Cederlund, M. (2019). Autism Mental Status Examination (AMSE): A valid instrument in the evaluation of pre-school children with suspected autism spectrum disorders? *Journal of Autism and Developmental Disorders*, 49(7), 2965–2979. <https://doi.org/10.1007/s10803-019-04012-1>
- Chiarotti, F., & Venerosi, A. (2020). Epidemiology of autism spectrum disorders: A review of worldwide prevalence estimates since 2014. *Brain Sciences*, 10(5), 274. <https://doi.org/10.3390/brainsci10050274>
- Cicchetti, D. V., Volkmar, F., Klin, A., & Showalter, D. (1995). Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. *Child Neuropsychology*, 1(1), 26–37. <https://doi.org/10.1080/09297049508401340>
- Conner, C. M., Cramer, R. D., & McGonigle, J. J. (2019). Examining the diagnostic validity of autism measures among adults in an outpatient clinic sample. *Autism in Adulthood*, 1(1), 60–68. <https://doi.org/10.1089/aut.2018.0023>
- Galdino, M. P., Pegoraro, L. F. L., Saad, L. O., Grodberg, D., & Celeri, E. H. R. V. (2020). Evidence of validity of the Autism Mental Status Examination (AMSE) in a Brazilian sample. *Journal of Autism and Developmental Disorders*, 50(7), 2320–2325. <https://doi.org/10.1007/s10803-018-3530-0>
- Gordon, M., & Miller, L. (2013). *Leiter International Performance Scale – Third Edition (Leiter-3)*. Stoelting Co.
- Grodberg, D., Siper, P., Jamison, J., Buxbaum, J. D., & Kolevzon, A. (2016). A simplified diagnostic observational assessment of autism spectrum disorder in early childhood. *Autism Research*, 9, 443–449. <https://doi.org/10.1002/aur.1523>
- Grodberg, D., Weinger, P. M., Halpern, D., Parides, M., Kolevzon, A., & Buxbaum, J. D. (2014). The Autism Mental Status Exam: Sensitivity and specificity using DSM-5 criteria for autism spectrum disorder in verbally fluent adults. *Journal of Autism and Developmental Disorders*, 44(3), 609–614. <https://doi.org/10.1007/s10803-013-1917-5>
- Grodberg, D., Weinger, P. M., Kolevzon, A., Soorya, L., & Buxbaum, J. D. (2012). Brief report: The Autism Mental Status Examination: Development of a brief autism-focused exam. *Journal of Autism and Developmental Disorders*, 42(3), 455–459. <https://doi.org/10.1007/s10803-011-1255-4>
- Irrázaval, M., López, I., Figueroa, C., et al. (2023). Adaptation and validation of the Autism Mental Status Examination (AMSE) in Chile: Seeking to reduce the diagnostic gap. *Andes Pediátrica*, 94(4), 475–484. <https://doi.org/10.32641/andespediatr.v94i4.4476>
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223. <https://doi.org/10.1023/A:1005592401947>
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule (2nd ed.)*. 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.
- Mullen, E. M. (1995). *Mullen Scales of Early Learning*. American Guidance Service Inc.
- Øien, R. A., Siper, P., Kolevzon, A., & Grodberg, D. (2020). Detecting autism spectrum disorder in children with ADHD and social disability. *Journal of Attention Disorders*, 24(7), 1078–1084. <https://doi.org/10.1177/1087054716642518>
- Øien, R. A., Vambheim, S. M., Hart, L., Nordahl-Hansen, A., Erickson, C., Wink, L., Eisemann, M. R., Shic, F., Volkmar, F. R., & Grodberg, D. (2018). Sex differences in children referred for assessment: An exploratory analysis of the Autism Mental Status Exam (AMSE). *Journal of Autism and Developmental Disorders*, 48(7), 2286–2292. <https://doi.org/10.1007/s10803-018-3516-y>
- Paula, C. S., Cukier, S., Cunha, G. R., Irrázaval, M., Montiel-Nava, C., García, R., Rosoli, A., Valdez, D., Bordini, D., Shih, A., Garrido, G., & Rattazzi, A. (2020). Challenges, priorities, barriers to care, and stigma in families of people with autism: Similarities and differences among six Latin American countries. *Autism*, 24(8), 2228–2242. <https://doi.org/10.1177/1362361320931229>
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview—Revised (ADI-R)*. Western Psychological Services.
- Schopler, E., Van Bourgondien, M., Wellman, G., & Love, S. (2010). *The Childhood Autism Rating Scale – Second Edition (CARS-2)*. Western Psychological Services.
- Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland Adaptive Behavior Scales (3rd ed.)*. Pearson.
- Surén, P., Saasen-Havdahl, A., Bresnahan, M., Hirtz, D., Hornig, M., Lord, C., Reichborn-Kjennerud, T., Schjølberg, S., Øyen, A.-S., Magnus, P., Susser, E., Lipkin, W. I., & Stoltenberg, C. (2019). Sensitivity and specificity of early screening for autism. *BJPsych Open*, 5(3), e41. <https://doi.org/10.1192/bjo.2019.34>
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale (4th ed.)*. Pearson.
- Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence (4th ed.)*. Pearson.
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children (5th ed.)*. Pearson.



Yang, S., Han, D., Zhou, H., Yang, C., Zhang, K., Chen, S., Yang, R., Cao, X., Grodberg, D., Zhao, X., & Kang, C. (2023). Validity and Cutoff Score of the Autism Mental Status Exam for an Autism Spectrum Disorder Diagnosis in Chinese Children. *Journal of autism and developmental disorders*, 53(12), 4822–4829. <https://doi.org/10.1007/s10803-022-05730-9>

Zeidan, J., Fombonne, E., Scora, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, 15(5), 778–790. <https://doi.org/10.1002/aur.2696>

Zhang, M., & Hudson, J. A. (2018). The development of temporal concepts: Linguistic factors and cognitive processes. *Frontiers in Psychology*, 9, 2451. <https://doi.org/10.3389/fpsyg.2018.02451>