



UNIVERSIDAD AUTÓNOMA DE ZACATECAS
UNIDAD ACADÉMICA DE DOCENCIA SUPERIOR
GRUPO DE INVESTIGACIÓN ESTUDIOS SOCIOEDUCATIVOS Y CULTURALES

e-ISSN: 2992-751X

Zacatecas, Zacatecas, México
Correo electrónico: iie_docencia@uaz.edu.mx
Disponible en <https://revistas.uaz.edu.mx/index.php/invinted>



Evaluación de la efectividad de los puntos de corte en la prueba RAVEN para la identificación de alumnos con aptitudes sobresalientes en México

Evaluation of the Effectiveness of Cut-off Points in the RAVEN Test for the Identification of Students with Gifted Abilities in Mexico

Luis Miguel Cortés Costilla
Secretaría de Educación Pública
ORCID: <https://orcid.org/0009-0002-7118-7389>

Enrique Hernández Arteaga
Universidad Autónoma de Tlaxcala
ORCID: <https://orcid.org/0000-0003-3929-6200>

Claudia Yolanda Ahumada-Ruíz
Secretaría de Educación Pública
ORCID: <https://orcid.org/0009-0008-0914-6926>

DOI: <https://doi.org/10.71770/18p7zq73>

Recibido: 30/10/2024 | Aceptado: 21/01/2025

RESUMEN

La baja identificación de alumnos con Aptitudes Sobresalientes (AS) en México es un desafío, especialmente en áreas rurales y suburbanas. Esto se debe, en parte, a la falta de instrumentos validados para el contexto educativo. Es importante considerar el punto de corte óptimo al momento de la identificación, dado que este podría estar establecido por encima de las capacidades de algunos estudiantes. Las evaluaciones tradicionales, con percentiles elevados, pudieran limitar la identificación adecuada de estos alumnos, limitando su identificación. Método: Participaron 130 estudiantes de escuelas públicas en Zacatecas, a quienes se les aplicó RAVEN. Para validar los resultados, se compararon los puntajes de RAVEN con los puntajes de CI Estimado mediante el WISC-IV, considerado el Gold Standard. Se analizaron diferentes puntos de corte (percentiles 85, 90, 95, y 98) para determinar cuál ofrece la mejor precisión diagnóstica. Resultados: El percentil 90 en RAVEN ofrece el mejor equilibrio entre exactitud y sensibilidad para identificar estudiantes AS con un CI estimado de ≥ 115 , mientras que el percentil 95 parece ser adecuado para identificar estudiantes con un CI superior (≥ 120 y ≥ 125). El percentil mostró mayor exactitud, pero baja sensibilidad, lo que podría excluir a estudiantes con potencial. Conclusiones: Dependiendo el objetivo al momento de identificación es fundamental determinar el perfil a considerar, dado que el percentil 90, detecta a más alumnos, pero con falsos positivos y, los percentiles más altos (95 y 98) son eficientes para identificar estudiantes AS con mayor potencial y, dejando fuera aquellos alumnos que se pudieran potencializar.

PALABRAS CLAVE: identificación, prueba de inteligencia no verbal, inclusión, evaluación educativa

ABSTRACT

The low identification of Gifted children (GC) in Mexico is a challenge, especially in rural and suburban areas. This is partly due to the lack of validated instruments for the educational context. It is important to consider the optimal cut-off point when identifying students, as this could be set above the capabilities of some students. Traditional assessments, with high percentiles, may limit the adequate identification of these students. Method: 130 students from public schools in Zacatecas participated, and the RAVEN test was administered. To validate the results, RAVEN scores were compared with Estimated IQ scores obtained through the WISC-IV, considered the Gold Standard. Different cut-off points (percentiles 85, 90, 95, and 98) were analyzed to determine which offers the best diagnostic accuracy. Results: The 90th percentile on RAVEN provides the best balance between accuracy and sensitivity for identifying GC students with an estimated IQ of ≥ 115 , while the 95th percentile seems suitable for identifying students with higher IQs (≥ 120 and ≥ 125). The higher percentile showed greater accuracy but lower sensitivity, potentially excluding students with potential. Conclusions: Depending on the objective of identification, it is essential to determine the profile to be considered, as the 90th percentile identifies more students, though with some false positives, and the higher percentiles (95 and 98) are efficient for identifying GC students with greater potential, potentially leaving out those who could be further developed.

KEYWORDS: Identification, Non-verbal intelligence test, Inclusion, Educational assessment

1. Introducción

La identificación precisa de estudiantes con aptitudes sobresalientes (AS) representa un desafío crítico para el sistema educativo actual en México, particularmente en contextos rurales y suburbanos. Diversos estudios han demostrado que la identificación e intervención educativa apropiada contribuye significativamente a su desarrollo académico, emocional y social, además de reducir significativamente dificultades académicas y sociales (Donate & Borges, 2018; McBee et al., 2016; Pedersen et al., 2023; Steenbergen-Hu et al., 2020; Wu et al., 2019). Este estudio tiene como propósito identificar el percentil óptimo del test RAVEN que ofrece la mayor capacidad predictiva para identificar a estudiantes AS en Zacatecas, México, lo que podría contribuir en delimitar ciertas lagunas metodológicas en el proceso de identificación a nivel nacional, sobre todo en áreas rurales y suburbanas.

Con ello, el objetivo de este estudio es establecer puntos de corte inclusivos y equitativos que fortalezcan las políticas educativas nacionales y promuevan una identificación oportuna y precisa de estudiantes AS, especialmente en zonas escolares con menos oportunidades de enriquecimiento educativo. Estos hallazgos no solo ofrecen una contribución relevante al sistema educativo en México, sino que también se suman al debate internacional en torno a la identificación de estudiantes AS, ayudando a reducir desigualdades en el acceso a la educación de calidad para este grupo.

La Secretaría de Educación Pública [SEP], en México define a las y los alumnos con "Aptitudes Sobresalientes (AS)", como aquellos individuos que pueden distinguirse claramente del grupo social y educativo en una o más de las siguientes áreas: Intelectual, Artístico, Psicomotriz, Socio-afectiva y Creativa" (SEP, 2022, p.12). Por otro lado, la Asociación Nacional para Niños Superdotados (NAGC, por sus siglas en inglés), menciona que son aquellos que demuestran niveles elevados de aptitud, definidos como una capacidad elevada para razonar y aprender o competencia medida por pruebas de rendimiento o desempeño documentado igual o superior al 10% en uno o más dominios, que incluyen áreas como matemáticas, música, lenguaje y habilidades sensoriomotoras (National Association for Gifted Children, 8 de febrero 2023).

Sin embargo, uno de los principales desafíos a los que se enfrentan las y los alumnos AS es el proceso de identificación (Renzulli & Reis, 2021). VanTassel-Baska (2000) señala que la identificación es un tema vital, dado que es importante planificar los servicios especializados para estudiantes AS. Donde los principales desafíos se refieren a la conceptualización adoptada, el modelo teórico y el enfoque de identificación. El primero se refiere a la definición operacional que se utiliza a nivel local, nacional e internacional (McClain & Pfeiffer, 2012). La segunda al modelo teórico adoptado por las autoridades educativas (SEP, 2022) y, finalmente al enfoque de identificación, ya sea: cuantitativo o cualitativo (Ryser, 2018).

Pero, ¿cómo son detectados las y los alumnos AS? La Encuesta Nacional de Educación para Alumnos con Altas Capacidades (Kurtz, 2019), en el que participaron profesionales en educación especial, madres y padres de niñas y niños con algún tipo de AS, muestran que el 54% de estudiantes son identificados mediante el Cognitive

Abilities Test (CogAT, Lohman, 2001), el 40% con alguna escala de inteligencia de Wechsler (Wechsler, 1949), el 34% con el Naglieri Nonverbal Ability Test (NNAT, Naglieri, 1997) y el 7% con las Matrices Progresivas de Raven (RAVEN, Raven, 1938). Sin embargo, estos sistemas de identificación en ocasiones se ven sesgados por la carga verbal y cultural que implican, dejando fuera a las y los estudiantes con pocas posibilidades de capital cultural y de zonas rurales o suburbanas (Lohman, 2005).

La literatura actual enfrenta una controversia sobre el método más eficaz para la identificación de estudiantes AS, dividiéndose entre dos enfoques principales: cuantitativo y cualitativo (Ryser, 2018). Estas discusiones sobre los enfoques de identificación se centran en las diversas definiciones e instrumentos utilizados. Algunos investigadores priorizan definiciones basadas en aptitudes cognitivas (Gülsen Erden et al., 2020), mientras que otros apoyan una definición centrada en las manifestaciones de aptitudes o talentos (Gagné, 2017). Krisel (2012) enfatiza la importancia de considerar ambos métodos para lograr una identificación más integral y equitativa de los estudiantes AS.

El método cuantitativo, considerado el enfoque predominante para la identificación de estudiantes AS a nivel mundial, se basa principalmente en los puntajes del coeficiente intelectual (CI) y las pruebas de rendimiento académico (National Association for Gifted Children & The Council of State Directors of Programs for the Gifted, 2015). Este método establece como criterio una puntuación por encima del percentil 90 (10%), con el objetivo de asegurar que los resultados sean significativos y no producto de variaciones aleatorias en la población.

Las escalas de inteligencia de Wechsler (Wechsler, 2007) son utilizadas frecuentemente para identificar a estudiantes AS. En la mayoría de los casos, se considera que un estudiante es AS cuando obtiene un puntaje de CI de al menos 130 (± 2 DE) (Gülsen Erden et al., 2020) o, de manera más inclusiva, de 115 (± 2 DE) (Pfeiffer & Blei., 2008). Esta población representa aproximadamente el 2.2% del total estudiantil (Holoher-Ertl et al., 2008). Diversas investigaciones han demostrado la eficacia de las escalas de Wechsler para identificar a alumnas y alumnos AS (Guignard et al., 2016; Gülsen Erden et al., 2020; Molinero et al., 2015). Además, se ha observado una correlación moderada entre el CI y el rendimiento académico, con coeficientes que oscilan entre 0.50 y 0.60 en educación primaria y entre 0.40 y 0.50 en la preparatoria (Guez et al., 2018).

Sin embargo, diferentes escalas o instrumentos de inteligencia pueden arrojar resultados distintos para un mismo estudiante. Por ejemplo, Johnsen (2021) señala que una prueba podría otorgar al estudiante un puntaje de 130 (percentil 98), mientras que en otra podría obtener una puntuación de 110 (percentil 75). A pesar de que ambas pruebas cumplen con parámetros psicométricos sólidos, pueden diferir en los comportamientos o habilidades que evalúan, las definiciones que utilizan, los métodos de administración o los errores estándar de medición. Por esta razón, Covarrubias-Pizarro (2018) subraya que la definición operativa debe ser considerada y analizada cuidadosamente por el sector educativo. Esto es crucial, dado que, si se utiliza un punto de corte estricto de CI de 130, solo alrededor del 1% al 3% de la población será identificado como estudiante AS. En cambio, al adoptar un término más flexible, como

"estudiantes con aptitudes y talentos excepcionales", y ser más inclusivos, se podría aumentar la tasa de incidencia a aproximadamente un 20%.

A lo largo de la historia, las pruebas de inteligencia de Wechsler han demostrado ser herramientas efectivas para identificar a estudiantes AS (Wechsler, 2007). Sin embargo, es importante señalar que el CI no es el único indicador de la inteligencia o del potencial de un individuo (Alsuwailimi, 2024). La inteligencia es un constructo multifactorial que abarca habilidades emocionales, creativas y prácticas, las cuales no siempre se capturan en una prueba de CI. Además, no todas y todos los estudiantes AS demuestran estas habilidades en su vida diaria (Renzulli & Reis, 2021).

El uso de puntajes de pruebas estandarizadas de rendimiento académico y del CI como únicos criterios para identificar a estudiantes AS ha sido objeto de críticas en el campo de la educación especial. Se menciona que estos instrumentos no consideran las definiciones más recientes, o que se centran exclusivamente en un constructo, como la inteligencia, dejando de lado habilidades sociales, comunicativas, de liderazgo, entre otras (Alsuwailimi, 2024; Hodges & Gentry, 2021). Además, existe la preocupación de que algunas pruebas estandarizadas puedan presentar sesgos raciales o culturales (Flynn, 2023).

Las escalas de Wechsler han sido particularmente criticadas, lo que ha llevado a algunas y algunos investigadores a implementar escalas o instrumentos de evaluación no verbales. Estas pruebas no verbales han ganado popularidad a nivel mundial, ya que se consideran más inclusivas para poblaciones cultural o lingüísticamente diversas, debido a su menor carga cultural y menor demanda lingüísticas (Lee et al., 2021). Asimismo, muchos estudiantes AS pueden pasar inadvertidos debido a desigualdades en las oportunidades educativas, entornos sociales, diferencias culturales, rasgos físicos o comportamentales, así como a conflictos con la nominación libre (Rothenbusch et al., 2016), donde el docente selecciona a las y los estudiantes con mayor habilidad o talento para ser valorado como alumno AS.

El CogAT (Lohman & Nicpon, 2012), el NNAT (Naglieri, 1997) y RAVEN (Raven, 2000) son instrumentos ampliamente utilizados como evaluaciones no verbales para identificar a estudiantes AS, siendo reconocidos por su capacidad para evaluar habilidades intelectuales en diversos grupos raciales y económicos (Kurtz, 2019). Estas evaluaciones se basan en tareas no verbales, lo que las hace más inclusivas para estudiantes de diferentes orígenes étnicos, logrando que estos sean identificados (Alsuwailimi, 2024)

Naglieri and Ford (2003) fueron pioneros en evaluar la efectividad del NNAT y el RAVEN en la identificación de estudiantes AS de diversos antecedentes sociales y demográficos, utilizando una muestra de más de 20,000 estudiantes. Los autores exploraron las diferencias entre grupos, y sus hallazgos revelaron que el NNAT identificó proporciones similares de estudiantes AS en cada grupo racial: el 2.5% de los estudiantes caucásicos, el 2.6% de los estudiantes afrodescendientes y el 2.3% de los estudiantes latinos obtuvieron puntajes igual o superior al percentil 98, un umbral típico requerido para acceder a los programas AS. Concluyeron que el NNAT es una herramienta efectiva para identificar un rango amplio de estudiantes, independientemente de su grupo racial y económico.

Por otro lado, Lohman (2005) cuestiona la idea de que el NNAT sea una herramienta equitativa de identificación, argumentando que los hallazgos de Naglieri y Ford no han sido replicados. Señala que la muestra de su estudio no refleja la realidad educativa, ya que ciertos subgrupos raciales estaban significativamente subrepresentados. Además, indica que menos del 6% de los estudiantes provenían de distritos urbanos, y que un porcentaje mayor de participantes afrodescendientes y latinos pertenecían a entornos socioeconómicos altos en contraste con sus contrapartes caucásicas. Lohman sostiene que:

Si bien las pruebas no verbales pueden parecer reducir el sesgo, cuando se utilizan de manera aislada, en realidad lo agravan, al pasar por alto tanto a estudiantes de minorías como a estudiantes mayoritarios que demuestran un rendimiento académico actual o que tienen el potencial de beneficiarse de una educación enriquecida (2005, p. 26).

Investigaciones posteriores han reforzado la crítica de Lohman, destacando correlaciones significativas entre la etnia de las y los estudiantes, su estatus socioeconómico y su rendimiento en el NNAT. Por ejemplo, un estudio realizado por Carman y Taylor (2010) analizó la afirmación de que el NNAT es una medida "culturalmente neutral" de la capacidad académica, utilizando una muestra de 2,000 estudiantes de preescolar en un distrito escolar del sur de Estados Unidos, donde el NNAT era obligatorio para la identificación de alumnas y alumnos AS. Los investigadores encontraron vínculos sustanciales entre la raza de los estudiantes, su origen socioeconómico y sus puntajes en el NNAT.

Un análisis de regresión múltiple demostró que los grupos minoritarios subrepresentados obtuvieron puntajes notablemente más bajos que sus compañeros. Además, las y los estudiantes de familias de nivel socioeconómico medio a alto tenían el doble de probabilidades de ser identificados como AS basándose en los puntajes del NNAT, en comparación con las y los estudiantes de familias de bajo nivel socioeconómico. Los investigadores concluyeron: "este análisis de los puntajes del NNAT muestra que, al menos en este distrito, los resultados del NNAT no son tan libres de diferencias de grupo como se publicita" (Carman y Taylor (2010) p. 79).

A pesar de la controversia, NNAT y CogAT han ganado popularidad en el ámbito educativo para identificar a estudiantes AS y al momento de tomar decisiones sobre su acceso a programas avanzados o la promoción escolar anticipada. Aunque ambos instrumentos son ampliamente reconocidos y utilizados, principalmente en los Estados Unidos, tanto en evaluaciones individuales como grupales (Carman, 2020), desafortunadamente no están disponibles para la población mexicana. En México, el único instrumento similar y accesible es la prueba RAVEN. Sin embargo, sus normas no han sido actualizadas.

En cuanto al RAVEN, Naglieri y Ford (2003) concluyeron que las normas desactualizadas del RAVEN provocaron una sobreidentificación de alumnas y alumnos AS en los primeros grados escolares, resaltando la necesidad urgente de actualizar los datos normativos. Resultados similares se observaron en estudios posteriores. Por ejemplo, Shaunessy (2004) y Lewis et al. (2007) indicaron que el RAVEN ha tendido a

sobreidentificar estudiantes, hipotetizando que es debido al Efecto Flynn. Este fenómeno, descrito por Flynn (2007), el cual sugiere que las normas de pruebas antiguas pueden inflar los puntajes, ya que el conocimiento general de la población aumenta con el tiempo, estimándose un incremento promedio de 3 puntos por cada 10 años transcurridos desde la publicación del instrumento.

Así, al usar pruebas como el RAVEN con normas desactualizadas puede llevar a que los puntajes altos no reflejen con precisión las habilidades del estudiantado en comparación con sus pares actuales. Además, los factores sociodemográficos influyen significativamente, ya que el puntaje de un estudiante está condicionado a la manera en que su entorno y las oportunidades se relacionan con la muestra utilizada para la estandarización de la prueba (Hoover, 2001).

A pesar de la falta de normas actualizadas para la población mexicana, el RAVEN sigue siendo ampliamente utilizado en México y Latinoamérica (Chávez-Soto et al., 2018; Navarro-Saldaña et al., 2022; Robles-Guerrero, 2022; Rodríguez Cervantes et al., 2022; Soto et al., 2014; Valadez-Sierra, 2012) debido a su facilidad de administración en entornos grupales, su simple corrección y su baja dependencia verbal, lo que lo convierte en una de las pruebas factibles (Valadez-Sierra, 2012) y es la novena prueba más utilizada en Estados Unidos (Kurtz, 2019).

Diversos estudios emplean el percentil 90 o superior en esta prueba como punto de corte para identificar a estudiantes AS (Rodríguez Cervantes et al., 2022). Un ejemplo de ello es el estudio realizado en Lima, Perú, por Navarro-Saldaña et al. (2022), que evaluó a 650 niños (51.53% mujeres y 47.84% hombres, con un rango de edades entre 11 y 17 años) utilizando el RAVEN para la detección de estudiantes AS. Los autores lograron identificar a 97 estudiantes, lo que representa el 14.2% de la muestra, demostrando así la efectividad del RAVEN en la identificación de estudiantes AS en esa población. Sin embargo, la alta tasa de detección pudiera mostrar la sobrestimación de alumnos AS.

De manera similar, Robles-Guerrero (2022) utilizó el RAVEN como prueba de detección para identificar a estudiantes AS en México. Este estudio, realizado con estudiantes de primer año de secundaria, se evaluó a un total de 162 participantes (79 hombres y 93 mujeres). El análisis identificó a 18 estudiantes que obtuvieron puntajes en el percentil 90 o superior, lo que representa el 12% de la muestra. Este hallazgo es consistente con estudios previos de Chávez-Soto et al. (2018) y Soto et al. (2014), quienes estiman que la prevalencia de estudiantes AS en educación básica oscila entre el 3% y el 20%. Como se ha señalado, el RAVEN puede considerarse una herramienta de evaluación no verbal confiable y válida, útil para evaluaciones tanto individuales como grupales en la identificación de estudiantes AS. Sus correlaciones con pruebas de inteligencia oscilan entre .50 y .80 (Sattler, 1992). Aunque actualmente no están disponibles puntajes estandarizados para la población mexicana del RAVEN, este instrumento sigue siendo ampliamente utilizado por profesionales de educación especial.

Este estudio se centra en el problema de cuál punto de corte sería el más óptimo para identificar estudiantes AS en México. Con el fin de desarrollar prácticas de identificación inclusivas y precisas que garanticen el reconocimiento y apoyo adecuados para estos estudiantes, el propósito de este trabajo es determinar qué

percentil del test RAVEN ofrece la mayor capacidad predictiva para identificar a alumnos con AS en Zacatecas, México, abordando así las lagunas metodológicas existentes en el proceso de identificación de estudiantes AS en el país. El objetivo es mejorar las prácticas de identificación en México, proponiendo ajustes en los puntos de corte que permitan una detección equitativa y representativa. Al ubicar estos hallazgos en el contexto más amplio de la identificación de estudiantes AS, esta investigación pretende ofrecer evidencia relevante para fortalecer las políticas educativas nacionales, promoviendo una identificación oportuna que permita a estos estudiantes recibir el apoyo necesario para su desarrollo pleno y, de esta forma, contribuir a reducir las desigualdades educativas en el país.

2. Metodología

Este es un estudio de tipo no experimental, descriptivo, transversal y, psicométrico. Las y los participantes fueron seleccionados a conveniencia de diferentes escuelas de la mancha urbana y de una comunidad del municipio de Tlaltenango de Sánchez Román perteneciente al estado de Zacatecas. De éstas; 4 fueron primarias, 1 secundaria técnica y 1 telesecundaria.

2.1. Participantes

Participaron 130 estudiantes mexicanos de habla hispana (54 mujeres, edad $M = 11,20$, $DE = 1,91$; 76 hombres, edad $M = 22,29$, $DE = 1,95$), tomados de una muestra de validación del test de RAVEN (en preparación). Las edades de las y los participantes oscilaron entre los 8 años y los 14 años y 11 meses (edad promedio = 11.25, $DE = 1,93$). Todas y todos estaban inscritos en escuelas públicas en el sur de Zacatecas, México. Las y los participantes fueron evaluados utilizando el test de RAVEN de forma grupal. El reclutamiento se realizó de agosto del 2022 a mayo del 2023.

Para determinar el Gold Standard, de la muestra de 130 participantes, se seleccionó un subgrupo de 30 participantes, que obtuvieron un puntaje de CI estimado ≥ 115 en la versión mexicana del WISC-IV (Wechsler, 2007), utilizando una versión abreviada y siguiendo el procedimiento de CI estimado descrito por Sattler (2009). Las y los participantes fueron agrupados según su CI estimado en cuatro rangos (≥ 115 , ≥ 120 , ≥ 125 , ≥ 130 , ver Tabla 1). Es importante mencionar que estos subgrupos no son excluyentes, lo que significa que las y los participantes con puntajes de CI más altos están incluidos en todos los subgrupos inferiores. Por ejemplo, una o un participante con un CI de 128 se clasificaría en los subgrupos de CI estimado ≥ 115 , $CI \geq 120$ y $CI \geq 125$. Cabe destacar que ningún participante alcanzó un CI estimado de ≥ 130 .

Tabla 1. Distribución de las y los participantes según su CI estimado

CI Estimado	≥ 115	≥ 120	≥ 125	≥ 130
N(Total de participantes)	30	25	9	0
Género(Mujeres, hombres)	14, 16	12, 13	6, 3	0, 0

Fuente: elaboración propia.

Los puntajes totales obtenidos de las y los 130 participantes del test de RAVEN se convirtieron en percentiles según su grupo de edad, basados en la validación local (en preparación). Los estudiantes fueron clasificados en subgrupos de percentiles tomando en consideración su desempeño en el test de RAVEN (ver Tabla 2). Es importante señalar que estos subgrupos basados en percentiles no son excluyentes; similar a lo mostrado con el CI estimado. Los datos de las y los participantes con puntajes por debajo del percentil 85 no se presentan en esta sección, pero fueron considerados en los análisis estadísticos.

Tabla 2. Distribución de los participantes según el percentil de RAVEN

RAVEN: Percentil	85	90	95	98
N(Total de participantes)	63	39	22	6
Género (Mujer, hombre)	30,33	20,19	11,11	1,5

Fuente: elaboración propia.

Los criterios de inclusión fueron: (a) asistir a una escuela pública; (b) consentimiento verbal por parte del estudiante; y (c) la ausencia de cualquier diagnóstico reportado de condiciones como sordera, ceguera, discapacidad intelectual o trastornos neurológicos asociados con daño cerebral.

2.2. Instrumentos

Matrices Progresivas de Raven (RAVEN, 2000): Este test mide el componente deductivo del factor g de Spearman. La característica esencial de la habilidad deductiva es "la capacidad para generar nuevos conceptos, mayormente no verbales" (Raven et al., 2000, p. 1). La versión más reciente de las Matrices Progresivas Estándar utiliza los mismos ítems que la versión de 1938 (Raven, 2000). El test consta de cinco series (A, B, C, D, E) de 12 problemas, cada uno siguiendo un tema común. En cada ítem, las y los estudiantes deben analizar los componentes de una matriz incompleta y seleccionar la opción que mejor la complete. La administración del test RAVEN toma aproximadamente 60 minutos. Dado que el test RAVEN tiene solo un nivel, todas y todos los estudiantes completaron la misma escala. Las respuestas fueron registradas en una hoja separada y corregidas manualmente.

CI Estimado: Se utilizó la Escala de Inteligencia de Wechsler para Niños-IV (WISC-IV, Wechsler, 2007). Se administraron dos subpruebas: Semejanzas y Matrices. En la subprueba de Semejanzas, se presentan dos palabras que el evaluador lee en voz alta, y el examinado debe explicar en qué se parecen. En la subprueba de Matrices, el participante observa una matriz incompleta y selecciona la parte faltante de entre cinco opciones de respuesta. Las puntuaciones directas de estas subpruebas se calcularon y luego se convirtieron en puntuaciones estándar con base en las normas de edad. Estas puntuaciones se combinaron para obtener el CI estimado. Sattler (2010), muestra que esta versión abreviada de dos tareas, es una de las mejores combinaciones de subpruebas para estimar el CI, con alta fiabilidad (.93) y validez (.89).

2.3. Procedimiento

La precisión diagnóstica se evaluó mediante estadísticas de eficiencia diagnóstica. Para el test RAVEN, se utilizaron puntos de corte en los percentiles 85, 90, 95 y 98 para toda la muestra, basados en los percentiles obtenidos de un estudio de validación con población mexicana (en preparación), considerando las normas por edad. Además, se emplearon puntuaciones de CI estimado de la versión mexicana del WISC-IV, con puntos de corte en 115, 120, 125 y 130.

Estos puntos de corte se seleccionaron debido a la falta de consenso entre las y los expertos en el campo de la educación especial para estudiantes AS, así como entre las normativas nacionales e internacionales, respecto al punto de corte preciso para definir a los estudiantes AS (Ryser, 2018). Esta falta de consenso, donde las consideraciones de desarrollo individual y la inclusión son esenciales, busca asegurar que los estudiantes AS no sean excluidos (Gagné, 2017; Renzulli & Reis, 2021). Para abordar este problema, se incluyeron cuatro puntos de corte diferentes, tomando como referencia las tablas del WISC-IV (Wechsler, 2007): CI 115 o percentil 85 (1 DE por encima de la media), CI 120 o percentil 90 (1 1/3 DE por encima de la media), CI 125 o percentil 95 (1 2/3 DE por encima de la media), y CI 130 o percentil 98 (2 DE por encima de la media). Estos valores proporcionan una gama de puntos de corte para identificar a las y los estudiantes AS, ofreciendo una evaluación más completa de cómo se desempeña el RAVEN en comparación con el WISC-IV.

En este estudio, las y los estudiantes AS se identificaron operacionalmente como aquellos que obtuvieron un CI estimado de ≥ 115 en el WISC-IV, considerado como el "Gold Standard" para la identificación precisa. Las puntuaciones percentilares del RAVEN se utilizaron como predictores de estas capacidades.

Para evaluar la precisión diagnóstica del RAVEN, se calcularon diferentes métricas entre ellas una matriz de confusión, es una disposición tabular 2x2 específica que permite visualizar el desempeño de un algoritmo. En este tipo de tabla de contingencia, con dos dimensiones ("real" y "predicción"), y conjuntos idénticos de "clases" en ambas dimensiones, de este análisis se obtienen cuatro tipos de resultados: verdaderos positivos (VP), falsos negativos (FN), falsos positivos (FP) y verdaderos negativos (VN), junto con las clasificaciones positivas y negativas. Además, se calcularon las siguientes métricas: Exactitud (Accuracy), la cual mide la proporción de alumnos correctamente clasificados (tanto positivas como negativas) sobre el total de alumnos. En este contexto, indica con qué frecuencia el RAVEN identifica correctamente a los estudiantes como AS o no-AS.

$$Exactitud = \frac{VP + VN}{Total}$$

La Precisión (Valor Predictivo Positivo), mide la proporción de predicciones positivas (estudiantes identificados como AS por el RAVEN) que son realmente correctas. La precisión es crucial cuando las consecuencias de los falsos positivos son significativas.

$$\text{Precisión} = \frac{VP}{\text{Total clasificados positivos}}$$

La Sensibilidad (Tasa de Verdaderos Positivos), corresponde a la proporción de verdaderos positivos (estudiantes que son genuinamente AS según el WISC-IV) que son correctamente identificados por el RAVEN. Una alta sensibilidad asegura que la mayoría de los estudiantes AS sean identificados correctamente, minimizando los falsos negativos.

$$\text{Sensibilidad} = \frac{VP}{\text{Total positivos}}$$

La Puntuación F1 (F1 Score), hace referencia a la media armónica entre la precisión y la sensibilidad. Esta métrica proporciona un equilibrio entre ambas, lo cual es útil cuando es necesario equilibrar el impacto de los falsos positivos y falsos negativos.

$$F - \text{Score} = 2x \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$$

Finalmente, para evaluar el rendimiento diagnóstico, realizamos análisis de curvas ROC (Receiver Operating Characteristic) por sus siglas en inglés. La curva ROC gráfica la sensibilidad (tasa de verdaderos positivos) frente a 1-especificidad (tasa de falsos positivos) a lo largo de una gama de posibles puntos de corte. El valor del Área Bajo la Curva (AUC) indica el rendimiento diagnóstico general de la prueba, siendo 1.0 una precisión perfecta, 0.5 predicción al azar, y un valor inferior a 0.5 indicando una predicción peor que el azar. El AUC es una medida ampliamente aceptada para comunicar el tamaño del efecto y la efectividad diagnóstica de una prueba (Swets, 1988).

3. Resultados

La Tabla 4 y la Figura 1 presentan la matriz de confusión, la cual detalla el desempeño de la prueba RAVEN en distintos puntos de corte de percentiles (85, 90, 95 y 98) para la identificación de estudiantes AS, tomando en cuenta los rangos de CI estimado (≥ 115 , ≥ 120 , ≥ 125). El desempeño de la clasificación se analiza mediante las tasas de VP, FP, FN y VN.

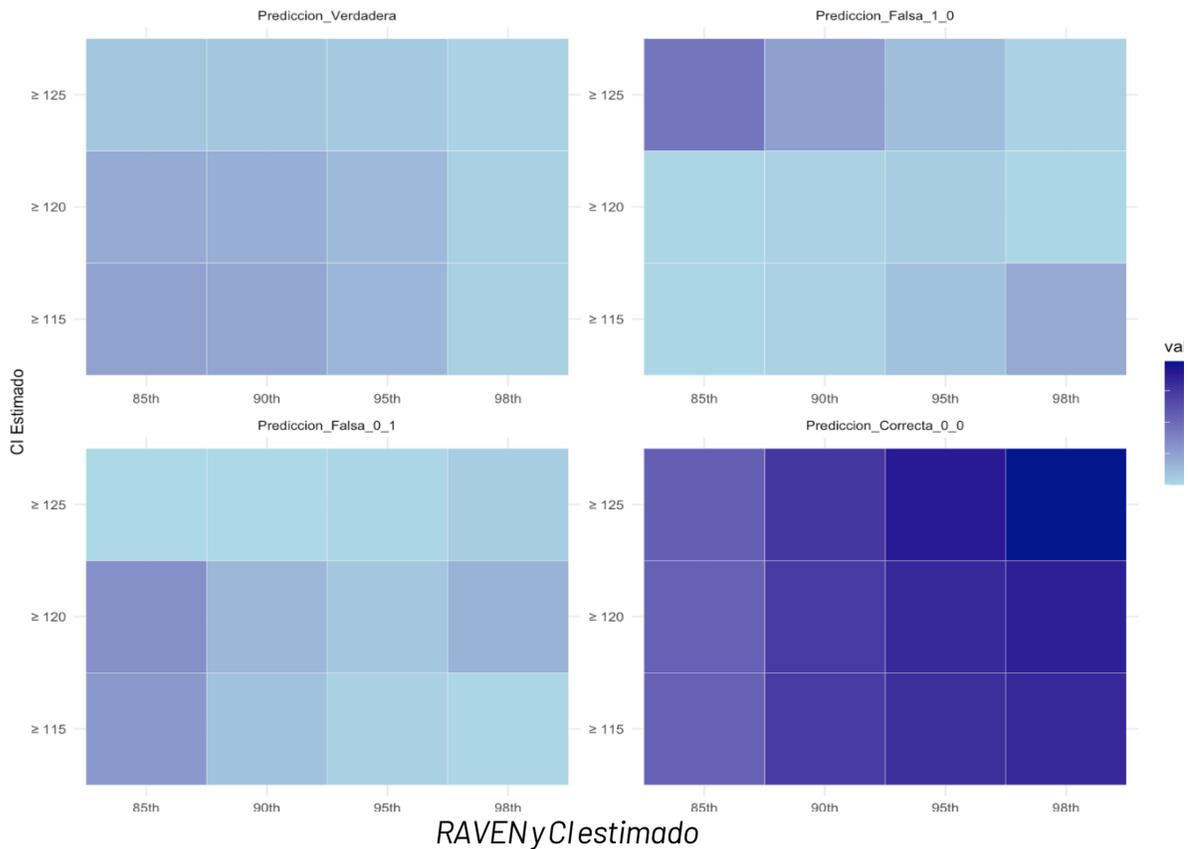
Evaluación de la efectividad de los puntos de corte en la prueba RAVEN para la identificación de alumnos con aptitudes sobresalientes en México

Tabla 4. Matriz de Confusión como predictor de CI estimado en los Percentiles del RAVEN

CI estimado WISC-IV	Percentil RAVEN	Verdaderos positivos(1,1)	Falsos positivos(1,0)	Falsos negativos(0,1)	Verdaderos negativos(0,0)
≥115	85th	29	1	34	66
	90th	27	3	12	88
	95th	18	12	4	96
	98th	5	25	1	99
≥120	85th	24	1	39	66
	90th	22	3	17	88
	95th	16	6	9	99
	98th	5	1	20	104
≥125	85th	9	54	0	67
	90th	9	30	0	91
	95th	8	14	1	107
	98th	3	3	6	118

Fuente: elaboración propia.

Figura 1 Distribución de predicciones Verdaderas y Falsas de acuerdo a los perfiles del



Fuente: elaboración propia.

Evaluación de la efectividad de los puntos de corte en la prueba RAVEN para la identificación de alumnos con aptitudes sobresalientes en México

Los resultados revelan que el número de VP en cada categoría de CI estimado (≥ 115 , ≥ 120 , ≥ 125) varía según el percentil del RAVEN. El mayor número de verdaderos positivos se encuentra en el percentil 90, con 22 detecciones en la categoría de CI estimado ≥ 120 , lo que sugiere una mayor precisión en la predicción para este rango. En contraste, el menor número de VP se observa en el percentil 98, en la categoría de CI estimado ≥ 125 , con solo 3 verdaderos positivos identificados.

Además, un número elevado de FP, particularmente en la categoría de CI estimado ≥ 125 en el percentil 85, donde se registran 54 falsos positivos. Esto podría sugerir una tendencia a sobre identificar a estudiantes con CI estimados bajos, lo que podría llevar a errores en la clasificación. En cuanto a los FN, un factor crucial a tener en cuenta al momento de elegir el percentil, como se puede observar en la categoría de CI estimado ≥ 125 en el percentil 85 no se reportan falsos negativos, lo que podría indicar una alta precisión en la identificación de estudiantes AS en ese rango.

La Tabla 5, resume cinco métricas de eficiencia diagnóstica: precisión, exactitud, sensibilidad (Recall), puntuación F1 (F1 Score) y AUC para los percentiles del RAVEN (85, 90, 95, 98) y los puntajes de CI estimado del WISC-IV (≥ 115 , ≥ 120 , ≥ 125 , ≥ 130). Estas métricas ofrecen distintas perspectivas sobre la precisión diagnóstica de la prueba RAVEN en la identificación de estudiantes AS.

Tabla 5. Métricas de eficiencia diagnóstica para los percentiles del RAVEN y CI estimado del WISC-IV

<i>CI estimado WISC-IV</i>	<i>Percentil RAVEN</i>	<i>Exactitud</i>	<i>Precisión</i>	<i>Recall</i>	<i>F1 Score</i>	<i>AUC</i>
≥ 115	85	0.730	0.460	0.966	0.623	0.813
	90	0.884	0.692	0.900	0.782	0.89
	95	0.876	0.818	0.60	0.69	0.78
	98	0.80	0.833	0.166	0.277	0.57
≥ 120	85	0.692	0.381	0.960	0.545	0.794
	90	0.846	0.564	0.880	0.687	0.85
	95	0.884	0.727	0.640	0.680	0.79
	98	0.838	0.833	0.200	0.322	0.59
≥ 125	85	0.584	0.142	1.00	0.250	0.776
	90	0.769	0.230	1.00	0.375	0.876
	95	0.884	0.363	0.88	0.516	0.884
	98	0.930	0.500	0.333	0.400	0.654

Es importante mencionar que ningún estudiante de nuestra muestra logró el punto de corte de CI estimado del WISC-IV de ≥ 130 , por lo que los valores cero (0) en las métricas corresponden a esta categoría. Limitando el análisis de prueba de eficiencia para este umbral de CI estimado.

Para el análisis del desempeño de la prueba mediante varias métricas de precisión diagnóstica (exactitud, precisión, sensibilidad, F1 Score y AUC, ver Tabla 5). Estos indicadores se calcularon para diferentes puntos de corte de percentiles del RAVEN (85, 90, 95 y 98) con el fin de ofrecer una visión integral del rendimiento de la prueba, se realizará el análisis considerando el CI estimado ≥ 115 , 120 y 125 .

3.1 La precisión diagnóstica para un CI estimado de ≥ 115

La métrica de exactitud mide la proporción de casos correctamente clasificados (tanto estudiantes con AS como sin ellas) sobre el total. Un valor alto en esta métrica refleja un mejor rendimiento general. El punto de corte del percentil 90 mostró la mayor exactitud (0.884), lo que indica que este corte clasificó correctamente al 88.4% de los estudiantes, sugiriendo que el percentil 90 es un marcador sólido para identificar a estudiantes con un CI ≥ 115 . Por otro lado, el percentil 85, aunque logró una alta sensibilidad, tuvo una exactitud más baja (0.730), lo que podría indicar una mayor cantidad de falsos positivos y, en consecuencia, un mayor riesgo de identificación errónea de estudiantes con AS.

En cuanto a la precisión, esta métrica indica la proporción de verdaderos positivos entre todas las predicciones positivas. Un valor alto de precisión refleja una menor tasa de falsos positivos. El punto de corte del percentil 95 obtuvo la mayor precisión (0.818), lo que sugiere que la mayoría de los estudiantes identificados AS eran efectivamente AS, minimizando los falsos positivos. Esto es crucial para reducir las identificaciones incorrectas. En contraste, el percentil 85 mostró una menor precisión debido al mayor número de falsos positivos.

La Sensibilidad, mide la proporción de estudiantes AS que fueron correctamente identificados. Un valor alto implica menos falsos negativos. La mayor sensibilidad se alcanzó con el punto de corte del percentil 85 (0.966), lo que indica que este corte identificó exitosamente a la mayoría de los estudiantes AS. No obstante, esta alta sensibilidad se acompañó de una menor precisión, lo que refleja un mayor número de falsos positivos.

La Puntuación F1 Score, la cual proporciona una medida equilibrada entre precisión y sensibilidad. El punto de corte del percentil 90 obtuvo la mayor puntuación F1 Score (0.782), lo que resalta su efectividad para balancear la identificación de estudiantes AS y minimizar las clasificaciones incorrectas. Este percentil ofrece una mayor exactitud que el percentil 85, aunque con una sensibilidad ligeramente menor, logrando un mejor equilibrio entre identificar estudiantes AS y evitar falsos positivos.

El AUC, representa la capacidad diagnóstica general del RAVEN a través de todos los puntos de corte. Un valor cercano a 1 indica un excelente rendimiento, mientras que un valor de 0.5 sugiere predicción aleatoria. El AUC fue más alto para el punto de corte del percentil 90 (0.89), lo que demuestra una fuerte precisión diagnóstica.

3.2 La precisión diagnóstica para un CI estimado de ≥ 120

Los resultados mostraron que el percentil 95 del RAVEN logró la mayor exactitud (0.884), siendo uno de los percentiles con mayor puntuación. Este nivel de exactitud sugiere una alta capacidad para diferenciar entre estudiantes por encima y por debajo

del umbral de CI estimado de 120. Referente a la precisión, el punto de corte del percentil 98 alcanzó la mayor precisión (0.833), lo que refleja una reducción de falsos positivos. Esto señala la alta capacidad para identificar con precisión a estudiantes con CI estimado más alto sin sobre identificar a estudiantes no-AS. En relación a la sensibilidad, el percentil 85 obtuvo la mayor sensibilidad (0.960), lo que sugiere que es eficaz para clasificar una gran proporción de estudiantes AS. Sin embargo, este alto nivel de sensibilidad conlleva a una disminución en la precisión, indicando la inclusión de más falsos positivos.

Así mismo, el F1 Score, en el percentil 90 mostró la mayor puntuación (0.687), lo que ofrece un enfoque equilibrado para la identificación de estudiantes AS. El percentil 95 también tuvo una puntuación F1 Score alta (0.680), reflejando un buen rendimiento para niveles moderados de AS. El AUC más alto fue para el punto de corte del percentil 90 (0.85), lo que respalda su uso para distinguir de manera confiable a estudiantes AS.

3.3 La precisión diagnóstica para un CI estimado de ≥ 125

El desempeño diagnóstico del test RAVEN para identificar a estudiantes con un CI estimado de ≥ 125 mostró que la mayor exactitud (0.930), se alcanzó con el punto de corte en el percentil 98, lo que indica una diferenciación sólida entre las y los estudiantes por encima y por debajo del umbral de CI 125. Esta alta precisión refleja una capacidad notable para identificar correctamente a estudiantes AS, aunque la menor sensibilidad en este percentil sugiere que algunas y algunos estudiantes realmente AS no fueron detectados.

La Precisión, el percentil que logró una mayor puntuación fue en el percentil 98 (0.500), equilibrando verdaderos y falsos positivos. Aunque esta precisión es moderada, refleja un balance razonable dado el umbral de CI más alto, donde la mitad de los estudiantes identificados como AS cumplían realmente con los criterios de CI. La Sensibilidad, en los percentiles 85 y 90 lograron una sensibilidad perfecta (1.00), pero esto fue a costa de una menor precisión (0.142 y 0.230, respectivamente), lo que pudiera indicar que, aunque casi todos los estudiantes AS fueron identificados, también se incluyeron muchos falsos positivos.

Esto sugiere que los percentiles más bajos no son prácticos al momento de identificar puntos de corte de CI más altos debido a la inclusión sustancial de estudiantes no-AS. El F1 Score, más alto se registró en el percentil 95 (0.516), lo que indica un balance razonable entre precisión y sensibilidad, a pesar de la mayor complejidad para distinguir a las y los estudiantes en este nivel. El AUC más alto se alcanzó en el percentil 95 (0.884), lo que sugiere que es el punto de corte más confiable para identificar a estudiantes AS.

3.4 La precisión diagnóstica para un CI estimado de ≥ 130

El análisis del desempeño diagnóstico del test RAVEN para identificar a estudiantes con un CI estimado de ≥ 130 , según la evaluación realizada con el WISC-IV, arrojó resultados inconclusos. En todos los puntos de corte del percentil RAVEN evaluados (85, 90, 95 y 98), ningún estudiante alcanzó o superó el punto de corte del CI de 130. Como resultado, los valores de exactitud, precisión, sensibilidad, F1 Score y AUC se registraron como cero. Esta falta de datos por encima del umbral de CI 130 indica que ninguno de los estudiantes de la muestra cumplió con este criterio, lo que impidió medir la eficiencia diagnóstica del RAVEN en este punto de corte.

4. Discusión

El presente estudio tuvo como objetivo el establecer cuál es el mejor punto de corte para identificar a alumnas y alumnos AS en México utilizando el Test de Matrices Progresivas RAVEN. Los hallazgos presentados en este artículo buscan ampliar la información disponible a nivel local, nacional y global, mostrando los resultados obtenidos de la población mexicana en una escala no verbal, considerando su contexto intercultural (Dietz, 2016), una variable importante a tener en cuenta, y poco explorada en esta población.

Cuando consideramos el CI estimado de ≥ 115 , encontramos que el percentil 90 del RAVEN ofrece el mejor balance de métricas de eficiencia diagnóstica, incluyendo alta exactitud, precisión y F1 Score, lo que lo convierte en un indicador confiable para identificar a estudiantes con un CI estimado de ≥ 115 . El percentil 85, aunque con alta sensibilidad, puede llevar a un número significativo de falsos positivos, afectando la precisión del proceso de identificación. En cambio, el percentil 95 ofrece mayor precisión, pero a costa de perder a algunos estudiantes AS, lo que indica una menor sensibilidad.

Así mismo, cuando tomamos un CI estimado de ≥ 120 , el punto de corte del percentil 95 del RAVEN proporciona el mejor equilibrio de métricas de eficiencia diagnóstica, incluyendo alta exactitud, una precisión razonable y una puntuación F1 score satisfactoria. Sin embargo, el percentil 90 demuestra una buena alternativa con una precisión ligeramente inferior pero una mayor sensibilidad, lo que lo hace adecuado para una identificación más amplia de estudiantes AS.

Aunado a esto, cuando tomamos un CI estimado de ≥ 125 , el punto de corte en el percentil 95 ofrece el mejor equilibrio de precisión diagnóstica para identificar a estudiantes, combinando alta exactitud general con un balance óptimo entre precisión y sensibilidad. El percentil 98 ofrece la mayor exactitud, pero a costa de una menor sensibilidad, mientras que los percentiles 85 y 90, aunque logran una sensibilidad perfecta, tienen una precisión más baja.

La sensibilidad y especificidad de una prueba no son valores absolutos. La precisión y exactitud de una prueba varían según la muestra y los puntos de corte seleccionados. La muestra utilizada en el presente estudio fue seleccionada aleatoriamente de diferentes escuelas locales, que tienen características demográficas específicas, incluidas etnia, nivel educativo de las madres y los padres y representación regional. Sin embargo, consideramos que se necesitan más estudios para explorar las propiedades psicométricas del RAVEN mediante estudios de validación cruzada que corroboran los hallazgos a favor de la precisión diagnóstica y la validez de criterio del RAVEN, en población mexicana, pero de otra región del país.

Además, los estudios de validación cruzada con diferentes muestras, pudieran dar luz a futuras investigaciones, donde la utilización de criterios alternativos para definir operativamente a los estudiantes AS. Sugerimos que el CI estimado es un indicador confiable y válido para identificar a estudiantes AS, pero esta característica no debe considerarse la única a tomar en cuenta, ya que el constructo abarca más dimensiones conceptuales y operativas que deben considerarse (Renzulli et al., 2021). Las cuales no fueron consideradas en este estudio.

Futuros estudios también podrían ampliar la presente investigación validando el RAVEN con otros constructos relacionados con las y los estudiantes AS, como el lenguaje, las matemáticas, aspectos socioemocionales y motrices, entre otros (Alsuwailimi, 2024). Esta no será una tarea fácil, ya que hay pocos "Gold Standards" establecidos en la literatura y existen falsas concepciones sobre los estudiantes AS (Renzulli et al., 2021). Pequeños estudios están utilizando evaluaciones desarrolladas por paneles de expertos en áreas como el aprendizaje, la creatividad, la motivación, el liderazgo y las artes (SEP, 2022). Sin embargo, aún se necesita mucho trabajo en el campo de la identificación de alumnos AS.

La sensibilidad, especificidad y clasificación correcta de una característica clínica específica en pruebas psicológicas y neuropsicológicas es reportada con baja frecuencia en estudios o manuales técnicos, a diferencia de la literatura médica y clínica (Sherman et al., 2020). Es importante destacar que la sensibilidad de una prueba solo puede aumentarse a expensas de la especificidad, y viceversa (Mariska, 2008). No hay manera de mejorar ambos parámetros simultáneamente. Por lo tanto, la puntuación seleccionada como punto de corte o percentil para representar la línea divisoria entre estudiantes AS y no-AS se considera arbitraria, especialmente porque los estudiantes AS están en constante evolución y abarcan desde una ausencia clara hasta una presencia notable (Redding & Grissom, 2021). En la práctica médica, la clasificación de una condición clínica suele centrarse en la detección temprana, ya que puede ser una cuestión de vida o muerte (Wilson et al., 1968). Sin embargo, en las ciencias sociales, la detección correcta y temprana puede no ser una prioridad, por lo que pudiera ser el motivo de su bajo reporte en los manuales técnicos.

El valor predictivo de una prueba se ve afectado tanto por el punto de corte o percentil seleccionado como por la tasa base de la condición de interés (Hajian-Tilaki,

2018), en este caso, las y los estudiantes AS. Cuando la tasa base es baja, el valor predictivo de un resultado negativo será mayor que el de un resultado positivo. En otras palabras, cuando la condición de interés está presente, un resultado positivo en la prueba generalmente no es útil para confirmar su presencia.

Por ejemplo, si consideramos a las y los estudiantes AS como un fenómeno dicotómico y la tasa base en la población escolar es del 5%, utilizando el percentil 90 como punto de corte en el RAVEN, podemos asumir que, de una población escolar de 10,000 estudiantes, 500 podrían considerarse AS. De estos, el RAVEN identificaría correctamente a 450 (90%) utilizando el percentil 90 como punto de corte. De manera similar, de los 9,500 estudiantes considerados no-AS, 8,360 (88%) obtendrían puntajes por debajo de este percentil en el RAVEN. Este tipo de análisis muestra que la sensibilidad, especificidad y punto de corte del RAVEN, o de cualquier instrumento, varían según la tasa de prevalencia y el punto de corte seleccionado. Es esencial que los profesionales consideren estos factores al identificar a estudiantes AS, asegurando que el proceso no sea meramente protocolario, sino que tenga un impacto real en la detección y desarrollo de estas y estos estudiantes.

5. Conclusiones

Los resultados presentados parecieran confirmar, la utilización del punto de corte en el percentil 90 del RAVEN para identificar a estudiantes AS, con CI estimados cercanos a los puntos de corte de 115 y 120, debido a que las métricas diagnósticas se muestran equilibradas y alta precisión general. Por lo que es indispensable que el especialista en el área defina sus parámetros y objetivos de identificación, es decir, a que beneficios accederán los alumnos identificados (Renzulli, 2014). Por otro lado, para los puntos de corte de CI estimado ≥ 125 , el punto de corte del percentil 95 proporciona un buen equilibrio entre precisión y sensibilidad, lo que lo convierte en un indicador confiable de AS. Sin embargo, la elección de este punto, pudiera dejar afuera a muchas y muchos alumnos potenciales o que se pudieran beneficiar de actividades con mayor demanda cognitiva. Los profesionales de la evaluación deben considerar diferentes variables para crear un ajuste continuo, tomando en consideración los contextos culturales mejorarán su precisión diagnóstica y utilidad.

Con estos resultados podemos afirmar que el RAVEN posee las propiedades psicométricas necesarias para identificar a estudiantes AS en los niveles de primaria y secundaria. La cual puede ser utilizada como herramienta de cribado grupal, dado que el RAVEN mostró evidencia convincente de validez de criterio, aunque se debe considerar el percentil a utilizar y analizar la posibilidad de combinarse con otros instrumentos para confirmar si un estudiante es AS. Reconocemos que el uso de una prueba grupal puede no detectar a todas y todos los estudiantes AS, dado que existen diversas variables de ruido que podrían influir en la aplicación.

Finalmente, esta investigación presenta las propiedades de detección del RAVEN en la población del sur de Zacatecas. Es importante mencionar que no consideramos factores interculturales, donde la diversidad cultural y las diferencias étnicas podrían influir significativamente en los resultados, como en estudios realizados en Estados Unidos con estudiantes asiático-americanos, afrodescendientes y latinos, entre otros (McClain & Pfeiffer, 2012; Pfeiffer & Blei., 2008; Pfeiffer & Petscher, 2008). Dado que todas las pruebas tienen cierta carga cultural, sería necesario comparar estos resultados con otras regiones culturales del país para determinar si se deben considerar sesgos culturales.

5.1 Limitaciones del estudio y futuras investigaciones

A pesar de los hallazgos presentados, este estudio presenta algunas limitaciones. Futuras investigaciones deberían considerar una muestra más diversa y explorar la integración de otros factores contextuales, regionales y culturales. También sería beneficioso realizar estudios longitudinales para observar cómo evolucionan los dominios cognitivos en esta población y su impacto en las dinámicas socio-culturales y educativas de la región.

La muestra fue pequeña y ningún estudiante alcanzó un CI estimado de 130 o más, lo que limita la evaluación de las dimensiones teóricas de las puntuaciones estandarizadas. Además, esperamos que los futuros estudios exploren el impacto de los factores socio-culturales y familiares, en la efectividad del RAVEN para determinar si la prueba es capaz de detectar a esta población.

Esta investigación resalta el potencial del RAVEN como prueba de identificación grupal para detectar a estudiantes AS. Aún se necesita más trabajo sobre mejores métodos de identificación que aborden la diversidad intercultural de la población mexicana y que diferencien eficazmente entre los métodos de identificación cuantitativos y cualitativos.

Cuando el RAVEN se usa como parte de una evaluación integral para detectar a estudiantes AS, parece prometedor su desempeño para identificar a estudiantes que de otro modo podrían quedar excluidos.

Agradecimientos

Gracias al Dr. Enrique Hernández, al Maestro Rutilio Nava, a la Maestra Claudia Ahumada, al Dr. José María Celaya.

Conflicto de intereses

Declaramos no tener conflicto de interés con ninguno de los actores del estudio.

Financiamiento

No se recibió ningún financiamiento.

Proyecto

Este artículo es parte del Proyecto de investigación desarrollado en el programa de doctorado en Educación de la Universidad Autónoma de Fresnillo.

Referencias

- Alsuwailimi, M. S. (2024). Identification of gifted students in different global contexts: literature review. *Journal of Special Education & Rehabilitation*(2314-8608), 17.
- Carman, C. A., & Taylor, D. K. (2010). Socioeconomic Status Effects on Using the Naglieri Nonverbal Ability Test (NNAT) to Identify the Gifted/Talented. *Gifted Child Quarterly*, 54(2), 75-84. <https://doi.org/10.1177/0016986209355976>
- Carman, C. A., Walther, C. A., & Bartsch, R. A. (2020). Differences in using the Cognitive Abilities Test (CogAT) 7 nonverbal battery versus the Naglieri Nonverbal Ability Test (NNAT) 2 to identify the gifted/talented. *Gifted Child Quarterly*, 64(3), 171- 191.
- Chávez-Soto, B. I., Zacatelco, R. F., & González, G. A. (2018). ¿Es efectiva la nominación del maestro en estudiantes sobresalientes? *Revista Educación y Desarrollo*(45), 25-35.
- Covarrubias-Pizarro, P., . (2018). Del concepto de aptitudes sobresalientes al de altas capacidades y el talento. *IE Revista de Investigación Educativa de la REDIECH*, 9(17), 53-67. <https://doi.org/https://www.redalyc.org/articulo.oa?id=521655454004>
- Donate, N., & Borges, Á. (2018). La influencia de la clase social en la identificación e intervención del alumnado de altas capacidades The influence of social class in the identification and intervention of students with high abilities. *Talincrea*, 4(8), 3-14. https://www.cucs.udg.mx/talineng/sites/default/files/adjuntos/04_08/04_influencia.pdf
- Flynn, A. S. (2023). Black Minds Matter: A Longitudinal Analysis of the Persistent Underrepresentation of Black Students in Gifted Education Programs. *Journal of Leadership, Equity, and Research*, 9(1), 6-20.
- Flynn, J. R. (2007). *What is intelligence?: Beyond the Flynn effect*. Cambridge University Press.
- Gagné, F. (2017). The Integrative Model of Talent Development (IMTD). In J. A. Plucker, A. N. Rinn, & M. C. Makel (Eds.), *From Giftedness to Gifted Education: Reflecting Theory in Practice*. Prufrock Press Inc. .

- Guez, A., Peyre, H., Le Cam, M., Gauvrit, N., & Ramus, F. (2018). Are high-IQ students more at risk of school failure?. *Intelligence*, 71, 32-40. <https://doi.org/10.1016/j.intell.2018.09.003>
- Guignard, J. H., Kermarrec, S., & Tordjman, S. (2016). Relationships between intelligence and creativity in gifted and non-gifted children. *Learning and Individual Differences*, 52, 209-215. <https://doi.org/10.1016/j.lindif.2015.07.006>
- Gülşen Erden , İbrahim Yiğit , Çelik, C., & Guzey, M. (2020). The diagnostic utility of the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) in identification of gifted children. *The Journal of General Psychology*, 143(3), 371-390. <https://doi.org/10.1080/00221309.2020.1862038>
- Hajian-Tilaki, K. (2018). The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. *Statistical Methods in Medical Research*, 27(8), 2374-2383. <https://doi.org/10.1177/0962280216680383>
- Hodges, J., & Gentry, M. (2021). Underrepresentation in Gifted Education in the Context of Rurality and Socioeconomic Status. *Journal of Advanced Academics*, 32(2), 135-159. <https://doi.org/10.1177/1932202x20969143>
- Holocher-Ertl, S., Kubinger, K. D., & Hohensinn, C. (2008). Identifying children who may be cognitively gifted: The gap between practical demands and scientific supply. *Psychology Science*, 50(2), 97.
- Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. . (2001). *Iowa tests of basic skills (ITBS) forms A, B, and C*. Riverside Publishing Company.
- Johnsen, S. K. (2021). Making decisions about placement. In S. K. Johnsen (Ed.), *Identifying gifted students: A practical guide*. Taylor & Francis.
- Krisel, S. (2012). Characteristics of gifted children as a guide to identification. In S. L. Hunsaker (Ed.), *Identification: The theory and practice of identifying students for gifted and talented education services*(pp. 75-97). Creative Learning Press.
- Kurtz, H. (2019). National Survey of Gifted Education. *EdWeek Research Center*.
- Lee, H., Karakis, N., Olcay Akce, B., Azzam Tuzgen, A., Karami, S., Gentry, M., & Maeda, Y. (2021). A Meta-Analytic Evaluation of Naglieri Nonverbal Ability Test: Exploring Its Validity Evidence and Effectiveness in Equitably Identifying Gifted Students. *Gifted Child Quarterly*, 65(3), 199-219. <https://doi.org/10.1177/0016986221997800>
- Lohman, D. F. (2005). Review of Naglieri and Ford (2003): Does the Naglieri Nonverbal Ability Test identify equal proportions of high-scoring White, Black, and

- Hispanic students? *Gifted Child Quarterly*(49), 19-28.
<https://doi.org/10.1177/001698620504900103>
- Lohman, D. F., & Nicpon, M. F. (2012). Ability testing & talent identification. In S. L. Hunsaker (Ed.), *Identification: The theory and practice of identifying students for gifted and talented education services*(pp. 287-335).
- Mariska, M. G. L., Jonathan J. Deeks, Constantine Gatsonis, et al. (2008). Systematic Reviews of Diagnostic Test Accuracy. *Annals of Internal Medicine*, 149(12), 889- 897.
<https://doi.org/10.7326/0003-4819-149-12-200812160-00008> %
- McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly*, 60(4), 258-278. <https://doi.org/10.1177/001698621665625>
- McClain, M. C., & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology*,(28), 59-88.
<https://doi.org/10.1080/15377903.2012.643757>
- Molinero, C., Mata, S., Calero, M. D., García-Martín, , & M. B., A.-C., A. . (2015). Usefulness of WISC-IV in determining intellectual giftedness. *The Spanish journal of psychology*, 18(E60). <https://doi.org/10.1017/sjp.2015.63>
- Naglieri, J., A., & Ford, D., Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, 47(2), 155-160.
- Naglieri, J. A. (1997). *Naglieri nonverbal ability test*. The Psychological Corporation.
- National Association for Gifted Children. (8 de Febrero 2023). *Definitions of giftedness*
<http://www.nagc.org/resources-publications/resources/definitions-giftedness>
- National Association for Gifted Children, & The Council of State Directors of Programs for the Gifted. (2015). 2014-2015 State of the states in gifted education: Policy and practice data. <https://docslib.org/doc/8149589/2014-2015-state-of-the-states-in-gifted-education-policy-and-practice-data>
- Navarro-Saldaña, G., Flores-Oyarzo, G., & González Navarro, M. G. (2022). Estudiantes con alta capacidad: explorando su distribución según tipo de establecimiento educativo. *Liberabit*, 28(1). <http://dx.doi.org/10.24265/liberabit.2022.v28n1.02>

- Pedersen, B., Makel, M. C., Rambo-Hernandez, K. E., Peters, S. J., & Plucker, J. (2023). Most Mathematics Classrooms Contain Wide-Ranging Achievement Levels. *Gifted Child Quarterly*, 67(3), 220-234. <https://doi.org/10.1177/00169862231166074>
- Pfeiffer, & Blei. (2008). Gifted identification beyond the IQ test: Rating scales and other assessment procedures. In S. I. Pfeiffer (Ed.), *Handbook of giftedness in children: Psychoeducational theory, research, and best practices*, (pp. 177-198).
- Pfeiffer, S. I., & Petscher, Y. (2008). Identifying Young Gifted Children Using the Gifted Rating Scales—Preschool/ Kindergarten Form. *Gifted Child Quarterly*, 52(1), 19-29. <https://doi.org/10.1177/0016986207311055>
- Raven, J. C. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, 41, 1-48.
- Redding, C., & Grissom, J. A. (2021). Do Students in Gifted Programs Perform Better? Linking Gifted Program Participation to Achievement and Nonachievement Outcomes. *Educational Evaluation and Policy Analysis*, 43(3), 520-544. <https://doi.org/10.3102/01623737211008919>
- Renzulli, J. S., & Reis, S. M. (2021). The Three Ring Conception of Giftedness: A Change in Direction from Being Gifted to the Development of Gifted Behaviors. In R. J. Sternberg & D. Ambrose (Eds.), *Conceptions of Giftedness and Talent* (pp. 335- 355). Springer International Publishing. https://doi.org/10.1007/978-3-030-56869-6_19
- Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., Westberg, K. L., & Reed, R. E. S. (2021). *Scales for Rating the Behavioral Characteristics of Superior Students: Renzulli Scales: Technical and Administration Manual*. Routledge.
- Robles-Guerrero, C. (2022). *ESTRATEGIAS DE ENRIQUECIMIENTO EXTRACURRICULAR PARA ALUMNOS CON APTITUD INTELECTUAL SOBRESALIENTE, EN UNA SECUNDARIA DE TEPOZTLÁN. UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS*. Morelos.
- Rodríguez Cervantes, C. J., Valadez Sierra, M. D. L. D., Verche, E., Soltero Avelar, R., , & González Betanzos, F. (2022). Funciones Ejecutivas en alta capacidad intelectual (ACI), trastorno por déficit de atención e hiperactividad (TDAH), doble excepcionalidad (ACI-TDAH) e inteligencia promedio. *Electronic Journal of Research in Educational Psychology*, 20(58). <https://doi.org/10.25115/ejrep.v20i58.4188>

- Rothenbusch, S., Zettler, I., Voss, T., Lösch, T., & Trautwein, U. (2016). Exploring reference group effects on teachers' nominations of gifted students. *Journal of Educational Psychology, 108*(6), 883. <https://doi.org/10.1037/edu0000085>
- Ryser, G., R. (2018). Qualitative and Quantitative Approaches to Assessment. In S. K. Johnsen(Ed.), *Identifying gifted students: A practical guide*. Prufrock Press Inc.
- SEP. (2022). *Atención educativa a estudiantes con aptitudes sobresalientes: preescolar, primaria y secundaria*. Secretaría de Educación Pública.
- Sherman, E. M., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology, 35*(6), 735-764.
- Soto, B. I. C., Ramírez, F. Z., & Tomasini, G. A. (2014). ¿ Quiénes son los alumnos con aptitud sobresaliente? Análisis de diversas variables para su identificación. . *Revista Electrónica "Actualidades Investigativas en Educación, 14*(2), 1-32.
- Steenbergen-Hu, S., Olszewski-Kubilius, P., & Calvert, E. (2020). The Effectiveness of Current Interventions to Reverse the Underachievement of Gifted Students: Findings of a Meta-Analysis and Systematic Review. *Gifted Child Quarterly, 64*(2), 132-165. <https://doi.org/10.1177/0016986220908601>
- Valadez-Sierra, M. D. L. D. (2012). Identificación y evaluación de niños superdotados y talentosos. In M. D. L. D. Valdez Sierra, J. B. Morejón, & M. A. Z. Berbena (Eds.), *Alumnos superdotados y talentosos: Identificación, evaluación e intervención. Una perspectiva para docentes*. Editorial El Manual Moderno.
- VanTassel-Baska, J. (2000). The on-going dilemma of effective identification practices in gifted education. *The Communicator, 31*(2), 39-41. <https://www.davidsongifted.org/gifted-blog/the-on-going-dilemma-of-effective-identification-practices-in-gifted-education/>
- Wechsler, D. (2007). *Escala Wechsler de inteligencia para niños-IV (WISC-IV)*. Editorial El Manual Moderno.
- Wilson, J. M. G., Jungner, G., & Organization, W. H. (1968). Principles and practice of screening for disease.
- Wu, V., East, P., Delker, E., Blanco, E., Caballero, G., Delva, J., Lozoff, B., & Gahagan, S. (2019). Associations Among Mothers' Depression, Emotional and Learning-Material Support to Their Child, and Children's Cognitive Functioning: A 16-Year

Evaluación de la efectividad de los puntos de corte en la prueba RAVEN para la identificación de alumnos con aptitudes sobresalientes en México

Longitudinal Study. *Child Development*, 90(6), 1952-1968.
<https://doi.org/https://doi.org/10.1111/cdev.13071>