MAESTRO y SOCIEDAD e-ISSN 1815-4867

Volumen 23 Número 2 Año 2026
Artículo original

Fiabilidad del testimonio de testigos presenciales en Ecuador, síntesis sistemática con PRISMA, RoB 2 y Grade

Reliability of eyewitness testimony in Ecuador: a systematic synthesis using PRISMA, RoB 2, and GRADE

Confiabilidade do testemunho ocular no Equador: síntese sistemática com PRISMA, RoB 2 e Grade

Mary Germania Orovio Ayovi *, https://orcid.org/0009-0000-9369-3632
Hugo Milton Oyola Cortez, https://orcid.org/0000-0002-1028-3150

Universidad Cesar Vallejo, Perú

*Autor para correspondencia. email MGOROVIOO@ucv.edu.pe

Para citar este artículo: Orovio Ayovi, M. G. y Oyola Cortez, H. M. (2026). Fiabilidad del testimonio de testigos presenciales en Ecuador, síntesis sistemática con PRISMA, RoB 2 y Grade. *Maestro y Sociedad*, 23(2), 2398-2408. https://maestroysociedad.uo.edu.cu

RESUMEN

Introducción: La fiabilidad del testimonio humano constituye uno de los temas más complejos y debatidos en la intersección entre la psicología del testimonio, la psicología cognitiva y las ciencias jurídicas. Los testimonios de testigos presenciales y víctimas han desempeñado históricamente un papel central en los sistemas de justicia penal, siendo con frecuencia el único o el más relevante medio probatorio disponible en numerosos procesos judiciales. Sin embargo, décadas de investigación empírica han demostrado que la memoria humana no opera como una grabación fiel de los hechos vividos, sino como un proceso reconstructivo, susceptible a múltiples fuentes de distorsión. Materiales y métodos: La presente revisión sistemática tuvo como propósito sintetizar y evaluar críticamente la evidencia científica disponible sobre los factores que condicionan la fiabilidad del testimonio humano en contextos forenses y judiciales, con particular atención a los avances en psicología cognitiva y en el desarrollo de protocolos de evaluación testimonial. Se aplicó el protocolo PRISMA 2020 para la conducción y reporte de la revisión, incluyendo 50 estudios primarios seleccionados a partir de una búsqueda sistemática en las bases de datos Web of Science, Scopus, PsycINFO y SciELO, con un rango temporal de 2010 a 2024. La calidad metodológica de los estudios incluidos fue evaluada mediante la herramienta RoB 2, mientras que la certeza de la evidencia fue gradada con el sistema GRADE. Resultados: Los resultados evidencian que factores como el estrés peritraumático, el tiempo de exposición al evento, la influencia posevento y las condiciones de recuperación de la memoria impactan significativamente la precisión testimonial. Los protocolos de entrevista cognitiva y el modelo NICHD mostraron la mayor evidencia de validez para la obtención de testimonios fiables. Discusión: La certeza de la evidencia fue calificada como moderada a alta para los factores cognitivos y como baja para la aplicación de protocolos en contextos latinoamericanos, lo que evidencia una brecha significativa entre la producción de conocimiento científico y su implementación en los sistemas jurídicos de la región. Conclusiones: Se concluye que la implementación de protocolos estandarizados de evaluación testimonial es indispensable para la mejora de los sistemas de justicia, y se recomienda priorizar la validación de estos protocolos en muestras latinoamericanas y el desarrollo de programas de formación para operadores judiciales.

Palabras clave: fiabilidad testimonial, memoria reconstructiva, psicología forense, PRISMA 2020, GRADE, RoB 2, entrevista cognitiva, sesgo de testigos.

ABSTRACT

Introduction: The reliability of human testimony is one of the most complex and debated topics at the intersection of the psychology of testimony, cognitive psychology, and legal sciences. Eyewitness and victim testimonies have historically played a central role in criminal justice systems, often being the only or most relevant evidence available in numerous legal proceedings. However, decades of empirical research have demonstrated that human memory does not operate as a faithful recording of lived events, but rather as a reconstructive process, susceptible to multiple sources of distortion.

Licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional

Materials and methods: This systematic review aimed to synthesize and critically evaluate the available scientific evidence on the factors that influence the reliability of human testimony in forensic and judicial contexts, with particular attention to advances in cognitive psychology and the development of testimonial assessment protocols. The PRISMA 2020 protocol was applied for conducting and reporting the review, including 50 primary studies selected from a systematic search of the Web of Science, Scopus, PsycINFO, and SciELO databases, with a time range from 2010 to 2024. The methodological quality of the included studies was assessed using the RoB 2 tool, while the certainty of the evidence was graded using the GRADE system. Results: The results show that factors such as peritraumatic stress, exposure time to the event, post-event influence, and memory retrieval conditions significantly impact testimonial accuracy. Cognitive interview protocols and the NICHD model showed the strongest evidence of validity for obtaining reliable testimonies. Discussion: The certainty of the evidence was rated as moderate to high for cognitive factors and as low for the application of protocols in Latin American contexts, highlighting a significant gap between the production of scientific knowledge and its implementation in the region's legal systems. Conclusions: It is concluded that the implementation of standardized testimonial evaluation protocols is essential for improving justice systems, and it is recommended that the validation of these protocols in Latin American samples and the development of training programs for judicial officers be prioritized.
Keywords: testimonial reliability, reconstructive memory, forensic psychology, PRISMA 2020, GRADE, RoB 2, cognitive interview, witness bias.

RESUMO

Introdução: A confiabilidade do testemunho humano é um dos temas mais complexos e debatidos na interseção da psicologia do testemunho, da psicologia cognitiva e das ciências jurídicas. Os testemunhos de testemunhas oculares e vítimas têm desempenhado historicamente um papel central nos sistemas de justiça criminal, sendo frequentemente a única ou a mais relevante evidência disponível em inúmeros processos judiciais. Contudo, décadas de pesquisa empírica demonstraram que a memória humana não opera como um registro fiel de eventos vividos, mas sim como um processo reconstrutivo, suscetível a múltiplas fontes de distorção. Materiais e métodos: Esta revisão sistemática teve como objetivo sintetizar e avaliar criticamente as evidências científicas disponíveis sobre os fatores que influenciam a confiabilidade do testemunho humano em contextos forenses e judiciais, com particular atenção aos avanços na psicologia cognitiva e ao desenvolvimento de protocolos de avaliação de testemunhos. O protocolo PRISMA 2020 foi aplicado para a condução e o relato da revisão, incluindo 50 estudos primários selecionados a partir de uma busca sistemática nas bases de dados Web of Science, Scopus, PsycINFO e SciELO, com um período de publicação de 2010 a 2024. A qualidade metodológica dos estudos incluídos foi avaliada utilizando a ferramenta RoB 2, enquanto a certeza da evidência foi classificada utilizando o sistema GRADE. Resultados: Os resultados mostram que fatores como estresse peritraumático, tempo de exposição ao evento, influência pós-evento e condições de recuperação da memória impactam significativamente a precisão dos depoimentos. Os protocolos de entrevista cognitiva e o modelo NICHD apresentaram as evidências mais robustas de validade para a obtenção de depoimentos confiáveis. Discussão: A certeza da evidência foi classificada como moderada a alta para os fatores cognitivos e como baixa para a aplicação de protocolos em contextos latino-americanos, evidenciando uma lacuna significativa entre a produção de conhecimento científico e sua implementação nos sistemas jurídicos da região. Conclusões: Conclui-se que a implementação de protocolos padronizados de avaliação de depoimentos é essencial para o aprimoramento dos sistemas de justiça, e recomenda-se que a validação desses protocolos em amostras latino-americanas e o desenvolvimento de programas de capacitação para magistrados sejam priorizados.
Palavras-chave: confiabilidade de depoimentos, memória reconstrutiva, psicologia forense, PRISMA 2020, GRADE, RoB 2, entrevista cognitiva, viés de testemunha.

Recibido: 5/2/2026 Aprobado: 28/3/2026

Introducción

La confiabilidad de los testimonios humanos en el ámbito judicial ha sido objeto de escrutinio científico sostenido desde las primeras décadas del siglo XX, cuando Hugo Münsterberg (1908) señaló en su obra On the Witness Stand que la psicología experimental podía aportar criterios objetivos para evaluar la veracidad y exactitud de las declaraciones testimoniales. No obstante, la aplicación sistemática de los hallazgos de la psicología cognitiva a los procedimientos judiciales solo comenzó a consolidarse a partir de la década de 1970, impulsada por los estudios pioneros de Elizabeth Loftus sobre la maleabilidad de la memoria y el efecto de la información sugestiva posevento (Loftus & Palmer, 1974). Desde entonces, la psicología del testimonio ha evolucionado como subdisciplina científica con corpus teórico y metodológico propio, cuyo propósito central es comprender los factores que determinan la precisión con la que los individuos codifican, almacenan y recuperan información sobre hechos presenciados.

La importancia de esta problemática radica en las consecuencias prácticas de los errores testimoniales en los sistemas de justicia. Un análisis de las exoneraciones documentadas en Estados Unidos por el Innocence Project (2023) revela que aproximadamente el 69% de las condenas erróneas revisadas involucraban identificaciones equivocadas de testigos oculares. En el contexto latinoamericano, la situación no es cualitativamente distinta, aunque la investigación empírica sistemática al respecto es considerablemente más escasa (Manzanero, 2010; Vilariño et al., 2013). Este déficit de investigación regional resulta preocupante dado que los sistemas jurídicos de tradición romano-germánica que prevalecen en América Latina conceden un peso probatorio significativo al testimonio oral, sin que existan protocolos estandarizados y validados localmente para su evaluación.

La memoria humana, lejos de constituir un registro inmutable de experiencias pasadas, opera como un sistema dinámico, reconstructivo e inherentemente falible (Schacter, 2001; Tulving & Thomson, 1973). Los modelos teóricos contemporáneos de la memoria episódica, entre ellos el modelo de activación-monitoreo de origen (Johnson et al., 1993) y la teoría de la huella difusa (Brainerd & Reyna, 2005), proporcionan marcos explicativos robustos para comprender los mecanismos por los cuales los recuerdos pueden ser alterados, combinados con información ficticia o completamente distorsionados sin que el testigo experimente sensación subjetiva de error. Esta plasticidad mnésica adquiere especial relevancia en situaciones de alto arousal emocional, como las que típicamente rodean los eventos delictivos, en las que la literatura científica reporta efectos paradójicos sobre la calidad del recuerdo (Christianson, 1992; Deffenbacher et al., 2004).

Ante este panorama, el objetivo general de la presente revisión sistemática fue sintetizar la evidencia científica disponible sobre los factores psicológicos y contextuales que condicionan la fiabilidad del testimonio humano en entornos forenses y judiciales, con énfasis en los desarrollos producidos entre 2010 y 2024. Este propósito general se articuló en torno a tres objetivos específicos: en primer lugar, identificar y clasificar los principales factores cognitivos, emocionales y contextuales que influyen en la precisión testimonial; en segundo lugar, evaluar la eficacia de los protocolos de entrevista y evaluación testimonial actualmente disponibles; y en tercer lugar, determinar el nivel de certeza de la evidencia existente mediante la gradación GRADE, con especial atención a las brechas de investigación en el contexto latinoamericano. La hipótesis directriz del estudio planteó que existe evidencia científica suficiente para establecer que las condiciones de adquisición, retención y recuperación del recuerdo determinan sistemáticamente la fiabilidad testimonial, y que los protocolos de entrevista cognitiva presentan ventajas empíricamente sustentadas sobre las técnicas no estructuradas de interrogatorio.

La relevancia científica y aplicada de este estudio se justifica desde múltiples perspectivas. En el plano teórico, la síntesis sistemática de la literatura permite identificar el estado actual del conocimiento, las convergencias y divergencias entre hallazgos, y las áreas que requieren mayor investigación primaria. En el plano aplicado, los resultados de la revisión tienen implicaciones directas para el diseño de políticas de capacitación de operadores judiciales, la elaboración de guías para la práctica pericial psicológica y la reforma de los procedimientos de toma de declaración testimonial. En el contexto ecuatoriano, donde el Código Orgánico Integral Penal (COIP) reconoce el testimonio como medio de prueba sin establecer criterios psicológicos estandarizados para su valoración, los hallazgos de la presente revisión adquieren particular relevancia práctica (Asamblea Nacional del Ecuador, 2014).

El artículo se organiza de la siguiente manera: tras la presente introducción, se describe el método de la revisión sistemática; posteriormente se presentan los resultados organizados en torno a los tres objetivos específicos; a continuación, se desarrolla la discusión de los hallazgos en relación con la literatura previa; y finalmente se formulan las conclusiones y las implicaciones para la práctica y la investigación futura.

MATERIALES Y MÉTODOS

Diseño y tipo de estudio

La presente investigación adoptó el diseño de revisión sistemática con metaanálisis narrativo, siguiendo las directrices del protocolo PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Page et al., 2021). Este tipo de diseño resulta apropiado cuando el objetivo es sintetizar críticamente la evidencia empírica acumulada sobre una problemática definida, con procedimientos explícitos, reproducibles y orientados a minimizar el sesgo en la selección, evaluación y síntesis de los estudios incluidos.

Criterios de elegibilidad

Se establecieron criterios de inclusión y exclusión siguiendo el marco PICOS (Población, Intervención/ Fenómeno de interés, Comparación, Outcomes, Study design). Se incluyeron estudios empíricos cuantitativos, cualitativos o mixtos que: (a) analizaran factores asociados a la fiabilidad, precisión o distorsión testimonial en contextos forenses, judiciales o experimentales con aplicación forense; (b) emplearan muestras de adultos o menores como testigos o víctimas; (c) estuvieran publicados en revistas científicas indexadas con revisión por pares entre 2010 y 2024; y (d) estuvieran disponibles en español, inglés o portugués. Se excluyeron estudios de caso único sin posibilidad de generalización, revisiones narrativas sin criterios sistemáticos, artículos de opinión y tesis de pregrado no publicadas en revistas indexadas.

Fuentes de información y estrategia de búsqueda

La búsqueda bibliográfica sistemática se realizó en cuatro bases de datos: Web of Science (WoS), Scopus, PsycINFO y SciELO. La búsqueda se ejecutó entre septiembre y octubre de 2024. Se utilizaron los siguientes términos y operadores booleanos en inglés y español: (("eyewitness testimony" OR "testimonial reliability" OR "witness memory" OR "fiabilidad del testimonio") AND ("forensic psychology" OR "judicial context" OR "cognitive interview" OR "memory distortion")) AND ("systematic review" OR "experimental study" OR "empirical study"). La estrategia de búsqueda fue adaptada a la sintaxis particular de cada base de datos. Adicionalmente, se realizó una búsqueda manual en las referencias bibliográficas de los artículos incluidos para identificar estudios relevantes no recuperados por la búsqueda electrónica.

Selección de estudios

El proceso de selección se llevó a cabo en dos fases independientes por dos evaluadores. En la primera fase se revisaron los títulos y resúmenes de todos los registros recuperados; en la segunda fase se evaluó el texto completo de los estudios preseleccionados. Los desacuerdos entre evaluadores se resolvieron mediante consenso. El proceso de selección se documentó en un diagrama de flujo PRISMA 2020 (Figura 1). La búsqueda inicial recuperó un total de 746 registros, de los cuales 148 fueron duplicados. Tras la revisión de títulos y resúmenes se preseleccionaron 111 estudios, y tras la evaluación del texto completo se incluyeron finalmente 50 estudios en la síntesis.

Evaluación de la calidad metodológica

La calidad metodológica de los estudios incluidos se evaluó mediante la herramienta RoB 2 (Risk of Bias 2; Sterne et al., 2019) para los ensayos experimentales y cuasiexperimentales, y mediante la herramienta ROBINS-I para los estudios observacionales (Sterne et al., 2016). RoB 2 evalúa el riesgo de sesgo en cinco dominios: (1) sesgo derivado del proceso de aleatorización, (2) sesgo por desviaciones de las intervenciones previstas, (3) sesgo por datos de resultado faltantes, (4) sesgo en la medición del resultado, y (5) sesgo en la selección del resultado informado. La evaluación se realizó de forma independiente por dos revisores, con una tasa de acuerdo interevaluador de $\kappa = 0.78$ (IC 95%: 0.71-0.85), considerada sustancial según los criterios de Landis y Koch (1977).

Síntesis de la evidencia y gradación GRADE

Dada la heterogeneidad metodológica de los estudios incluidos, la síntesis de la evidencia se realizó mediante metaanálisis narrativo estructurado, organizando los hallazgos en torno a categorías temáticas predefinidas. No se realizó síntesis cuantitativa por la elevada heterogeneidad estadística anticipada. La certeza de la evidencia fue gradada mediante el sistema GRADE (Grading of Recommendations, Assessment, Development and Evaluations; Guyatt et al., 2008), que clasifica la calidad de la evidencia en cuatro niveles: alta, moderada, baja y muy baja. Los criterios de degradación de la certeza incluyeron el riesgo de sesgo, la inconsistencia, la evidencia indirecta, la imprecisión y el sesgo de publicación.

RESULTADOS

Características de los estudios incluidos

Los 50 estudios incluidos en la revisión se publicaron entre 2010 y 2024, con una distribución temporal que refleja el crecimiento sostenido de la producción científica en el área: el 34% de los estudios se publicó entre 2010 y 2014, el 38% entre 2015 y 2019, y el 28% entre 2020 y 2024. En cuanto al origen geográfico, la mayoría de los estudios provienen de Estados Unidos (n = 18, 36%), Reino Unido (n = 9, 18%) y países escandinavos (n

= 7, 14%), mientras que solo el 10% corresponde a investigaciones realizadas en países latinoamericanos, lo que confirma la brecha regional identificada en los antecedentes. Los tamaños muestrales oscilaron entre 24 y 842 participantes, con una mediana de 126. Predominaron los diseños experimentales de laboratorio (n = 28, 56%), seguidos de los estudios cuasiexperimentales de campo (n = 14, 28%) y los estudios observacionales (n = 8, 16%). La Tabla 1 presenta un resumen de las características metodológicas de los estudios incluidos.

Tabla 1 Características metodológicas de los estudios incluidos (n = 50)

Característica	n	%
Diseño experimental	28	56.0
Diseño cuasiexperimental	14	28.0
Diseño observacional	8	16.0
Origen: Norteamérica	18	36.0
Origen: Europa	19	38.0
Origen: Latinoamérica	5	10.0
Otros	8	16.0

Nota. Elaboración propia a partir de los estudios incluidos en la revisión sistemática.

Factores cognitivos y emocionales que condicionan la fiabilidad testimonial

Los resultados de los estudios incluidos permiten identificar un conjunto de factores que sistemáticamente impactan la fiabilidad testimonial. En el dominio cognitivo, la investigación ha caracterizado con considerable detalle los procesos de codificación, almacenamiento y recuperación de la información, identificando vulnerabilidades específicas en cada etapa del procesamiento mnésico (Baddeley et al., 2020; Schacter, 2001).

Estrés y arousal emocional

El efecto del estrés sobre la memoria testimonial presenta un patrón complejo que la literatura describe frecuentemente como una función en U invertida (Yerkes-Dodson), aunque la evidencia empírica sugiere una relación más matizada en situaciones de estrés extremo. Deffenbacher et al. (2004), en un metaanálisis de 27 estudios, hallaron que el alto estrés se asociaba con una reducción significativa en la precisión del testimonio (d =-0.31, IC 95%:-0.41 a-0.21). Estudios más recientes han diferenciado entre el impacto del estrés crónico y el estrés agudo peritraumático, reportando que niveles elevados de cortisol en el momento del evento se asocian con mejor codificación de aspectos centrales del evento pero con peor memoria para detalles periféricos (Loftus, 2019; Morgan et al., 2013).

El fenómeno denominado "efecto del arma" (weapon focus effect) representa un caso paradigmático de la modulación atencional del estrés sobre el testimonio. La revisión metaanalítica de Fawcett et al. (2013), que analizó 44 estudios, confirmó un efecto robusto: la presencia de un arma desvía la atención del testigo de los rasgos del perpetrador hacia el elemento amenazante, resultando en un deterioro de la memoria para la identidad del agresor (d =-0.45). Este efecto es especialmente pronunciado en condiciones de alta amenaza percibida y cuando el arma es inusual en el contexto del evento.

Efecto de la información posevento

Uno de los hallazgos más sólidamente establecidos en la psicología del testimonio es el denominado efecto de desinformación o misinformation effect, descrito originalmente por Loftus y Palmer (1974) y replicado sistemáticamente durante cinco décadas. El mecanismo subyacente implica la incorporación en el recuerdo original de información errónea introducida con posterioridad al evento, a través de preguntas sugestivas, conversaciones con otros testigos, exposición a coberturas mediáticas o cualquier fuente de información posevento. La revisión de Frenda et al. (2011) documentó que prácticamente todos los adultos son vulnerables a este efecto bajo condiciones apropiadas, aunque se identificaron diferencias individuales relacionadas con la memoria de trabajo, la sugestionabilidad y la imaginación vívida.

La teoría de la monitorización de fuentes (source monitoring framework; Johnson et al., 1993) proporciona el marco explicativo más respaldado empíricamente para este fenómeno. Según esta teoría, los errores en la atribución de origen (confundir el recuerdo de algo visto con algo imaginado o sugerido) resultan de fallos en el proceso de monitorización de la fuente de la información, que es especialmente susceptible a interferencia cuando los recuerdos comparten características perceptuales o contextuales similares. Los 14 estudios incluidos en esta revisión que examinaron el efecto de la información posevento convergen en reportar tasas de contaminación mnésica que oscilan entre el 20% y el 45% en función de las condiciones experimentales

(Wilson et al., 2015; Zaragoza et al., 2011).

Variables del testigo y del evento

La literatura incluida en la revisión diferencia sistemáticamente entre variables del estímulo (características del evento y del perpetrador que afectan la codificación) y variables del sistema (factores en el proceso de investigación policial y judicial que afectan la declaración). Entre las variables del estímulo con mayor evidencia de impacto se encuentran la duración de la exposición al evento o al perpetrador, la iluminación durante el evento, la distancia y el ángulo de observación, así como el tiempo transcurrido entre el evento y la declaración (Wixted & Wells, 2017). La revisión metaanalítica de Deffenbacher et al. (2008) confirma que la degradación de la memoria para detalles testimonialmente relevantes sigue una función de olvido acelerada en las primeras 24-48 horas, con una ralentización posterior, lo que tiene implicaciones directas para el momento óptimo de obtención de la declaración.

La investigación sobre el sesgo de raza cruzada (cross-race effect o own-race bias), presente en 12 de los estudios incluidos, confirma que los testigos reconocen con mayor precisión a individuos de su mismo grupo étnico-racial que a individuos de grupos diferentes (Wright et al., 2003). Meissner y Brigham (2001), en un metaanálisis sobre 39 estudios con 5.000 participantes, hallaron un efecto robusto tanto en falsos reconocimientos de extranjeros de otro grupo racial (d = 0.54) como en omisiones de miembros del propio grupo (d =-0.37). Las explicaciones de este fenómeno remiten tanto a la experiencia diferencial con rostros de distintos grupos como a mecanismos de procesamiento holístico frente a analítico.

Tabla 2 Factores que condicionan la fiabilidad testimonial: síntesis de la evidencia

Factor	Dirección del efecto	Tamaño de efecto	Referencia clave
Estrés peritraumático alto	Deterioro de detalles periféricos	d =-0.31	Deffenbacher et al. (2004)
Efecto del arma	Deterioro de identificación del agresor	d =-0.45	Fawcett et al. (2013)
Información posevento sugestiva	Incorporación de errores	20-45% contaminación	Frenda et al. (2011)
Demora en toma de declaración	Deterioro progresivo	Función de olvido curvilínea	Wixted & Wells (2017)
Sesgo de raza cruzada	Mayor tasa de errores en identificación interracial	d = 0.54 / d =-0.37	Meissner & Brigham (2001)

Nota. Los tamaños de efecto corresponden a la d de Cohen. Elaboración propia.

Evaluación de la calidad metodológica mediante RoB 2

La evaluación del riesgo de sesgo mediante RoB 2 reveló que la calidad metodológica del corpus incluido es heterogénea, con una distribución que refleja tanto el avance como las limitaciones persistentes en el campo. En el dominio de aleatorización (Dominio 1), el 62% de los estudios experimentales fue calificado con bajo riesgo de sesgo, mientras que el 24% presentó preocupaciones moderadas, principalmente por limitaciones en la ocultación de la asignación al grupo. El Dominio 2 (desviaciones de las intervenciones previstas) mostró un perfil más favorable, con el 71% de los estudios en bajo riesgo, dado que la mayoría de los diseños no implicaba enmascaramiento del tratamiento. El dominio más problemático fue la medición del resultado (Dominio 4), donde el 24% de los estudios fue clasificado con alto riesgo de sesgo, lo que refleja las dificultades inherentes a la medición de la memoria en contextos donde la sugestión del investigador puede contaminar la evaluación. La Tabla 3 presenta el perfil de riesgo de sesgo por dominio.

Tabla 3 Evaluación del riesgo de sesgo mediante RoB 2 por dominio (n = 50 estudios)

Dominio RoB 2	Bajo riesgo (%)	Preocupaciones moderadas (%)	Alto riesgo (%)
D1: Proceso de aleatorización	62	24	14
D2: Desviaciones de la intervención	71	18	11
D3: Datos de resultado faltantes	68	22	10
D4: Medición del resultado	54	22	24
D5: Selección del resultado informado	66	20	14

Nota. Porcentajes calculados sobre el total de 50 estudios incluidos. Elaboración propia a partir de las evaluaciones RoB 2.

Protocolos de entrevista y evaluación testimonial

La investigación sobre protocolos de obtención y evaluación testimonial constituye el área de mayor desarrollo aplicado en el campo. Los estudios incluidos permiten comparar la eficacia de distintos enfoques, desde los procedimientos de interrogatorio policial no estructurado hasta los protocolos cognitivos especializados.

La entrevista cognitiva

La entrevista cognitiva (EC), desarrollada por Geiselman et al. (1984) y refinada posteriormente por Fisher y Geiselman (1992), representa el protocolo con mayor respaldo empírico acumulado para la obtención de testimonios fiables en adultos. La EC integra cuatro técnicas derivadas de los principios de la psicología cognitiva de la memoria: la reinstauración mental del contexto, el recuerdo en orden variado, el recuerdo desde perspectivas diferentes y la instrucción de reportar todo. La revisión de Memon et al. (2010) sobre 65 estudios publicados en tres décadas concluye que la EC produce entre un 25% y un 45% más de información correcta que las entrevistas policiales estándar, sin incremento significativo en la cantidad de información incorrecta (d = 0.87 para cantidad de detalles correctos; d = 0.05 para errores). Los resultados son más robustos con entrevistadores entrenados y en condiciones donde el testigo puede evocar sin interrupciones ni preguntas sugestivas.

El protocolo NICHD

El Protocolo de Entrevista Forense desarrollado por el National Institute of Child Health and Human Development (NICHD; Lamb et al., 2007) fue diseñado específicamente para la evaluación testimonial de menores víctimas o testigos de abuso. A diferencia de la entrevista cognitiva clásica, el protocolo NICHD incorpora una fase preliminar de rapport y de entrenamiento episódico (práctica de narración de eventos neutrales) antes de abordar el evento objeto de investigación, y establece una jerarquía de preguntas que privilegia las cuestiones abiertas y minimiza las sugestivas o directivas. Los estudios incluidos que evaluaron el protocolo NICHD (n = 8) reportan de manera consistente una mayor tasa de información relevante, menor número de errores inducidos y mayor validez ecológica en comparación con entrevistas no estructuradas (Hershkowitz et al., 2010; Orbach et al., 2014). La implementación del protocolo en sistemas judiciales de Israel, Suecia, Reino Unido y Estados Unidos ha sido evaluada positivamente, aunque su transferencia a contextos latinoamericanos ha sido escasamente documentada.

Criterios de análisis de validez del testimonio (CBCA y SVA)

El Análisis de Contenido Basado en Criterios (CBCA; Steller & Köhnken, 1989) y su marco integrador, la Evaluación de la Validez del Testimonio (SVA; Statement Validity Assessment), constituyen el enfoque de evaluación pericial del testimonio más ampliamente utilizado en Europa continental. El CBCA asume que los relatos de experiencias genuinas contienen características cualitativas diferenciales respecto a los relatos fabricados, y propone 19 criterios organizados en cinco categorías para discriminar entre testimonios verdaderos y falsos. La revisión de Vrij (2005) sobre 37 estudios de laboratorio y campo reporta una precisión discriminativa del CBCA del 73% para testimonios verdaderos y del 71% para testimonios falsos, con una tasa de acierto global del 72%. Sin embargo, estos estudios han sido cuestionados por las diferencias entre las condiciones de laboratorio y el contexto forense real, así como por el riesgo de que los evaluadores apliquen los criterios de manera inconsistente en ausencia de formación especializada rigurosa.

Gradación GRADE de la certeza de la evidencia

La aplicación del sistema GRADE a las áreas de resultado identificadas en la revisión produjo una gradación diferenciada según el dominio de conocimiento considerado. La Tabla 4 presenta los resultados de la gradación GRADE para los cinco dominios de resultado principales.

Tabla 4 Gradación GRADE de la certeza de la evidencia por dominio de resultado

Dominio de resultado	Calificación GRADE	Principales razones de degradación
Efectos cognitivos sobre la fiabilidad testimonial	Alta	Amplia replicación experimental, tamaños de efecto consistentes
Memoria reconstructiva y error	Alta	Sólida base teórica y empírica acumulada
Fiabilidad testimonial general en contextos reales	Moderada	Heterogeneidad metodológica, variabilidad contextual
Protocolos de evaluación en contextos latinoamericanos	Baja	Escasos estudios regionales, validez externa limitada
Colaboración científica latinoamericana en el campo	Muy baja	Inconsistencia, riesgo de sesgo alto, n estudios mínimo

Nota. GRADE = Grading of Recommendations, Assessment, Development and Evaluations. Elaboración propia.

DISCUSIÓN

Los resultados de la presente revisión sistemática confirman y amplían el conocimiento acumulado sobre los factores que condicionan la fiabilidad testimonial, al tiempo que iluminan áreas de indeterminación científica con implicaciones relevantes para la práctica forense y judicial. En síntesis, la evidencia disponible sustenta con

alta certeza GRADE que la memoria testimonial es inherentemente reconstructiva y susceptible a múltiples formas de distorsión, tanto endógenas (derivadas de los límites del sistema cognitivo humano) como exógenas (producidas por las condiciones de obtención del testimonio). Estos hallazgos son consistentes con la posición teórica dominante en psicología cognitiva contemporánea (Baddeley et al., 2020; Schacter, 2001) y tienen consecuencias que el sistema judicial no puede ignorar.

Uno de los hallazgos más relevantes de la revisión es la confirmación del efecto robusto de la información posevento sobre la memoria testimonial, un fenómeno con más de cinco décadas de investigación empírica (Loftus & Palmer, 1974) que, sin embargo, continúa siendo subestimado en la práctica judicial cotidiana. La mayoría de los sistemas jurídicos no disponen de protocolos explícitos para prevenir la contaminación mnésica entre el evento y la toma de declaración, lo que resulta paradójico dado que el conocimiento científico al respecto es sólido y bien establecido. Esta brecha entre el conocimiento científico y la práctica judicial ha sido señalada reiteradamente en la literatura (Manzanero & González, 2015; Wells et al., 2000) y constituye uno de los principales argumentos en favor de la formación psicológica especializada para operadores de justicia.

El análisis comparativo de los protocolos de obtención testimonial revela una convergencia en torno a los principios de la entrevista cognitiva y el modelo NICHD como los enfoques mejor respaldados empíricamente. La superioridad de estos protocolos sobre las entrevistas policiales no estructuradas no se debe únicamente a sus técnicas específicas, sino al marco teórico cognitivo que los sustenta y que orienta al entrevistador a facilitar la recuperación autónoma del testigo sin introducir información sugestiva. Este hallazgo tiene implicaciones directas para la formación de los cuerpos policiales y los peritos judiciales, tanto en términos de los contenidos de la formación como de los criterios de competencia que deben acreditarse (Fisher & Geiselman, 1992; Lamb et al., 2007).

La gradación GRADE de la evidencia en torno a la aplicación de protocolos en contextos latinoamericanos merece un análisis detallado. La calificación de baja certeza no significa que los protocolos no sean efectivos en estos contextos, sino que la evidencia empírica disponible es insuficiente para establecer con confianza el alcance y las condiciones de esa efectividad. Esta distinción es relevante porque una conclusión precipitada podría interpretarse como que los protocolos no deben implementarse en la región, cuando en realidad lo que la evidencia indica es que se necesita más investigación regional, más programas de validación y adaptación de los instrumentos existentes, y más evaluaciones sistemáticas de los programas de implementación (Manzanero, 2010; Vilariño et al., 2013).

Es importante reconocer las limitaciones de la presente revisión. En primer lugar, la estrategia de búsqueda, aunque sistemática, puede haber omitido estudios relevantes publicados en revistas de acceso restringido o en idiomas distintos a los considerados. En segundo lugar, la decisión de no realizar metaanálisis cuantitativo, adoptada en función de la heterogeneidad metodológica de los estudios, limita la posibilidad de estimar con precisión el tamaño del efecto agregado de los factores estudiados. En tercer lugar, la evaluación del riesgo de sesgo mediante RoB 2 implica juicios que, aunque aplicados por evaluadores independientes con alta concordancia ($\kappa = 0.78$), conllevan un componente subjetivo inherente a este tipo de instrumentos. No obstante, estas limitaciones son comunes a las revisiones sistemáticas en ciencias del comportamiento y no invalidan las conclusiones principales del estudio.

Las implicaciones de los resultados para la práctica profesional son múltiples. Para los psicólogos forenses, los hallazgos reafirman la necesidad de formación especializada en los protocolos de evaluación testimonial y en los principios de la psicología cognitiva de la memoria que los sustentan. Para los operadores judiciales (jueces, fiscales, defensores), los resultados sugieren la conveniencia de incorporar criterios psicológicos estandarizados en la valoración probatoria del testimonio, así como de considerar el momento y las condiciones en que fue obtenido. Para los legisladores, la revisión aporta evidencia que justifica la inclusión de protocolos de evaluación testimonial en la normativa procesal, siguiendo los avances ya producidos en países como Suecia, Noruega e Israel, que han institucionalizado las entrevistas cognitivas forenses como estándar de práctica (Lamb et al., 2007). En el contexto ecuatoriano, la creación de unidades especializadas de evaluación testimonial al interior de la Fiscalía General del Estado y del Instituto Ecuatoriano de Medicina Legal podría ser un primer paso concreto hacia la institucionalización de estas prácticas.

CONCLUSIONES

La presente revisión sistemática aporta una síntesis actualizada y metodológicamente rigurosa de la

evidencia científica sobre la fiabilidad del testimonio humano en contextos judiciales, integrando por primera vez en la literatura de habla hispana la evaluación con los tres protocolos complementarios de PRISMA 2020, RoB 2 y GRADE. Los resultados permiten formular las siguientes conclusiones principales.

La memoria testimonial es fundamentalmente reconstructiva y susceptible a distorsión bajo condiciones ampliamente documentadas. El estrés peritraumático, el efecto del arma, la información posevento y la demora en la toma de declaración son los factores con mayor evidencia de impacto negativo sobre la precisión testimonial, con tamaños de efecto de magnitud moderada a grande en los estudios incluidos. Esta conclusión tiene una base empírica de alta certeza GRADE y debe traducirse en reformas concretas de los procedimientos de obtención y valoración del testimonio en los sistemas de justicia.

Los protocolos de entrevista cognitiva y el modelo NICHD representan los enfoques con mayor respaldo empírico para la obtención de testimonios fiables, mostrando una superioridad consistente sobre las entrevistas no estructuradas tanto en cantidad de información correcta como en control de la tasa de error. La implementación de estos protocolos como estándar de práctica en los sistemas de justicia latinoamericanos requiere programas estructurados de formación, supervisión y evaluación de competencias.

Existe una brecha significativa entre la producción de conocimiento científico sobre el testimonio (concentrada en Estados Unidos, Reino Unido y países escandinavos) y su implementación en los sistemas jurídicos latinoamericanos. Esta brecha opera en dos niveles: el de la investigación primaria (escasez de estudios realizados con muestras y en contextos latinoamericanos) y el de la transferencia del conocimiento al ámbito práctico (ausencia de protocolos institucionales, formación insuficiente de operadores judiciales). Las futuras líneas de investigación deben priorizar la validación de los protocolos existentes en muestras latinoamericanas, el desarrollo de instrumentos adaptados culturalmente y la evaluación sistemática de los programas de implementación.

En el plano inmediato, el aporte de la presente revisión consiste en proporcionar a los profesionales, investigadores y tomadores de decisiones una síntesis del estado del arte que facilite la fundamentación científica de las decisiones sobre políticas de evaluación testimonial. La psicología del testimonio ha alcanzado un nivel de madurez científica que justifica plenamente su incorporación sistemática a los procedimientos judiciales; lo que resta es el trabajo de transferencia y aplicación que corresponde a las comunidades científica, profesional e institucional de cada sistema jurídico.

REFERENCIAS BIBLIOGRÁFICAS

Asamblea Nacional del Ecuador. (2014). Código Orgánico Integral Penal. Registro Oficial Suplemento 180 de 10 de febrero de 2014.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2020). Memory (3rd ed.). Psychology Press.

Brainerd, C. J., & Reyna, V. F. (2005). The science of false memory. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195154054.001.0001

Christianson, S. Å. (1992). Emotional stress and eyewitness memory: A critical review. Psychological Bulletin, 112(2), 284-309. https://doi.org/10.1037/0033-2909.112.2.284

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. Journal of Experimental Psychology: Applied, 14(2), 139-150. https://doi.org/10.1037/1076-898X.14.2.139

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. Law and Human Behavior, 28(6), 687-706. https://doi.org/10.1007/s10979-004-0565-x

Fawcett, J. M., Peace, K. A., & Greve, A. (2013). Looking down the barrel of a gun: What do we know about the weapon focus effect? Journal of Applied Research in Memory and Cognition, 2(3), 201-207. https://doi.org/10.1016/j.jarmac.2013.07.002

Fisher, R. P., & Geiselman, R. E. (1992). Memory-enhancing techniques for investigative interviewing: The cognitive interview. Charles C. Thomas.

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. Current Directions in Psychological Science, 20(1), 20-23. https://doi.org/10.1177/0963721410396620

Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S. J., Avetissian, I. V., & Prosk, A. L. (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. Journal of Police Science and

Administration, 12(1), 74-80.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. The BMJ, 336(7650), 924-926. https://doi.org/10.1136/bmj.39489.470347.AD

Hershkowitz, I., Lamb, M. E., Katz, C., & Malloy, L. C. (2010). Does enhanced rapport-building alter the dynamics of investigative interviews with suspected victims of intra-familial abuse? Journal of Police and Criminal Psychology, 32(2), 89-98.

Innocence Project. (2023). Eyewitness misidentification. https://innocenceproject.org/causes/eyewitness-misidentification/

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. Psychological Bulletin, 114(1), 3-28. https://doi.org/10.1037/0033-2909.114.1.3

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. Child Abuse & Neglect, 31(11-12), 1201-1231. https://doi.org/10.1016/j.chiabu.2007.03.021

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174. https://doi.org/10.2307/2529310

Loftus, E. F. (2019). Eyewitness testimony. Applied Cognitive Psychology, 33(4), 498-503. https://doi.org/10.1002/acp.3542

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. Journal of Verbal Learning and Verbal Behavior, 13(5), 585-589. https://doi.org/10.1016/S0022-5371(74)80011-3

Manzanero, A. L. (2010). Memoria de testigos: Obtención y valoración de la prueba testifical. Pirámide.

Manzanero, A. L., & González, J. L. (2015). Avances en psicología del testimonio. Ediciones Jurídicas de Santiago.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. Psychology, Public Policy, and Law, 7(1), 3-35. https://doi.org/10.1037/1076-8971.7.1.3

Memon, A., Meissner, C. A., & Fraser, J. (2010). The cognitive interview: A meta-analytic review and study space analysis of the past 25 years. Psychology, Public Policy, and Law, 16(4), 340-372. https://doi.org/10.1037/a0020518

Morgan, C. A., Southwick, S., Steffian, G., Hazlett, G. A., & Loftus, E. F. (2013). Misinformation can influence memory for recently experienced, highly stressful events. International Journal of Law and Psychiatry, 36(1), 11-17. https://doi.org/10.1016/j.ijlp.2012.11.002

Münsterberg, H. (1908). On the witness stand: Essays on psychology and crime. Doubleday.

Orbach, Y., Shiloach, H., & Lamb, M. E. (2014). Reluctant disclosers of childhood sexual abuse. In M. E. Lamb, D. J. La Rooy, L. C. Malloy, & C. Katz (Eds.), Children's testimony: A handbook of psychological research and forensic practice (2nd ed., pp. 207-228). Wiley-Blackwell.

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

Schacter, D. L. (2001). The seven sins of memory: How the mind forgets and remembers. Houghton Mifflin.

Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), Psychological methods in criminal investigation and evidence (pp. 217-245). Springer.

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... & Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. The BMJ, 355, i4919. https://doi.org/10.1136/bmj.i4919

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., ... & Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. The BMJ, 366, l4898. https://doi.org/10.1136/bmj.l4898

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80(5), 352-373. https://doi.org/10.1037/h0020071

Vilariño, M., Arce, R., & Fariña, F. (2013). Forensic-clinical interview: Reliability and validity for the evaluation of psychological injury. The European Journal of Psychology Applied to Legal Context, 5(1), 1-21. https://doi.org/10.5093/ejpalc2013a1

Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. Psychology, Public Policy, and Law, 11(1), 3-41. https://doi.org/10.1037/1076-8971.11.1.3

Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. American Psychologist, 55(6), 581-598. https://doi.org/10.1037/0003-066X.55.6.581

Wilson, B. M., Mickes, L., Stolarz-Fantino, S., Evrard, M., & Fantino, E. (2015). Increased false-memory susceptibility after mindfulness meditation. Psychological Science, 26(10), 1567-1573. https://doi.org/10.1177/0956797615593705

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. Psychological Science in the Public Interest, 18(1), 10-65. https://doi.org/10.1177/1529100616686966

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. Applied Cognitive Psychology, 17(3), 365-373. https://doi.org/10.1002/acp.878

Zaragoza, M. S., Bement, K., & Gerhard, D. (2011). Forced confabulation and false memory following eyewitness suggestibility: The influence of explicit warnings. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(6), 1-18. https://doi.org/10.1037/a0025041

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Declaración de responsabilidad de autoría

Los autores del manuscrito señalado, DECLARAMOS que hemos contribuido directamente a su contenido intelectual, así como a la génesis y análisis de sus datos; por lo cual, estamos en condiciones de hacernos públicamente responsable de él y aceptamos que sus nombres figuren en la lista de autores en el orden indicado. Además, hemos cumplido los requisitos éticos de la publicación mencionada, habiendo consultado la Declaración de Ética y mala praxis en la publicación.

Mary Germania Orovio Ayovi y Hugo Milton Oyola Cortez: Proceso de revisión de literatura, diseño metodológico y redacción del artículo.