Ciencias y Saberes

ISSN 2960-8430

PRUEBAS DE VALIDEZ EN EVALUACIÓN NEUROPSICOLÓGICA FORENSE EN ADULTOS: REVISIÓN SISTEMÁTICA

Crhystelle Elena Ortiz Vidal
crhystelle.ortiz.53@est.ucacue.edu.ec
ORCID: https://orcid.org/0009-0003-5483-8693
Maestría en Psicología Forense
Universidad Católica de Cuenca - Ecuador

Erick Gabriel Valarezo Guzmán
erick.valarezo@ucacue.edu.ec
ORCID: https://orcid.org/0000-0003-2913-9228
Maestría en Psicología Forense
Universidad Católica de Cuenca - Ecuador

Recibido: 23/03/26
Aceptado: 22/04/26
Publicado: 01/05/26

RESUMEN

Las evaluaciones neuropsicológicas en contextos forenses requieren herramientas que permitan valorar la credibilidad del desempeño cognitivo y de los síntomas informados. En este marco, las pruebas de validez del rendimiento (PVT) y las pruebas de validez de los síntomas (SVT) se emplean para identificar respuestas no creíbles o clínicamente improbables. El objetivo de este estudio fue identificar y sintetizar la evidencia sobre las principales PVT y SVT utilizadas en población adulta en contextos forenses, médico-legales y clínico-forenses. Para ello, se realizó una revisión sistemática conforme a la guía PRISMA 2020, efectuando búsquedas en PubMed, Scopus y Web of Science para el periodo 2015-2026 y se incluyeron 14 estudios. Los hallazgos muestran el uso frecuente de instrumentos como el Test of Memory Malingering (TOMM), el Word Memory Test (WMT), el Reliable Digit Span (RDS), el Groningen Effort Test (GET) y el IOP-29, además de indicadores embebidos y enfoques multimétodo. En conjunto, los resultados sugieren que la integración de varios indicadores de validez fortalece la interpretación de los hallazgos en la evaluación neuropsicológica forense.

Palabras clave: neuropsicología forense, validez del rendimiento, validez de síntomas, simulación, evaluación neuropsicológica.

VALIDITY TESTS IN FORENSIC NEUROPSYCHOLOGICAL ASSESSMENT OF ADULTS: A SYSTEMATIC REVIEW

ABSTRACT

Neuropsychological assessments in forensic settings require tools that allow for the evaluation of the credibility of cognitive performance and reported symptoms. In this context, performance validity tests (PVT) and symptom validity tests (SVT) are used to identify responses that are implausible or clinically unlikely. The objective of this study was to identify and synthesize the evidence regarding the main PVTs and SVTs used in adult populations in forensic, medico-legal, and clinical-forensic contexts. To this end, a systematic review was conducted in accordance with the PRISMA 2020 guidelines, with searches performed in PubMed, Scopus, and Web of Science for the period 2015–2026, and 14 studies were included. The findings show the frequent use of instruments such as the Test of Memory Malingering (TOMM), the Word Memory Test (WMT), the Reliable Digit Span (RDS), the Groningen Effort Test (GET), and the IOP-29, in addition to embedded indicators and multi-method approaches. Taken together, the results suggest that the integration of various validity indicators strengthens the interpretation of findings in forensic neuropsychological assessment.

Keywords: forensic neuropsychology, performance validity, symptom validity, malingering, neuropsychological assessment.

Correo principal para contacto: crhystelle.ortiz.53@est.ucacue.edu.ec

1. INTRODUCCIÓN

La evaluación neuropsicológica forense es un proceso especializado cuyo propósito es ofrecer evidencia científica acerca del funcionamiento cognitivo y psicológico de una persona dentro de procedimientos judiciales. A diferencia de la evaluación clínica convencional, se lleva a cabo en entornos adversariales, donde los resultados pueden influir directamente en decisiones legales relacionadas con la responsabilidad penal, discapacidad o indemnizaciones económicas (Neal et al., 2022). En este contexto, el perito psicólogo debe basar sus conclusiones en datos confiables, procedimientos estandarizados y criterios metodológicos estrictos, dada la gran repercusión jurídica de sus hallazgos (Sweet et al., 2021).

Uno de los mayores retos en esta área es la posible distorsión intencional del desempeño cognitivo o de los síntomas que se reportan. Esto puede manifestarse mediante simulación, exageración o la creación deliberada de déficits, motivada generalmente por incentivos externos propios del ámbito medicolegal (Young, 2014). La detección de estas respuestas no creíbles es fundamental, ya que afecta directamente la validez de las conclusiones periciales.

En respuesta a este desafío, la documentación científica ha desarrollado diversas herramientas diseñadas para evaluar la credibilidad, tanto del rendimiento como de los síntomas durante la evaluación neuropsicológica. Estas técnicas se dividen principalmente en dos categorías: pruebas de validez del rendimiento (Performance Validity Tests, PVT) y pruebas de validez de los síntomas (Symptom Validity Tests, SVT). Las PVT permiten identificar patrones de desempeño incompatibles con un esfuerzo auténtico durante la evaluación cognitiva (Boone, 2007), mientras que las SVT valoran la plausibilidad, coherencia y veracidad de los síntomas declarados por el evaluado (Giromini et al., 2022). Actualmente, la implementación sistemática de estas pruebas es considerada un elemento esencial en la evaluación neuropsicológica forense, ya que mejora la validez de la interpretación de los resultados y reduce el riesgo de errores diagnósticos o conclusiones incorrectas (Heilbronner et al., 2009; Slick et al., 1999).

En las últimas décadas, la investigación en torno a la validez del rendimiento y de los síntomas ha tenido un crecimiento significativo en el campo de la neuropsicología forense. Diversos autores han indicado que el uso combinado de múltiples indicadores de validez, incluyendo pruebas independientes e indicadores embebidos, puede aumentar la precisión diagnóstica y facilitar la distinción entre déficits cognitivos genuinos y respuestas relacionadas con simulación o falta de esfuerzo (Larrabee, 2012). Sin embargo, la evidencia disponible muestra una considerable heterogeneidad en términos de instrumentos empleados, contextos de aplicación, poblaciones estudiadas y criterios de interpretación, lo que dificulta la comparación directa entre estudios y la consolidación de prácticas basadas en evidencia.

Dentro de este escenario, las revisiones sistemáticas son una herramienta metodológica clave para sintetizar de forma organizada y transparente la evidencia científica existente sobre un fenómeno determinado. Estas revisiones permiten integrar los hallazgos de múltiples investigaciones, evaluar la coherencia de los resultados y detectar lagunas en el conocimiento. La declaración PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) ofrece directrices ampliamente

aceptadas para la elaboración y presentación de revisiones sistemáticas en ciencias de la salud y sociales (Page et al., 2021).

A pesar del crecimiento constante en la bibliografía de este ámbito, sigue siendo necesaria una integración sistemática de la evidencia reciente acerca del uso de pruebas de validez del rendimiento y de los síntomas en contextos forenses, médico-legales y clínico-forenses donde existen incentivos externos, especialmente, en población adulta. Por ello, el presente estudio tuvo como finalidad identificar y recopilar la evidencia científica disponible sobre las principales pruebas de validez del rendimiento y de los síntomas aplicadas en evaluaciones neuropsicológicas realizadas a adultos en dichos contextos, mediante una revisión sistemática de artículos publicados entre 2015 y 2026.

2. ESTRATEGIAS METODOLÓGICAS / MATERIALES Y MÉTODOS

Este trabajo se llevó a cabo como una revisión sistemática de la documentación científica. La metodología se ajustó a las directrices establecidas por la declaración PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), garantizando así la transparencia y rigurosidad en los procesos de búsqueda, selección y síntesis de los estudios incluidos (Page et al., 2021).

Con el propósito de asegurar la relevancia y coherencia de la evidencia recopilada, se definieron criterios claros de inclusión. Para esto, se consideraron estudios empíricos originales cuantitativos, cualitativos y mixtos, que aborden el uso, validez o eficacia de las pruebas de validez del rendimiento (PVT) y/o de los síntomas (SVT) en evaluaciones neuropsicológicas realizadas en ámbitos forenses o clínico-forenses. Asimismo, se incorporaron investigaciones de validación de instrumentos, estudios observacionales y correlacionales vinculados a la detección de rendimiento o sintomatología simulada o no fiable. Además, se incluyeron documentos de consenso de alta relevancia para el área, con el fin de contextualizar y complementar la evidencia empírica recopilada.

Los estudios debían centrarse en población adulta (≥18 años) evaluada en contextos periciales, médico-legales o en situaciones con incentivos externos relevantes (responsabilidad penal, litigios civiles, daño personal, discapacidad o accidentes). De esta manera, se incluyeron publicaciones provenientes de diferentes países, sin restricción geográfica. Además, se consideraron artículos publicados entre enero de 2015 y enero de 2026, disponibles en inglés o español; se priorizó el acceso a texto completo para su evaluación cuando fue posible, no obstante, se excluyeron editoriales, cartas al editor, revisiones narrativas, tesis, capítulos de libros e informes técnicos no revisados por pares. Asimismo, se excluyeron estudios realizados exclusivamente en contextos clínicos sin relevancia forense o sin presencia de incentivos externos, así como aquellos que abordaron las pruebas de validez de manera tangencial.

La identificación de los estudios se realizó mediante búsquedas sistemáticas en bases de datos científicas internacionales especializadas en psicología, neuropsicología y ciencias de la salud. En particular, se consultaron las bases de datos PubMed, Scopus y Web of Science, seleccionadas por su amplia cobertura y relevancia para la literatura en evaluación neuropsicológica y contexto forense.

Adicionalmente, se llevó a cabo una búsqueda manual mediante revisión de las listas de referencias de los estudios incluidos, con el fin de identificar investigaciones potencialmente relevantes no recuperadas en la búsqueda electrónica. La última actualización de la búsqueda bibliográfica se realizó en enero de 2026. La estrategia de búsqueda se diseñó de manera estructurada, mediante la combinación de términos libres y, cuando estuvieron disponibles, descriptores controlados procedentes de los tesauros MeSH (Medical Subject Headings) y del Thesaurus of Psychological Index Terms de la American Psychological Association.

Los términos de búsqueda se estructuraron en tres categorías conceptuales: (1) evaluación neuropsicológica forense, (2) validez del rendimiento y de los síntomas y (3) distorsión de la respuesta en contextos médico-legales. Estos términos se combinaron utilizando los operadores booleanos AND y OR, ajustando las ecuaciones para adecuarlas a la sintaxis particular de cada base de datos. Asimismo, se aplicaron filtros de idioma (inglés y español) y periodo de publicación (enero de 2015 a enero de 2026), cuando dichas opciones estuvieron disponibles. Las estrategias utilizadas en cada base de datos fueron las siguientes:

PubMed: ("forensic neuropsychology"[tiab] OR forensic[tiab]) AND ("performance validity test"[tiab] OR PVT[tiab] OR "symptom validity test"[tiab] OR SVT[tiab] OR malingering[tiab])

Scopus: (TITLE-ABS-KEY("forensic neuropsychology" OR forensic) AND TITLE-ABS-KEY("performance validity test" OR PVT OR "symptom validity test" OR SVT OR malingering))

Web of Science: (TS=("forensic neuropsychology" OR forensic) AND TS=("performance validity test" OR PVT OR "symptom validity test" OR SVT OR malingering))

En este contexto, el proceso de selección de los estudios se llevó a cabo de acuerdo con las etapas establecidas por la declaración PRISMA 2020 (Page et al., 2021), incluyendo las fases de identificación, cribado y elegibilidad. En una primera fase, los registros recuperados a partir de las búsquedas en las bases de datos fueron exportados a un gestor bibliográfico, donde se realizó la identificación y eliminación de duplicados. Posteriormente, dos revisores evaluaron de manera independiente los títulos y resúmenes de los registros identificados, con el fin de determinar su pertinencia en función de los criterios de inclusión y exclusión previamente establecidos. Los estudios que cumplieron con estos criterios fueron seleccionados para la evaluación a texto completo.

En la fase de elegibilidad, los artículos potencialmente relevantes fueron analizados en texto completo para confirmar su adecuación metodológica y temática. Los estudios que no cumplieron con los criterios establecidos fueron excluidos en esta etapa. Las discrepancias entre los revisores durante cualquiera de las fases del proceso fueron resueltas mediante discusión y consenso. Finalmente, el proceso completo de identificación, cribado, elegibilidad e inclusión de los estudios fue documentado mediante un diagrama de flujo PRISMA, en el que se detalla el número de registros

identificados, los estudios excluidos en cada fase y el número final de investigaciones incluidas en la revisión.

Para la extracción de datos, se diseñó una matriz estructurada de los estudios incluidos. Las variables registradas comprendieron: autor(es), año de publicación, país, diseño metodológico, características de la muestra, contexto de evaluación (forense, médico-legal o clínico-forense), tipo de pruebas empleadas (pruebas de validez del rendimiento [PVT] y/o de los síntomas [SVT]), modalidad de aplicación (pruebas independientes o indicadores embebidos) y principales hallazgos relacionados con la detección de desempeño o sintomatología no creíble.

Finalmente, la mencionada extracción fue realizada de manera independiente por dos revisores. Posteriormente, los datos obtenidos fueron comparados y las discrepancias se resolvieron mediante discusión y consenso. Este procedimiento permitió garantizar la consistencia en la recopilación de la información y facilitó la síntesis comparativa de los estudios.

3. RESULTADOS

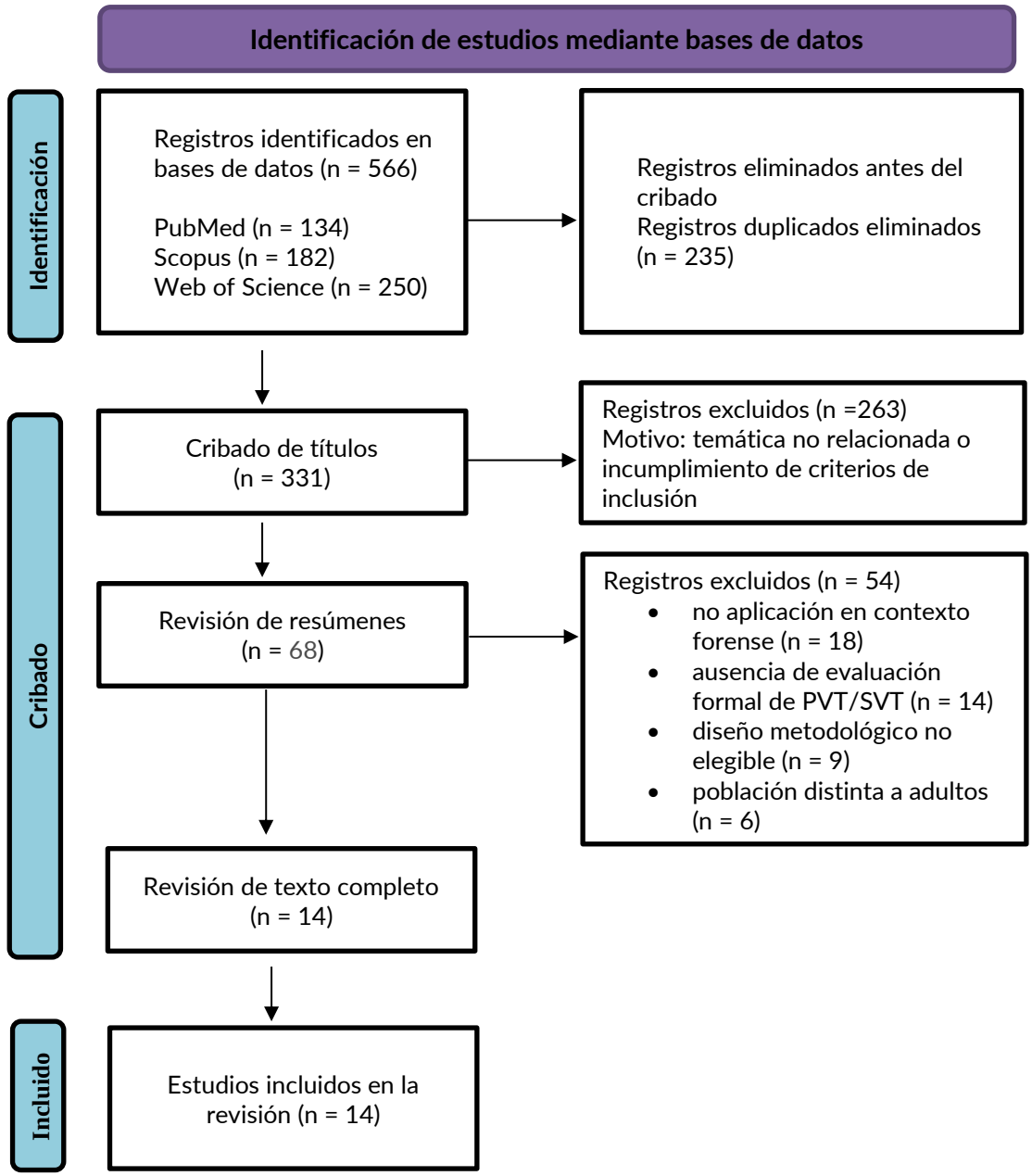
El proceso de búsqueda bibliográfica permitió identificar un total de 566 registros en las bases de datos consultadas (PubMed, Scopus y Web of Science). Tras la eliminación de 235 registros duplicados, se obtuvieron 331 estudios únicos para la fase inicial de cribado por título. Durante la revisión de títulos, 263 registros fueron excluidos por no estar directamente relacionados con la temática de estudio, ya que abordaban principalmente evaluaciones neuropsicológicas en contextos clínicos generales o poblaciones no forenses, sin analizar específicamente pruebas de validez del rendimiento o de los síntomas.

Como resultado de esta etapa, 68 artículos fueron seleccionados para la revisión de resúmenes. De estos, 54 estudios fueron excluidos debido a que no describían de manera explícita la aplicación de pruebas de validez del rendimiento o de los síntomas, no se desarrollaban en contextos forenses, médico-legales o clínico-forenses con incentivos externos, presentaban diseños metodológicos no elegibles o no correspondían a población adulta.

Los 14 estudios restantes fueron evaluados a texto completo y cumplieron con los criterios de inclusión previamente establecidos, por lo que fueron incorporados en la presente revisión sistemática. El proceso completo de identificación, cribado, elegibilidad e inclusión de los estudios se presenta en la Figura 1

Figura 1

Diagrama de flujo del proceso de selección de estudios según la guía PRISMA 2020.



Fuente: PRISMA (2020).

La calidad metodológica y el riesgo de sesgo de los estudios incluidos fueron evaluados mediante un enfoque estructurado basado en dominios relevantes para estudios de validación y observacionales, tomando como referencia los criterios propuestos en herramientas como QUADAS-2 para estudios diagnósticos. En este sentido, se tuvieron en cuenta varios aspectos para la evaluación de los estudios: (a) el diseño de la investigación, (b) el tamaño y la representatividad de la muestra, (c) el control de variables relevantes, (d) la validez y estandarización de los instrumentos empleados y (e) la adecuación de los análisis estadísticos realizados.

Cada estudio fue analizado de forma independiente por los revisores, quienes lo clasificaron según el riesgo de sesgo (bajo, moderado o alto) basado en el grado de cumplimiento de los criterios metodológicos. Este procedimiento facilitó una valoración más sistemática y transparente de la calidad de la evidencia.

En términos generales, los estudios incluidos mostraron una calidad metodológica que varió entre moderada y alta. La mayoría correspondió a investigaciones de validación, comparativas u observacionales que emplearon instrumentos estandarizados y procedimientos analíticos adecuados, lo que indica una base metodológica sólida en conjunto. Sin embargo, algunos trabajos presentaron limitaciones importantes, principalmente relacionadas con tamaños muestrales pequeños, posibles sesgos en la selección de participantes y la ausencia de estándares de referencia claros.

Asimismo, los estudios de caso y algunos diseños observacionales presentaron mayor riesgo de sesgo debido a la falta de control experimental y a la limitada generalización de sus resultados. La heterogeneidad en los diseños metodológicos y en los contextos de evaluación, también constituye un factor que limita la comparabilidad directa entre los estudios. La síntesis del riesgo de sesgo de los estudios incluidos se presenta en la Tabla 1.

Tabla 1

Valoración cualitativa de la solidez metodológica de los estudios incluidos.

Autor (Año)	Tipo de estudio	Riesgo de sesgo	Justificación
Bosso et al. (2022)	Estudio de caso	Alto	Diseño de caso único, sin grupo de comparación ni control de variables.
Sweet et al. (2021)	Documento de consenso	Moderado	Basado en consenso de expertos sin datos empíricos directos.
Laera et al. (2025)	Estudio de validación	Bajo	Diseño robusto, tamaño muestral adecuado y uso de criterios diagnósticos.
Nussbaum et al. (2023)	Estudio de validación	Bajo	Uso de estándares de referencia y metodología estructurada.
Liu et al. (2024)	Estudio comparativo	Bajo	Diseño comparativo con control adecuado y análisis estadístico sólido
Erdodi (2023)	Modelo estadístico	Bajo	Enfoque multivariado con adecuada validación metodológica.
Nauta et al. (2022)	Estudio observacional	Moderado	Posible sesgo de selección y ausencia de control experimental.
Fuermaier et al. (2024)	Estudio comparativo	Bajo	Diseño controlado con adecuada diferenciación de grupos.
Teßmann et al. (2025)	Estudio de validación	Bajo	Reporte de sensibilidad/especificidad y diseño adecuado.
Varela et al. (2022)	Desarrollo y validación	Bajo	Diseño psicométrico robusto con validación interna.
Messa et al. (2022)	Estudio observacional	Moderado	Tamaño muestral reducido y posible sesgo en población específica.

Schroeder et al. (2021)	Estudio de tasas base	Bajo	Tamaño muestral amplio y análisis estadístico sólido.
Rai et al. (2024)	Estudio de tasas base	Bajo	Gran tamaño muestral y control de variables demográficas.
Henry (2025)	Estudio de validación	Moderado	Limitaciones en tamaño muestral y generalización.

Fuente: autoría propia.

Los 14 estudios incluidos en la revisión presentan una diversidad de diseños metodológicos y contextos de aplicación dentro del ámbito de la neuropsicología forense, medicolegal y clínico-forense. La mayoría corresponde a estudios empíricos de tipo observacional, de validación de instrumentos y análisis de tasas base de invalidez, orientados a examinar la utilidad de las pruebas de validez del rendimiento (PVT) y de los síntomas (SVT). En términos geográficos, los estudios se desarrollaron principalmente en Norteamérica y Europa, incluyendo investigaciones realizadas en Estados Unidos, Canadá, Italia, Austria, Alemania, Países Bajos y Australia, además de un estudio realizado en China. Esta distribución refleja una concentración de la investigación en países con tradición consolidada en neuropsicología forense.

Sin embargo, el tamaño muestral mostró una alta variabilidad, desde estudios de caso único hasta investigaciones con muestras amplias superiores a 3000 participantes. En general, los estudios se centraron en población adulta evaluada en contextos forenses civiles, médico-legales y clínico-forenses con incentivos externos, incluyendo litigantes por lesiones personales, solicitantes de discapacidad o jubilación anticipada, y pacientes evaluados en contextos clínicos con potencial ganancia secundaria. Las principales características metodológicas de los estudios incluidos se presentan en la Tabla 2.

Tabla 2

Características de los estudios incluidos en la revisión sistemática.

Autor (Año)	País	Diseño	Muestra (n)	Población evaluada	Contexto de evaluación	Instrumentos evaluados	Objetivo del estudio
Bosso et al. (2022)	Italia	Estudio de caso	1	Adulto con sospecha de simulación cognitiva.	Forense	PVT + evaluación psiquiátrica y neuropsicológica	Analizar la relación entre simulación cognitiva y psicopatología.
Sweet et al. (2021)	EE.UU.	Documento de consenso	N/A	Evidencia sobre evaluación de validez en neuropsicología.	Clínico-forense	PVT y SVT	Actualizar lineamientos sobre evaluación de validez.
Laera et al. (2025)	Italia	Estudio de validación diagnóstica	114	Solicitantes de renovación de licencia.	Forense civil	IOP-29, IOP-M	Examinar la especificidad del IOP en contextos de alto riesgo.
Nussbaum et al. (2023)	Canadá	Estudio de validación	99	Sobrevivientes de accidente vehicular.	Médico-legal	TSI-2 ATR + SVT + PVT	Validar la escala ATR como SVT.

Liu et al. (2024)	China	Estudio comparativo	420	Pacientes con traumatismo craneoencefálico severo.	Forense	WAIS, SCL-90, prueba de memoria forzada, P3	Evaluar un indicador embebido para la detección de simulación.
Erdodi (2023)	Canadá	Modelo estadístico	452	Muestra clínico-forense mixta.	Clínico-forense	Indicadores de validez embebidos (EVI) + PVT.	Validar el Erdodi Index como modelo combinado.
Nauta et al. (2022)	Países Bajos	Estudio observacional	99	Pacientes con esclerosis múltiple.	Clínico	PVT	Analizar tasas de desempeño subóptimo.
Fuermaier et al. (2024)	Alemania	Estudio comparativo	180	Adultos con TDAH.	Clínico-forense	CAARS, SRSI, RDS	Comparar indicadores de validez de síntomas.
Teßmann et al. (2025)	Austria	Estudio de validación	132	Solicitantes de jubilación anticipada	Forense civil	GET, WMT, RDS	Evaluar la utilidad del GET como PVT atencional
Varela et al. (2022)	EE.UU.	Estudio de desarrollo y validación	183	Adultos en evaluación neuropsicológica.	Clínico-forense	NAB Embedded Validity Indicator	Desarrollar un indicador embebido en la NAB.
Messa et al. (2022)	Australia	Estudio observacional	23	Adultos con discapacidad intelectual.	Clínico-forense	Indicadores de validez embebidos múltiples.	Examinar tasas de falsos positivos en discapacidad intelectual.
Schroeder et al. (2021)	EE.UU.	Estudio de tasas base	242	Pacientes con incentivo por discapacidad (SSD).	Clínico	Múltiples PVT	Determinar tasas base de invalidez.
Rai et al. (2024)	EE.UU.	Estudio de tasas base	3297	Adultos en contexto medicolegal.	Forense	10 PVT	Analizar el efecto de la edad en tasas de fallo.
Henry (2025)	EE.UU.	Estudio de validación	100	Litigantes por lesión personal.	Forense civil	D-KEFS CWIT	Evaluar el CWIT como PVT embebido.

Fuente: autoría propia.

Como se puede apreciar, los estudios incluidos reportan el uso de múltiples pruebas de validez del rendimiento (PVT) y de los síntomas (SVT), tanto en formato independiente como integradas como indicadores embebidos dentro de baterías neuropsicológicas. En varios estudios, se describe la utilización combinada de diferentes indicadores de validez, incluyendo pruebas independientes, indicadores embebidos y modelos multivariados.

Entre los instrumentos identificados se encuentran pruebas ampliamente utilizadas como el Test of Memory Malingering (TOMM), el Word Memory Test (WMT), el Reliable Digit Span (RDS) y el Groningen Effort Test (GET), así como medidas embebidas como el D-KEFS Color-Word Interference Test (CWIT) y diversos indicadores derivados de instrumentos clínicos como el SCL-90. Asimismo, se reportan

instrumentos orientados a la validez de síntomas, como el TSI-2 Atypical Response Scale (ATR) y el Inventario de Problemas (IOP-29/IOP-M), junto con modelos combinados como el Erdodi Index.

Los estudios abarcan diversos contextos de evaluación, incluyendo litigios por lesiones personales, evaluaciones de discapacidad, jubilación anticipada, seguridad social, accidentes vehiculares y traumatismo craneoencefálico, entre otros escenarios médico-legales y clínico-forenses con incentivos externos. La síntesis de las principales pruebas de validez identificadas, junto con su clasificación, tipo de indicador, contexto de aplicación e indicadores psicométricos reportados, se presenta en la Tabla 3.

Tabla 3

Resumen de las pruebas de validez del rendimiento y de los síntomas identificadas.

Autor (Año)	Tipo de prueba	Instrumento	Tipo de indicador	Contexto de aplicación	Indicadores psicométricos	Utilidad en evaluación
Bosso et al. (2022)	PVT	Pruebas de elección forzada + batería neuropsicológica.	Independiente	Forense	No reporta métricas formales.	Diferenciación entre simulación cognitiva y psicopatología.
Sweet et al. (2021)	Marco conceptual	Evaluación multimétodo de validez (PVT/SVT).	No aplica	Clínico-forense	Recomendaciones basadas en evidencia.	Orientación para la interpretación de la validez.
Laera et al. (2025)	SVT	IOP-29 / IOP-M	Independiente	Forense civil	Alta especificidad reportada	Detección de sobreinformación
Nussbaum et al. (2023)	SVT	TSI-2 ATR	Embebido	Medicolegal (accidente vehicular).	Sens. .35–.53 / Esp. .92–.95	Identificación de invalidez sintomática.
Liu et al. (2024)	SVT (indicador embebido)	Indicador derivado del SCL-90	Embebido	Forense (TCE severo)	Asociación significativa con simulación	Apoyo en la detección de simulación
Erdodi (2023)	Modelo combinado de PVT.	Erdodi Index	Modelo multivariado	Clínico-forense	Incremento en precisión diagnóstica reportado.	Integración de múltiples indicadores de validez.
Nauta et al. (2022)	PVT	Múltiples PVT	Independiente	Clínico (EM, con implicaciones médico-legales)	Tasas de fallo reportadas	Identificación de desempeño subóptimo
Fuermaier et al. (2024)	SVT / PVT	CAARS, SRSI, RDS	Embebidos e independientes.	Clínico-forense (TDAH)	Diferenciación entre perfiles válidos e inválidos.	Complemento en evaluación de TDAH adulto.
Teßmann et al. (2025)	PVT	GET	Independiente	Forense civil	Sens. 71–77% / Esp. 82–89%	Detección de esfuerzo no creíble.
Varela et al. (2022)	PVT (indicador embebido)	NAB Embedded Validity Indicator	Embebido	Clínico-forense	Alta especificidad reportada	Indicador embebido en

						batería neuropsicológica.
Messa et al. (2022)	Indicadores embebidos múltiples.	Indicadores de validez embebidos.	Embebidos	Clínico-forense (discapacidad intelectual).	Análisis de tasas de falsos positivos.	Interpretación de validez en poblaciones con discapacidad.
Schroeder et al. (2021)	PVT	Múltiples PVT	Independiente	Clínico (incentivo por discapacidad).	46% tasa de invalidez con incentivo.	Influencia de incentivos externos en desempeño.
Rai et al. (2024)	PVT	10 PVT	Independientes y embebidos.	Medicolegal	Incremento de tasas de fallo con la edad (BR Fail).	Ajuste por variables demográficas.
Henry (2025)	PVT (indicador embebido)	D-KEFS CWIT	Embebido	Forense civil	Sens. .65 / Esp. .90	Identificación de funcionamiento no creíble.

Fuente: autoría propia.

4. DISCUSIÓN

El objetivo de esta revisión sistemática fue identificar y sintetizar la evidencia científica acerca de las pruebas de validez del rendimiento (Performance Validity Tests, PVT) y de validez de los síntomas (Symptom Validity Tests, SVT) empleadas en evaluaciones neuropsicológicas de población adulta en contextos forenses, médico-legales y clínico-forenses con incentivos externos. En conjunto, los resultados indican que estas herramientas son esenciales para detectar respuestas no creíbles, aunque su desempeño diagnóstico varía según el tipo de prueba, el contexto evaluativo y las características del grupo estudiado.

Comparativamente, los estudios muestran que los PVT independientes, como el Test of Memory Malingering (TOMM) y el Word Memory Test (WMT), presentan mayor solidez psicométrica y una especificidad más consistente que muchos indicadores embebidos. Por su parte, medidas como el Reliable Digit Span (RDS) y otros índices derivados de pruebas clínicas exhiben mayor variabilidad en su capacidad discriminativa, especialmente, en muestras clínicas complejas o con comorbilidad neurológica. En cuanto a los SVT, la literatura revisada muestra una heterogeneidad importante en sensibilidad, atribuible en parte a la naturaleza del autoinforme y a la influencia de variables contextuales y personales (Sweet et al., 2021; McWhirter et al., 2020).

Estos hallazgos concuerdan con los modelos teóricos sobre simulación en contextos forenses, según los cuales los incentivos externos (indemnizaciones, litigios civiles o la determinación de responsabilidad penal) pueden modular el comportamiento del evaluado y afectar la credibilidad de los resultados (Larrabee, 2012; Marcopulos et al., 2024). Los PVT parten de la premisa que incluso personas con deterioro neurológico genuino suelen rendir adecuadamente en tareas simples diseñadas para detectar desempeño inválido; por ello, puntuaciones anómalamente bajas pueden sugerir esfuerzo insuficiente o manipulación intencional. De manera complementaria, los SVT permiten identificar patrones atípicos, sobre reporte e

inconsistencias en la sintomatología informada, lo que favorece la detección de exageración o simulación (Rogers, 2018; Giromini et al., 2022).

Un hallazgo especialmente relevante es que el uso combinado de múltiples indicadores de validez (incluyendo PVT independientes, indicadores embebidos y SVT) mejora la precisión diagnóstica y reduce el riesgo de clasificaciones erróneas. Este enfoque multimétodo es coherente con recomendaciones internacionales actuales y resulta particularmente útil en escenarios forenses complejos, donde depender de un solo indicador puede conducir a interpretaciones insuficientes o inestables (Sherman et al., 2020; Sweet et al., 2021; Neal et al., 2022). Además, algunos estudios sugieren que esta estrategia también es prometedora en población hispanohablante, aunque la evidencia disponible todavía es limitada y requiere mayor validación transcultural (Aguilar et al., 2021).

No obstante, la revisión evidencia una notable variabilidad en el rendimiento de los instrumentos, atribuible a factores como el nivel educativo, las diferencias culturales y lingüísticas, la presencia de patología neurológica genuina y la sofisticación de las estrategias de simulación. Estos elementos pueden modificar tanto la sensibilidad como la especificidad de las pruebas y aumentar el riesgo de falsos positivos o falsos negativos cuando no se interpretan dentro de un análisis clínico-forense integral.

Desde una perspectiva práctica, los resultados respaldan la incorporación sistemática de múltiples indicadores de validez en evaluaciones neuropsicológicas con implicaciones médico-legales. En particular, la combinación de al menos dos PVT independientes junto con un SVT puede fortalecer la validez de las conclusiones periciales, aportar mayor solidez a la argumentación técnico-científica y reducir el riesgo de inferencias erróneas en la toma de decisiones judiciales.

Entre las limitaciones del estudio destacan la heterogeneidad metodológica de los artículos revisados, la diversidad de diseños y la variabilidad en los criterios utilizados para definir invalidez del rendimiento o de los síntomas. Asimismo, la mayoría de los estudios procede de contextos anglosajones, lo que restringe la generalización de los hallazgos a otras poblaciones, en especial, hacia contextos latinoamericanos en los que todavía son insuficientes las normas y validaciones locales de varios instrumentos. La ausencia de un metaanálisis cuantitativo también limita la estimación precisa del tamaño del efecto.

Para futuras investigaciones, se recomienda promover estudios que examinen la validez, utilidad clínica y aplicabilidad transcultural de estos instrumentos en contextos diversos, así como diseños longitudinales y multimétodo. Igualmente, el desarrollo de modelos integrativos que combinen varios indicadores de validez podría mejorar la precisión diagnóstica y optimizar los protocolos de evaluación neuropsicológica en contextos médico-legales.

5. CONCLUSIONES / CONSIDERACIONES FINALES

Los resultados de esta revisión sistemática indican que las pruebas de validez del rendimiento (Performance Validity Tests, PVT) y las pruebas de validez de los síntomas (Symptom Validity Tests, SVT) son herramientas clave para evaluar la

credibilidad del desempeño cognitivo y del reporte sintomático en evaluaciones neuropsicológicas realizadas en adultos dentro de contextos forenses, medicolegales y clínico-forenses con incentivos externos.

La evidencia recopilada muestra que diversos instrumentos, tanto independientes como embebidos, se emplean de forma recurrente en estos ámbitos, destacándose pruebas como el Test of Memory Malingering (TOMM), el Word Memory Test (WMT), el Reliable Digit Span (RDS), el Groningen Effort Test (GET) y el IOP-29, así como la aplicación de modelos combinados de validez. Además, varios estudios refieren el uso de enfoques multimétodo que integran múltiples indicadores de validez en una misma evaluación.

Estos hallazgos apuntan a que la incorporación de múltiples indicadores de validez favorece una mejor interpretación de los resultados neuropsicológicos y disminuye el riesgo de conclusiones incorrectas en contextos con implicaciones legales. Sin embargo, la heterogeneidad metodológica observada en los estudios, junto con las diferencias en contextos de aplicación y características de las muestras, limita la posibilidad de generalizar estos resultados.

Por lo tanto, se recomienda la incorporación sistemática de PVT y SVT como parte de protocolos de evaluación neuropsicológica en contextos con incentivos externos, así como el desarrollo de futuras investigaciones que fortalezcan la evidencia empírica, especialmente, en poblaciones y contextos subrepresentados, como los países latinoamericanos.

6. REFERENCIAS

Aguilar, C., Bailey, C., Karyadi, K. A., Kinney, D. I., & Nitch, S. R. (2021). Performance validity testing in hospitalized monolingual Spanish-speaking forensic patients. *Applied Neuropsychology: Adult*, *30(3)*, 314–320. https://doi.org/10.1080/23279095.2021.1970555

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. Guilford Press.

Bosso, T., Vischia, F., Keller, R., Vai, D., Imperiale, D., & Vercelli, A. (2022). A case report and literature review of cognitive malingering and psychopathology. *Frontiers in Psychiatry*, *13*, 981475. https://doi.org/10.3389/fpsyt.2022.981475

Erdodi, L. A. (2023). Multivariate models of performance validity: The Erdodi Index captures the dual nature of non-credible responding. *Assessment*, *30(5)*, 1467–1485. https://doi.org/10.1177/10731911221101910

Fuermaier, A. B. M., Tucha, L., Merten, T., Fathollah Gol, M., & Tucha, O. (2024). Symptom validity testing in adults with clinically diagnosed ADHD: Comparison of the Conners' Adult ADHD Rating Scale (CAARS) and the Self-Report Symptom Inventory (SRSI). *Journal of Clinical and Experimental Neuropsychology*, *46(8)*, 693–706. https://doi.org/10.1080/13803395.2024.2411365

Giromini, L., Young, G., & Sellbom, M. (2022). Assessing negative response bias using self-report measures: New articles, new issues. *Psychological Injury and Law*, *15*, 1–21. https://doi.org/10.1007/s12207-022-09444-2

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23(7)*, 1093–1129. https://doi.org/10.1080/13854040903155063

Laera, D., Pignolo, C., Barbara, G., Carucci, M., Giromini, L., Erdodi, L. A., Pasqualini, S., Lorenzoni, A., Zennaro, A., & Chiloiro, D. (2025). An Inventory of Problems (IOP) study of symptom and performance validity in a sample of driver's license renewal or reinstatement applicants. *Psychological Injury and Law*, *18*, 155–164. https://doi.org/10.1007/s12207-025-09530-1

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18(4)*, 625–631. https://doi.org/10.1017/S1355617712000240

Liu, C., Lu, Q., Rao, G., Chen, X., Liang, M., & Liu, Z. (2024). Malingering assessment after severe traumatic brain injury in forensic psychology with a potential embedded symptom validity indicator of Symptom Checklist 90. *Frontiers in Psychology*, *15*, 1320636. https://doi.org/10.3389/fpsyg.2024.1320636

Marcopulos, B. A., Kaufmann, P., & Patel, A. C. (2024). Forensic neuropsychological assessment. *Behavioral Sciences & the Law*, *42(4)*, 265–277. https://doi.org/10.1002/bsl.2656

McWhirter, L., Ritchie, C. W., Stone, J., & Carson, A. (2020). Performance validity test failure in clinical populations: A systematic review. *Journal of Neurology, Neurosurgery & Psychiatry*, *91(9)*, 945–952. https://doi.org/10.1136/jnnp-2020-323776

Messa, I., Holcomb, M., Lichtenstein, J. D., Tyson, B. T., Roth, R. M., & Erdodi, L. A. (2022). They are not destined to fail: A systematic examination of scores on embedded performance validity indicators in patients with intellectual disability. *Australian Journal of Forensic Sciences*, *54(5)*, 664–680. https://doi.org/10.1080/00450618.2020.1865457

Nauta, I. M., Bertens, D., van Dam, M., Huiskamp, S., Driessen, S., Geurts, J. J. G., Uitdehaag, B. M. J., Fasotti, L., Hulst, H. E., de Jong, B. A., & Klein, M. (2022). Performance validity in outpatients with multiple sclerosis and cognitive complaints. *Multiple Sclerosis Journal*, *28(4)*, 642–653. https://doi.org/10.1177/13524585211025780

Neal, T. M. S., Martire, K. A., Johan, J. L., Mathers, E. M., & Otto, R. K. (2022). The law meets psychological expertise: Eight best practices to improve forensic

psychological assessment. *Annual Review of Law and Social Science*, *18*, 169–192. https://doi.org/10.1146/annurev-lawsocsci-050420-010148

Nussbaum, S. H., Ales, F., Giromini, L., Watson, M., & Erdodi, L. (2023). Cross-validating the Atypical Response Scale of the TSI-2 in a sample of motor vehicle collision survivors. *Psychological Injury and Law*, *16*, 351–370. https://doi.org/10.1007/s12207-023-09487-z

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, *n71*. https://doi.org/10.1136/bmj.n71

Rogers, R. (2018). *Clinical assessment of malingering and deception (4th ed.)*. Guilford Press.

Sherman, E. M. S., Slick, D. J., Iverson, G. L., & Erdodi, L. A. (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. https://doi.org/10.1093/arclin/acaa019

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction. *The Clinical Neuropsychologist*, *13(4)*, 545–561. https://doi.org/10.1076/clin.13.4.545.1744

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., & Rohling, M. L. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *35(6)*, 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Teßmann, J., Fuermaier, A. B. M., Dandachi-FitzGerald, B., & Lehrner, J. (2025). The utility of an attention-based performance validity test in a sample of Austrian early retirement claimants. *Psychological Injury and Law*, *18*, 284–299. https://doi.org/10.1007/s12207-025-09544-9

Varela, J. L., Ord, A. S., Phillips, J. I., Shura, R. D., & Sautter, S. W. (2022). The development and validation of the embedded validity indicator for the Neuropsychological Assessment Battery. *Archives of Clinical Neuropsychology*, *37*, 133–145. https://doi.org/10.1093/arclin/acab025

Young, G. (2014). *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court*. Springer.