

Estudios de replicación, pre-registros y ciencia abierta en Psicología

Eduardo GARCÍA-GARZÓN

Óscar LECUONA

Universidad Autónoma de Madrid (España)

Guillermo V. CARBAJAL

Instituto de Neurociencia de Castilla y León, Salamanca (España)

Instituto Biomédico de Salamanca (España)

Resumen

En los últimos años, la Psicología ha sufrido una crisis de confianza. Esta crisis, marcada por la evidencia de que una gran parte de los efectos científicos publicados en la literatura no eran replicables, tiene varias causas: primero, la existencia de sistema inadecuado de incentivos en la carrera investigadora; segundo, un uso extendido de prácticas cuestionables de investigación; tercero, la ausencia de métodos de control de dichas prácticas y buena praxis investigadora (principalmente, pre-registros, estudios de réplica y protocolos de ciencia abierta).

Este artículo presenta, de forma novedosa en castellano, una revisión exhaustiva de dichas causas y de las soluciones propuestas por la comunidad científica. Por último, se ofrece una propuesta de adaptación de estas soluciones al currículum académico de estudiantes de Psicología.

Para ello, se introduce un sistema de aprendizaje basado en proyectos de investigación mediante estudios de réplica que incluyen protocolos de pre-registro y uso de ciencia abierta.

Abstract

Over the last years, Psychology has experienced a “confidence crisis”. This crisis, prompted by the inability to replicate a significant proportion of scientific published effects, has several causes: First, the existence of an inadequate system of incentives to progress in the academic career. Second, the extensive use of questionable research practices. Third, the limited use of good scientific praxis methods that could control the later (i.e., pre-registered and replication studies and open science frameworks).

This article features an exhaustive review in Spanish of both, the aforementioned causes and solutions proposed by the scientific community as of yet. Lastly, a proposal for introducing these solutions in the academic curriculum of psychology students is presented.

Accordingly, a project-based learning built on replication studies, including pre-registered protocols and open science usage, is proposed.

Dirección de los autores: Departamento de Psicología Social y Metodología, Facultad de Psicología. c/ Iván Pavlov, 6. 28049 Madrid. *Correo electrónico:* eduardo.garciag@uam.es

Agradecimientos: Los autores quieren agradecer a Miguel Vadillo (Universidad Autónoma de Madrid), David Santos (Universidad Autónoma de Madrid) y Juan Ramón Barrada (Universidad de Zaragoza) sus correcciones, sugerencias e ideas en la preparación de este artículo.

Financiación: Este trabajo ha recibido financiación del Ministerio de Educación, Ciencia y Deporte [ayuda para Formación del Profesorado Universitario FPU15/03246] y del programa *Erasmus Mundus* de la Comisión Europea [ayuda para la formación en el extranjero EM-ACN *Mobility Fellowship*].

Recibido: diciembre de 2017. *Aceptado:* marzo de 2018.

La publicación en revistas de alto impacto de estudios científicos cuestionables exaltó una controversia preexistente sobre los criterios de calidad en la producción científica (Gelman y Loken, 2014). Este debate, activo durante varias décadas (Gelman, 2016; Santos, 2013), encontró en Daryl Bem¹ (2011) su punto álgido: la publicación en una revista de alto impacto en psicología social de una serie de hallazgos polémicos (i.e., existencia de percepción extrasensorial) en un estudio aparentemente legítimo y que cumplía los estándares metodológicos exigidos (incluyendo nueve experimentos diferentes). Mientras que su adecuación técnica (método, diseño, análisis) fue pronto cuestionada -y sus conclusiones finalmente refutadas (Wagenmakers, Wetzels, Borsboom y van der Maas, 2011)- este estudio se convirtió en el perfecto ejemplo de las deficiencias en el proceso de publicación científica.

Bem (2011) marca el comienzo un proceso de escrutinio de publicaciones incluyendo resultados cuestionables, que culminó con la presentación de los resultados del *Reproducibility Project: Psychology (Open Science Collaboration, 2015)*: un esfuerzo conjunto de más de 270 investigadores para replicar 100 estudios de efectos publicados en revistas de alto impacto, se encontró que la tasa de replicabilidad era sólo del 25% en psicología social y del 50% en psicología cognitiva (*Open Science Collaboration, 2015*). Estos resultados se sumaron las conclusiones en la misma dirección del *Many Labs Project* (Klein, Ratliff, Vianello *et al.*, 2014) o la polémica relativa a las “posturas de poder” (Ranehill, Dreber, Johannesson, Leiberg y Weber, 2015; Simmons y Simonsohn, 2017), entre muchos otros (Gelman, 2016). Esta situación instauró con un clima generalizado de escepticismo sobre la calidad de la investigación psicológica. Así, se gesta una “crisis de confianza” (Ranehill *et al.*, 2015) dentro de la comunidad científica donde se pone en duda un sistema de publicación (Ioannidis, 2005; John, Loewenstein y Prelec, 2012) que se demuestra incapaz de discernir entre resultados sistemáticos y anecdóticos.

Causas de la crisis de la replicabilidad

Las causas de esta crisis pueden agruparse en factores contextuales e individuales. Los factores contextuales incluyen la presión para publicar resultados científicos positivos y novedosos (“publica o perece”) (van Dijk, Manor, y Carey, 2014), fomentada por unos criterios de evaluación de la calidad investigadora basados en ratios de publicación por año y la imposibilidad de publicar estudios con resultados nulos (Ioannidis, 2006). La consecuencia directa es la aparición de

sesgos en el proceso de publicación científica (Simonsohn, Nelson, y Simmons, 2014): la publicación sistemática de estudios dependientes de su resultado, donde los artículos publicados sólo incluyen resultados confirmatorios de las hipótesis presentadas. En Psicología, este tipo de resultados representan el 90% del volumen de publicación (Fanelli, 2010), distorsionando la evidencia científica real en el campo (Song, Hooper, y Loke, 2013).

Esta situación fomenta que los investigadores tomen decisiones que maximicen la probabilidad de encontrar resultados significativos (más fácilmente publicables) en sus investigaciones. Estos factores individuales son conocidos como *prácticas cuestionables de investigación* (PCI) (Fiedler y Schwarz, 2016; John *et al.*, 2012). En este artículo, centraremos la atención sobre estos factores individuales, pero los lectores son referidos a las referencias para encontrar una mayor discusión sobre los factores contextuales previamente nombrados.

Prácticas cuestionables de investigación

Las prácticas cuestionables de investigación (PCI) son decisiones tomadas durante el proceso de producción científica (John *et al.*, 2012; Wicherts, Veldkamp, Augusteijn *et al.*, 2016) que buscan maximizar la aparición de resultados significativos (Ioannidis, 2005) en detrimento de su validez (Fiedler y Schwarz, 2016). Son producto de los “grados de libertad del investigador” (Simmons, Nelson y Simonsohn, 2011; Gelman y Loken, 2013), es decir, la flexibilidad para tomar decisiones legítimas (por ejemplo, escoger entre dos métodos de análisis) que conducen a conclusiones radicalmente diferentes (Barrada, 2017). Conviene remarcar que las PCI no incluyen prácticas ilegítimas, como la falsificación de datos u otros fraudes científicos.

Lamentablemente, las PCI están muy extendidas en la comunidad científica, independientemente de su lugar de origen y área de investigación (Agnoli, Wicherts, Veldkamp, Albiero y Cubelli, 2017; Fiedler y Schwarz, 2016; John *et al.*, 2012). De forma preocupante, ciertas PCI han mostrado una prevalencia de hasta el 75% (por ejemplo, informar selectivamente estudios con resultados significativos; John *et al.*, 2012). Asimismo, existen pocos currículum académicos que traten de forma directa qué constituye una PCI y qué alternativas de buenas prácticas científicas existen. Por ello, se describen a continuación qué PCI existen en función del momento en que son realizadas (tabla 1).

PCI del marco teórico e hipótesis

No distinguir entre investigación confirmatoria y exploratoria.

Cada investigación debería especificar si las hipótesis presentadas son confirmatorias (por ejemplo, si el

1. Un excelente texto que recoge en detalle la historia y reacciones de los investigadores implicados en esta polémica se puede encontrar en: <https://slate.com/health-and-science/2017/06/daryl-bem-proved-esp-is-real-showed-science-is-broken.html>.

Tabla 1. Prácticas cuestionables de investigación (PCI) relativas a las distintas fases del proyecto de investigación (VD: Variable Dependiente, VI: Variable Independiente).

Fase	Práctica cuestionable de investigación
Marco Teórico e Hipótesis	No distinguir entre investigación confirmatoria y exploratoria.
	No clarificar la dirección y magnitud esperada de los efectos.
Diseño de Investigación	Recogida y operacionalización múltiple VD y VI sin justificación u hipótesis previa.
	No especificar el plan de muestreo y no realizar análisis de potencia estadística.
	Falta de asignación aleatoria a participantes.
	Corregir, codificar o descartar datos durante la fase de recogida de datos por experimentadores no ciegos.
Análisis de Datos	No especificar el tratamiento de valores perdidos y las técnicas de pre-procesamiento de variables.
	No realizar de pruebas estadísticas para comprobar la presencia de casos atípicos (<i>outliers</i>) y supuestos estadísticos de las pruebas.
	Aumentar el tamaño muestral hasta encontrar un resultado significativo.
	Probar varias alternativas de análisis de datos hasta encontrar una relación significativa para probar una hipótesis (<i>p-hacking</i>).
Presentación de resultados	Presentar resultados exploratorios como confirmatorios (<i>HARKing</i>).
	Interpretar y presentar incorrectamente información relativa a pruebas estadísticas.
	No permitir la reproducibilidad de los resultados.
	No ofrecer código/datos en abierto.

investigador tiene hipótesis direccionales basadas en una teoría previa) o exploratorias (por ejemplo, hipótesis no direccionales). Constituye una PCI el presentar hipótesis exploratorias como confirmatorias tras observar los resultados de los análisis, tanto en las conclusiones (ver *HARKing*) como en la introducción (*SHARKing*; Hollenbeck y Wright, 2017).

No clarificar la dirección y magnitud de los efectos esperados.

Si una hipótesis de investigación tiene un carácter confirmatorio, es necesario distinguir la dirección y magnitud del efecto esperado. Esta elección es clave para los cálculos de potencia estadística en un marco frecuentista (Button, Ioannidis, Mokrysz *et al.*, 2013) o la definición de distribuciones previas en un análisis bayesiano (Wagenmakers, Wetzels, Borsboom y van der Maas, 2012).

PCI del diseño de investigación

Recogida y operacionalización múltiple de variables dependientes e independientes sin justificación u hipótesis previa.

El investigador debería justificar cada medida recogida y utilizada, así como su operacionalización, especificando los niveles de medida escogidos (Steegeen, Tuerlinckx, Gelman y Vanpaemel, 2016). PCI comunes incluyen el seleccionar diversas variables dependientes e independientes o recoger la misma variable de diferentes modos sin una justificación teórica, facilitando la búsqueda a posteriori de resultados significativos (ver *p-hacking*).

No especificar el plan de muestreo y no realizar análisis de potencia estadística.

La representatividad de la muestra debe estar teóricamente y estadísticamente justificada. Dentro de un marco de inferencia frecuentista, el tamaño de la muestra debe ser evaluado mediante análisis de potencia estadística (Button *et al.*, 2013). El fallar en asegurar dicha potencia supone un aumento de los errores tipo I y sesga los resultados, imposibilitando la detección real de la existencia o no de los efectos bajo prueba (Cárdenas y Arancibia, 2014). Este problema es común en Psicología, donde la potencia media de los estudios no solo es muy baja, sino que no ha mejorado en los últimos 50 años (Szucs e Ioannidis, 2017).

Falta de asignación aleatoria a participantes.

Los procedimientos de asignación aleatoria deben ser aplicados en las condiciones experimentales, evitando sesgos de asignación y la actuación de variables extrañas. Si estos procedimientos no son utilizados, el investigador debe explicitar las razones tras esa decisión y qué potenciales variables extrañas que puedan influir en los resultados.

Corregir, codificar o descartar datos durante la fase de recogida de datos por experimentadores no ciegos.

Los investigadores deben evitar el introducir variables que permitan eliminar participantes en base a sus puntuaciones en dicha variable. El uso de estrategias de ciego, no sólo en la asignación de participantes, sino también en el tratamiento y codificación de los datos, evita sesgos en su procesamiento. Estas técnicas incluyen el introducir ruido sistemático en los datos o sustituir la denominación de las variables por un investigador ciego previo a los análisis (Klein y Roodman, 2005).

PCI en el análisis de datos

No especificar el tratamiento de valores perdidos y técnicas de pre-procesamiento de variables.

El no identificar el tipo de valores perdidos existentes en los datos y proponer un tratamiento inadecuado de los

mismos (por ejemplo, eliminar participantes con valores perdidos) puede invalidar las conclusiones obtenidas (Little y Rubin, 2002). Igualmente, cualquier transformación de los datos para poder cumplir con los supuestos de una prueba estadística (por ejemplo, transformación logarítmica) debe estar justificada a priori, pese a que estas prácticas están seriamente cuestionadas en la actualidad (Field y Wilcox, 2017).

No realizar pruebas estadísticas para comprobar la presencia de casos atípicos (outliers) y los supuestos estadísticos de las pruebas.

Cada prueba estadística lleva asociada una serie de supuestos (o análisis de sensibilidad en inferencia bayesiana o análisis de ajuste de modelo) que deben ser cumplidos para que las conclusiones obtenidas puedan ser válidas. Otra PCI es no aportar información de la presencia de *outliers* y casos influyentes en los análisis (Fields y Wilcox, 2017).

Aumentar el tamaño muestral hasta encontrar un resultado significativo.

Esta PCI es conocida como “la regla de parada”: Analizar los datos progresivamente mientras se recoge muestra hasta encontrar un efecto significativo (Rouder, 2014), momento en el que se detiene dicha recogida. Esta práctica incrementa la tasa de error tipo I, ya que altera la potencia de las pruebas estadísticas. Las alternativas bayesianas no sufren esta limitación (Rouder, 2014).

Probar varias alternativas de análisis de datos hasta encontrar una relación significativa para probar una hipótesis (p-hacking).

El *p-hacking* (esto es, la búsqueda de resultados significativos) se basa en la realización de diversos análisis y pruebas estadísticas hasta encontrar un resultado significativo (Head, Holman, Lanfear, Kahn y Jennions, 2015), incluyendo: (a) modificar los participantes incluidos o excluidos en un análisis, (b) modificar las variables introducidas o excluidas en un modelo, (c) cambiar los niveles de las variables involucradas, (d) cambiar el nivel de significatividad (por ejemplo, de 0'05 a 0'01 o viceversa) entre análisis, (e) redondear incorrectamente los valores *p* (por ejemplo, de 0'051 a 0'05), (f) realizar comparaciones múltiples sin aplicar las correcciones apropiadas, (g) transformar los datos e(h) incluir en el estudio solamente aquellos análisis que tengan resultados significativos y positivos para la hipótesis. Esta última PCI, altamente relacionada con los sesgos de publicación, es quizá la más extendida y su prevalencia estimada es del 82% en Psicología (Francis, 2014).

PCI en la presentación de resultados

Presentar resultados exploratorios como confirmatorios (HARKing)

HARKing (del inglés *Hypothesize After Results are Known* o Hipotetizar Después de Conocer los Resultados)

se define como la presentación de una hipótesis que sólo se ha incluido en el artículo después de conocer los resultados de un análisis (Kerr, 1998). Junto al *p-hacking*, esta es la PCI más común (Francis, 2014).

Interpretar y presentar incorrectamente información relativa a pruebas estadísticas.

Una PCI común en Psicología (Lilienfeld, Sauvigné, Lynn, Cautin, Latzman y Waldman, 2015) es la interpretación incorrecta de estadísticos de contraste, valores *p* e intervalos de confianza (Gelman y Stern, 2006; Morey, Hoekstra, Rouder, Lee y Wagenmakers, 2016; Wasserstein y Lazar, 2016). Esta PCI presenta una alta prevalencia, tanto a nivel internacional (Nuijten, Hartgerink, Assen, Epskamp y Wicherts, 2016) como nacional (Pardo, Garrido, Ruiz y San Martín, 2007). Herramientas como *statcheck* (www.statcheck.io) pueden comprobar de forma automática la presencia de algunos de estos errores (Nuijten *et al.*, 2016). La interpretación de los *p*-valores ha recibido una amplia atención, llevando a la *American Statistical Association* a intentar clarificar el significado y las grandes restricciones de su uso como método inferencial (Wasserstein y Lazar, 2016), aunque con limitado éxito (Matthews, Wasserstein y Spiegelhalter, 2017). Ante esta situación, se han propuesto diversas técnicas para mejorar la inferencia clásica, como los test de severidad (Lakens, 2017), el uso de intervalos de confianza (Cumming, 2014) o la adopción de alternativas de análisis bayesianas (Morey, Hoekstra, Rouder, Lee y Wagenmakers, 2016).

No permitir la reproducibilidad de los resultados.

Los autores deben facilitar la reproducción fiel de sus análisis en posteriores estudios de réplica. Si el autor no puede incluir esta información en el manuscrito final, se recomienda proveerla mediante un repositorio público o institucional o justificar la ausencia de la misma (por ejemplo, cuando se aplican limitaciones para proteger la confidencialidad de los participantes).

No ofrecer código o datos en abierto.

Unida a la anterior, esta PCI imposibilita la reproducibilidad o futura replicación de los resultados. Deben proveerse mediante repositorios abiertos los datos y código necesarios para reproducir los análisis presentados. En caso de que los autores no los hicieran públicos, deberían incluirse los motivos en el manuscrito final.

Soluciones a la crisis de la replicabilidad

Para minimizar la incidencia de las PCI, se han propuesto de forma generalizada dos iniciativas complementarias: (1) el pre-registro de estudios científicos y (2) la política de ciencia abierta.

Pre-registro de proyectos científicos

Esta práctica consiste en registrar los protocolos de investigación en documentos donde el investigador refleja todas las decisiones relativas a sus hipótesis, operacionalización de variables, recogida de datos y plan de análisis antes de que la investigación comience. Una vez escritos, los autores envían estos documentos a una revista para que sean revisados por pares, sugiriendo cambios que la calidad científica del proyecto y disminuir los grados de libertad del investigador (Van't Veer y Giner-Sorolla, 2016). Una vez aceptados, la revista adquiere un compromiso de publicación de los resultados. Así, tanto revistas como autores pueden centrarse en asegurar la calidad de las investigaciones, en lugar de buscar resultados significativos (King, Dablander, Jakob *et al.*, 2016).

Para su correcto funcionamiento, es importante asegurar que los pre-registros restrinjan al máximo la posibilidad de cometer PCI. Otra recomendación es asegurar que los pre-registros de estudios de réplica conduzcan a realizar réplicas exactas de los estudios (*versus* réplicas conceptuales), que permitan comparar directamente los estudios originales y de los intentos de réplica sin que otros factores (por ejemplo, manipulaciones de variables diferentes) puedan influir en los resultados (Petty y Cacioppo, 2016).

¿Cómo realizar un pre-registro?

Existen en la actualidad (junio de 2017) al menos 54 revistas en Psicología que cuentan con políticas editoriales favorables con pre-registros (Gonzales y Cunningham, 2015) incluyendo revistas de alto prestigio como *Cortex*, *Behavioural Neuroscience*, *Frontiers in Psychology* o *Psychological Science*. Por otro lado, la *Open Science Framework* (OSF, *Center for Open Science*) es la plataforma libre más común de publicación de pre-registros, que ha desempeñado un papel fundamental en su promoción y aprendizaje y donde los autores pueden encontrar guías y recomendaciones para su redacción.

Ciencia abierta

La ciencia abierta (*Open Science*, en inglés) es una corriente dentro del movimiento por el conocimiento libre que defiende una ciencia transparente y accesible. El conocimiento libre contaría con cuatro tipos de libertades (León Rojas, 2005): (1) libertad de adquisición y uso, (2) libertad de adaptarse libremente a las necesidades del adquiridor, (3) libertad de compartirse con otras personas, y (4) libertad de mejorar dicho conocimiento y compartir las versiones adaptadas y mejoradas con la comunidad.

El proyecto europeo *Facilitate Open Science Training for European Research* (FOSTER) propone una taxonomía de tipos de ciencia abierta con las siguientes libertades de acceso a (1) las contribuciones y productos científicos, (2)

los datos utilizados, (3) los protocolos y procedimientos reproducibles, como las listas de comandos o *scripts*.

Otros componentes de ciencia abierta relativos a factores contextuales incluyen la evaluación por pares abierta (la Iniciativa de Apertura de Revisores por Pares, o *PRO Initiative*), las medidas de impacto alternativas, las directrices TOP para las revistas (*Transparence and Openness Promotion; Center for Open Science*, 2017) o el depósito de manuscritos científicos previos a su publicación, o *pre-prints*. Esta última práctica implica una mayor transparencia, además de una mayor celeridad para registrar oficialmente un trabajo científico. Otra iniciativa interesante es el reconocimiento público de autores que siguen principios de ciencia abierta en sus estudios mediante el uso de insignias (*open science badges*).

Existe asimismo un alto interés político en el fomento de ciencia abierta a nivel institucional. Algunas de las instituciones con mayor impacto social que han optado por estas políticas incluyen la Organización para la Cooperación y el Desarrollo Económico (OECD, 2015), la Organización Mundial de la Salud (Barbour, Chinnock, Cohe y Yamey, 2006), o la Comisión Europea (*European Commission*, 2016).

McKiernan, Bourne, Brown *et al.* (2016) han observado los beneficios de operar bajo protocolos de ciencia abierta: las publicaciones abiertas (1) obtienen más citas; (2) tienen más difusión (doble de menciones en redes sociales, visualizaciones en bases de datos) y (3) tienen incrementalmente mayor factor de impacto (aproximándose al nivel de revistas restringidas de alto impacto).

¿Cómo practicar la ciencia abierta?

Siguiendo la taxonomía del proyecto FOSTER, el acceso libre puede garantizarse publicando las contribuciones científicas en: (a) revistas abiertas (un directorio puede encontrarse en: doaj.org), o (b) revistas de alto impacto que permitan acceso libre (por ejemplo, *PLoS One*). Los datos y protocolos pueden publicarse mediante multitud de herramientas, destacando *Open Science Framework* (OSF) o *GitHub* (github.com). Respecto a los *pre-prints*, en Psicología destaca *PsyArXiv*, integrado en OSF.

La ciencia abierta asume que no siempre será legítimo desarrollar esta práctica investigadora con el máximo nivel de apertura. Su práctica constituye un continuo que recorrería desde el nivel básico de facilitar el libre acceso a las contribuciones tras su publicación, al nivel más avanzado de publicar solicitudes de financiación, protocolos y *scripts* de investigación.

Una solución pedagógica

La crisis de replicabilidad ha evidenciado la necesidad de acometer reformas en la carrera investigadora. Estos cambios deben implementarse en los distintos estadios de la misma, siguiendo una doble estrategia:

1. Por un lado, mostrar los beneficios de utilizar pre-registros y ciencia abierta a los investigadores en Psicología, extendiendo su uso como estándares de investigación. Cabe recalcar la necesidad por parte de las agencias de evaluación de la calidad investigadora de favorecer y valorar dichos esfuerzos. Iniciativas estatales como el portal Recolecta (www.recolecta.fecyt.es) ya facilitan la adopción de estos métodos. En este punto, la implicación de doctorandos y directores de tesis es clave para evitar que la crisis de replicabilidad se extienda. Algunas acciones conllevarían realizar estudios de réplicas como parte integral de las tesis doctorales, pre-registrando este tipo de estudios (Everett y Earp, 2015).
2. Por otro lado, deben promoverse en Grados y Másteres en Psicología el uso de estas herramientas. Diferentes asignaturas pueden informar de cómo la crisis de replicabilidad ha afectado a su dominio de conocimiento, instruyendo además de forma práctica y accesible en el uso de pre-registros y ciencia abierta. La formación deficitaria en metodología sienta la base para la proliferación de las PCI, el uso incorrecto de terminología de investigación (Lilienfeld *et al.*, 2015) y de las técnicas de análisis estadístico (Hardin, Hoerl, Horton *et al.*, 2015).

Experiencias previas han estudiado los efectos de incluir en el currículo de grado proyectos de replicación (Frank y Saxe, 2012; Hawkins, Smith, Au *et al.*, 2018;) y familiarizarse con las plataformas de ciencia abierta (por ejemplo, OSF). El uso de estas estrategias (1) facilitaron la adquisición de conceptos estadísticos de alta relevancia –y tradicionalmente obviados en el currículo académico– como la potencia estadística, alternativas de inferencia (por ejemplo, el análisis bayesiano) o el meta-análisis (Frank, 2016) y (2) promovieron un mayor espíritu crítico respecto a la literatura científica.

En España, la adaptación de los trabajos de fin de Grado y Máster o de trabajos en asignaturas donde se requiera realizar un proyecto de investigación a estudios de replicación con pre-registro se presenta como una alternativa sólida a los trabajos tradicionalmente realizados. Para ello, se propone un marco docente que permita la integración de estas experiencias con los potenciales contenidos de estas asignaturas (tabla 2).

El primer punto sería presentar a los estudiantes la situación actual, los problemas evidenciados por la crisis y las soluciones propuestas. Es necesario que entiendan porqué la replicación y el pre-registro son herramientas fundamentales para el avance científico. Después, se propondrá a los estudiantes un estudio de replicación pre-registrado como alternativa a los trabajos tradicionales de investigación. A la hora de escoger qué replicar, el profesor ha de asegurarse que los estudios originales deben incluirla suficiente infor-

Tabla 2. Desarrollo de proyectos de investigación basados en replicación y pre-registros.

Fase	Contenido
<i>Introducción a la crisis de la replicabilidad</i>	Antecedentes formales Prácticas Cuestionables de Investigación Estudios de replicación y pre-registro
<i>Selección de estudio a replicar</i>	Selección de estudio basado en importancia, viabilidad e información disponible Lectura crítica de materiales relativos a la investigación original Selección de ensayo/experimento a replicar
<i>Preparación documentos pre-registro</i>	Diseño de investigación basado en plantilla de pre-registro Mejoras al diseño original y controles alternativos Revisión del pre-registro Presentación de materiales en repositorios abierto
<i>Realizar la investigación</i>	Adaptación de materiales Recogida de datos Análisis de datos Preparación del manuscrito Actualización de los materiales, código y datos en un repositorio abierto
<i>Revisión del manuscrito</i>	Revisión de la calidad del manuscrito desarrollado Diferencias entre la investigación original, el protocolo de pre-registro diseñado y el manuscrito final.

mación para que la réplica sea válida (Hawkins, Smith, Au *et al.*, 2018). La OSF mantiene un listado de estudios replicados con materiales en abierto en los que el docente puede inspirarse.

La siguiente fase incluiría el desarrollo de protocolos de pre-registros, pudiendo ser utilizado como método de evaluación continua previa a la recogida de datos. Esta fase ilustrará cómo su uso limita los grados de libertad del investigador y previene las PCI. Una vez los protocolos pre-registrados tengan la aprobación del docente, pueden depositarse en un repositorio abierto, acostumbrando a los estudiantes a estas prácticas de ciencia abierta. Aunque la OSF es la elección predilecta, repositorios institucionales o las plataformas de docencia *on-line* (por ejemplo, Moodle) pueden ser sustitutos funcionales si la barrera lingüística supone un problema.

Finalmente, los estudiantes ejecutarían la réplica. Esta parte del proyecto mostrará diferentes formas de evaluar la evidencia científica y cómo las réplicas pueden llevar a conclusiones diferentes del estudio original sin invalidar los resultados de ninguno de los dos. Los resultados finales y manuscritos de investigación pueden ser publicados en plataformas abiertas como *pre-prints* en el repositorio utilizado.

Conclusiones

La crisis de replicabilidad ha provocado una revisión del proceso y la evidencia científica tanto dentro como fuera de la comunidad científica. Diferentes áreas de la Psicología han afrontado un duro periodo donde las teorías científicas que las sustentaban eran puestas en duda. Por otro lado, muchos investigadores han sufrido represalias por exponer estas debilidades. Pero, gracias a sus aportaciones, durante la última década la Psicología ha contribuido a un proceso de evolución de la práctica científica, donde las dudas sobre el modelo de producción científica (por ejemplo, preparación de las investigaciones, criterios de evaluación, o el propio modelo de publicación científica), han generado alternativas para una ciencia más robusta, reproducible y colaborativa.

En España, cada vez más investigadores optan por participar en estudios de replicabilidad o facilitan el acceso a sus publicaciones y recursos de investigación. Sin embargo, la implantación de estas acciones es aún muy limitada. Por ello, es primordial involucrar tanto a investigadores como a estudiantes en los cambios que la Psicología ha de realizar a través de la formación en replicación, pre-registro y ciencia abierta. Sólo así evitaremos que una nueva generación de psicólogos cometa los mismos errores, generando a su vez una visión crítica clave para el futuro de nuestra disciplina.

Referencias

- Agnoli, F., Wicherts, J.M., Veldkamp, C.L.S., Albiero, P. y Cubelli, R. (2017). [Questionable research practices among italian research psychologists](#). *PLoS ONE* 12 (3): e0172792 [DOI: 10.1371/journal.pone.0172792].
- Barbour, V., Chinnock, P., Cohe, B. y Yamey, G. (2006). [The impact of open access upon public health](#). *Bulletin of the World Health Organization*, 84 (5), 339-340.
- Barrada, J.R. (2017). [Grados de libertad del investigador: tamaños del efecto](#). *Un nombre al azar: Universidad, psicología, educación e investigación*, 16 de mayo.
- Bem, D.J. (2011). [Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect](#). *Journal of Personality and Social Psychology*, 100 (3), 407-425 [DOI: 10.1037/a0021524].
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J. y Munafò, M.R. (2013). [Power failure: why small sample size undermines the reliability of neuroscience](#). *Nature Reviews Neuroscience*, 14 (5), 365-376 [DOI: 10.1038/nrn3475].
- Cárdenas, J.M. y Arancibia, H.M. (2014). [Potencia Estadística y Cálculo del Tamaño del Efecto en G*Power: Complementos a las Pruebas de Significación Estadística y su Aplicación en Psicología](#). *Salud y Sociedad. Revista Latinoamericana de Psicología de la Salud y Social*, 5 (2), 201-224 [DOI: 10.22199/S07187475.2014.0002.00006].
- Cumming, G. (2014). [The New Statistics: Why and How](#). *Psychological Science*, 25 (1), 7-29 [DOI: 10.1177/0956797613504966].
- European Commission (2016). *Open Science (Open Access)* <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access>.
- Everett, J.A.C. y Earp, B.D. (2015). [A tragedy of the \(academic\) commons: interpreting the replication crisis in psychology as a social dilemma for early-career researchers](#). *Frontiers in Psychology*, 6, 1152. [DOI: 10.3389/fpsyg.2015.01152].
- Fanelli, D. (2010). ["Positive" Results Increase Down the Hierarchy of the Sciences](#). *PLoS ONE*, 5 (4), e10068 [DOI: 10.1371/journal.pone.0010068].
- Fiedler, K. y Schwarz, N. (2016). [Questionable Research Practices Revisited](#). *Social Psychological and Personality Science*, 7 (1), 45-52 [DOI: 10.1177/1948550615612150].
- Field, A.P. y Wilcox, R.R. (2017). [Robust statistical methods: a primer for clinical psychology and experimental psychopathology researchers](#). *Behaviour Research and Therapy*, 98, 19-38 [DOI.org/10.1016/j.brat.2017.05.013].
- Francis, G. (2014). [The frequency of excess success for articles in Psychological Science](#). *Psychonomic Bulletin & Review*, 21 (5), 1180-1187 [DOI: 10.3758/s13423-014-0601-x].
- Frank, M. C. y Saxe, R. (2012). [Teaching Replication](#). *Perspectives on Psychological Science*, 7 (6), 600-604 [DOI: 10.1177/1745691612460686].
- Gelman, A. (2016). [What has happened down here is the winds have changed](#). *Statistical Modeling, Causal Inference, and Social Science*, 21/09/
- Gelman, A., y Loken, E. (2013). [The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time](#). Department of Statistics, Columbia University.
- Gelman, A. y Loken, E. (2014). [The Statistical Crisis in Science](#). *American Scientist*, 102 (6), 460 [DOI: 10.1511/2014.111.460].
- Gelman, A. y Stern, H. (2006). [The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant](#). *The American Statistician*, 60 (4), 328-331 [DOI: 10.1198/000313006X152649].
- Gonzales, J.E., y Cunningham, C.A. (2015). [The promise of pre-registration in psychological research. Encouraging a priori research and decreasing publication bias](#). *APA Psychological Science Agenda*, agosto de 2015.
- Hardin, J., Hoerl, R., Horton, N.J., Nolan, D., Baumer, B., Hall-Holt, O., Murrell, P., Peng, R., Roback, P., Temple Lang, D. y Ward, M.D. (2015). [Data Science in Statistics Curricula: Preparing Students to "Think](#)

- [with Data](#)". *The American Statistician*, 69 (4), 343-353 [DOI: 10.1080/00031305.2015.1077729].
- Hawkins, R.X.D., Smith, E.N., Au, C., Arias, J.M., Capatano, R., Hermann, E., Keil, M., Lampinen, A., Raposo, S., Reynolds, J., Salehi, S., Salloum, J., Tan, J. y Frank, M.C. (2018). [Improving the Replicability of Psychological Science Through Pedagogy](#). *Advances in Methods and Practices in Psychological Science*, 1 (1), 7-18 [DOI: 10.31234/osf.io/p73he].
- Head, M.L., Holman, L., Lanfear, R., Kahn, A.T. y Jennions, M.D. (2015). [The Extent and Consequences of P-Hacking in Science](#). *PLOS Biology*, 13 (3), e1002106 [DOI: 10.1371/journal.pbio.1002106].
- Hollenbeck, J.R., y Wright, P.M. (2017). [Harking, Sharking, and Tharking: Making the Case for Post Hoc Analysis of Scientific Data](#). *Journal of Management*, 43 (1), 5-18 [DOI: 10.1177/0149206316679487].
- Ioannidis, J. P. A. (2005). [Why Most Published Research Findings Are False](#). *PLOS Medicine*, 2 (8), e124 [DOI: 10.1371/journal.pmed.0020124].
- Ioannidis, J.P.A. (2006). [Journals Should Publish All "Null" Results and Should Sparingly Publish "Positive" Results](#). *Cancer Epidemiology, Biomarkers & Prevention*, 15 (1), 186 [DOI: 10.1158/1055-9965.EPI-05-0921].
- John, L.K., Loewenstein, G. y Prelec, D. (2012). [Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling](#). *Psychological Science*, 23 (5), 524-532 [DOI: 10.1177/0956797611430953].
- Kerr, N.L. (1998). [HARKING: Hypothesizing After the Results are Known](#). *Personality and Social Psychology Review*, 2 (3), 196-217 [DOI: 10.1207/s15327957pspr0203_4].
- King, M., Dablander, F., Jakob, L., Agan, M.L.F., Huberl, F., Haslbeck, J.M.B. y Brecht, K.F. (2016). [Registered Reports for Student Research](#). *Journal of European Psychology Students*, 7 (1), 20-23 [DOI: 10.5334/jeps.401].
- Klein, J.R. y Roodman, A. (2005). [Blind Analysis in Nuclear and Particle Physics](#). *Annual Review of Nuclear and Particle Science*, 55, 141-163 [DOI: 10.1146/annurev.nucl.55.090704.151521].
- Klein, R.A., Ratliff, K.A., Vianello, M., Adams, R.B., Bahník, Š., Bernstein, M. J., ... y Nosek, B.A. (2014). [Investigating Variation in Replicability](#). *Social Psychology*, 45 (3), 142-152 [DOI: 10.1027/1864-9335/a000178].
- Lakens, D. (2017). [Equivalence Tests: A Practical Primer for \$t\$ Tests, Correlations, and Meta-Analyses](#). *Social Psychological and Personality Science*, 8 (4) 355-362 [DOI: 10.1177/1948550617697177].
- León Rojas, J.M. (2005). [Liberalidad del conocimiento desde la cesión de derechos de propiedad intelectual](#). En *Encuentro Internacional sobre Conocimiento Libre. II Conferencia Internacional de Software Libre* (pp. 190-200). Mérida: Junta de Extremadura.
- Lilienfeld, S.O., Sauvigné, K.C., Lynn, S.J., Cautin, R.L., Lutzman, R.D. y Waldman, I.D. (2015). [Fifty psychological and psychiatric terms to avoid: a list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases](#). *Frontiers in Psychology*, 6 : 1100 [DOI : 10.3389/fpsyg.2015.01100].
- Little, R.J.A. y Rubin, D.B. (2002). *Statistical Analysis with Missing Data (2nd Ed.)*. Hoboken, N.J: Wiley-Interscience.
- Matthews, R., Wasserstein, R. y Spiegelhalter, D. (2017). [The ASA's p-value statement, one year on](#). *Significance*, 14 (2), 38-41 [DOI: 10.1111/j.1740-9713.2017.01021.x].
- McKiernan, E.C., Bourne, P.E., Brown, C.T., Buck, S., Kenall, A., Lin, J., ... y Yarkoni, T. (2016). [How open science helps researchers succeed](#). *eLife*, 5: e16800 [DOI: 10.7554/eLife.16800].
- Morey, R.D., Hoekstra, R., Rouder, J.N., Lee, M.D. y Wagenmakers, E.-J. (2016). [The fallacy of placing confidence in confidence intervals](#). *Psychonomic Bulletin & Review*, 23 (1), 103-123 [DOI: 10.3758/s13423-015-0947-8].
- Nuijten, M.B., Hartgerink, C.H.J., van Assen, M.A.L.M., Epskamp, S. y Wicherts, J.M. (2016). [The prevalence of statistical reporting errors in psychology \(1985-2013\)](#). *Behavior Research Methods*, 48 (4), 1205-1226 [DOI: 10.3758/s13428-015-0664-2].
- OECD (2015). [Making Open Science a Reality](#). *OECD Science, Technology and Industry Policy Papers*, 25. París: OECD [DOI: 10.1787/5jr2f963zs1-en].
- Open Science Collaboration (2015). [Estimating the reproducibility of psychological science](#). *Science*, 349 (6251), aac4716 [DOI: 10.1126/science.aac4716].
- Pardo, A., Garrido, J., Ángel Ruiz, M. y San Martín, R. (2007). [La interacción entre factores en el análisis de varianza: errores de interpretación](#). *Psicothema*, 19(2), 343-349.
- Petty, R.E. y Cacioppo, J.T. (2016). [Methodological choices have predictable consequences in replicating studies on motivation to think: Commentary on Ebersole et al. \(2016\)](#). *Journal of Experimental Social Psychology*, 67, 86-87 [DOI: 10.1016/j.jesp.2015.12.007].
- Ranehill, E., Dreber, A., Johannesson, M., Leiber, S., Sul, S. y Weber, R.A. (2015). [Assessing the Robustness of Power Posing: No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women](#). *Psychological Science*, 26 (5), 653-656 [DOI: 10.1177/0956797614553946].
- Rouder, J.N. (2014). [Optional stopping: No problem for Bayesians](#). *Psychonomic Bulletin & Review*, 21 (2), 301-308 [DOI: 10.3758/s13423-014-0595-4].
- Santos, D. (2013). [La crisis en la psicología social contemporánea: el fenómeno del priming](#). *Revista electrónica de Psicología Social "Poiésis"*, 25 [http://www.funlam.edu.co/revistas/index.php/poiesis/article/view/636/605].

- Simmons, J.P., Nelson, L.D. y Simonsohn, U. (2011). [False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant](#). *Psychological Science*, 22 (11), 1359-1366 [DOI: 10.1177/0956797611417632].
- Simmons, J.P. y Simonsohn, U. (2017). [Power Posing: P-Curving the Evidence](#). *Psychological Science*, 28 (5), 687-693 [DOI: 10.1177/0956797616658563].
- Simonsohn, U., Nelson, L.D. y Simmons, J.P. (2014). [P-curve: A key to the file-drawer](#). *Journal of Experimental Psychology: General*, 143 (2), 534-547 [DOI: 10.1037/a0033242].
- Song, F., Hooper, L. y Loke, Y. (2013). [Publication bias: what is it? How do we measure it? How do we avoid it?](#) *Open Access Journal of Clinical Trials*, 5, 71-81 [DOI: 10.2147/OAJCT.S34419].
- Steenen, S., Tuerlinckx, F., Gelman, A. y Vanpaemel, W. (2016). [Increasing Transparency Through a Multiverse Analysis](#). *Perspectives on Psychological Science*, 11 (5), 702-712 [DOI:10.1177/1745691616658637].
- Szucs, D. y Ioannidis, J.P.A. (2017). [Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature](#). *PLOS Biology*, 15 (3), e2000797 [DOI: 10.1371/journal.pbio.2000797].
- Van't Veer, A.E. y Giner-Sorolla, R. (2016). [Pre-Registration in Social Psychology - A Discussion and Suggested Template](#). *Journal of Experimental Social Psychology*, 67, 2-12 [DOI: 10.1016/j.jesp.2016.03.004].
- van Dijk, D., Manor, O. y Carey, L.B. (2014). [Publication metrics and success on the academic job market](#). *Current Biology*, 24 (11), 516-517 [DOI: 10.1016/j.cub.2014.04.039].
- Wagenmakers, E., Wetzels, R., Borsboom, D. y van der Maas, H.L.J. (2011). [Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem \(2011\)](#). *Journal of Personality and Social Psychology*, 100 (3), 426-432 [DOI: 10.1037/a0022790].
- Wasserstein, R.L. y Lazar, N.A. (2016). [The ASA's Statement on p-Values: Context, Process, and Purpose](#). *The American Statistician*, 70 (2), 129-133 [DOI: 10.1080/00031305.2016.1154108].
- Wicherts, J.M., Veldkamp, C.L.S., Augusteijn, H.E.M., Bakker, M., van Aert, R.C.M., y van Assen, M.A.L.M. (2016). [Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking](#). *Frontiers in Psychology*, 7, 1832 [DOI:10.3389/fpsyg.2016.01832].

