

Falacias estadísticas en investigación psicológica y cómo enseñar la sabiduría podría ayudar

Statistical Fallacies in Psychological Research and how teaching wisdom may help

Carlos Ruiz-Matuk^a ORCID: 0000-0003-2681-4953

Recibido: 8/02/2023 • Aprobado: 25/04/2023

Cómo citar: Ruiz-Matuk, C. (2023). Falacias estadísticas en investigación psicológica y cómo enseñar la sabiduría podría ayudar. *Ciencia y Educación*, 7(2), 89–102. <https://doi.org/10.22206/cyed.2023.v7i2.pp89-102>

Resumen

El propósito principal de este artículo es defender la posición de que algunas falacias alrededor de la aplicación de la estrategia *frecuentista*, representada por la tradición de la prueba estadística de hipótesis nula, han sido confundidas con las soluciones sustantivas en la investigación. Por otro lado, se intenta ilustrar cómo esta confusión incide en la crisis de la replicación (Costello & Watts, 2022) y, además, exponer cómo un investigador, dada sus propias limitaciones, debe enfrentar los retos que representan todas las estrategias estadísticas que están a su alcance. Se advierte sobre la alta probabilidad de que estas estrategias estén sujetas a la avaricia cognoscitiva que caracteriza a los seres humanos. Finalmente, la enseñanza de la sabiduría se ofrece como una alternativa para reducir dichos sesgos.

Palabras clave: análisis estadísticos; falacias lógicas; sabiduría; enseñanza; psicología.

Abstract

The main purpose of this article is to defend the position that some fallacies surrounding the application of the frequentist strategy, represented by the tradition of statistical test of null hypothesis, have been confused with substantive solutions in research. On the other hand, we try to illustrate how this confusion affects the replication crisis (Costello & Watts, 2022) and also to expose how a researcher, given his own limitations, must face the challenges represented by all the statistical strategies that are within his reach. It warns about the high probability that these strategies are subject to the cognitive miserliness that characterizes human beings. Finally, the teaching of wisdom is offered as an alternative to reduce such biases.

Keywords: Statistical analysis; Logical Fallacies, Wisdom; Teaching; Psychology.

^a Universidad Iberoamericana, República Dominicana. Correo-e: c.ruiz4@unibe.edu.do



Introducción

Muchos críticos señalan que los psicólogos son notoriamente deficientes en el uso de las estadísticas, debido en parte a que nunca se les enseñó de manera adecuada los análisis estadísticos y los utilizan como una forma de inferencia automática basada en *rituales* (Gigerenzer, 2004; 2018). Aún más, otros han argumentado que lo que se asume en la investigación psicológica como práctica de la inferencia estadística está infestada de problemas en sí misma, en especial se enfatiza la presencia de falacias alrededor de la aplicación de la estrategia *frecuentista* representada por la tradición de la prueba estadística de hipótesis nula (Borg et al., 2023; Eden & Inan, 2022; Filipini & Vinceti, 2022; Hassler, 2022; Madjarova et al., 2022; Mayo & Hand, 2022; Oberoi & Atri, 2021; Rogers, 2022).

Objetivo

En este artículo se plantea que, frente a las malas prácticas señaladas en el uso de las estadísticas, la estrategia a utilizar por los psicólogos debe responder a la pregunta de investigación, sin importar la tradición escogida, y que, además de evitar los ritos, se deben enfrentar las limitaciones que representan cada una de estas estrategias, dadas las mismas restricciones que las capacidades de procesamiento de la información permita. En tal sentido, aunque algunas de estas estrategias estadísticas han sido asociadas a la incurrancia de falacias, ellas no deben interpretarse como repudiables. Una alternativa antídoto a la reincidencia de estas falacias sería la enseñanza de sabiduría en los planes de estudio de las carreras que dentro de su entrenamiento en la investigación incluyan la estadística.

La tradición frecuentista y las falacias

La tradición frecuentista de la prueba de significación de la hipótesis nula (NHST, por sus siglas en inglés para *Null Hypothesis Significance Testing*) ha sido el resultado de un matrimonio de conveniencia

(Rodríguez, 2016) del enfoque de Fisher y la teoría de decisión de Neyman-Pearson. Debido a que en los libros de texto de estadística se asume como un procedimiento formal para el manejo de los errores en la toma de decisiones a partir de los resultados de una investigación, se ignora que es un híbrido entre estos dos enfoques (el enfoque de Fisher y el de Neyman-Pearson), considerados por algunos como esencialmente contrarios (Gigerenzer, 2004). En primer lugar, Fisher solo argumentó sobre la existencia del error Tipo I, es decir, rechazar la hipótesis nula cuando esta es verdadera. Pearson y Neyman introducen el concepto de error Tipo II, que consiste en aceptar la hipótesis nula cuando esta es falsa, y a partir de ello, el concepto de potencia estadística (que no es más que la probabilidad de evitar el error Tipo II).

Ha existido un gran debate por muchos años relacionado con lo que se ha denominado la *crisis de la significación estadística* (Carver, 1978; Cohen, 1994; 1995; Thompson, 2004). Por ejemplo, Kline (2004; 2013) ha identificado cinco falacias sobre los valores de p ("The Big Five": Las Cinco Grandes) y otras doce falacias relacionadas con la toma de decisión de la significación estadística.

Falacias sobre los valores de p

Debe entenderse que la abreviación p se refiere a la probabilidad condicional de la frecuencia relativa de los resultados. Es decir, es la probabilidad de resultados específicos o más extremos, dado que la hipótesis nula es verdadera, la muestra es aleatoria y otras suposiciones se cumplen. Aquí es útil adoptar la perspectiva frecuentista en la que la probabilidad se ve como la probabilidad de un resultado sobre eventos repetibles en condiciones constantes (control de la varianza sistemática), exceptuando la casualidad (error de muestreo). Desde este punto de vista, la probabilidad no se aplica directamente a un evento único o discreto, sino que esta probabilidad se basa en la frecuencia relativa esperada en un gran número de ensayos, a largo plazo. Esto, en primer lugar, parece

ser contraintuitivo, quizá por la exigencia de permanecer en un estado de incertidumbre aun cuando ya se ha tomado una decisión. De aquí, quizás surgen las falacias en el razonamiento estadístico.

Las cinco falacias (o los Cinco Grandes) sobre el concepto de p (Kline, 2004; 2011; 2013; Tabri & Elliott, 2014), son las siguientes:

- *La falacia de las probabilidades contra el azar*, o la falsa creencia de que p indica la probabilidad de que un resultado particular ocurrió por casualidad (es decir, debido a un error de muestreo). Se debe entender que p se calcula para un rango de resultados, la mayoría no observados, y no para un solo resultado. Además, p se calcula asumiendo que la hipótesis nula (H_0) ya es verdadera, por lo que la probabilidad de que el error de muestreo sea la única explicación que ya se considera sería 1.0 (es decir, perfecta). Por lo tanto, es una conclusión inválida asumir que p es la medición de la probabilidad de error de muestreo. Este error es cometido por autores que incluso critican el uso de p para la toma de decisiones en la investigación (Sullivan & Feinn, 2012) y ha sido encontrado en un 89 % de una muestra de 30 libros de introducción a la psicología (Cassidy et al., 2019).
- *La falacia del error local Tipo I* para el caso, por ejemplo, en que $p < .05$ y $\alpha = .05$ (es decir, H_0 es rechazada), se dice que la probabilidad de que la decisión de rechazar la hipótesis nula sea un error de Tipo I, es inferior al 5 %. Esta creencia es falsa porque cualquier decisión particular de rechazar H_0 es correcta o incorrecta, por lo que no se asocia a ninguna probabilidad del error (que no sea 0 o 1.0). Solo con una replicación suficiente se podría determinar si la decisión de rechazar H_0 en un estudio en particular fue correcta o no. Esta falacia (Shine, 1980) puede estar detrás de las decisiones incorrectas de muchos investigadores al interpretar una p (probabilidad a posteriori) como el α (probabilidad a priori).
- *La falacia de la probabilidad inversa* es la falsa creencia de que p es la probabilidad de que la hipótesis nula sea verdadera. Es la creencia de que la hipótesis nula es verdadera o falsa dados los datos obtenidos [$p (H_0 | D)$]. Este error se deriva de olvidar que los valores de p son probabilidades de datos bajo la consideración de que la hipótesis nula es verdadera, y no al revés. Falk y Greenbaum (1995) se han referido a esta forma de razonamiento deductivo como la *ilusión de la prueba probabilística por contradicción* y luego fue conocida como la *falacia del condicional transpuesto* (Ziliak & McCloskey, 2008). Algunos autores han sugerido que este error puede ser superado utilizando intervenciones educativas (Cumming, 2013; Kalinowski et al., 2008).
- *La falacia de la hipótesis de investigación válida* (Carver, 1978) es la falsa creencia de que $1-p$ es la probabilidad de que la hipótesis alternativa sea verdadera. La cantidad $1-p$ es una probabilidad, pero es solo la probabilidad de obtener un resultado aún menos extremo bajo H_0 que el realmente encontrado.
- *La falacia de replicación* se refiere a la creencia de que el nivel de significación es el complemento ($1-p$) de la probabilidad de replicación del estudio. Por ejemplo, si los resultados de un estudio arrojan una p de .05, existe .95 de probabilidad de que en los siguientes estudios se encuentre los mismos resultados. Se ha sugerido que una buena explicación de la “crisis de replicación” en la que se encuentra todavía se debe al uso de esta ilusión (Gigerenzer, 2018).
- De estas cinco falacias sobre el concepto de p , algunos autores plantean que las más importantes y comunes son la falacia de replicación y la falacia de la probabilidad inversa (Maxwell et al., 2015; 2018). Como Kline (2004; 2013) ha insinuado, las probabilidades (mal percibidas) a favor de la hipótesis del investigador son tan atractivas

que lo único que falta es concluir que los resultados también deben ser importantes, porque son estadísticamente significativos.

Falacias al tomar decisiones sobre la hipótesis nula (H_0)

Los Cinco Grandes no son el final de las distorsiones cognitivas en las pruebas de significación. También, es posible llegar a varias conclusiones falsas después de decidir rechazar o no rechazar H_0 . Kline (2013) advierte sobre 12 falacias más al tomar decisiones acerca de la hipótesis nula:

- *La falacia de la magnitud*, o la falsa creencia de que los valores *bajos de p* indican grandes efectos. Este error está intrínsecamente relacionado con la ignorancia de la relación entre tamaño del efecto y el de la muestra con los valores de p . Si se tiene una muestra grande, efectos muy pequeños pueden resultar significativos con valores de p muy bajos. Este error se ha identificado en estudiantes de psicología (Kühberger et al., 2015), pero no es exclusivo en esta población ni en esta área de estudio (Hentschke & Stüttgen, 2011).
- *La falacia de significación* se expresa en la creencia del investigador de que el rechazo de la hipótesis nula (H_0) confirma la hipótesis alterna (H_1). Esta falsa idea en realidad se basa en dos errores cognitivos. En primer lugar, la decisión de rechazar la H_0 en un solo estudio no implica que se haya probado la H_1 . En segundo lugar, incluso si la hipótesis estadística H_1 es correcta, ello no significa que la hipótesis sustantiva detrás de H_1 sea correcta. Detrás de este error está la falacia del razonamiento silogístico conocido como la *afirmación del consecuente* (Popper, 1959). Algunos estudios han encontrado la manifestación de esta falacia, no solo en académicos de la psicología (Badenes-Ribera et al., 2015; 2016), sino también en ciencias de la salud (Silva-Ayçaguer et al., 2010).
- *La falacia de causalidad* por su parte es precisamente esa tendencia a asumir que la significación estadística consiste en el descubrimiento de un mecanismo causal encubierto. Además de establecer la línea temporal entre los eventos estudiados, el investigador tiene que descartar aquellas hipótesis alternativas que pueden también explicar los resultados. Las hipótesis estadísticas frente a las sustantivas no solo difieren en sus niveles de abstracción, sino que también tienen diferentes implicaciones tras el rechazo de H_0 . Si H_0 y H_1 reflejan meramente la estadística, hay poco que hacer después de rechazar H_0 excepto la replicación. Pero si H_1 refleja una hipótesis científica, el trabajo comienza justo después de rechazar H_0 . Siguiendo en cierto modo el principio Duhem-Quine, parte de la tarea consiste en evaluar hipótesis sustantivas en competencia que también son compatibles con la hipótesis estadística H_1 . Si no se pueden descartar explicaciones alternativas, la confianza en la hipótesis original debe moderarse.
- *La falacia cero y la falacia de equivalencia*, siendo la primera como la falsa creencia de que no rechazar la hipótesis nula, significa que el tamaño del efecto es cero. Aunque puede ser verdad, esto no se puede decir a partir del resultado en una muestra, especialmente si esta muestra no garantiza una buena potencia. Lo único que puede inferirse es que la decisión de no rechazar una hipótesis nula sería un error de Tipo II. Un error Tipo II, puede ser el resultado de una muestra pequeña, un mal diseño, procedimientos no estandarizados, así como usar medidas poco confiables. La falacia de equivalencia ocurre cuando no rechazar $H_0: \mu_1 = \mu_2$ se interpreta como diciendo que las poblaciones de las cuales se supone se extrajeron las muestras, son equivalentes. Esto es incorrecto porque incluso si las medias de las poblaciones son iguales, las distribuciones pueden diferir de otras maneras, como la variabilidad o la forma de la distribución.

- La *falacia de la calidad y falacia del éxito* no son más que las creencias de que obtener significación estadística confirma la calidad del diseño experimental además de indicar que el estudio ha sido exitoso. En realidad, un diseño pobre puede conducir al error Tipo I, y no conseguir significación estadística puede ser el resultado de la práctica correcta de la ciencia.
- De la falacia del éxito se puede intuir la *falacia del fracaso*. Esta falacia es la creencia errónea de que no conseguir resultados estadísticamente significativos es un fracaso. Esto deja de lado el hecho de que la hipótesis substantiva puede ser falsa *per se* y que encontrar datos que no la confirman no necesariamente resultan de una baja potencia o métodos inapropiados.
- La *falacia de la reificación* es la creencia errónea de que el resultado de un estudio no se considera replicado si se rechaza H_0 en el primer estudio, pero no en el segundo. Existe evidencia de replicación aun cuando los estudios rechacen o no la hipótesis nula. El error está en no considerar la potencia, el tamaño de la muestra, y el tamaño del efecto en los diferentes estudios.
- La *falacia de la objetividad* es el mito detrás de considerar que las pruebas de significación son el único método objetivo de prueba de hipótesis. Tampoco es el único marco para probar hipótesis. Por ejemplo, la estimación bayesiana es una alternativa a las pruebas de significación.
- La *falacia de la santificación* se refiere al pensamiento dicotómico sobre los valores p continuos. Si $\alpha = .05$, por ejemplo, $p = .049$ versus $p = .051$ son prácticamente idénticos en términos de resultados de la prueba. Sin embargo, un investigador puede hacer un gran problema con el primero, pero ignorar el segundo. También hay evidencia del *efecto acantilado*, que se refiere a una disminución abrupta en el grado de confianza de que existe un efecto para una p justo más alta que .05 (Nelson et al., 1986).
- La *falacia de la robustez* es la creencia de que las pruebas estadísticas paramétricas clásicas son robustas contra valores extremos o violaciones de supuestos sobre las distribuciones de las que se extraen las muestras de estudio, aun cuando estas son pequeñas y no probabilísticas. Este es el segundo mito presentado por Hunter y May (1993) sobre el uso de pruebas de hipótesis paramétricas y no paramétricas, induciendo a los investigadores a usar las estadísticas paramétricas. El hecho de que la mayoría de los artículos no asegure a los lectores que los resultados son confiables, desatendiendo las recomendaciones (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008; Appelbaum et al., 2018; Cooper, 2020) es parte de la crisis de replicación de la que hablaremos más adelante.

Es importante recalcar que a pesar de las grandes críticas de Kline (2004; 2013) a los procedimientos reconocidos como propios del ritual NHST, sus recomendaciones incluyen la aplicación de los mismos en investigaciones exploratorias, usando información sobre la potencia, si la hipótesis nula es justificable y esta se auxilia de otros resultados. Esto último ha sido apoyado por muchos autores e instituciones de la psicología (Aiken et al., 2008; APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008; Appelbaum et al., 2018; Maxwell et al., 2015; 2018; Washburn, Carsel et al., 2018; Washburn, Hanson et al., 2018; Wilkinson & Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1999).

La crisis de la replicación y la estadística

Si se quiere reconocer la verdadera relación entre el valor de p y la probabilidad de una replicación exitosa y no cometer la falacia de la replicación,

sería útil recordar los resultados impactantes del reporte en *Science* sobre las replicaciones de 100 estudios de psicología en tres revistas científicas importantes (Open Science Collaboration, 2015). El indicador de la crisis es que hay una proporción muy baja; tan solo 36 % de las investigaciones con resultados en verdad significativos respecto de las que originalmente se catalogaron como tales. Algunos autores han sostenido que si Fisher tuviera la oportunidad de evaluar la supuesta crisis, probablemente argumentaría que la ciencia procede, como debe ser, en una forma de autocorrección (Shrout & Rodgers, 2018; Tackett et al., 2017). Si los resultados se replicaran, y no se comete la falacia del éxito, también deberían someterse a escrutinio. Todas las formas de replicación son informativas, tanto si se replican efectos estadísticamente significativos como no significativos.

La respuesta de la estadística a los problemas relacionados con la crisis de replicación puede abarcar cuatro aspectos (Shrout & Rodgers, 2018):

1. El uso de análisis de potencia refinados para los estudios de replicación.
2. El rol del metaanálisis para entender la variación de la replicación.
3. La promesa del análisis bayesiano para entender la variación en la replicación.
4. Los métodos de remuestreo.

Bajo la influencia de Cohen (1988), los psicólogos se han dado cuenta de la necesidad de considerar la potencia en el diseño de sus estudios, antes de recopilar datos. Debido a que los cálculos de potencia son más relevantes al realizarse antes de recopilar y examinar los datos, debe mostrarse cómo las estimaciones del *tamaño del efecto* se derivaron de investigaciones y teorías previas, esto puede evitar la invención de datos o plagios (Wilkinson & Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1999).

Una alternativa recomendada (Maxwell et al., 2015) ha sido el que los psicólogos no se enfoquen más en réplicas únicas, y en su lugar utilicen estudios múltiples para obtener un rango de nuevos tamaños de efectos. De esta manera, al completarse múltiples estudios y ser reportados, el foco puede girarse hacia la estimación de las distribuciones de tamaños de efectos en lugar de decisiones binarias (Schauer, 2022). Esto lleva a considerar la importancia del metaanálisis, el cual no solo permite esta estimación del promedio de los efectos y su variación, sino que también ayuda en la exploración sustantiva de los factores que moderan estos efectos.

Adicionalmente, está la promesa del análisis bayesiano. Fisher rechazaba al principio la perspectiva bayesiana, pero luego apreció más la teoría de la probabilidad inversa. La probabilidad inversa se refiere a la evaluación de la probabilidad de una hipótesis, dados los datos. En contraste, los procedimientos frecuentistas NHST evalúan la probabilidad de los datos, dada la hipótesis nula. En un nivel conceptual, un bayesiano combina el conocimiento previo (una distribución previa) con los datos que llegan por el conocimiento revisado (una distribución posterior). Específicamente, un bayesiano toma los resultados anteriores, los combina con uno o más esfuerzos de replicar esos resultados, y llega a una opinión actualizada (Austin et al., 2002; Kruschke & Liddell, 2018; Ortega & Navarrete, 2017; Rice & Krakauer, 2022). Esto permite darle plausibilidad a la hipótesis nula. Mientras que, en la postura frecuentista, la hipótesis nula es rechazada o no, en los métodos bayesianos se apoya una aproximación más formal de la evaluación directa de dicha hipótesis. Wagenmakers et al. (2018) dan una explicación bastante clara de la cuantificación de la evidencia para la ausencia de un efecto relativo a uno hipotetizado usando el factor de Bayes.

Por último, los métodos de *remuestreo* prueban hipótesis usando los datos para construir distribuciones muestrales empíricas, en lugar de distribuciones mues-

trales teóricas (Shrout & Rodgers, 2018). El objetivo del remuestreo es tomar una decisión inferencial, igual que el de una prueba estadística paramétrica cualquiera. La diferencia está en cómo se logra el objetivo (Beasley & Rodgers, 2009). Estos métodos, así como los de *validación cruzada*, siguen la perspectiva de los estudios de replicación. Por ejemplo, los métodos de validación cruzada usan muestras múltiples dentro de un estudio singular; la primera muestra (la muestra de estimación o entrenamiento) es usada para identificar modelos en un contexto aproximadamente exploratorio, mientras que una segunda muestra (la muestra de calibración) es usada para investigar en un sentido confirmatorio la hipótesis.

Los problemas en las nuevas aproximaciones

Los métodos complementarios que han surgido no son necesariamente sustitutos de la posición frecuentista, ni tampoco están exentos de problemas en sus interpretaciones.

En primer lugar, se descuida la posibilidad de varios sesgos al hacer análisis de potencia tomando el tamaño del efecto como un valor fijo y conocido. Por ejemplo, el sesgo de publicación y la variación de muestreo. La tendencia a publicar resultados significativos propicia un sesgo en la publicación de los efectos, y por su lado, el impacto del tamaño del efecto sobre la potencia no es lineal (Head et al., 2015). La gran prevalencia del *sesgo de selección* (como ha sido denominado) subraya la importancia de adoptar mejores prácticas de investigación en la búsqueda de una mayor transparencia, preinscripción y registros de los informes de investigación, en especial para los economistas y psicólogos. Aunque no están exentos, por ejemplo, los estudios de las ciencias de la medicina (Bartoš et al., 2022).

Por otro lado, los procedimientos bayesianos requieren una idea problemática para el cerebro humano: la probabilidad de una hipótesis (también denominada *probabilidad de evento único*). Se han descrito varios

conceptos cognitivos estrechamente relacionados con las probabilidades de un solo evento, y su capacidad de intercambio en la mente humana da como resultado “ilusiones cognitivas”, aparentes déficits en el razonamiento sobre la incertidumbre (Anderson, 1998). Las aproximaciones convencionales del estudio del razonamiento bayesiano en psicología no han estado exentas de críticas (Mandel, 2014). Además, se ha señalado la dificultad que posee el ser humano para interpretar el teorema de Bayes si no es bajo una explicación de frecuencias (Gigerenzer, 1994).

Por su lado, los métodos de remuestreo usualmente consideran el contexto de investigación en el que se recolectan los datos para ser fijados, y así, la fuente de variación es ignorada (Shrout & Rodgers, 2018). Y al igual que todas las estadísticas que estiman los parámetros de la población, si el proceso de muestreo es defectuoso, como datos faltantes o valores atípicos, o una muestra que no es representativa de la población, la inferencia se verá comprometida (Beasley & Rodgers, 2009).

En definitiva, las interpretaciones de la estadística, sin importar su enfoque, al parecer, no están exentas de los errores de razonamiento del ser humano y su tendencia a la avaricia cognoscitiva (Stanovich, 2018b; Toplak et al., 2014). La literatura acerca de las posibilidades de reducir los sesgos cognoscitivos (o *diabiasing*) ha sido desalentadora (Kahneman, 2011). Dichos esfuerzos por reducirlos a través de la enseñanza del pensamiento crítico han sido, al menos, muy dependientes del dominio que se enseña (Willingham, 2008), aun cuando estos esfuerzos son considerados entre los legados potencialmente más perdurables de la psicología para la promoción del bienestar humano (Lilienfeld et al., 2009).

Pero siguen apareciendo nuevas propuestas para reducir nuestros sesgos cognoscitivos. Por ejemplo, ha surgido una que se ha enfocado en el *ruido* (Kahneman et al., 2021). El concepto de ruido no es más que la variabilidad no deseada en el juicio

humano. Aunque estos autores siguen asumiendo la importancia del sesgo cognoscitivo en la toma de decisiones, esta perspectiva contrasta con el enfoque de la psicología individual. Es decir, se enfatiza la función grupal bajo el uso de juicios promedios que permitan identificar a las personas mejores predictoras, y así mismo se incluye el uso de guías explícitas que parecen anular el juicio individual. Esta modalidad ha sido bastante fructífera en términos de la discusión intelectual que ha provocado, pero ha sido considerada algo controversial (Kahneman et al., 2022).

Una alternativa que ha surgido en los últimos 30 años es la búsqueda de la sabiduría (Sternberg, 2001; 2003; Sternberg & Hagen, 2019). El concepto de sabiduría, que parece ser consensuado, incluye humildad intelectual, reconocimiento de diferentes puntos de vista, apreciación del contexto dentro del cual se desarrolla el problema, sensibilidad a la posibilidad de cambio en las relaciones sociales, reconocimiento de la probabilidad de múltiples resultados de un conflicto, un punto de vista sobre el problema que trasciende el yo y preferencia por el compromiso en la resolución de puntos de vista opuestos (Baltes & Smith, 2008; Grossmann, 2017a; 2017b; Santos et al., 2017; Vervaeke & Ferraro, 2013).

Aunque los psicólogos se han interesado recientemente en el método empírico de investigar la naturaleza y función de la sabiduría, los filósofos han examinado durante mucho tiempo la idea (Baltes & Smith, 2008). Comprender las formas más efectivas de descubrir, evaluar y hacer uso del conocimiento adquirido depende de encontrar y analizar formas confiables de medir los parámetros de sabiduría (Brienza et al., 2018; Grossmann, Weststrate, Ardel et al., 2020; Grossmann, Weststrate, Ferrari et al., 2020).

En la “Ética a Nicómaco”, Aristóteles recomienda la adquisición de la *phronesis* para alcanzar una buena vida. Afirma que esta nos dará la capacidad de pensar de manera más creativa y crítica, lo que mejorará

nuestra capacidad de actuar con eficacia. Además, explica que el hombre de la *phronesis* sabe mejor que nadie lo que es mejor hacer y lo que es conveniente evitar bajo cualquier conjunto dado de condiciones. A la luz del conocimiento limitado, esta cualidad implica juicio, toma de decisiones y elección de un curso de acción. Por lo tanto, se introduce la idea de la sabiduría práctica.

La sabiduría práctica incluye recurrir a múltiples fuentes de conocimiento y perspectivas, permanecer receptivo a la retroalimentación, ser transparente en la comunicación y comprender la interconexión de los problemas morales y sociales para alcanzar soluciones aceptables. También requiere una conciencia de los propios prejuicios y limitaciones, al tiempo que reconoce la complejidad humana. De esta manera, la sabiduría práctica permite a las personas interactuar responsablemente con el mundo y crear acciones significativas en la vida (Brienza et al., 2018; Grossmann, Weststrate, Ardel et al., 2020; Grossmann, Weststrate, Ferrari et al., 2020).

Un investigador sabio puede aceptar entonces la incertidumbre detrás de las relaciones probabilísticas de los datos porque reconoce la probabilidad de múltiples resultados y prefiere comprometerse en la resolución de puntos de vista opuestos o pensamiento reflexivo. En tal sentido, se ha encontrado una asociación negativa entre la tendencia a recurrir a sesgos con la humildad intelectual (Bowes et al., 2021), una facultad propia de la sabiduría.

Entonces, indudablemente, se necesita de sabiduría a la hora de llegar a conclusiones sobre los resultados de los análisis de nuestras investigaciones. En especial, se requiere del componente sobre el pensamiento reflexivo y la humildad intelectual (Grossmann et al., 2016; Grossmann & Dorfman, 2019; Hammond, 2007; Pratiwi & Ariawan, 2020). Es posible que se puedan superar muchas falacias en la medida en que se aplica el pensamiento crítico (West et al., 2008), aunque podría necesitarse cierta capacidad cognoscitiva (West & Stanovich, 2003).

Una información que puede ser desalentadora es que la humildad intelectual, ese aspecto propio de la sabiduría que puede ayudar a reducir los sesgos en la interpretación de los resultados de la investigación, aparenta ser poco común en la comunidad científica, probablemente debido a su independencia de las capacidades cognitivas (Stanovich et al., 2013; Stanovich & West, 2007; 2008a; 2008b). Sin embargo, no todo está perdido, existe luz al final del túnel. La sabiduría puede ser enseñada, aprendida o desarrollada (Alhosseini & Ferrari, 2019; Huynh & Grossmann, 2020; Stanovich, 2001; 2018a; Sternberg, 2001; 2003; Sternberg & Hagen, 2019). Al parecer, solo se necesita aprender el distanciamiento del ego (*ego-decentering*) utilizando ileísmo (hablar en tercera persona de uno mismo) para que el razonamiento sea más sabio (Grossmann, 2017a; Grossmann et al., 2020; Grossmann & Dorfman, 2019; Kross & Grossmann, 2012).

Conclusión

En este artículo se han identificado algunas de las falacias que el investigador debe evitar a la hora de realizar e interpretar los resultados de un estudio. Especialmente, se han señalado los sesgos que se pueden presentar al aplicar la estadística llamada frecuentista que utiliza el denominado modelo de la prueba de hipótesis nula. Estas falacias están detrás de la crisis de replicación que ha surgido en los últimos años. Las propuestas que han surgido en los últimos tiempos han ayudado un poco a desenmarañar las problemáticas de la estrategia Fisheriana, pero no han resuelto varios de los problemas que surgieron.

Así, se propuso que en cualquier perspectiva asumida deben tomarse en consideración las limitaciones del procesamiento de la información y la tendencia del ser humano a la avaricia cognoscitiva. Enseñar la sabiduría al interesado en la carrera de investigación es una alternativa que puede dar resultados positivos. Parece ser ambicioso, pero si dicha enseñanza se enfoca en el distanciamiento del

yo (*ego-decentering*) podría resultar en un investigador menos vulnerable a las falacias y demandas de cualquier estrategia empleada en el análisis estadístico.

Referencias

- Aiken, L. S., West, S. G. & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist*, 63(1), 32-50. <https://doi.org/10.1037/0003-066x.63.1.32>
- Alhosseini, F. & Ferrari, M. (2019). Effects of Causal Attribution and Implicit Mind-Set on Wisdom Development. *International Journal of Aging & Human Development*, 90(4), 319-336. <https://doi.org/10.1177/0091415019836098>
- Anderson, J. L. (1998). Embracing Uncertainty: The Interface of Bayesian Statistics and Cognitive Psychology. *Conservation Ecology*, 2(1), 2. <https://doi.org/10.5751/es-00043-020102>
- APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? [Review of Reporting standards for research in psychology: Why do we need them? What might they be? by A. P. and C. B. W. G. on J. A. R. Standards]. *American Psychologist*, 63(9), 839-851. <https://doi.org/10.1037/0003-066x.63.9.839>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M. & Rao, S. M. (2018). Journal Article Reporting Standards for Quantitative Research in Psychology: The APA Publications and Communications Board Task Force Report. *American Psychologist*, 73(1), 3-25. <https://doi.org/10.1037/amp0000389>

- Austin, P. C., Brunner, L. J. & Hux, J. E. (2002). Bayeswatch: an overview of Bayesian statistics: An introduction to Bayesian statistics. *Journal of Evaluation in Clinical Practice*, 8(2), 277-286. <https://doi.org/10.1046/j.1365-2753.2002.00338.x>
- Badenes-Ribera, L., Frias-Navarro, D., Iotti, B., Bonilla-Campos, A. & Longobardi, C. (2016). Misconceptions of the p-value among Chilean and Italian Academic Psychologists. *Frontiers in Psychology*, 7, 1247. <https://doi.org/10.3389/fpsyg.2016.01247>
- Badenes-Ribera, L., Frías-Navarro, D., Monterde-I-Bort, H. & Pascual-Soler, M. (2015). Interpretation of the p value: A national survey study in academic psychologists from Spain. *Psicothema*, 27(3), 290-295. <https://doi.org/10.7334/psicothema2014.283>
- Baltes, P. B. & Smith, J. (2008). The Fascination of Wisdom: Its Nature, Ontogeny, and Function. *Perspectives on Psychological Science*, 3(1), 56-64. <https://doi.org/10.1111/j.1745-6916.2008.00062.x>
- Bartoš, F., Maier, M., Wagenmakers, E.-J., Nippold, F., Doucouliagos, H., Ioannidis, J. P. A., Otte, W. M., Sladekova, M., Fanelli, D. & Stanley, T. D. (2022). *Footprint of publication selection bias on meta-analyses in medicine, economics, and psychology*. <http://arxiv.org/abs/2208.12334>
- Beasley, W. H. & Rodgers, J. (2009). Resampling methods. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 362-386). Sage.
- Borg, D. N., Barnett, A. G., Caldwell, A. R., White, N. M. & Stewart, I. B. (2023). The bias for statistical significance in sport and exercise medicine. *Journal of Science and Medicine in Sport*, 26(3), 164-168. <https://doi.org/10.1016/j.jsams.2023.03.002>
- Bowes, S. M., Costello, T. H., Lee, C., McElroy-Heltzel, S., Davis, D. E. & Lilienfeld, S. O. (2021). Stepping Outside the Echo Chamber: Is Intellectual Humility Associated With Less Political Myside Bias? *Personality and Social Psychology Bulletin*, 48(1), 150-164. <https://doi.org/10.1177/0146167221997619>
- Brienza, J. P., Kung, F. Y. H., Santos, H. C., Bobocel, D. R. & Grossmann, I. (2018). Wisdom, Bias, and Balance: Toward a Process-Sensitive Measurement of Wisdom-Related Cognition. *Journal of Personality and Social Psychology*, 115(6), 1093-1126. <https://doi.org/10.1037/pspp0000171>
- Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48(3), 378-399.
- Cassidy, S. A., Dimova, R., Giguère, B., Spence, J. R. & Stanley, D. J. (2019). Failing Grade: 89% of Introduction-to-Psychology Textbooks That Define or Explain Statistical Significance Do So Incorrectly. *Advances in Methods and Practices in Psychological Science*, 2(3), 233-239. <https://doi.org/10.1177/2515245919858072>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997-1003. <https://doi.org/10.1037/0003-066x.49.12.997>
- Cohen, J. (1995). The earth is round ($p < .05$): Rejoinder. *American Psychologist*, 50(12), 1103. <https://doi.org/10.1037/0003-066x.50.12.1103>
- Cooper, H. (2020). *Reporting quantitative research in psychology: How to meet APA style journal article reporting standards*. American Psychological Association.
- Costello, F. & Watts, P. (2022). How to Tell When a Result Will Replicate: Significance and Replication in Distributional Null Hypothesis Tests. *ArXiv*. <https://doi.org/10.48550/arxiv.2211.02609>

- Cumming, G. (2013). The New Statistics. *Psychological Science*, 25(1), 7-29. <https://doi.org/10.1177/0956797613504966>
- Eden, A. B. & Inan, N. G. (2022). Common Misconceptions and Misunderstandings in Magic Cut-Off for Significance: P-Value. *Proceedings of the 4th International Conference on Statistics: Theory and Applications*. <https://doi.org/10.11159/icsta22.125>
- Falk, R. & Greenbaum, C. W. (1995). Significance Tests Die Hard: The Amazing Persistence of a Probabilistic Misconception. *Theory & Psychology*, 5(1), 75-98. <https://doi.org/10.1177/0959354395051004>
- Filippini, T. & Vinceti, S. R. (2022). The role of statistical significance testing in public law and health risk assessment. *Journal of Preventive Medicine and Hygiene*, 63(1), E161-E165. <https://doi.org/10.15167/2421-4248/jpmh2022.63.1.2394>
- Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129-161). John Wiley & Sons.
- Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, 33(5), 587-606. <https://doi.org/10.1016/j.socec.2004.09.033>
- Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science*, 1(2), 198-218. <https://doi.org/10.1177/2515245918771329>
- Grossmann, I. (2017a). Wisdom and How to Cultivate It. *European Psychologist*, 22(4), 233-246. <https://doi.org/10.1027/1016-9040/a000302>
- Grossmann, I. (2017b). Wisdom in Context. *Perspectives on Psychological Science*, 12(2), 233-257. <https://doi.org/10.1177/1745691616672066>
- Grossmann, I., Dorfman, A. (2019). Wise Reasoning in an Uncertain World. In R. Sternberg, H. Nusbaum & J. Glück, (eds.). *Applying Wisdom to Contemporary World Problems* (pp. 51-79). Springer International Publishing. https://doi.org/10.1007/978-3-030-20287-3_3
- Grossmann, I., Dorfman, A., Oakes, H., Santos, H. C., Vohs, K. D. & Scholer, A. A. (2020). Training for Wisdom: The Distanced-Self-Reflection Diary Method. *Psychological Science*, 32(3), 381-394. <https://doi.org/10.1177/0956797620969170>
- Grossmann, I., Gerlach, T. M. & Denissen, J. J. A. (2016). Wise Reasoning in the Face of Everyday Life Challenges. *Social Psychological and Personality Science*, 7, 611-622. <https://doi.org/10.1177/1948550616652206>
- Grossmann, I., Weststrate, N. M., Ardel, M., Brienza, J. P., Dong, M., Ferrari, M., Fournier, M. A., Hu, C. S., Nusbaum, H. C. & Vervaeke, J. (2020). The Science of Wisdom in a Polarized World: Knowns and Unknowns. *Psychological Inquiry*, 31(2), 103-133. <https://doi.org/10.1080/1047840x.2020.1750917>
- Grossmann, I., Weststrate, N. M., Ferrari, M. & Brienza, J. P. (2020). A Common Model Is Essential for a Cumulative Science of Wisdom. *Psychological Inquiry*, 31(2), 185-194. <https://doi.org/10.1080/1047840x.2020.1750920>
- Hammond, K. R. (2007). *Beyond Rationality: The Search for Wisdom in a Troubled Time*. Oxford University Press.
- Hassler, U. (2022). When More Is Less: Pitfalls of significance testing. *ArXiv*. <https://doi.org/10.48550/arxiv.2211.11814>
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T. & Jennions, M. D. (2015). The Extent and Consequences of P-Hacking in Science. *PLoS Biology*, 13(3), e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
- Hentschke, H. & Stüttgen, M. C. (2011). Computation of measures of effect size for neuroscience data sets. *European Journal of Neuroscience*, 34(12), 1887-1894. <https://doi.org/10.1111/j.1460-9568.2011.07902.x>

- Hunter, M. A. & May, R. B. (1993). Some myths concerning parametric and nonparametric tests. *Canadian Psychology / Psychologie Canadienne*, 34(4), 384-389. <https://doi.org/10.1037/h0078860>
- Huynh, A. C. & Grossmann, I. (2020). A pathway for wisdom-focused education. *Journal of Moral Education*, 49(1), 9-29. <https://doi.org/10.1080/03057240.2018.1496903>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar.
- Kahneman, D., Krakauer, D. C., Sibony, O., Sunstein, C. & Wolpert, D. (2022). An exchange of letters on the role of noise in collective intelligence. *Collective Intelligence*, 1(1). <https://doi.org/10.1177/26339137221078593>
- Kahneman, D., Sibony, O. & Cass R. Sunstein. (2021). *Noise: A Flaw in Human Judgment*. William Collins.
- Kalinowski, P., Fidler, F. & Cumming, G. (2008). Overcoming the Inverse Probability Fallacy. *Methodology*, 4(4), 152-158. <https://doi.org/10.1027/1614-2241.4.4.152>
- Kline, R. B. (2004). *Beyond significance testing. Reforming data analysis methods in behavioral research*. APA Books.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). The Guilford Press.
- Kline, R. B. (2013). *Beyond Significance Testing: Statistics Reform in the Behavioral Sciences*. American Psychological Association. <http://www.jstor.org/stable/j.ctv1chrvvt>
- Kross, E. & Grossmann, I. (2012). Boosting wisdom: Distance from the self enhances wise reasoning, attitudes, and behavior. *Journal of Experimental Psychology: General*, 141(1), 43-48. <https://doi.org/10.1037/a0024158>
- Kruschke, J. K. & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin and Review*, 25, 178-206. <https://doi.org/10.3758/s13423-016-1221-4>
- Kühberger, A., Fritz, A., Lerner, E. & Scherndl, T. (2015). The significance fallacy in inferential statistics Psychology. *BMC Research Notes*, 8. <https://doi.org/10.1186/s13104-015-1020-4>
- Lilienfeld, S. O., Ammirati, R. & Landfield, K. (2009). Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspectives on Psychological Science*, 4(4), 390-398. <https://doi.org/10.1111/j.1745-6924.2009.01144.x>
- Madjarova, S. J., Williams, R. J., Nwachukwu, B. U., Martin, R. K., Karlsson, J., Ollivier, M. & Pareek, A. (2022). Picking apart p values: common problems and points of confusion. *Knee Surgery, Sports Traumatology, Arthroscopy*, 30(10), 3245-3248. <https://doi.org/10.1007/s00167-022-07083-3>
- Mandel, D. R. (2014). The psychology of Bayesian reasoning. *Frontiers in Psychology*, 5, 1144. <https://doi.org/10.3389/fpsyg.2014.01144>
- Maxwell, S. E., Delaney, H. D. & Kelley, K. (2018). *Designing Experiments and analyzing data: A model comparison perspective*. Routledge.
- Maxwell, S. E., Lau, M. Y., Howard, G. S. & Scott, E. (2015). Is psychology suffering from a replication crisis? *American Psychologist*, 70, 487-498. <https://psycnet.apa.org/doi/10.1037/a0039400>
- Mayo, D. G. & Hand, D. (2022). Statistical significance and its critics: practicing damaging science, or damaging scientific practice? *Synthese*, 200(3), 220. <https://doi.org/10.1007/s11229-022-03692-0>
- Nelson, N., Rosenthal, R. & Rosnow, R. L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *American Psychologist*, 41(11), 1299-1301. <https://doi.org/10.1037/0003-066x.41.11.1299>
- Oberoi, S. S. & Atri, M. (2021). When and Where P-Value is Required: Misnomers in Clinical Statistical Significance. *International Journal of Scientific Research*, 10(1), 1-2. <https://doi.org/10.36106/ijsr/9300808>

- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Ortega, A. & Navarrete, G. (2017). Bayesian Hypothesis Testing: An Alternative to Null Hypothesis Significance Testing (NHST) in *Psychology and Social Sciences*. <https://www.intechopen.com/chapters/56653>
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson.
- Pratiwi, I. M. & Ariawan, V. A. N. (2020). *Introducing local wisdom by numbering and reading*. 1521. <https://doi.org/10.1088/1742-6596/1521/3/032013>
- Rice, K. M. & Krakauer, C. A. (2022). Three-Decision Methods: A Sensible Formulation of Significance Tests—and Much Else. *Annual Review of Statistics and Its Application*, 10(1), 525-546. <https://doi.org/10.1146/annurev-statistics-033021-111159>
- Rodríguez, E. (2016). Estadística y psicología: análisis histórico de la inferencia estadística. En P. J. Ortega. (Ed.), *Disertaciones de Enerio sobre temas diversos: Filosofía, Ciencia, Psicología y Estadística* (pp. 160-166). FUNGLODE.
- Rogers, B. D. (2022). *Null Hypothesis Statistical Testing: A Survey of the History, Critiques, and Alternative Methodologies*. University of South Carolina Scholar Commons Scholar Commons.
- Santos, H. C., Huynh, A. C. & Grossmann, I. (2017). Wisdom in a complex world: A situated account of wise reasoning and its development. *Social and Personality Psychology Compass*, 11(10), e12341. <https://doi.org/10.1111/spc3.12341>
- Schauer, J. M. (2022). *Avoiding Questionable Research Practices in Applied Psychology*. 301-342. https://doi.org/10.1007/978-3-031-04968-2_14
- Shine, L. C. (1980). The Fallacy of Replacing an a Priori Significance Level With an a Posteriori Significance Level. *Educational and Psychological Measurement*, 40(2), 331-335. <https://doi.org/10.1177/001316448004000207>
- Shrout, P. E. & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*, 69(1), 487-510. <https://doi.org/10.1146/annurev-psy-122216-011845>
- Silva-Ayçaguer, L. C., Suárez-Gil, P. & Fernández-Somoano, A. (2010). The null hypothesis significance test in health sciences research (1995-2006): statistical analysis and interpretation. *BMC Medical Research Methodology*, 10(1), 44. <https://doi.org/10.1186/1471-2288-10-44>
- Stanovich, K. E. (2001). The Rationality of Educating for Wisdom. *Educational Psychologist*, 36(4), 247-251. https://doi.org/10.1207/s15326985ep3604_3
- Stanovich, K. E. (2018a). Perceiving rationality correctly. *Psychonomic Bulletin & Review*, 25, 793-826.
- Stanovich, K. E. (2018b). Miserliness in human cognition: the interaction of detection, override and mindware. *Thinking & Reasoning*, 24(4), 423-444. <https://doi.org/10.1080/13546783.2018.1459314>
- Stanovich, K. E. & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking and Reasoning*, 13(3), 225-247. <https://doi.org/10.1080/13546780600780796>
- Stanovich, K. E. & West, R. F. (2008a). On the Relative Independence of Thinking Biases and Cognitive Ability. *Journal of Personality and Social Psychology*, 94(4), 672-695. <https://doi.org/10.1037/0022-3514.94.4.672>
- Stanovich, K. E. & West, R. F. (2008b). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking and Reasoning*, 14(2), 129-167. <https://doi.org/10.1080/13546780701679764>
- Stanovich, K. E., West, R. F. & Toplak, M. E. (2013). Myside Bias, Rational Thinking, and Intelligence. *Current Directions in Psychological Science*, 22(4), 259-264. <https://doi.org/10.1177/0963721413480174>

- Sternberg, R. J. (2001). Why Schools Should Teach for Wisdom: The Balance Theory of Wisdom in Educational Settings. *Educational Psychologist*, 36(4), 227-245. https://doi.org/10.1207/s15326985ep3604_2
- Sternberg, R. J. (2003). *Wisdom, Intelligence, and Creativity Synthesized*. Cambridge University Press.
- Sternberg, R. J. & Hagen, E. S. (2019). Teaching for Wisdom. In R. J. Sternberg & J. Glück (Eds.), *The Cambridge Handbook of Wisdom*. Cambridge University Press. <https://doi.org/10.1017/9781108568272.018>
- Sullivan, G. M. & Feinn, R. (2012). Using Effect Size—or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3), 279-282. <https://doi.org/10.4300/jgme-d-12-00156.1>
- Tabri, N. & Elliott, C. M. (2014). *Principles and Practice of Structural Equation Modeling* (4th ed., Vol. 1). The Guilford Press. <https://doi.org/10.15353/cgjsc.v1i1.3787>
- Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., Oltmanns, T. F. & Shrout, P. E. (2017). It's Time to Broaden the Replicability Conversation: Thoughts for and From Clinical Psychological Science. *Perspectives on Psychological Science*, 12(5), 742-756. <https://doi.org/10.1177/1745691617690042>
- Thompson, B. (2004). The “significance” crisis in psychology and education. *Journal of Socio-Economics*, 33(5), 607-613. <https://doi.org/10.1016/j.socec.2004.09.034>
- Toplak, M. E., West, R. F. & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2), 147-168. <https://doi.org/10.1080/13546783.2013.844729>
- Vervaeke, J. & Ferraro, L. (2013). The Scientific Study of Personal Wisdom, From *Contemplative Traditions to Neuroscience*. 21-51. https://doi.org/10.1007/978-94-007-7987-7_2
- Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N. & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin and Review*, 25, 35-57. <https://doi.org/10.3758/s13423-017-1343-3>
- Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M., Sun, J., Prims, J. P., Mueller, A. B., Melton, Z. J., & Carsel, T. S. (2018). Why do some psychology researchers resist adopting proposed reforms to research practices? A description of researchers' rationales. *Advances in Methods and Practices in Psychological Science*, 1(2), 166–173. <https://doi.org/10.1177/2515245918757427>
- West, R. F. & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory and Cognition*, 31, 243-251. <https://doi.org/10.3758/bf03194383>
- West, R. F., Toplak, M. E. & Stanovich, K. E. (2008). Heuristics and Biases as Measures of Critical Thinking: Associations with Cognitive Ability and Thinking Dispositions. *Journal of Educational Psychology*, 100(4), 930-941. <https://doi.org/10.1037/a0012842>
- Wilkinson, L. & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604. <https://doi.org/10.1037/0003-066x.54.8.594>
- Willingham, D. T. (2008). Critical Thinking: Why Is It So Hard to Teach? *Arts Education Policy Review*, 109(4), 21-32. <https://doi.org/10.3200/aep.109.4.21-32>
- Ziliak, S. T. & McCloskey, D. N. (2008). *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. University of Michigan Press.