



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

FACULTAD DE CIENCIAS HUMANAS Y
SOCIALES

**Evaluación de la fiabilidad y validez de las escalas de
inteligencia en castellano**

Autora: Victoria Recalde Gamarra

Director: José Manuel Caperos Montalbán

Madrid

2021/2022

Resumen

El desarrollo de los instrumentos para evaluar la inteligencia surgió en el siglo XX, a partir de ahí surgen numerosas teorías sobre la inteligencia, así como escalas para su medición con el objeto de estudiar la inteligencia humana, y medirla para identificar qué tanto de aleja o se aproxima a la media de la inteligencia poblacional. Este trabajo tiene como objetivo evaluar la fiabilidad y validez de las escalas de inteligencia editadas al castellano, para ello se realizó una revisión bibliográfica, analizando aquellos estudios de validación para las escalas WAIS-IV, RIAS y Batería III de Aprovechamiento y se comprobó que son válidos para hacer una correcta estimación de la inteligencia. También se ha estudiado la población a la que puede ser aplicable, nos hemos centrado en la población infantil, adulta y una evaluación continuada. Para las tres escalas los estudios demuestran una buena evidencia de fiabilidad y validez.

Palabras claves: fiabilidad, validez, inteligencia, propiedades psicométricas, WAIS-IV, RIAS, Batería III Aprov.

Abstract

The evolution of instruments to assess intelligence arose in the 20th century, from which numerous theories about intelligence emerge, as well as scales for its measurement with the aim of studying human intelligence and quantifying it to identify how far it moves away from or approaches the average of population intelligence. The main objective of this study is to evaluate the reliability and validity of the intelligence scales published in Spanish, for this, a bibliographic review was carried out, analyzing those validation studies for the WAIS-IV, RIAS and Battery III scales of Achievement and to verify through a bibliographic review if they are valid to make a correct assessment of intelligence, the population to which it can be applicable has also been studied. We have focused on the child and adult population and an ongoing evaluation. For all three scales, the studies show good evidence of reliability and validity.

Key words: reliability, validity, intelligence, psychometric properties, WAIS-IV, RIAS, Batería III Aprov.

Contenido

Definición y modelos de inteligencia.....	3
Pruebas de medición de la inteligencia.....	5
Principales pruebas para la medición de la inteligencia en España.....	7
Método.....	8
Estrategia de búsqueda.....	8
Variables.....	9
Resultados.....	10
Resumen de los estudios encontrados y escalas analizadas.....	10
Fiabilidad como consistencia interna.....	14
Fiabilidad como estabilidad temporal.....	15
Evidencias de validez en relación con la estructura interna.....	16
Evidencias de validez convergente.....	17
Evidencias de validez de criterio.....	18
Ámbito de aplicación de cada escala.....	20
Discusión.....	21
Conclusiones.....	24
Referencias.....	25

Definición y modelos de inteligencia

El estudio sobre la inteligencia ha sido un tema característico del siglo XX, aunque uno de los primeros estudios planteados acerca de la inteligencia lo realizó Galton a finales del siglo XIX. Muchos investigadores movidos por el interés de conocer la inteligencia humana hicieron grandes avances a este respecto y aportaron teorías que ayudaron a comprender el funcionamiento intelectual de la persona, cada uno con una visión propia de la capacidad cognitiva, de sus componentes y de la base que lo sostienen, por lo que podemos concluir que la unificación de un concepto de la inteligencia y su medida sigue siendo un debate en la actualidad (Drozdick et al., 2018).

De forma genérica podríamos partir de la definición de Eysenck (1972) que conceptualiza la inteligencia como la capacidad de superar dificultades ante situaciones nuevas. A continuación, se abordarán los distintos modelos y conceptualizaciones realizados a lo largo de este siglo XX.

Alfred Binet publica la primera medida de inteligencia en 1905 y plantea en ella un modelo de inteligencia, definida como la capacidad del hombre que le permite percibir el mundo en su totalidad, analizando los elementos constitutivos (color, forma, distancia) y relacionándolos entre sí. Para él, un elemento percibido exige múltiples funciones que se componen de la comprensión, la memoria, la imaginación, el juicio y la palabra.

Binet (1903) concreta el término de la inteligencia en cuatro procesos: la comprensión, la invención, la dirección y la censura. Señala que para que un elemento percibido pase a formar parte del conocimiento, ha de seguir estos procesos y en el mismo orden: primero, el objeto percibido requiere de la *comprensión*, que le da un sentido para encaminarse al resultado final: quedarse con lo esencial; luego pasa por el proceso de la *invención*, aquí al elemento constitutivo percibido y comprendido, se le añade otro elemento de características similares; la complejidad de este trabajo requiere un tercer proceso que es la *dirección*, criterio que permite mantener el objetivo final sin desviarse. Y, por último, pasa por el proceso de *censura*, que acepta o rechaza otras ideas o elementos que quieren incorporarse al objeto comprendido, y lo hace según se ajusten o no al objetivo planteado.

Stern (1914) conceptualiza la inteligencia como la capacidad general que un individuo posee para adaptarse, de manera consciente, a determinadas situaciones y problemas. Es el primero en considerar la inteligencia del niño integrando la edad mental (rendimiento en las pruebas de inteligencia) con la edad cronológica, para resumir en términos cuantitativos un valor que

represente la capacidad intelectual del niño, así implementa el término de Cociente Intelectual y complementa la escala de inteligencia de Binet. El valor adquirido en esa prueba se compara con un índice de inteligencia que él mismo desarrolla y permite apreciar en qué categoría se encuentra según el valor de CI alcanzado; este índice (CI 130-CI 20) consta de las siguientes categorías: genio, superior, normal, retardo mental limítrofe o *borderline*, retardo mental leve, moderado, grave y profundo.

Charles Spearman (1931) conocido por su desarrollo de la Teoría Bifactorial para estudiar la inteligencia: sostiene que, si dos habilidades están correlacionadas, es porque dependen de un factor común *g*, y de factores *s* específicos de cada habilidad. El factor *g* considera que es una energía disponible en el cerebro y que difiere cuantitativamente de un individuo a otro, y viene determinado por la genética. En cambio, los factores *s* son habilidades especializadas para cada serie de actividades: verbales, mecánicas, etc.

Para Lewis Terman (1932), la inteligencia es la capacidad de pensar de manera abstracta. Fue el primero en plantear que la inteligencia cambia con la edad. Sostiene que el CI aumenta a medida que el niño va creciendo, llega a su puntuación más alta en la adultez y vuelve a bajar en la vejez, por el deterioro cognitivo propio de la edad. También demuestra en sus estudios que el ambiente social influye en el desarrollo de la capacidad cognitiva, así como la estimulación temprana que los niños reciben.

Thurstone (1924) fue el primero de los psicólogos que estudió la psicometría en los Estados Unidos, considera la inteligencia como la capacidad mental que difiere de una persona a otra, aunque el nivel de desarrollo sea igual, es decir, el mismo nivel de inteligencia pero diferentes maneras de resolución de problemas, que viene determinada por la capacidad mental de cada uno; a esto le denomina capacidades especiales o incapacidades si fuera en sentido negativo, así desarrolla la Teoría Multifactorial para el estudio de las distintas capacidades primarias del hombre.

Cattell es otro autor conocido en el ámbito de la psicometría, sigue el modelo factorial y jerárquico de Spearman y Thurstone, y propone un modelo genetista de la inteligencia (Cattell y Baggaley, 1956), que se refleja en las habilidades cognitivas y se basa en dos factores principales: la inteligencia fluida y la cristalizada. La fluida hace referencia a habilidades no verbales (disposiciones biológicas), y la cristalizada se basa en el conocimiento adquirido desde la experiencia en la interacción con el ambiente. Más adelante Horn redefine la teoría de

la inteligencia cristalizada y fluida y es lo que se conoce en la actualidad (Horn y Cattell, 1966).

David Wechsler (1939) sigue el modelo de inteligencia de Cattell con algunos matices, considera la inteligencia como una capacidad añadida o global para actuar con un propósito, para pensar con la razón y para interactuar de manera efectiva con el medio que le rodea. Es una capacidad añadida, porque consta de varias habilidades o elementos, y es global, porque comprende como un todo el comportamiento individual; no se valora por las habilidades, sino por la capacidad reflejada en su conducta habitual. Su propuesta de medida de inteligencia es considerada uno de los instrumentos más desarrollados a nivel educativo.

Eysenck (1973) sostiene que el concepto de inteligencia surge de la observación de cómo las personas resuelven problemas simples y complejos, y también de la manera de aprender cosas que requieren esfuerzo. Por tanto, para él es una aptitud que se puede reflejar en la práctica y se deduce de la conducta observada, utilizando reglas científicas de procedimiento experimental. Su teoría se apoya en el componente genético y la influencia del ambiente en el CI de una persona.

Sternberg (1987) propone una idea de inteligencia que se relaciona con el comportamiento adaptativo dirigido a un fin. Desarrolla la Teoría Triárquica, en la que considera tres elementos como componentes de la inteligencia (analítica, creativa y práctica) y se valora por la manera de interiorizar la información; estos elementos se explican mediante tres aspectos: el mundo interno de la persona, la experiencia personal -haciendo énfasis en su historia de aprendizaje-, y el contexto. Los componentes de esta teoría coinciden con los dos componentes que Cattell desarrolla en la propuesta de medida de la inteligencia: las habilidades fluidas, que se corresponden con el razonamiento inductivo-deductivo en la resolución de problemas, y la inteligencia cristalizada, que se corresponde con la experiencia y la historia de aprendizaje, con el cómo interioriza y maneja la información añadida en su proceso de adquisición de conocimientos (Sternberg, 1990).

Pruebas de medición de la inteligencia

Binet y Simon (1948) diseñaron la primera prueba psicométrica para la medición de la inteligencia en niños. El instrumento constaba de 30 ítems de dificultad creciente, y la puntuación de los sujetos se interpretaba en base a los valores normativos de su edad. Los resultados permitían la clasificación jerárquica de los sujetos y la detección de problemas del

desarrollo, indicando quién está más adelantado a lo esperable para su edad y quién más atrasado. Se establece esta medición en función del nivel del desarrollo mental del niño.

A partir de los resultados de Simon y Binet, Terman Lewis (1925) desarrolla una prueba incorporando el concepto de Cociente Intelectual desarrollado por Stern (1949). La prueba resultante, denominada de Stanford-Binet, da sus resultados a partir de la puntuación ponderada asignada a cada uno de los factores de la capacidad cognitiva valorada, y los integra en un valor proporcional que define dicho cociente, es decir, las puntuaciones en las distintas dimensiones estudiadas se reflejan en un valor que es el CI total.

Unos años después, Spearman desarrolló una medida de la inteligencia siguiendo su teoría bifactorial: factor g y factor e . El factor g es un factor cuantitativo único y constante de la inteligencia, y está presente en todos los procesos cognitivos de una misma persona, sin embargo, varía de un individuo a otro. El factor e también es un factor cuantitativo, pero variable entre las distintas habilidades de la misma persona y diferente de un individuo a otro, por lo tanto, es propio de cada habilidad, y no depende ni del factor g ni de otros factores e .

Siguiendo el modelo bifactorial de Spearman, Cattell, en el año 1983 publica una medida de inteligencia que propone estudiar la relación de los factores g y e . Para entender esta relación, se necesita comprender el significado del factor que funciona como parámetro, y se determina por su desempeño en el test. También requiere la capacidad de predicción en la composición factorial de una tarea nueva. Y sostiene que a las capacidades mentales primarias subyacen otras capacidades intelectuales, que podrían explicar las relaciones entre capacidades, enfocando un concepto de inteligencia más psicológico que estadístico, distanciándose así de la propuesta de interpretación de Spearman.

Wechsler publica un test que en la actualidad es el más utilizado y revisado, considera que las habilidades añadidas y globales son sensibles a la medición, aunque sostiene que el resultado no es la mera suma de esas habilidades, porque el producto final, es decir, la conducta, depende de la combinación y configuración de cada habilidad; no es el número lo que determina la conducta sino la forma como interactúan entre sí. También hay unos elementos importantes que son las motivaciones y los impulsos que influyen en la conducta. Wechsler considera que cada conducta requiere determinadas habilidades y, que un exceso sería poco eficiente. Esta medida de la inteligencia se publica en el año 1955 y consta de dos escalas: una verbal y otra manipulativa, cuyo resultado valora el CI total de la persona.

Sternberg (1993) en su propuesta de medida de la inteligencia aborda las relaciones entre las variables que considera factores. Sostiene que, dependiendo del nivel de procesamiento en el que se ejecuta dicha comparación, será más elevado el nivel de relación entre la capacidad evaluada y el tiempo de comparación, que tiene que ver con el tiempo que le lleva al sujeto a acceder a la información requerida. Hace una categorización de los procesos con las siguientes distinciones: 1) Distinción entre los procesos ejecutivos y no ejecutivos. 2) Distinción entre los procesos que intervienen en el aprendizaje y los que intervienen en la ejecución de lo ya aprendido. 3) Distinción entre procesos controlados y procesos automáticos. 4) Y, por último, distinción entre diversos niveles de aprendizaje. Se considera su instrumento de medida como una herramienta que complementa a los tradicionales y factoriales.

Principales pruebas para la medición de la inteligencia en España

En España comenzó a utilizarse las escalas de inteligencia en contexto sanitario y educativo, en adelante presentamos las características de las pruebas más utilizadas (Muñiz y Fernández-Hermida, 2010).

La Escala de Inteligencia de Wechsler para adultos (WAIS-IV): es uno de los instrumentos más estudiados y utilizados en Europa y América (Drozdick et al., 2018). Presenta puntuaciones complejas que estudian cuatro áreas del funcionamiento cognitivo (comprensión verbal, razonamiento perceptivo, velocidad de procesamiento y memoria de trabajo), también ofrece el valor de la aptitud intelectual general conocida como CI. Este instrumento está adaptado a la población infantil y se denomina Test de Inteligencia de Wechsler para Niños (WISC-V) en su quinta edición ofrece una valoración de la capacidad de inteligencia en cinco dimensiones, y una valoración general de la inteligencia con identificación de puntos fuertes y débiles del sujeto.

Test de Inteligencia No Verbal (TONI- 4): es una escala que suelen utilizar cuando un sujeto o un grupo presenta dificultades para responder a pruebas tradicionales. Evalúa la inteligencia general a través del razonamiento abstracto y resolución de problemas, se puede hacer de manera verbal y no verbal (Brown et al., 2019).

Evaluación Factorial de las Aptitudes Intelectuales (EFAI): esta escala permite identificar posibles sesgos acuñando el término *estilos de respuestas* en los resultados, esto permite evaluar la coherencia en las respuestas. Pertenece a un grupo de instrumentos llamados de estudios factoriales, y evalúa la capacidad del sujeto para resolver problemas complejos, la flexibilidad intelectual y los procesos lógicos de deducción e inducción (Santamaría et al., 2005).

Matrices progresivas de Raven (2): es un instrumento que valora la aptitud cognitiva mediante formas no verbales, es decir, se trabaja con figuras geométricas que son identificables y no hace falta dar la respuesta de forma oral o escrita; evalúa la capacidad de razonar y resolver problemas complejos; para realizar esta prueba no es importante el nivel educativo (COP, 2020).

RIAS: Escala de Inteligencia de Reynolds: Reynolds (2009), propone un instrumento de medida que estudia tanto la capacidad intelectual como la memoria, se desarrolla en seis escalas: adivinanza, para medir el nivel cultural; categoría, para evaluar el razonamiento abstracto no verbal; analogía verbal, para medir el vocabulario y el razonamiento analítico y verbal; figuras incompletas, para ver su nivel de razonamiento no verbal; memoria verbal, para valorar la memoria a nivel verbal, contextualizada y a corto plazo, y por último, memoria no verbal, para apreciar la memoria inmediata sin componentes verbales, se utiliza esta escala como medida screening de identificación grupal de niños con problemas de aprendizaje.

Objetivo

El objetivo del presente estudio es analizar las propiedades psicométricas de las distintas escalas empleadas para la evaluación de la inteligencia en castellano. Para ello se evaluará:

- 1) Su fiabilidad en términos de estabilidad temporal y consistencia interna.
- 2) Las evidencias de validez en relación con el contenido, a la estructura interna de la prueba y su relación con terceras variables.
- 3) Las características poblacionales en las que existe evidencias positivas de validación.

Método

Estrategia de búsqueda

Se ha realizado una búsqueda sistemática de estudios de validación de las escalas de inteligencia en la base de datos publicados en la *American Psychological Association* (Psycinfo), la página web del Colegio Oficial de Psicólogos y, por último, se realizaron búsquedas manuales en el buscador *Google Scholar*. La estrategia de búsqueda se basó en el descriptor “intelligence” AND "RIAS" OR "Batería III Aprov" OR "WAIS-IV y en las siguientes ecuaciones de búsquedas: “intelligence and measure”, “evaluation and test” y “standard of intelligence” AND "reliability" OR "validity" OR "validation" OR "psychometrics" OR "psychometric properties" AND "Spain" OR "Spanish

Los criterios de inclusión fueron el idioma y el año en el que está publicado el test y su adaptación, hemos dado preferencia al año en que está publicada la validación de los estudios e incluimos el idioma, dando prioridad al inglés, idioma utilizado por la comunidad científica que ofrece cierto rigor a la investigación. Se han localizado 15 estudios para analizar las distintas variables para las tres escalas de inteligencia: RIAS, WAIS-IV y Batería III de Aprovechamiento de Woodcock.

Variables

Como variables de los estudios se ha considerado en primer lugar la fiabilidad de los instrumentos, entendida como la precisión con la que se mide el rasgo psicológico. La fiabilidad se ha analizado tanto en base a la consistencia interna, como a la estabilidad temporal de los resultados. La consistencia interna es el grado en que los ítems se relacionan entre sí, es decir, la medida en que cada ítem mide lo mismo que el resto de los ítems de la escala. La medida más frecuentemente empleada de consistencia interna es el coeficiente Alpha de Cronbach. La estabilidad temporal es el grado en que la escala mantiene su capacidad de medida al valorar un constructo latente en un sujeto, esto se comprueba al aplicar la prueba en distintos momentos y obtener el mismo resultado. La forma más frecuente de evaluarla es a través la técnica test-retest, que consiste en calcular la relación de Pearson entre las puntuaciones de ambos momentos (Prieto y Delgado, 2010).

En segundo lugar, se han recopilado las evidencias de validez de los instrumentos, entendida como el grado en que la teoría y la evidencia empírica apoyan la interpretación de las puntuaciones de un test para un uso concreto (Abad et al., 2011). La validez se ha analizado en base a la estructura interna, y a las relaciones del test con otras variables tanto de forma convergente y discriminante y la validez de criterio de la escala. La estructura interna permite verificar si los ítems concuerdan con el modelo propuesto del constructo que se mide, es decir, mide la relación de los ítems con los componentes de la prueba para ver si se ajustan al constructo a medir. La validez convergente es la medida en que una escala se relaciona con otras escalas que miden el mismo constructo; se espera una relación positiva cuando las escalas miden las mismas variables. La validez discriminante hace referencia al estudio de la escala con variables ajenas al constructo, se espera una baja relación si la escala es válida. Finalmente, la validez de criterio permite observar la relación entre la puntuación en la escala que se emplea y una variable diferente pero relevante para el constructo que mide, por ejemplo, averiguar la relación que hay entre el nivel educativo de los padres con la inteligencia de los

hijos. Se calcula mediante el Coeficiente de correlación de Pearson y el grado de relación determina si el criterio utilizado es válido para la escala (Ríos y Well, 2014).

Por último, se han estudiado aquellas poblaciones en las que las escalas han sido validadas teniendo en cuenta fundamentalmente el sexo, la edad, el nivel educativo o la región de procedencia, pero también cualquier variable mencionada en los estudios que se haya considerado relevante en este ámbito.

Resultados

Resumen de los estudios encontrados y escalas analizadas

Se han seleccionado tres escalas de inteligencia (WAIS-IV, RIAS y Batería III Aprov) para analizar sus propiedades psicométricas en términos de fiabilidad y validez. Para ello hemos identificado 15 artículos que analizan esas variables de las diferentes escalas, a continuación, se resume los resultados de los estudios realizados y analizados (Tabla 1).

Los estudios de validación del WAIS-IV reflejan coherencia entre los resultados de la aplicación en su versión original y la aplicación en la versión en castellano. Coinciden en una buena fiabilidad y validez en las distintas variables analizadas en general. Se observan las evidencias de validez en tres fuentes: estructura interna (Fuica et al., 2016; Rosas et al., 2014) en ambos estudios se encontraron una buena estructura interna; con respecto a la validez en relación con otras variables, en dos estudios analizados (Pluck, 2020; Rosas et al., 2014) se muestran validez convergente analizando otras escalas de inteligencias como el FIX y la Tarea de decisión léxica en español. Finalmente, la validez de criterio se refleja en estudios realizados para analizar la influencia de la urbanidad/ruralidad en las habilidades cognitivas (Fuica et., al 2012) y la relación de rendimiento académico con la inteligencia (López et al., 2016).

En lo referente a la fiabilidad, los estudios demuestran que este instrumento es fiable para evaluar a la población en general y también para la población clínica (Abad et al., 2016; Dasí et al., 2021; Rosas et al., 2014).

La escala de Inteligencia de Reynolds (RIAS) que es una prueba que mide la capacidad intelectual en un extenso rango de edad, de 3 a 94 años, los escasos estudios de validez amparan la fiabilidad y validez de la prueba. En cuanto a las evidencias de validez en relación con otras variables, algunos estudios sobre la validez convergente confirmaron la idoneidad de esta escala reflejando una buena relación con escalas que pretenden medir también la capacidad intelectual, estas escalas analizadas son WISC-IV (Edwards y Paulin, 2007), WJ III (Krach et

al., 2009) y McCarthey (Montañez, 2022). Con respecto a la validez de criterio, Dombrowski y Mrazik (2008) estudiaron la relación entre las dificultades de aprendizaje que presenta un sujeto con su capacidad intelectual, los resultados reflejan buena validez.

Finalmente, las evidencias de fiabilidad estudiados mediante la consistencia interna (Gygi et al., 2019; Santamaría et al., 2009) muestran una buena fiabilidad y el estudio de la estabilidad mediante Test-retest (Santamaría et al., 2009) también ofrece una buena evidencia de la fiabilidad de esta prueba.

La batería III de Aprovechamiento de Woodcock es un conjunto de pruebas que mide la capacidad cognitiva del individuo, su fiabilidad y validez se muestran en los estudios realizados para validar su escala y los resultados demuestran ser adecuados para medir la inteligencia. Tiene una buena estructura interna (Diamantopoulou et al., 2012) y para ver las evidencias de validez en relación con otros criterios, se ha investigado la relación del rendimiento académico con la inteligencia (Rosselli et al., 2001) y la influencia del idioma en el CI (Sandilos et al., 2015) cuyos resultados son positivos, así también se refleja esta idoneidad de la escala en el estudio de validez convergente comparando las puntuaciones de esta escala con el WISC-IV donde se observa una relación positiva.

Tabla 1*Descripción de las características de las muestras estudiadas en los estudios primarios*

Estudios	Nivel educativo	Sexo %	Edad %	Población
			Adolescentes	
			17%	
Rosas et al., 2014	Bajo 33%	Mujer 50%	Adultos	Chile
	Medio 35%		51%	
	Alto 31%	Mayores		
		32%		
	Primarios			
	26%			
Santamaría, 2009	Secundarios	Mujeres:	Niños 61%	España
	26%	51%	Adultos 28%	
	Bachillerato	Varones:	Mayores 11%	
	20%	49%		
	Superiores			
			6-7 años 19%	
Diamantopoulou et al., 2012	1° y 2° EP 34%	Niños	7-8 años 15%	España
	3° y 4° EP 31%	51%	8-10 años 15%	
	5 y 6° EP 35%	Niñas	11-12 años 15%	
		49%	11-12 años 15%	
			12-13 años 17%	
Rosselli et al., 2001	1° y 2° 29%	Niños 49%	6-7 años 29 %	Colombia
	3° y 4° 41%	Niñas 51%	8-9 años 41%	

	5° y 6° 30%		10-11 años 30%	
	Básico 32%			
Fuica et al., 2014	Medio 43%	-	-	Chile
	Superior 34%			
	EA 1 9%			
	EA 2 22%	Hombres		
Abad et al., 2016	EA 3 58%	49%	16-69 años	España
	EA 4 11%	Mujeres 51%		
		Grupo control		
		Hombres		
		25%		
		Mujeres 75%		
Dasí et al., 2021		Grupo	18-65 años	España
		experimento		
		Hombres		
		48%		
		Mujeres 52%		
				EE. UU 68%
		Hombres		Afroamericano
Edwards y Paulin, 2007	Infantil- 5° de EP	54%	6-12 años	19%
		46%		Hispanoamericano
				13%
				EE. UU 33%
		Hombres		Dinamarca 13%
Gygi et al., 2019		50%	3-99 años	Alemania 28%
		Mujeres 50%		España 26%

	Estudios universitarios	Hombres		
Krach et al., 2009	83%	20%	18-56 años	EE. UU
	Estudios superiores 17%	Mujeres 80%		
	Formación profesional 22%	Hombres		
Pluck, 2020		49%	Adultos	Ecuador
	Post grados 78%	Mujeres 51%		
		Hombres		
Sandilos et al., 2016	Infantil	44%	4-5 años	EE. UU
		Mujeres 56%		
Dombrowski y Mrazik, 2008	Primaria y secundaria		Niños y adolescentes	Canadá
	FP de grados medios y estudios universitarios	Hombres		
López et al. 2016		33%	17-65 años	América latina
	Párvulo 1%	Mujeres 67%		
	Pre-jardín 7%			
	Jardín 37%	Hombres		
Montañez y Beltrán, 2022		89%	4-6 años	Colombia
	Transición 47%	Mujeres 11%		
	Primero 3%			
	Sin estudio 4%			

Fiabilidad como consistencia interna

Las evidencias de fiabilidad de la Escala total del WAIS- IV, según el Coeficiente Alpha de Cronbach es de 0,94, y reflejan una excelente consistencia interna (Rosas et al., 2014). Las evidencias de fiabilidad para las subpruebas oscilan entre 0,62-0,93, siendo Cancelación el

subtest que tiene menor valoración y que mide el Índice de Velocidad de Procesamiento de esta escala, esto se realiza analizando el comportamiento de subtest, dimensiones analizadas y escala total (Rosas et al., 2014). Otro estudio donde podemos ver la consistencia interna de esta escala lo realizó Abad et al. (2016) con un valor de $\omega=0,79-0,66$ es aceptable su fiabilidad en este estudio.

En cuanto a la fiabilidad de la Escala de Inteligencia de Reynolds (RIAS), la valoración según estudios de Santamaría (2009) es $\alpha=0,94$, lo que manifiesta una buena consistencia interna, es decir, los distintos ítems de la escala son adecuados para hacer una primera evaluación de la capacidad cognitiva de los sujetos, en las diferentes dimensiones estudiadas de la estructura cognitiva en esta escala.

Gygi et al. (2019) también analiza la fiabilidad de esta escala mediante Omega, el valor es $\omega=0,70$ encontrándose en un nivel aceptable, pero bajo, reflejando una evidencia de consistencia interna baja, pero dentro de lo normal.

Las evidencias de fiabilidad de Batería III Aprovechamiento de Woodcock, se analizan en los estudios realizados por Diamantopoulou et al. (2012), se observa un $\alpha=0,91$, lo que manifiesta una buena consistencia interna, reflejando así la idoneidad de los distintos ítems para valorar la capacidad intelectual de los niños.

Fiabilidad como estabilidad temporal

Como se observa en la tabla 2, se analiza la estabilidad temporal del WAIS-IV aplicando la técnica Test-Retest y se relacionan sus índices en valores entre $r=0,81$ (IMT: Índice de Memoria de Trabajo) y $r=0,94$ para CIT (Cociente Intelectual Total), estas evidencias ponen de manifiesto la capacidad de la prueba en cuanto a la fiabilidad.

Las evidencias encontradas en estabilidad temporal para la Escala de Inteligencia RIAS, las puntuaciones de Test-retest de 14 días entre intervalos de aplicación oscilan por valoración en Índices entre $r=0,88$ (Índice de Memoria de General) y $r=0,95$ (Índice de Inteligencia General); estos valores indican una buena estabilidad que aporta fiabilidad a la escala total.

En Batería III PROV observamos el valor de estabilidad temporal en los resultados provenientes de la aplicación de la técnica Test-retest con siete días de intervalo entre aplicación y los resultados reflejan una correlación de $r\geq 0,93$, lo que demuestra una buena evidencia de fiabilidad en estabilidad temporal.

Tabla 2*Fiabilidad de las Escalas*

	Estudio	Consistencia interna		Test-retest
		Alpha	Omega	Menos de 6 meses
WAIS- IV	Rosas et al. 2016	0,94	---	0,81-0,94
WAIS- IV	Dasí et al. 2021	0,80	---	---
WAIS- IV	Abad et al. 2016	---	0,79-0,66	---
RIAS	Santamaría, 2009	0,94	---	0,88-0,95
RIAS	Gygi et al. 2019	---	0,70	---
Batería III Aprov	Diamantopoulou et al. 2012	0,91	---	0,93

Nota: WAIS: Escala de inteligencia de Wechsler. RIAS: Test de Inteligencia de Reynolds. Batería III APROV: Escala de aprovechamiento de Woodcock.

Evidencias de validez en relación con la estructura interna

Respecto a la estructura interna del WAIS-IV como se observa en la tabla 3, y que se evalúa mediante el Análisis Factorial Confirmatorio, en Rosas et al. (2014) observamos que las puntuaciones de cada subprueba con relación a los cuatro factores están entre $r=0,57$ y $r=0,87$. En conclusión, se justifica la organización propuesta en esta escala, que sigue un modelo factorial de segundo orden con cuatro dimensiones cognitivas, que funcionan como primer orden (Comprensión Verbal, Razonamiento Perceptual, Memoria de Trabajo y Velocidad de Procesamiento), y éstas contribuyen a la valoración del factor “g” de la escala total (CI).

Por último, Abad et al. (2016) valora el índice de ajuste al modelo propuesto mediante RMSEA cuyo valor es de 0,04, lo que podemos considerarlo una buena evidencia.

La estructura interna del RIAS, se ajusta al modelo factorial de segundo orden como factor “g” y dos factores de primer orden: inteligencia verbal e inteligencia no verbal, se observa en el estudio de Gygi et al. (2021) con un valor RMSEA de 0,31 en la que consideramos un buen ajuste del modelo propuesto.

Batería III Aprox corresponde a un modelo bifactorial en el que se confirma el estudio de un factor general y dos factores de segundo orden que se explicitan en las puntuaciones de las cuatro dimensiones especiales estudiadas (cálculo, resolución de problemas, conceptos y series), se valoró mediante el análisis factorial confirmatorio, con una valoración significativa $RMSEA=0,10$.

Evidencias de validez convergente

WAIS-IV + FIX. Se realizó un estudio de validación de la escala con la batería de Evaluación de la Inteligencia Fluida para valorar la validez convergente del WAIS-IV y se encontró una buena correlación de $r=0,62$, lo que refleja una relación positiva entre las dos escalas que miden la misma variable con similares dimensiones de la inteligencia.

FIX: Evaluación de Inteligencia Fluida: es un test de screening que permite la evaluación de la inteligencia fluida, se utiliza como alternativa a las escalas de medidas tradicionales. Cuenta con 20 ítems y sirve para medir de manera rápida la capacidad intelectual, ahorrando tiempo (Rosas et al., 2012).

WAIS-IV + Tarea de decisión léxica en español. Pluck (2020) analiza la validez convergente de esta escala estudiando la relación de una escala que mide variables similares y encontró una relación $r=0,71$, lo que refleja una buena relación entre las dos escalas que miden la inteligencia.

RIAS y WISC- IV. Según estudios de Edwards y Paulin (2007) se validó la escala de RIAS en término de validez convergente aplicando paralelamente la Escala de Inteligencia de Wechsler para niños (WISC- IV) al observarse una correlación de $r=0,90$ entre ambas escalas de inteligencia, se concluye que aporta validez a dicha escala.

RIAS y WJ III (Woodcock Johnson). Krach et al. (2009) realizaron un estudio de validación de la escala de RIAS en términos de validez convergente, es decir, estudiando la relación de esta escala con otra escala similar, y encontró una relación $r=0,86$ que refleja una buena evidencia de validez para esta escala.

RIAS y McCarthy. Montañez y Beltrán (2022) estudiaron la relación entre la escala de inteligencia RIAS con la escala McCarthy y encontraron un $r=0,86$ que refleja una buena relación entre las dos escalas.

Batería III PROV y WISC- IV. Se estudió la relación de esta escala con el WISC-IV y presenta una relación significativa de $r=0,86$, esto refleja que las dos escalas miden variables similares, es decir, la inteligencia (Diamantopoulou et al., 2012).

Evidencias de validez de criterio

Al evaluar la validez de criterio del WAIS-IV, analizando el estudio de las habilidades cognitivas en relación con la población a la que se aplica, se encontró una relación positiva en la que se observa que, la ubicación urbana o rural, influye en el rendimiento de la valoración intelectual. Se encontró que hay una diferencia significativa de 14 puntos a favor de la población proveniente de zona urbana en comparación con la rural, es decir, los provenientes de zonas urbanas obtenían puntuaciones más altas que los provenientes de zonas rurales.

Siguiendo con la evaluación de la validez de criterio, López et al. (2016) estudia la relación de la atención como capacidad cognitiva con la inteligencia y descubrió que hay una relación significativa entre ambas variables de $r=0,5$ reflejando una buena evidencia de validez de criterio.

Se analiza la validez de criterio del RIAS estudiando la relación entre la inteligencia y el rendimiento académico, y se observa una relación de entre $r=0,41$ a $r=0,73$ para las diferentes dimensiones de la inteligencia que mide RIAS, sin embargo, no se aporta una valoración global de la escala. Se demostró que los sujetos con rendimiento alto en matemáticas, lectura, escritura y lenguaje obtienen puntuaciones más altas en la prueba de la inteligencia, sin embargo, los sujetos que tienen bajos rendimientos no presentan una valoración CI por debajo de la norma.

Un estudio de Rosselli et al. (2001) investigó la relación de las pruebas de Batería III Aprox, con el rendimiento académico con el fin de predecir mediante estas puntuaciones las expectativas de rendimiento en el futuro del sujeto y de describir el perfil neuropsicológico de niños hispanohablantes con Dificultades de Aprendizaje. Se observó que hay relaciones significativas entre las subpruebas de Batería III Aprox y el rendimiento en algunas asignaturas académicas, refleja que entre el subtest Cálculos y la asignatura de Ciencias hay una relación significativa, así como el subtest de Idiomas se relaciona con el desempeño escolar en lenguaje. Los valores de la relación de Pearson oscilan para estas pruebas en $r=0,16-0,31$.

Sandilos et al. (2015) analiza la relación entre el idioma con la inteligencia, es decir, si personas que hablan varios idiomas dialecto, éstas influyen en la capacidad intelectual y vemos que la diferencia es estadísticamente significativa ($p < .05$).

Tabla 3

Evidencias de validez de las escalas de inteligencia en relación con la estructura interna y a su relación con otras variables

Escala	Estudio	Estructura interna		Validez convergente	Validez de criterio
		r Pearson	RMSEA	r Pearson	r Pearson
WAIS- IV	Rosas et al. 2016	0,62-0,94	---	0,79 +FIX	---
WAIS- IV	Fuica et al. 2012	---	---	---	SD= 10 Ruralidad-urbanidad
WAIS- IV	Pluck, 2020	---	---	0,71 +SLVT	---
WAIS- IV	López, 2016	---	---	---	0,47 Atención
WAIS- IV	Abad et al. 2016	---	0,040- 0,046	---	0,35 Rendimiento académico
RIAS	Edwards y Paulin, 2007	---	---	0,90 + WISC-IV	---
RIAS	Krach et al. 2009	---	---	0,72-0,86 RIAS+WJ III	---
RIAS	Montañez, 2022	---	---	0,86 +McCarthy	---
RIAS	Dombrowski et al. 2008	---	---	---	0,41-0,73

					Dificultad de aprendizaje
RIAS	Gygi et al. 2019		0,31		
Batería III Aprov	Diamantopoulou et al. 2012	---	0,10	0,86 +WISC-IV	---
Batería III Aprov	Sandilos et al. 2015	---	---	---	p>0.05 Idiomas
Batería III Aprov	Rosselli et al. 2001	---	---	---	0,16-0,31 Rendimiento académico

Ámbito de aplicación de cada escala

Se analizaron tres escalas de diferentes aplicaciones como se describe en la tabla 4. La Escala de Inteligencia de Wechsler (WAIS-IV) valora la inteligencia de adultos en Chile comprendidos entre 16 y 90 años; la Escala de Inteligencia de Reynolds evalúa la inteligencia de toda la población española desde los infantes hasta los mayores, por último, la Batería de Aprovechamiento de Woodcock validados para su uso en niños de seis a 13 años.

Tabla 4*Población Aplicada*

Pruebas.	Ámbito de aplicación
WAIS- IV	La Escala WAIS-IV fue validada para la población chilena de personas de entre 16-90 años, tanto para hombres como para mujeres y tiene una representatividad de la población para dicho país.
RIAS	Población española de 2.065 sujetos, en donde el 51% son mujeres y 49% hombres de entre 2 y 94 años, con una participación representativa de todas las zonas del país (norte, sur, este y centro) y variabilidad en los niveles de estudios alcanzados (estudios primarios, secundarios y superiores). Por lo que la escala es apta para medir la inteligencia de la población estudiada
Batería III APROV	La prueba fue validada analizando la muestra de 424 niños de seis a 13 años provenientes de la región de Murcia con una muestra seleccionada de ocho colegios urbanos y no urbanos con nivel socioeconómico bajo, medio y algo.

Discusión

Las evidencias de fiabilidad demostradas en los diferentes estudios de las tres escalas reflejan ser adecuadas para medir la inteligencia general de la población. Sin embargo, un estudio sobre la fiabilidad de la escala WAIS-IV indica que, si bien el modelo factorial es debidamente explicado, al comparar las puntuaciones del factor *g* con la inteligencia fluida, los resultados son similares, por lo que ven la necesidad de definir mejor dicho concepto para mejorar la precisión de la escala con la teoría que lo apoya (Abad et al., 2016). Sabemos que la inteligencia fluida hace relación a la disposición genética o herencia biológica de la inteligencia y el factor *g* es la capacidad general del sujeto, por lo que al resultar muy similares las puntuaciones, desde esta perspectiva, se podría reducir la inteligencia al modelo genetista de la inteligencia de Galton, superado desde hace tiempo.

Otro dato importante con respecto a la fiabilidad de esta escala con relación a poblaciones clínicas, reflejada en el estudio de Dasí et al. (2021) en el que se diseña una versión corta de la

escala para pacientes con Esquizofrenia, estos autores resaltan la necesidad de más estudios y adaptaciones a este respecto, ya que consideran que hay algunas patologías que influyen en la capacidad cognitiva tanto por la enfermedad como por los efectos secundarios del tratamiento, de todos los estudios analizados en este trabajo esta versión del WAIS-IV es el que menor valor de fiabilidad obtuvo, aunque igualmente sea fiable.

Con respecto a la estructura interna de las escalas, hemos visto que las evidencias de validez van de moderados a excelentes para las diferentes escalas. Llama la atención que en varios estudios presentan los valores de la estructura interna por dimensiones estudiadas, pero no de la escala entera y en lo que se refiere a la escala de RIAS, los estudios son bien escasos, se considera importante un investigación en este ámbito ya que esta escala es una de las únicas que miden la inteligencia humana como un continuo, no hay versiones para niños ni para adultos, por lo que estudiar el comportamiento de los ítems para cada etapa del ciclo evolutivo supone un desafío y un valor para la escala, ya que su estructura interna es adecuada, aunque varía un poco para algún rango de edad. Casi todos los autores que hacen estudios de validación de la escala WAIS-IV concuerdan en que en esta versión han mejorado mucho la fiabilidad y validez de la escala, es decir, demuestran un interés y un preocupación por las propiedades psicométricas de la escala, no solo se enfocan en la incorporación de los nuevos avances producidos en el campo de la inteligencia, como lo hicieron esta vez enriqueciendo el modelo CHC (Cattell-Horn-Carroll) que es la base de la estructura factorial jerárquica que sigue Wechsler en su escala, aunque como hemos explicitado anteriormente, es posible que se necesite aclarar la definición de la inteligencia fluida para diferenciarlo mejor del factor g.

Las evidencias de validez convergente demuestran que estas pruebas de inteligencia arrojan puntuaciones similares a otras pruebas que también pretenden medir la inteligencia, son varios los estudios que se realizan a este respecto y cabe mencionar que la escala preferida para realizar un estudio de validez convergente es el WAIS, sea en su versión para adultos como para niños, hemos visto que según las estadísticas el WAIS es uno de los instrumentos más utilizados en Europa y América para evaluar la capacidad cognitiva, en general las relaciones son positivas.

La validez de criterio es una de las variables que resulta bastante útil, no solo para la confirmación de la validez de la escala sino también para ver el comportamiento de ciertas variables relacionadas con la inteligencia humana. Esto de cara a intervenciones es de vital importancia, tanto para predecir, como para prevenir y mejorar la habilidad cognitiva. Rosas et al. (2016) en su estudio sobre la relación de la zona de procedencia (urbanidad/ruralidad) con la

inteligencia, comprobó que aquellas personas que provienen de zona rural y que además el nivel educativo de los padres son bajos presentaban rendimientos más bajos que aquellos que provienen de la urbanidad, y también que los sujetos que provienen de zonas urbanas con un nivel educativo alto de los padres presentaban un rendimiento significativamente más alto que la media de poblaciones rurales, en ello reivindican la necesidad del acceso a la educación y de una educación de calidad para todos, sin tener en cuenta la zona de procedencia. El rendimiento académico también es una de las variables que despiertan bastante interés en los investigadores, curiosamente en las tres escalas se estudiaron el rendimiento académico para ver la validez en relación con otras variables de sus pruebas y en los tres encontraron que las relaciones son significativas, los sujetos que presentaban menor rendimiento académico también tenían un CI menor. Por otra parte, Sandilos et al. (2015) llevaron a cabo una investigación con el objetivo de ver la relación del idioma (bilingüismo) español-inglés en las habilidades cognitivas y llegaron a la conclusión de que aplicar una escala de inteligencia en un idioma que no es el primer idioma del sujeto, influye en el rendimiento en las valoraciones del CI, por ello ven la necesidad de o bien realizar una estandarización mixta que represente la realidad de los sujetos o bien mejorar la precisión en la dificultad de los diferentes ítems que se muestran más sensibles a estos sesgos.

La mayoría de las investigaciones respecto a la validez de las escalas centran su atención en la población en general, sin embargo, se echan en falta más estudios para poblaciones clínicas, por ejemplo, para sujetos con discapacidad física o dificultad motora, el único estudio se realizó para pacientes con esquizofrenia (Dasí et al., 2021) y también faltan más estudios para poblaciones especiales como personas con TDAH, TEA, Asperger, discalculia... Es verdad que varios estudios investigan la relación de la dificultad de aprendizaje y el rendimiento académico que pueden deberse a estas circunstancias, pero se podría precisar los estudios a este respecto.

El presente trabajo tuvo como objetivo principal evaluar la fiabilidad y validez de las escalas de inteligencia editadas en castellano, la limitación principal que se ha encontrado es que son pocos los estudios de validación realizados utilizando una muestra española o latinoamericana, o a veces los estudios utilizaban muestras muy pequeñas, por lo que, como analizamos test editados al español utilizamos estudios con muestras de EE. UU y de otros países de habla no hispana pero que los sujetos analizados son bilingües y aplicaron un test de adaptación en español, así se pudo enriquecer este trabajo no solo analizando muestras del idioma de

procedencia sino cómo se comportan estas escalas adaptadas con sujetos que no viven ni en España ni en Latinoamérica, pero hablan el español y son evaluados en ese idioma.

Conclusiones

- Hemos encontrado evidencias adecuadas de fiabilidad para las tres escalas, tanto en relación con la consistencia interna como con su estabilidad temporal, por tanto, concluimos que estas escalas miden con precisión la dimensión cognitiva evaluada, que es la inteligencia.
- Respecto a la estructura interna, algunos autores ofrecen una valoración global de la escala y otros una valoración por dimensión analizada en la escala, ambas evidencias reflejan una apropiada validez, aunque se pueden observar más variabilidad en las valoraciones por dimensiones o factores estudiados, identificando así qué factores o qué dimensiones son más o menos explicados.
- En la validez convergente se observa que cada escala evaluada, para su estudio, se sirve de otras escalas similares para valorar en qué medida están midiendo la inteligencia, y todos los estudios demuestran una adecuada validez convergente, ya que las puntuaciones son similares.
- Por último, tenemos la validez de criterio, en este sentido los autores han estudiado la relación de la escala de inteligencia con otras variables, como el contexto y la atención en el caso del WAIS-IV, el rendimiento académico en RIAS y Batería III APROV y además el idioma para ésta última escala. Se ha demostrado que estas variables están relacionadas con la inteligencia, y suponen un factor de riesgo o protección para los sujetos.
- Finalmente identificamos la población a la que puede ser aplicable estas escalas; se resalta en los estudios el interés de representatividad de la muestra, estudian grupos provenientes de todo el país, separados en distintas zonas, se aseguran la proporcionalidad entre hombres y mujeres, adultos y niños y provenientes de diferentes niveles de estudios. Esto hace que estas escalas sean extrapolables para la población estudiada. También resaltamos la presentación de algunas escalas adaptadas a la población infantil y adultos; midiendo de manera verbal y no verbal; para poblaciones clínicas o no, esto hace que las evaluaciones sean más específicas teniendo en cuenta el desarrollo y madurez adquirido.

Referencias

- Abad, F. J., Sorrel, M. A., Román, F. J., & Colom, R. (2016). The relationships between WAIS-IV factor index scores and educational level: A bifactor model approach. *Psychological Assessment, 28*(8), 987–1000.
<https://doi.org/10.1037/pas0000228>
- Abad, F.J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Síntesis.
- Ballester-Plané, J., Laporta-Hoyos, O., Macaya, A., Póo, P., Meléndez-Plumed, M., Vázquez, É., Delgado, I., Zubiaurre-Elorza, L., Narberhaus, A., Toro-Tamargo, E., Russi, M. E., Tenorio, V., Segarra, D., & Pueyo, R. (2016). Measuring intellectual ability in cerebral palsy: The comparison of three tests and their neuroimaging correlates. *Research in Developmental Disabilities, 56*, 83–98. <https://doi.org/10.1016/j.ridd.2016.04.009>
- Binet, A. (1903). Suggestions from unconscious stimuli (continued)--Time measurements. In *Alterations of personality*. (pp. 286–291). D Appleton & Company.
<https://doi.org/10.1037/11049-017>
- Binet, A. (1983). La inteligencia su medida y educación. *Journal for the Study of Education and Development, 22*, 115–120.
<https://dialnet.unirioja.es/servlet/articulo?codigo=668635>
- Binet, A., & Simon, T. (1948). The development of the Binet-Simon Scale, 1905-1908. In W. Dennis (Ed.), *Readings in the history of psychology* (pp. 412–424). Appleton-Century-Crofts. <https://doi.org/10.1037/11304-047>
- Brown., L., Sherbenou, R., & Johnsen, S. (2019). Evaluación del Test de Inteligencia no Verbal (Toni-4). *COP*. <https://www.cop.es/uploads/PDF/2019/TONI-4.pdf>
- Cattell, R. B., & Baggaley, A. R. (1956). The Objective Measurement of Attitude Motivation: Development and Evaluation of Principles and Devices 1. *Journal of Personality, 24*(4), 401-423.
- Dasí, C., Fuentes-Durá, I., Ruiz, J. C., & Navarro, M. (2021). Four-subtest short-form of the WAIS-IV for assessment of patients diagnosed with schizophrenia. *Revista de psiquiatría y Salud Mental, 14*(3), 139–147.
<https://doi.org/10.1016/j.rpsmen.2019.12.003>

- Diamantopoulou, S., Pina, V., Valero-Garcia, A. V., González-Salinas, C., & Fuentes, L. J. (2012). Validation of the Spanish Version of the Woodcock-Johnson Mathematics Achievement Tests for Children Aged 6 to 13. *Journal of Psychoeducational Assessment, 30*(5), 466–477. <https://doi.org/10.1177/0734282912437531>
- Dombrowski, S. C., & Mrazik, M. (2008). Test Review of RIAS: Reynolds Intellectual Assessment Scales. *Canadian Journal of School Psychology, 23*(2), 223–230. <https://doi.org/10.1177/0829573508324458>
- Edwards, O., & Paulin, R. (2007). Referred Students' Performance on the Reynolds Intellectual Assessment Scales and the Wechsler Intelligence Scale for Children—Fourth Edition. *Journal of Psychoeducational Assessment, 25*(4), 334–340. <https://doi.org/10.1177/0734282907300453>
- Extracto del capítulo V de libro de A. Binet, *Les idées modernes sur les enfants*. Paris, Flammarion, 1973, pp. 65-113. Publicación original en 1911. © de esta traducción, *Infancia _y Aprendizaje, 1983*.
- Eysenck, H. (1973). *The measurement of intelligence*. MTP.
- Fuica, P., Lira, J., Alvarado, K., Araneda, C., Lillo, G., Miranda, R., Tenorio, M., & Pérez-Salas, C. (2014). Habilidades Cognitivas, Contexto Rural y Urbano: Comparación de Perfiles WAIS-IV en Jóvenes. *Terapia Psicológica, 32*(2), 143–152.
- Gygi, J. T., Ledermann, T., Grob, A., Rudaz, M., & Hagmann-von Arx, P. (2019). The Reynolds Intellectual Assessment Scales: Measurement Invariance across Four Language Groups. *Journal of Psychoeducational Assessment, 37*(5), 590–602.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253–270. <https://doi.org/10.1037/h0023816>
- Krach, S. K., Loe, S. A., Jones, W. P., & Farrally, A. (2009). Convergent Validity of the Reynolds Intellectual Assessment Scales (RIAS) Using the Woodcock-Johnson Tests of Cognitive Ability, Third Edition (WJ-III) with University Students. *Journal of Psychoeducational Assessment, 27*(5), 355–365.
- López, E., Steiner, A. J., Hardy, D. J., IsHak, W. W., & Anderson, W. B. (2016). Discrepancies between bilinguals' performance on the Spanish and English versions of the WAIS

- Digit Span task: Cross-cultural implications. *Applied Neuropsychology: Adult*, 23(5), 343–352. <https://doi.org/10.1080/23279095.2015.1074577>
- Meili, R. (1972) *Intelligence*. En Eysenck, H., & Würzburg, W. (eds.), *Encyclopedia of psychology*. (Vol.2, pp. 144-147). Search Press.
- Montañez Romero, M. A., & Beltrán de la Rosa, E. E. (2022). Validez concurrente: Escala de Inteligencia Reynolds Rías y McCarthy para la evaluación del desarrollo cognitivo en niños. *Revista Iberoamericana de Diagnóstico y Evaluación-e Avaliação Psicológica*, 1(62), 29–36. <https://doi.org/10.21865/RIDEP62.1.03>
- Mülberger, A. (2015). Origen y evolución del test de inteligencia. *Investigación y Ciencia*. <https://www.investigacionyciencia.es/revistas/mente-y-cerebro/la-neurociencia-del-futuro-641/origen-y-evolucion-del-test-de-inteligencia-13330>
- Norman, G. (2014). When I say... reliability. *Medical Education*, 48(10), 946–947. <https://doi.org/10.1111/medu.12511>
- López, E., Steiner, A. J., Hardy, D. J., IsHak, W. W., & Anderson, W. B. (2016). Discrepancies between bilinguals' performance on the Spanish and English versions of the WAIS Digit Span task: Cross-cultural implications. *Applied Neuropsychology: Adult*, 23(5), 343–352. <https://doi.org/10.1080/23279095.2015.1074577>
- Pluck, G. (2020). A lexical decision task to measure crystallized-verbal ability in Spanish. *Revista Latinoamericana de Psicología*, 52(1), 1–10. <https://doi.org/10.14349/rlp.2020.v52.1>
- Reynolds, C. (2009). *RIAS (Escala de inteligencia de Reynolds)* (R. Kamphaus, Ed.). TEA.
- Rosas, R., Tenorio, M., Pizarro, M., Cumsill, P., & Bosch, A. (2014). Estandarización de la Escala Wechsler de inteligencia para adultos- IV edición en Chile. *Psykhé*, 23, 1–18. <https://doi.org/10.7764/psykhe.23.1.529>
- Rosselli, M., Ardila, A., Bateman, J. R., & Guzman, M. (2001). Neuropsychological Test Scores, Academic Performance, and Developmental Disorders in Spanish-Speaking Children. *Developmental Neuropsychology*, 20(1), 355–373. https://doi.org/10.1207/S15326942DN2001_3
- Sandilos, L. E., Lewis, K., Komaroff, E., Hammer, C. S., Scarpino, S. E., Lopez, L., Rodriguez, B., & Goldstein, B. (2015). Analysis of bilingual children's performance on the English and Spanish versions of the Woodcock-Muñoz Language Survey-R

- (WMLS-R). *Language Assessment Quarterly*, 12(4), 386–408.
<https://doi.org/10.1080/15434303.2015.1100198>
- Santamaría, P. (2009). Adaptación española de la Escala de Inteligencia de Reynolds. *TEA ediciones*. <https://web.teaediciones.com/Ejemplos/RIST.pdf>
- Santamaría, P., Arribas, D., Pereña, J., & Seisedos, N. (2005). Evaluación del test EFAI. *COP*. <https://www.cop.es/uploads/PDF/EFAI.pdf>
- Spearman, C. (1931). Our need of some science in place of the word “intelligence.” *Journal of Educational Psychology*, 22(6), 401–410. <https://doi.org/10.1037/h0070599>
- Spearman, C. (1955). *Las habilidades del hombre: su naturaleza y medición*. Paidós.
- Stern, W. (1914). Estimation and Testing of Finer Gradations of Intelligence (Method of Ranks). In *The psychological methods of testing intelligence*. (Vol. 13, pp. 109–144). Warwick & York. <https://doi.org/10.1037/11067-003>
- Stern, W. (1925). Theory of Constancy of Intelligence (H. Klüver, Trad.). *The Psychological Clinic*, 16, 110–118. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5149947/>
- Stern, W. (1949). *El cociente de inteligencia*. En W. Dennis (Ed.), *Lecturas en psicología general* (págs. 338–341). Prentice-Hall, Inc. <https://doi.org/10.1037/11352-048>
- Sternberg, R. (1987). Concepciones de la inteligencia. En *Inteligencia humana I. La naturaleza de la inteligencia y su medición*. (1.^a ed., Vol. 1, pp. 17–51). Paidós.
- Sternberg, R. (1990). *Más allá del cociente intelectual*. Editorial Desclee de Brouwer.
- Terman, L. (1932). *The measurement of intelligence* (3.^a ed.). The Riverside Press Limited.
- Termann, L., & Merril, M. (1979). *Medida de la inteligencia* (Revisado ed.). Espasa Calpe.
- Thurstone, L. L. (1924). A Definition of Intelligence. In *The nature of intelligence*. (pp. 156–163). Harcourt Brace & Company. <https://doi.org/10.1037/11388-011>
- Wechsler, D. (1939). *The measurement of adult intelligence*. Williams & Wilkins Co.
<https://doi.org/10.1037/10020-000>
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale--Fourth Edition (WAIS-IV)*. APA PsycTests. <https://doi.org/10.1037/t15169-000>
- William Stern (1871–1938) Psychologist. (2018). *Human Intelligence*.
<https://www.intelltheory.com/stern.shtml>