



# Anuario de Psicología Jurídica 2025

<https://journals.copmadrid.org/apj>



## ForenPsy: un Banco Estandarizado de Testimonios Ficticios de Testigos para la Investigación en Psicología Experimental y Judicial

Mario Álvarez, Naroa Martínez, Ujué Agudo y Helena Matute

*Departamento de Psicología, Universidad de Deusto, Bilbao, España*

### INFORMACIÓN DEL ARTÍCULO

*Historia del artículo:*  
Recibido 20 May 2024  
Aceptado 4 October 2024

*Palabras clave:*  
Testimonio  
Banco de datos  
Psicología forense  
Metodología  
Ciencia abierta

### RESUMEN

Para realizar experimentos controlados y replicables simulando juicios hemos creado ForenPsy, el primer banco de testimonios estandarizado y abierto en español. ForenPsy incluye nueve historias (tres de cada tipo de delito: homicidio, amenazas y allanamiento) con 14 testimonios cada una (siete de inocencia y siete de culpabilidad), lo que hace un total de 126. Trescientos participantes respondieron dos preguntas sobre cada testimonio: una sobre si el testimonio indicaba inocencia o culpabilidad, que muestra que el índice de acuerdo con el valor esperado fue .85, y otra sobre el grado de culpabilidad que atribuían a cada testimonio, que fue significativamente inferior para los testimonios de inocencia que para los de culpabilidad, lo que indica que los estímulos funcionaron adecuadamente. ForenPsy, así como las normas de los testimonios, está disponible en OSF y puede utilizarse y mejorarse de manera colaborativa para realizar experimentos replicables simulando contextos judiciales.

### ForenPsy: A standardized bank of fictitious witness statements for research in experimental and forensic psychology

### ABSTRACT

In order to conduct controlled and replicable experiments simulating legal trials, we present ForenPsy, the first standardized and open testimonial database in Spanish. We curated nine cases (three for each type of crime: homicide, threats, and burglary), with 14 testimonies each (seven suggesting innocence, seven suggesting guilt), resulting in 126 testimonies. Three hundred participants responded two questions for each testimony: one about whether the testimony suggested innocence or guilt, which showed that the index of agreement between the expected judgment and the actual judgment was .85, and another rating the degree of guilt suggested by each testimony, which yielded significantly lower values for innocence testimonies than for guilt testimonies, thereby showing that the testimonials worked properly. ForenPsy, as well as the standards for testimonies, are available at the OSF, so that it can be used and improved by researchers interested in conducting controlled and replicable experiments simulating judicial environments.

En entornos experimentales es frecuente el uso y la manipulación de estímulos (por ejemplo, dibujos, fotografías o palabras) para poder llevar a cabo las investigaciones. En ocasiones estos estímulos son diseñados por los propios investigadores para un experimento o una investigación en concreto, lo que requiere un examen cuidadoso de las propiedades de los estímulos para su uso en la investigación, ya que diferentes procesos cognitivos pueden verse afectados. Por ejemplo, en tareas de denominación que utilizan dibujos como estímulos y que son ampliamente utilizadas en diversas áreas de investigación e intervención en psicología, lingüística, educación o medicina se ha

demostrado que las características de los estímulos influyen en el procesamiento perceptivo, la activación de la información semántica, la selección léxica, la recuperación de la palabra y la planificación motora (ver Bonin et al., 2015; Humphreys y Riddoch, 2006; Levelt et al., 1999; Riddoch y Humphreys, 2001; Roelofs y Ferreira, 2019). Además, los estímulos de creación propia suelen ser a menudo específicos del experimento en cuestión, lo que hace difícil poder usar esos mismos estímulos en experimentos de otros laboratorios. Esta práctica dificulta la replicabilidad, la generalización de resultados y su transferencia a otros contextos. Es importante por ello contar con

Para citar este artículo: Álvarez, M., Martínez, N., Agudo, U. y Matute, E. (2025). ForenPsy: un banco estandarizado de testimonios ficticios de testigos para la investigación en psicología experimental y judicial. *Anuario de Psicología Jurídica*, 35, 113-119. <https://doi.org/10.5093/apj2025a9>

Financiación. Esta investigación se ha realizado en el marco del proyecto de investigación PID2021-126320NB-I00 financiado por la Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033) y por FEDER "Una manera de hacer Europa", así como el proyecto IT1696-22 del Departamento de Educación, Universidades e Investigación del Gobierno Vasco, concedidos a la Dra. Helena Matute, y del proyecto predoctoral de Formación de Personal Investigador con referencia BES-2017-081118 concedido a Mario Álvarez. Correspondencia: [matute@deusto.es](mailto:matute@deusto.es) (H. Matute).

ISSN: 1133-0740/© 2025 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

estudios metodológicos rigurosos en los que se diseñen y analicen bancos de estímulos considerando las posibles variables que influyen en su utilización.

Una importante línea de investigación que ha proporcionado estímulos bien controlados la constituyen los estudios normativos. Las normas ofrecen información e índices sobre aquellas variables importantes que influyen cuando se emplean dichos estímulos. Esto se hace con el fin de tener un mayor control de las variables a evaluar y de las variables extrañas que pueden contaminar los resultados. Así, el personal investigador y profesional de una amplia gama de disciplinas puede beneficiarse del uso de los bancos y bases de datos con estímulos estandarizados que ofrecen los estudios normativos. Por ejemplo, en psicología experimental es habitual emplear bases de datos psicolingüísticas para estímulos de texto (Duchon et al., 2013, para palabras en español) y bancos de imágenes estandarizados para estímulos visuales tales como dibujos (Martínez et al., 2020; Snodgrass y Vandervort, 1980), fotografías (Brodeur et al., 2010; Brodeur et al., 2014), objetos en 3D (p. ej., Peeters, 2018) y rostros con expresiones emocionales (Ekman y Friesen, 1976; Lundqvist et al., 1998). Los estudios normativos tienen, por tanto, gran repercusión en la metodología de la investigación, siendo un aspecto determinante para garantizar el rigor dado el control de variables que posibilitan, además de permitir la replicabilidad y generalización de los resultados, así como su uso en distintos contextos.

Un área de investigación que puede beneficiarse de los estudios normativos para la creación de estímulos es el ámbito de la psicología forense. Sin embargo, según nuestro conocimiento, hasta la fecha no existen muchos bancos estandarizados de estímulos para su uso en la investigación en este ámbito. Este tipo de estímulos podría resultar muy útil a la hora de realizar experimentos que simulen por ejemplo un contexto judicial. Si bien es cierto que en psicología forense es habitual emplear a modo de estímulos experimentales casos legales en formato de audio, audiovisual y de texto (Kassin y Sommers, 1997; Lassiter et al., 2002; Sealy y Cornish, 1973), según nuestro conocimiento hay pocos bancos de estímulos estandarizados (véase por ejemplo Lloyd et al., 2019 para una base de datos de detección de engaños compuesta por vídeos y Nolan et al., 2009 para una base de datos de grabaciones auditivas para investigación forense). Por este motivo, el uso de estímulos no estandarizados se observa repetidamente en diversos experimentos, lo que hace probable que se hereden los posibles problemas que pudiera generar un conjunto de estímulos no controlado de unos estudios a otros. A continuación, revisamos algunos de los estudios en el área de psicología forense que utilizan este tipo de estímulos no estandarizados.

En la investigación previa es común, por ejemplo, el uso de conjuntos de estímulos auditivos grabados por actores utilizando transcripciones de juicios reales (Furnham, 1986; Pennington, 1982; Sealy y Cornish, 1973; Wells et al., 1985). Por ejemplo, los estímulos realizados por Sealy y Cornish (1973) constan de dos juicios reales (uno de ellos un caso de robo y el otro de violación) grabados en audio por actores. En su investigación, Sealy y Cornish mostraron que variables como la edad, el sexo (entendido como sexo cromosómico) o la profesión de los testigos, entre otras variables, podían afectar a la decisión tomada en un juicio por los miembros de un jurado, lo que implicaba la necesidad de controlar al máximo la influencia de estas variables en futuras investigaciones. Este mismo conjunto de estímulos auditivos fue empleado posteriormente por Pennington (1982) en otra investigación. En sus experimentos buscaba conocer el efecto del orden de presentación de los testimonios de los testigos, encontrando también un efecto en la decisión de los jurados en función del momento de presentación de los testimonios. Por su parte Wells et al. (1985) utilizaron estímulos auditivos similares a los mencionados anteriormente para investigar también el efecto del orden en el que los abogados de la defensa y de la acusación daban su primer discurso sobre la decisión tomada por el jurado, viendo que el abogado que exponía en primer lugar influía más en la decisión del jurado. Furnham

(1986) llevó a cabo otra investigación que empleó estímulos auditivos utilizando cintas de audio sobre un caso real expuesto por Weld y Roff (1938). En su investigación encontraron un efecto de recencia en los veredictos de los miembros del jurado, de modo que los estímulos más recientes eran los que más influían en el veredicto. Contrariamente a estos trabajos realizados con estímulos no estandarizados, la base de estímulos auditivos de Nolan et al. (2009) constaba de una serie de grabaciones de 100 hombres de entre 18 y 25 años de edad en las que simulaban varias situaciones: una entrevista policial, una conversación con un cómplice, la lectura de un pasaje y la de una serie de frases, lo que proporcionaba unos materiales auditivos más uniformes para la investigación forense.

En experimentos más recientes en psicología forense es frecuente la utilización de vídeos a modo de estímulos en lugar de audios. Sin pretender exponer una relación exhaustiva de estos experimentos, describimos a continuación algunos de ellos. Por ejemplo, Lassiter et al. (2002) llevaron a cabo un experimento en el que grabaron en vídeo a un grupo de actores y a un juez jubilado simulando un juicio basado en un caso real. Comprobaron que el punto de vista de la grabación del vídeo afectaba a la decisión del jurado. Los dos grupos del experimento sólo diferían en el punto de vista de la grabación: un grupo podía ver al interrogador y el otro grupo no podía ver al interrogador. Los resultados del experimento mostraron que en el grupo en el que no se podía ver al interrogador el veredicto de culpabilidad era emitido por un 35% más de miembros que en el otro grupo, lo que sugiere que mostrar al interrogador en el estímulo de vídeo puede afectar a la reducción del veredicto de culpabilidad. En el caso de vídeos también hay bases de estímulos creadas para uso libre, como la de Lloyd et al. (2019). Esta base de datos consta de una serie de vídeos (320 en total) grabados por 80 personas (20 hombres blancos y 20 negros, 20 mujeres negras y 20 blancas), en la que cada uno grababa cuatro vídeos hablando sobre sus relaciones personales (una verdad positiva, una verdad negativa, una mentira positiva y una mentira negativa). Esta base de datos se diseñó para investigaciones de detección de mentiras con posibles sesgos raciales y de género. Otro ejemplo del uso de vídeos es la investigación publicada por Orcutt et al. (2001) sobre la detección de mentiras en niños por parte de un jurado.

También hay muchos trabajos que emplean texto para la investigación judicial. Así, Kassin y Sommers (1997) diseñaron un material consistente en un documento de texto que narraba un juicio llamado *State vs. Givens* [el Estado contra Givens] con cuatro versiones que diferían en la cantidad de información que proporcionaban. El material preparado por Kassin y Sommers se usó también en la investigación de Costabile y Klein (2005) para mostrar un efecto de orden de presentación de los testimonios en el juicio. El efecto de orden se conseguía cuando un testimonio resultaba más importante si se presentaba en último lugar, es decir, se observaba un efecto de recencia congruente con los resultados de Furnham (1986), mencionado anteriormente, con material de tipo auditivo. Otros ejemplos en los que se usó material escrito son el estudio de English et al. (2005), que usaron material escrito para comprobar si la presentación de información previa podía afectar a los expertos legales, el estudio de Rassin (2020), que comprobó que la información no probatoria podía afectar a la decisión de jueces, o el estudio de Ask y Granhag (2007), en el que usaron material escrito para suscitar tristeza o ira en investigadores forenses y comprobar si estas emociones influían en su juicio.

Por tanto, es habitual el uso de estímulos de audio, audiovisual y de texto de casos legales reales o ficticios para llevar a cabo investigaciones en psicología forense. Sin embargo, hasta donde sabemos, son escasos los bancos de estímulos estandarizados y no hemos localizado ningún estudio normativo con datos sobre aquellas variables relevantes que influyen cuando se emplean dichos estímulos en español. Esta información acerca del funcionamiento de los estímulos es determinante, dado que, tal y como hemos comentado,

las publicaciones previas han mostrado que hay muchas variables relacionadas con el uso de los estímulos que afectan a los resultados de la investigación. La literatura previa pone de relieve la necesidad de diseñar estímulos controlados, pues falta material estandarizado en el contexto jurídico.

Por los motivos comentados, esta investigación tuvo como objetivo desarrollar ForenPsy, un banco de estímulos estandarizado y de acceso abierto, que consta de 126 testimonios ficticios en español, relativos a tres tipos de delitos y organizados en nueve historias o casos. Además, se ofrecen datos normativos sobre su funcionamiento como testimonios de inocencia o de culpabilidad recogidos en una muestra de voluntarios hispanohablantes.

## Método

### Participantes

La estandarización de los testimonios se realizó con población general, ya que teníamos interés en poder ofrecer datos más cercanos a un contexto real con un posible jurado popular. En el estudio participaron 300 voluntarios (46.7% mujeres, 50.7% hombres, 2.7% otro), con edades comprendidas entre los 18 y los 64 años ( $M = 31.2$ ,  $SD = 9.6$ ). No se encontraron diferencias de género en ninguna de las variables medidas. Los sujetos se reclutaron mediante la plataforma Prolific. Utilizando los filtros que proporciona Prolific solicitamos participantes que estuvieran registrados en la plataforma que tuvieran el español como lengua materna. Se compensó su tiempo económicamente a través de la plataforma. Prolific recomienda que el pago mínimo sea de 6 libras la hora, de modo que como la participación en nuestro estudio duraba unos 12 minutos recibieron 1.2 libras por su participación. No se excluyó del estudio ningún participante.

### Materiales

El banco de testimonios ForenPsy consta de 9 historias o casos, agrupados en tres tipos de delito (tres de homicidios, tres de amenazas y tres de allanamiento), con 14 testimonios por cada caso (7 de inocencia y 7 de culpabilidad), lo que totaliza 126 testimonios estandarizados.

Los tipos de delitos se eligieron por ser algunos de los que permiten la presencia de un jurado popular en España (*Ley Orgánica 5/1995 del Tribunal del Jurado*, de 22 de mayo). Este criterio permitió que la investigación alcanzara la mayor validez ecológica posible a fin de poder aportar un mayor control metodológico a los experimentos o investigaciones que deseen usar el banco de testimonios ForenPsy en el futuro, especialmente en los casos en que sea necesaria la simulación de un juicio.

Para cada uno de los tres tipos de delito mencionados creamos tres historias ficticias (nueve en total) con información sobre un caso en el que una persona había sido acusada de cometer un delito. Para cada una de estas historias creamos 14 testimonios de supuestos testigos llamados a declarar. Como se ha indicado, la mitad de los testimonios buscaban llegar a veredictos de inocencia y la otra mitad de culpabilidad. Es decir, los investigadores diseñaron los testimonios de modo que fueran indicativos de inocencia o de culpabilidad (que denominaremos "juicio esperado"). Los dos tipos de testimonios (inocencia o culpabilidad, de acuerdo al "juicio esperado") se diseñaron considerando que la culpabilidad de una persona se refiere a su responsabilidad en la comisión de un delito, entendiendo delito como las acciones y omisiones dolosas o imprudentes penadas por la ley (*Ley Orgánica 10/1995 del Código Penal*, de 23 de noviembre). La inocencia de una persona se entendió como la falta de culpabilidad.

Con respecto al objetivo de aportar mayor rigor metodológico, para la elaboración de ForenPsy establecimos una serie de criterios,

teniendo en cuenta los resultados de la investigación previa que emplea estímulos de este tipo en psicología forense. Así, consideramos importante controlar experimentalmente las variables sociodemográficas de los testigos (Sealy y Cornish, 1973) y el orden de presentación de los testimonios (Costabile y Klein, 2005; Wells et al., 1985), dado que son variables que, como hemos comentado, pueden ejercer una gran influencia en los juicios.

Con el fin de evitar la identificación de ciertas variables demográficas tanto del acusado como de los testigos, utilizamos en el primer criterio únicamente las iniciales de las personas mencionadas en las diferentes historias o casos, nunca el nombre completo. Variables demográficas tales como el género, la clase social o la raza pueden obtenerse del nombre completo (Johnson y Kirk, 2020; Zwysen, 2021), lo que a su vez puede sesgar la toma de decisiones, como demostraron por ejemplo Sealy y Cornish (1973) en jurados populares, así como estudios en otras áreas que emplean nombres en texto (ver por ejemplo Bertrand y Mullainathan, 2004 para procesos de selección de personal).

En cuanto al orden, dado que la investigación previa ha mostrado que influye en la presentación de los estímulos (Costabile y Klein, 2005; Furnham, 1986) en investigaciones sobre juicios, las diferentes historias o casos, así como los testimonios dentro de cada historia, se presentaron en orden aleatorio para reducir así al máximo la posible influencia del orden de presentación.

Por último, para la preparación de los textos utilizados en la redacción de las historias y los testimonios se tuvo en cuenta una serie de criterios psicolingüísticos, como la longitud y la legibilidad, que podían influir en su comprensión, dado que ForenPsy utiliza material escrito. Específicamente, las 9 historias se diseñaron teniendo en cuenta la longitud (media de palabras = 302.67, rango = 266-357) y la legibilidad ( $M = 46.88$ , rango = 42.78-56.37), para lo que se considera un nivel de dificultad de comprensión muy fácil de acuerdo con los criterios de Fernández (1959). A su vez, para la creación de cada uno de los testimonios se tuvo también en cuenta la longitud (media de palabras = 17.2, rango = 9-32) y la legibilidad ( $M = 54.4$ , rango = 25.2-118), mostrando por tanto un nivel de dificultad de comprensión aceptable también para los testimonios, según los criterios de Fernández (1959).

Además, se valoraron los testimonios antes de su utilización por tres juristas independientes, a los que se pidió que indicaran en una escala de 1 a 5 tanto la claridad de cada testimonio (comprensión del texto), siendo 1 = *muy poco claro* y 5 = *muy claro*, como su coherencia, es decir, si el testimonio tenía relación lógica con el tipo de delito (homicidio, amenazas y allanamiento) y el tipo de testimonio según lo esperado por los investigadores (inocencia o de culpabilidad), siendo 1 = *muy poco coherente* y 5 = *muy coherente*. Los testimonios que recibieron valoraciones medias por debajo de 3 puntos en alguna de las dos preguntas (18 testimonios de los 126) se modificaron de acuerdo con los comentarios indicados por los expertos juristas para mejorar su claridad y/o su coherencia antes de iniciar el trabajo de recogida de datos con los participantes del estudio.

### Procedimiento

Los participantes se sirvieron de un cuestionario online en el que se presentaba a cada participante tres historias (una por cada tipo de delito: homicidio, amenazas y allanamiento) del total de nueve historias del banco ForenPsy. Se les pidió que leyera cada una de las tres historias con sus 14 testimonios. El orden en el que se presentaron las tres historias a cada participante y el orden de los testimonios dentro de cada historia fue aleatorio. Tras leer cada testimonio, debían responder dos preguntas. En la primera se les pedía, basándose solamente en la descripción de la historia y el testimonio que acababan de leer, que indicaran si consideraban que el testimonio indicaba

**Tabla 1.** Índices de acuerdo entre el juicio esperado y el juicio emitido en cada testimonio e índice de acuerdo medio de cada tipo de testimonio por historia

Tipo de testimonio	Testimonio	Hom.1	Hom.2	Hom.3	Ame.1	Ame.2	Ame.3	All.1	All.2	All.3
Inocencia	T.1	.94	.73	.69	.72	.86	.93	.97	.96	.99
	T.2	.93	.76	.87	.77	.93	.83	.73	.88	1.00
	T.3	.88	.86	.89	.77	.93	.83	.73	.88	1.00
	T.4	.88	.86	.89	.77	.88	.90	.98	.94	.99
	T.5	.88	.86	.89	.70	.88	.90	.98	.94	.99
	T.6	.94	.94	.96	.70	.89	.96	.98	.96	.92
	T.7	.79	.97	.91	.65	.89	.96	.98	.96	.92
	Total	.89 (.053)	.85 (.087)	.87 (.085)	.71 (.047)	.89 (.026)	.90 (.055)	.91 (.121)	.93 (.036)	.97 (.036)
Culpabilidad	T.8	.94	.90	.83	.65	.71	.91	.72	.72	.66
	T.9	.94	.90	.83	.77	.86	.89	.74	.76	.80
	T.10	.87	.87	.83	.71	.86	.89	.74	.76	.80
	T.11	.87	.87	.83	.71	.84	.85	.77	.74	.69
	T.12	.85	.85	.80	.76	.84	.85	.77	.74	.69
	T.13	.85	.85	.80	.76	.80	.88	.84	.76	.78
	T.14	.96	.95	.85	.95	.80	.88	.84	.76	.78
	Total	.90 (.048)	.88 (.036)	.82 (.018)	.80 (.094)	.81 (.053)	.88 (.022)	.77 (.048)	.75 (.016)	.74 (.060)

Nota. Hom. = homicidio; Ame = amenazas; All. = allanamiento; T. = testimonio; Total: media, DE entre paréntesis.

inocencia o culpabilidad en la persona acusada, siendo las únicas dos respuestas posibles (“inocencia” o “culpabilidad”). En la segunda pregunta debían indicar, utilizando una escala tipo Likert del 0 al 10, el grado de culpabilidad que indicaba cada testimonio, siendo 0 = *totalmente de inocencia* y 10 = *totalmente de culpabilidad*.

La primera variable dependiente calculada para cada testimonio fue el índice de acuerdo de cada testimonio y la proporción de participantes que emitieron un juicio (“juicio emitido”) acorde con el esperado por los investigadores al diseñar ese testimonio (“juicio esperado”) en la pregunta dicotómica (inocencia o culpabilidad) con respecto al total de participantes. En segundo lugar se calculó el grado medio de culpabilidad que indica el testimonio, expresado como el valor medio asignado por los participantes en la escala tipo Likert en la segunda pregunta. Los valores bajos en esta variable indican que los participantes asignaron al testimonio una puntuación de inocencia y los valores altos indican que asignaron una valoración de culpabilidad. Esta medida se calculó independientemente de que el juicio emitido por el participante coincidiera o no con el juicio esperado por los investigadores.

El comité de ética de la Universidad de Deusto aprobó el procedimiento usado en el presente estudio. El consentimiento informado de los participantes fue recogido por escrito.

## Resultados

Para las respuestas de cada testimonio analizamos los siguientes datos: (a) el índice de acuerdo existente entre el “juicio esperado” por los investigadores y el “juicio emitido” por los participantes (inocencia o culpabilidad) y (b) la puntuación promedio en culpabilidad en la escala tipo Likert, desde 0 (*totalmente de inocencia*) hasta 10 (*totalmente de culpabilidad*).

Sobre el índice de acuerdo entre el juicio esperado y el emitido, los testimonios en su conjunto obtuvieron un acuerdo medio de .85, lo que indica que el 85% de los participantes emitieron un juicio de inocencia o culpabilidad que concordaba con lo esperado por los investigadores. Atendiendo al tipo de delito, encontramos que el índice de acuerdo en los testimonios de homicidio fue de .87, en los testimonios de amenazas de .83 y en los de allanamiento de .85.

Si los analizamos por tipo de testimonio, los de inocencia mostraron un índice de acuerdo de .88 y los de culpabilidad de .82. Además, observamos el índice de acuerdo de cada testimonio de forma individual y encontramos que 89 de los 126 testimonios (70.63%) obtuvieron un acuerdo del .80 o superior y 31 de 126 testimonios (24.6%) un acuerdo entre el .70 y el .80. Por último, 6 de los 126 testimonios (4.76%) obtuvieron un acuerdo entre el .60 y el .70.

**Tabla 2.** Puntuaciones promedio de culpabilidad y desviaciones estándar para cada testimonio y puntuación media de cada tipo de testimonio por historia

Tipo de testimonio	Testimonio	Hom.1	Hom.2	Hom.3	Ame.1	Ame.2	Ame.3	All.1	All.2	All.3
Inocencia	T.1	2.81 (2.46)	4.35 (2.39)	3.72 (2.42)	3.65 (2.41)	3.54 (2.25)	2.71 (2.22)	1.73 (2.19)	1.92 (2.23)	1.61 (2.20)
	T.2	3.7 (2.42)	4.01 (2.42)	3.08 (2.06)	3.49 (2.42)	3.03 (2.06)	3.85 (1.87)	3.33 (2.51)	3.52 (2.05)	2.40 (2.01)
	T.3	3.60 (2.26)	3.63 (2.00)	3.11 (1.92)	4.09 (2.50)	3.41 (2.12)	3.25 (2.16)	1.96 (2.09)	2.65 (2.01)	1.90 (2.05)
	T.4	3.12 (2.05)	2.99 (2.15)	2.98 (1.79)	4.01 (2.46)	3.27 (1.95)	2.92 (1.94)	1.76 (2.20)	2.37 (2.11)	2.34 (2.32)
	T.5	3.37 (2.44)	3.24 (1.86)	3.27 (1.95)	4.27 (2.23)	3.22 (2.01)	3.21 (2.09)	2.64 (2.23)	2.41 (2.08)	2.39 (1.83)
	T.6	3.45 (2.37)	3.50 (1.73)	3.38 (2.21)	5.37 (2.86)	3.60 (2.08)	3.89 (1.95)	2.58 (2.29)	2.94 (2.21)	2.63 (2.05)
	T.7	3.86 (2.51)	2.37 (2.19)	3.24 (2.03)	3.86 (1.94)	2.80 (2.18)	3.22 (2.40)	3.00 (2.01)	3.62 (2.26)	2.49 (2.10)
	Total	3.34 (2.36)	3.44 (2.11)	3.25 (2.05)	4.11 (2.40)	3.27 (2.09)	3.29 (2.09)	2.43 (2.22)	2.78 (2.14)	2.25 (2.08)
Culpabilidad	T.8	7.44 (1.56)	6.90 (1.91)	6.27 (2.11)	5.88 (2.36)	5.77 (2.40)	6.81 (1.88)	5.56 (2.41)	5.39 (2.14)	5.23 (1.83)
	T.9	6.91 (2.04)	6.98 (1.95)	6.70 (2.10)	5.46 (2.00)	6.21 (2.17)	6.09 (1.95)	6.07 (2.30)	5.67 (2.19)	5.79 (2.05)
	T.10	6.35 (2.16)	6.54 (1.85)	6.28 (1.88)	6.27 (2.53)	6.72 (2.10)	5.74 (2.14)	5.89 (2.21)	5.69 (2.44)	5.02 (2.05)
	T.11	7.90 (1.70)	7.26 (1.92)	6.38 (2.01)	7.39 (1.73)	6.02 (2.14)	6.49 (1.98)	6.45 (2.22)	5.63 (2.04)	5.37 (1.93)
	T.12	7.19 (1.98)	6.13 (1.80)	5.68 (2.22)	6.88 (2.01)	5.64 (2.16)	7.02 (2.11)	4.49 (2.35)	5.64 (2.06)	5.04 (2.14)
	T.13	6.07 (2.09)	5.83 (2.04)	4.81 (2.24)	6.15 (1.91)	6.59 (2.21)	5.74 (1.88)	5.40 (2.40)	4.19 (2.29)	5.96 (2.17)
	T.14	6.28 (1.96)	6.98 (1.88)	5.94 (1.90)	5.52 (1.81)	5.88 (2.15)	6.44 (1.95)	5.03 (2.52)	5.58 (2.08)	6.19 (1.94)
	Total	6.88 (1.93)	6.66 (1.91)	6.01 (2.07)	6.22 (2.05)	6.12 (2.19)	6.33 (1.98)	5.56 (2.34)	5.40 (2.18)	5.51 (2.02)

Nota. Hom. = homicidio; Ame = amenazas; All. = allanamiento; T. = testimonio; Total: media, DE entre paréntesis.

Los índices de acuerdo entre juicio esperado y emitido para cada uno de los tipos de testimonios se muestran en la [Tabla 1](#). Como se aprecia, en todos los casos hay un índice de acuerdo superior al .70, siendo los testimonios de inocencia de la historia 1 de amenazas los más bajos, con un .71 y los de la historia 3 de allanamiento los que obtienen el índice de acuerdo más alto, con un .97.

Las puntuaciones promedio de culpabilidad de cada testimonio en la escala de 0 a 10 pueden verse en la [Tabla 2](#). En todas las historias se observa que los testimonios diseñados para indicar inocencia muestran puntuaciones promedio entre 2.25 y 4.11, mientras que los testimonios diseñados para indicar culpabilidad muestran puntuaciones promedio entre 5.40 y 6.88, lo que indica que, tal y como se esperaba, los participantes interpretaron los testimonios de inocencia con puntuaciones más bajas que los testimonios de culpabilidad. Para comprobar si estas diferencias eran significativas realizamos una prueba *t* para muestras relacionadas, que mostró que, efectivamente, las puntuaciones promedio de culpabilidad de los testimonios de inocencia en la escala Likert fueron significativamente menores que las de los testimonios de culpabilidad,  $t(299) = -34.1, p < .001, d = -1.97$ .

Finalmente, para comprobar si las diferencias entre los diferentes tipos de testimonio (inocencia o culpabilidad) se mantenían en cada tipo de delito, también realizamos pruebas *t* para muestras relacionadas comparando los tipos de testimonio para cada tipo de delito. Se encontraron diferencias significativas entre los testimonios de inocencia y los de culpabilidad en las historias de homicidios,  $t(299) = -31.59, p < .001, d = -1.82$ , amenazas,  $t(299) = -26.65, p < .001, d = -1.54$ , y allanamiento,  $t(299) = -28.04, p < .001, d = -1.62$ , lo cual indica que en todos los casos los testimonios producían diferencias significativas en función de si se habían diseñado para generar un veredicto de inocencia o de culpabilidad, por lo que podemos concluir que han funcionado según lo previsto. Por otra parte, no se hicieron comparaciones entre testimonios de diferentes tipos de delitos (homicidio vs. allanamiento vs. amenazas) ya que no se disponía de ninguna predicción al respecto.

## Discusión

El trabajo describe ForenPsy, un nuevo banco de testimonios ficticios de acceso abierto y estandarizado, para su uso en investigación psicológica en el ámbito forense. Está compuesto por nueve historias de tres tipos de delito (tres de homicidios, tres de amenazas y tres de allanamiento), con 14 testimonios por cada historia (7 de inocencia y 7 de culpabilidad), lo que hace un total de 126 testimonios estandarizados. ForenPsy utiliza una muestra de adultos hispanohablantes e incluye normas para cada testimonio acerca del índice de acuerdo con respecto al juicio esperado (inocencia o culpabilidad) así como la puntuación promedio de culpabilidad de cada testimonio en una escala de 0 a 10, en la que 0 supone que el testimonio indica claramente inocencia y 10 indica claramente culpabilidad.

Para la construcción de los estímulos hemos seguido una serie de criterios rigurosos encaminados a reducir la influencia de posibles sesgos ya descritos en la literatura previa derivados de variables sociodemográficas de los testigos ([Sealy y Cornish, 1973](#)) y del orden de presentación de los testimonios ([Costabile y Klein, 2005](#); [Wells et al., 1985](#)). El criterio empleado para prevenir un posible sesgo derivado de variables sociodemográficas fue el uso de iniciales en los testimonios, en lugar de nombres propios completos, que podrían dar lugar a sesgos de género, clase social y raza ([Johnson y Kirk, 2020](#); [Zwysen, 2021](#)). El trabajo no pretende indagar en esas variables sociodemográficas, como en los bancos diseñados por [Nolan et al. \(2009\)](#) o [Lloyd et al. \(2019\)](#), sino que estas variables influyan lo menos posible en el resultado. Además, el orden de presentación de las historias y los testimonios se estableció de manera aleatoria, para neutralizar así

también la posible influencia del orden de presentación sobre la estandarización del banco de estímulos que presentamos, de modo que en el futuro sea posible manipular estas variables según los objetivos de diferentes investigaciones y de ese modo analizar su influencia, partiendo del estándar normalizado que presentamos.

Debido a que empleamos materiales escritos como estímulos, hemos tenido en cuenta también en el diseño de los estímulos, variables psicolingüísticas que pueden afectar a la dificultad de comprensión, como la longitud y la legibilidad de las diferentes historias y testimonios. Por último, tres expertos juristas valoraron el banco ForenPsy completo en cuanto a la claridad de los testimonios y su coherencia con el tipo de delito y los testimonios que lo necesitaron fueron actualizados de acuerdo a las recomendaciones de los expertos.

Los resultados muestran un elevado grado de acuerdo entre el juicio esperado por los investigadores (culpabilidad o inocencia) y el juicio emitido por los participantes, así como unas puntuaciones de culpabilidad adecuadas en las escalas tipo Likert según el tipo de testimonio, con valores promedio significativamente menores para los testimonios de inocencia que para los testimonios de culpabilidad, lo que sugiere que el banco de estímulos funciona adecuadamente y puede ser utilizado en investigaciones futuras. Se ofrecen las normas de cada uno de los 126 testimonios en cuanto a su índice de acuerdo y su puntuación de culpabilidad.

ForenPsy se presenta por tanto como un material adecuado para la investigación debido a tres características fundamentales: el procedimiento riguroso seguido en su construcción, los datos obtenidos acerca de su comportamiento y la publicación de los datos obtenidos y los materiales empleados en abierto. Este tipo de estudios normativos abiertos es muy valorado en psicología, tratándose de un área fructífera de investigación de la influencia de estímulos como dibujos, fotografías o palabras en el estudio de diferentes procesos cognitivos que pueden verse afectados en diferentes áreas (véase por ej., [Martínez et al., 2020](#)).

Contar en el ámbito de la investigación jurídica con un banco abierto y estandarizado de testimonios es importante para poder llevar a cabo investigaciones utilizando instrumentos estandarizados, a fin de mejorar el control experimental y la replicabilidad de la investigación en el área, ya que en los trabajos revisados se ha observado que era habitual que cada equipo de investigación utilizara sus propios estímulos que luego no eran utilizados por otros equipos, o que cuando eran utilizados solían modificarse según las necesidades de cada investigación.

[Sealy y Cornish \(1973\)](#), por ejemplo, utilizaron el audio de una transcripción teatralizada de un juicio real. Tanto [Pennington \(1982\)](#) como [Wells et al. \(1985\)](#) utilizaron esas mismas transcripciones no estandarizadas. Aunque esta herramienta era reutilizable, no resultaba versátil, ya que solo tenían un caso con el que trabajar y estaba limitado por los testimonios proporcionados en el caso real. Lo mismo sucedió con el material creado por [Weld y Roff \(1938\)](#). Aunque presenta la ventaja de poder conocer con certeza el veredicto alcanzado, utilizar un caso real como base resulta limitante a la hora de tratar de trasladar esa herramienta a otros casos, situaciones o contextos.

Algunos instrumentos sí han sido diseñados para ser ligeramente modificados en función del veredicto que se quiere inducir. Es el caso del instrumento diseñado por [Kassin y Sommers \(1997\)](#), que presentaba varias versiones del caso a fin de modificar la decisión de los participantes, o el caso Jason Wells ([Simon et al., 2004](#)) que resulta similar a nuestro instrumento, con varias piezas de información que pueden inducir inocencia o culpabilidad. Sin embargo, aunque estas herramientas sean útiles, según nuestro conocimiento, no hay publicados datos a día de hoy acerca de su funcionamiento, por lo que no es posible controlar sus potenciales efectos extraños en los experimentos. Otros instrumentos cuidadosamente diseñados, como los de [Lloyd et al. \(2019\)](#) o [Nolan et al. \(2009\)](#), se han diseñado con un objetivo distinto a la investigación presente, ya que emplean material audiovisual y se centran en variables sociodemográficas, como la raza y el género.

Las normas y el banco de testimonios presentados en este trabajo son de acceso abierto y están a disposición de los investigadores y profesionales que los necesiten en el enlace de OSF proporcionado al final del artículo, para que puedan utilizarlo en sus propias investigaciones. Los investigadores que deseen usar el banco ForenPsy pueden consultar las puntuaciones correspondientes a cada testimonio y controlar de esta forma los posibles efectos del uso de los diferentes estímulos en su estudio o valorar qué testimonios son más adecuados para el propósito de su investigación y cuáles menos, en función de si desean utilizar unos testimonios de inocencia o de culpabilidad más ambiguos o más rotundos (según los índices de acuerdo y puntuación promedio de culpabilidad; véase por ejemplo [Agudo et al., 2024](#)).

Además, al ser ForenPsy un banco de acceso abierto y gratuito, creemos interesante destacar que cualquier persona puede contribuir en el futuro a mejorarlo, aumentar el tamaño de la base de datos (pudiendo ampliar la muestra y/o el número de estímulos), realizar nuevos análisis y publicar nuevas normas que pudieran resultar de interés a la comunidad científica en el futuro. El banco se plantea, por tanto, como un primer paso para crear un gran banco de datos colaborativo, abierto a todos los investigadores que deseen contribuir a su desarrollo, de manera que se pueda mejorar la replicabilidad y la comparación entre experimentos en el contexto de la psicología forense, algo que creemos que resulta muy necesario si tenemos en cuenta la crisis de replicabilidad que ha sufrido la ciencia, en particular la psicología, en los últimos años ([Camerer et al., 2018](#)). A través del mencionado enlace de OSF que se muestra al final del artículo es posible acceder también a una versión preliminar y más sencilla del banco de datos que utilizamos anteriormente en algunas investigaciones de nuestro equipo de investigación (e.g., [Agudo et al., 2024](#)), y esperamos que también se acceda en un futuro a las nuevas versiones que tanto nosotros mismos como aquellos otros investigadores interesados quieran enlazar a través de la plataforma de acceso abierto OSF.

Por último, creemos conveniente resaltar que además de que puede ampliarse y mejorarse en un futuro, este banco de testimonios puede contribuir, no solo a la investigación en psicología forense, sino también a otras muchas áreas de investigación relacionadas, tales como la investigación básica en recuerdos falsos (véase p. ej., [McCloskey y Zaragoza, 1985](#); [Stark et al., 2010](#)), la investigación sobre la influencia de creencias previas sobre los juicios y toma de decisiones ([Vicente et al., 2023](#)) o sobre la tendencia a saltar a la conclusión antes de analizar toda la evidencia disponible ([Moreno-Fernández et al., 2023](#)). Dado que se trata de materiales textuales flexibles que narran historias estandarizadas, pueden utilizarse lógicamente también en cualquier investigación de psicología que requiera de materiales textuales estandarizados en formato de historias cortas.

### Extended Summary

Upon reviewing previous experiments conducted in the judicial context, we observed that most experiments used instruments and materials that were not standardized for uniform use by different groups of researchers. In addition, most experiments and materials were prepared originally in English, so differences could arise when using those materials in Spanish-speaking countries. To address these concerns, we decided to develop an instrument in Spanish that could be used in judicial research in Spanish-speaking countries. To this end, we designed ForenPsy, a standardized bank of testimonies in Spanish that can be used in experiments by any group of independent researchers, as it is freely available on the Open Science Framework.

This new instrument encompasses three types of crimes (homicide, threats, and burglary), each included in three different stories or cases, with 14 testimonies for each story (7 for innocence and 7 for guilt). The total 126 testimonies underwent rigorous methodological

criteria to equalize their length and readability, meaning all had similar lengths and were similarly understandable. Possible biases based on race and gender were also controlled for by using initials instead of full names for both victims and defendants. To avoid unwanted effects due to presentation order, cases and testimonies were presented in a random order. Additionally, their coherence and clarity were previously assessed by three legal experts.

Three hundred Spanish-speaking adults (46.7% women, 50.7% men, 2.7% other) aged between 18 and 64 years old ( $M = 31.2$ ,  $SD = 9.6$ ) participated in the study through the Internet. These participants were asked to read three out of the nine stories each (one for each type of crime) with their corresponding 14 testimonies per story. For each testimony, participants were instructed to indicate whether they thought it suggested the innocence or guilt of the defendant. Furthermore, participants were requested to rate on a Likert scale (0-10) the degree to which the testimony pointed to innocence or guilt, with 0 = *completely indicating innocence* and 10 = *completely indicating guilt*.

Thus, our experiment comprised two dependent variables. First, the agreement index for each testimony, which was the percentage of participants whose verdict for that testimony matched the verdict expected by the researchers (either guilt or innocence), and second, the average guilt score for each testimony, defined as the mean response on the 0-10 Likert scale.

The results showed that 70.63% of testimonies obtained an agreement index of .80 or more, 24.6% between .70 and .80, and 4.76% obtained an agreement index between .60 and .70. These results indicate that the testimonies exhibit moderate to excellent agreement indexes and can be reliably used in future research. Regarding average guilt scores, we found that the average scores in each story ranged from 2.25 to 4.11 for testimonies designed as innocence testimonies, while they ranged from 5.40 to 6.88 for guilt designed testimonies. Paired sample *t*-tests comparing average scores between different types of testimonies showed that these differences were statistically significant.

The purpose of this research was to elaborate an instrument that could be used by researchers in experiments with standardized stimuli in judicial contexts. ForenPsy meets all these criteria. Firstly, the stimuli are standardized. We show whether each of them suggests innocence or guilt, which ones should be preferred as a function of whether guilt or innocence testimonies are needed, which ones are more ambiguous in case this is needed according to the goals of the research, which ones show a higher percentage of agreement. They have undergone rigorous methodological criteria, so, for example, they are equally understandable and do not present biases associated with full names that could induce sociodemographic biases. The potential effect of presentation order has also been controlled for.

Additionally, ForenPsy was designed in Spanish, so it can be used in Spanish-speaking countries. Last but not least, ForenPsy is open and freely available on the Open Science Framework, so anyone can use it. The raw data from this research are also freely available, so further analysis, comparisons, and improvements can be contributed by any researcher conducting forensic research in psychology.

As this research is freely available on the Open Science Framework, we consider it is a very important first step in building a large database to which any researcher can contribute with new testimonies, experiments, data, analysis, and improvements. Therefore, ForenPsy is an open testimonial bank that we hope will continue to grow and evolve through collective effort.

In conclusion, this research offers the first open bank of testimonies in Spanish and enables researchers not only to conduct their own forensic psychology research but also to contribute to improving this bank in the future. Indeed, the open format of this standardized bank should facilitate the collaborative efforts needed for scientific advancement and replicability.

## Conflicto de Intereses

Los autores de este artículo declaran que no tienen ningún conflicto de intereses.

## Nota

Los materiales y los datos de esta investigación están disponibles en abierto en el Open Science Framework, a través de este enlace: <https://osf.io/fk2z5/>

## Referencias

- Agudo, U., Liberal, K.G., Arrese, M. y Matute, H. (2024). The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications*, 99, Article 1. <https://doi.org/10.1186/s41235-023-00529-3>
- Ask, K. y Granhag, P. A. (2007). Hot cognition in investigative judgments: The differential influence of anger and sadness. *Law and Human Behavior*, 31(6), 537-551. <https://doi.org/10.1007/s10979-006-9075-3>
- Bertrand, M. y Mullainathan, S. (2003). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991-1013. <https://doi.org/10.3386/w9873>
- Bonin, P., Gelin, M., Laroche, B., Méot, A. y Bugaiska, A. (2015). The "how" of animacy effects in episodic memory. *Experimental Psychology*, 62(6), 371-384. <https://doi.org/10.1027/1618-3169/a000308>
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T. y Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5(5), Artículo e10773. <https://doi.org/10.1371/journal.pone.0010773>
- Brodeur, M. B., Guérard, K. y Bouras, M. (2014). Bank of standardized stimuli (BOSS) Phase II: 930 new normative photos. *PLoS ONE*, 9(9), Artículo e106953. <https://doi.org/10.1371/journal.pone.0106953>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B., Pfeiffer, T., Altmeld, A., Buttrick, N., Cahn, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T. y Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637-644. <https://doi.org/10.1038/s41562-018-0399-z>
- Costabile, K. A. y Klein, S. B. (2005). Finishing strong: Recency effects in juror judgments. *Basic and Applied Social Psychology*, 27(1), 47-58. [https://doi.org/10.1207/s15324834baspp2701\\_5](https://doi.org/10.1207/s15324834baspp2701_5)
- Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A. y Carreiras, M. (2013). Espal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, 45(4), 1246-1258. <https://doi.org/10.3758/s13428-013-0326-1>
- Ekman, P. y Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press. <https://www.paulekman.com/es/producto/pictures-of-facial-affect-pofa/>
- Englich, B., Mussweiler, T. y Strack, F. (2005). The last word in court--A hidden disadvantage for the defense. *Law and Human Behavior*, 29(6), 705-722. <https://doi.org/10.1007/s10979-005-8380-7>
- Fernández, J. (1959). Medidas sencillas de lecturabilidad. *Consigna*, 214, 29-32.
- Furnham, A. (1986). The robustness of the recency effect: Studies using legal evidence. *The Journal of General Psychology*, 113(4), 351-357. <https://doi.org/10.1080/00221309.1986.9711045>
- Humphreys, G. W. y Riddoch, M. J. (2006). Features, objects, action: The cognitive neuropsychology of visual object processing, 1984-2004. *Cognitive Neuropsychology*, 23(1), 156-183. <https://doi.org/10.1080/02643290542000030>
- Johnson, S. K. y Kirk, J. F. (2020). Dual-anonymization yields promising results for reducing gender bias: A naturalistic field experiment of applications for hubble space telescope time. *Publications of the Astronomical Society of the Pacific*, 132(1009), Artículo 034503. <https://doi.org/10.1088/1538-3873/ab6cec>
- Kassin, S. M. y Sommers, S. R. (1997). Inadmissible testimony, instructions to disregard, and the jury: Substantive versus procedural considerations. *Personality and Social Psychology Bulletin*, 23(10), 1046-1054. <https://doi.org/10.1177/01461672972310005>
- Lassiter, G. D., Geers, A. L., Handley, I. M., Weiland, P. E. y Munhall, P. J. (2002). Videotaped interrogations and confessions: A simple change in camera perspective alters verdicts in simulated trials. *Journal of Applied Psychology*, 87(5), 867-874. <https://psycnet.apa.org/doi/10.1037/0021-9010.87.5.867>
- Levett, W. J. M., Roelofs, A. y Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38. <https://doi.org/10.1017/S0140525X99001776>
- Ley Orgánica 5/1995 del Tribunal del Jurado (22 de mayo, 1995), *Boletín Oficial del Estado*, 122, 15001-15021. <https://www.boe.es/eli/es/lo/1995/05/22/5>
- Ley Orgánica 10/1995 del Código Penal (23 de noviembre, 1995). *Boletín Oficial del Estado*, 281, 33987-34058. <https://www.boe.es/eli/es/lo/1995/11/23/10>
- Lloyd, E. P., Deska, J. C., Hugenberg, K., McConnell, A. R., Humphrey, B. T. y Kunstman, J. W. (2019). Miami University deception detection database. *Behavior Research Methods*, 51, 429-439. <https://doi.org/10.3758/s13428-018-1061-4>
- Lundqvist, D., Flykt, A. y Öhman, A. (1998) *The Karolinska Directed Emotional Faces (KDEF)*. Department of Neurosciences, Karolinska Hospital, Stockholm. <https://kdef.se/>
- Martinez, N., Matute, H. y Goikoetxea, E. (2020). PicPsy: A new bank of 106 photographs and line drawings with written naming norms for Spanish-speaking children and adults. *PLOS ONE*, 15(9), Artículo e0238976. <https://doi.org/10.1371/journal.pone.0238976>
- McCloskey, M. y Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114(1), 1-16. <https://doi.org/10.1037/0096-3445.114.1.1>
- Moreno-Fernández, M. M., Blanco, F. y Matute, H. (2021). The tendency to stop collecting information is linked to illusions of causality. *Scientific Reports*, 11(1), Artículo 3942. <https://doi.org/10.1038/s41598-021-82075-w>
- Nolan, F., McDougall, K., de Jong, G. y Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law*, 16(1), 31-57. <https://doi.org/10.1558/ijsl.v16i1.31>
- Orcutt, H. K., Goodman, G. S., Tobey, A. E., Batterman-Faunce, J. M. y Thomas, S. (2001). Detecting deception in children's testimony: Factfinders' abilities to reach the truth in open court and closed-circuit trials. *Law and Human Behavior*, 25(4), 339-372. <https://doi.org/10.1023/A:1010603618330>
- Peeters, D. (2018). A standardized set of 3-D objects for virtual reality research and applications. *Behavior Research Methods*, 50(3), 1047-1054. <https://doi.org/10.3758/s13428-017-0925-3>
- Pennington, D. C. (1982). Witnesses and Their testimony: Effects of ordering on juror verdicts. *Journal of Applied Social Psychology*, 12(4), 318-333. <https://doi.org/10.1111/j.1559-1816.1982.tb00868.x>
- Rassin, E. (2020). Context effect and confirmation bias in criminal fact finding. *Legal and Criminological Psychology*, 25(2), 80-89. <https://doi.org/10.1111/lcrp.12172>
- Riddoch, M. J. y Humphreys, G. W. (2001). Object recognition. En B. Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 45-74). Psychology Press.
- Roelofs, A. y Ferreira, V. S. (2019). The architecture of speaking. En Hagoort, P. (Ed.), *Human language: From genes and brains to behavior* (pp. 35-50). MIT Press
- Sealy, A. P. y Cornish, W. R. (1973). Jurors and their verdicts. *The Modern Law Review*, 36(5), 496-508. <https://doi.org/10.1111/j.1468-2230.1973.tb01381.x>
- Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *The University of Chicago Law Review*, 71(2), 511-586. <http://www.jstor.org/stable/1600674>
- Snodgrass, J. G. y Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174-215. <https://doi.org/10.1037/0278-7393.6.2.174>
- Stark, C. E. L., Okado, Y. y Loftus, E. F. (2010). Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms. *Learning & Memory*, 17(10), 485-488. <https://doi.org/10.1101/lm.1845710>
- Vicente, L., Blanco, F. y Matute, H. (2023). I want to believe: Prior beliefs influence judgments about the effectiveness of both alternative and scientific medicine. *Judgment and Decision Making*, 18, Artículo e1. <https://doi.org/10.1017/jdm.2022.3>
- Weld, H. P. y Roff, M. (1938). A study in the formation of opinion based upon legal evidence. *The American Journal of Psychology*, 51(4), 609-628. <https://doi.org/10.2307/1415696>
- Wells, G. L., Wrightsman, L. S. y Miene, P. K. (1985). The timing of the defense opening statement: Don't wait until the evidence is. *Journal of Applied Social Psychology*, 15(8), 758-772. <https://doi.org/10.1111/j.1559-1816.1985.tb02272.x>
- Zwysen W. (2021) *Performance pay across Europe: Drivers of the increase and the link with wage inequality*. ETUI Research Paper-Working Paper No. 06. <https://doi.org/10.2139/ssrn.3884809>

