

Importancia del tamaño del efecto en el análisis de datos de investigación en psicología

Ronald W. Castillo-Blanco
Universidad Nacional Mayor de San Marcos, Lima, Perú

Alberto A. Alegre
Cerebrum-Perú, Lima

Recibido: 15 de junio de 2015 / Aprobado: 7 de julio de 2015

Inicialmente se discute el método tradicional de análisis inferencial en metodología cuantitativa (test de significancia de la hipótesis nula) haciendo ver sus limitaciones e imposibilidad de expresar la significancia práctica de los resultados. Se muestra el tamaño del efecto como una medida del grado en el que se presenta un determinado fenómeno de interés, medida cuyo reporte es solicitado explícitamente por la American Psychological Association (APA) en sus tres últimos manuales de publicación. Finalmente, se revela la ausencia de estudios que den cuenta de la aplicación del tamaño del efecto en el contexto local.

tamaño del efecto / significancia estadística / significancia práctica

The role of the Effect Size Report in Data Analysis in Psychology Research

In this essay we discuss the traditional method of inferential analysis in quantitative methodology (null hypothesis significance testing) and demonstrate its limitations and inability to provide a measure for the practical significance of the results. The effect size is presented as a quantitative measure of the strength of a phenomenon, and this report has been explicitly requested by the American Psychological Association (APA) in the last three publication manuals. Finally, we show that in our local context, there is a lack of studies describing the application of the effect size.

effect size / statistical significance / practical significance

Correos electrónicos: Ronald.castilloblanco@gmail.com; aalegrebravo@gmail.com

En psicología, desde un enfoque de investigación cuantitativo, se utilizan las herramientas de la ciencia estadística. Su importancia recae en el manejo de la variabilidad de las puntuaciones, lo que es un reto en las investigaciones, ya que estas dependen de las diferencias individuales de los sujetos y de los errores de medición (Myers y Well, 2003). Desde la década de 1940 el test de significancia de la hipótesis nula (NHST, por sus siglas en inglés) es el método de inferencia estadística más usado y difundido en el análisis de datos (Kline, 2004); sin embargo, este método es controversial y ha sido criticado tanto por su uso e interpretación inadecuados como por su lógica y concepción misma (Nickerson, 2000; Schmidt y Hunter, 1997).

Una interpretación errónea, pero frecuente, es la que se tiene cuando se da a los resultados del análisis inferencial con el NHST el calificativo de “significativo” o “no significativo”. Es más, cuando se consideran las anteriores atribuciones como evidencia de importancia o no de un determinado resultado, las conclusiones se tornan erróneas y lejanas al verdadero significado de la evidencia empírica. Así, por ejemplo, al analizar la diferencia en el clima familiar según el género en un determinado grupo, algunos investigadores suelen reportar: $t(210)=12.56, p<.001$, concluyendo: “dado que $p<.05$, entonces se tiene diferencias significativas en el clima familiar según el género en ...”, a lo que mediante la sola lectura del valor

p se le da la interpretación errónea de “diferencia importante”.

La medida estadística que evalúa de forma coherente la magnitud de un aspecto de interés de un estudio cuantitativo, y que, por ende, facilita la valoración de su importancia práctica, es el tamaño del efecto (Kirk, 1996). La presentación e interpretación del tamaño del efecto como una medida de la significancia práctica de los resultados no es solo un aspecto de sentido común en la investigación con metodología cuantitativa en psicología, sino también un objeto de exigencia de importantes instituciones académicas internacionales, como la American Psychological Association (APA) y la American Educational and Research Association (AERA).

El presente artículo tiene por objetivo contextualizar el tamaño del efecto a partir de las dificultades del método tradicional de análisis inferencial de datos (NHST), mostrar sus conceptualizaciones y describir sus aplicaciones, como una medida de la significancia práctica de los resultados y necesaria para los metaanálisis.

EL TEST DE SIGNIFICANCIA ESTADÍSTICA DE LA HIPÓTESIS NULA (NHST)

Orígenes y crecimiento en su aplicación

El método para obtener conclusiones mediante resultados de estadística inferencial, que actualmente es de gene-

ral aplicación, proviene de los aportes de Ronald Fisher y Jerzy Neyman con Egon Pearson (Cowles, 2001; Kline, 2004; Lehmann, 2011). Los anteriores representan dos líneas que contribuyen a lo que se denomina el procedimiento clásico de estadística inferencial: el test de significancia de la hipótesis nula. Este es un procedimiento usado ampliamente en psicología, ciencias de la salud, educación, etc.

El método del NHST empezó a hacerse de aplicación generalizada en psicología desde la década de 1940, experimentando un incremento rápido de su aplicación (Kline, 2004). Este aumento en el uso del test de significancia estadística en artículos de revistas especializadas en psicología puede revisarse con más detalle en el estudio de Hubbard y Ryan (2000), donde se tomó una muestra de 8001 artículos publicados entre los años 1911 y 1998 de una muestra aleatoria de 12 revistas de la American Psychological Association (APA). El porcentaje de aplicación del método de NHST fue de alrededor de 17 % entre los años 1911 y 1929, luego se da un incremento hacia el 50 % alrededor de 1940. Este aumento continuó, con un 80 % para 1960, y excediendo el 90 % hacia 1970.

Controversia del NHST

A pesar de su extendido uso, el método del NHST es controversial (Nickerson, 2000). En las últimas décadas, un am-

plio número de autores han criticado el exceso y las interpretaciones erróneas de artículos que con este método han sido publicados en la investigación psicológica (Hubbard y Ryan, 2000). Por ejemplo, desde los años sesenta se tiene a Rozeboom (1960) y Bakan (1966); en los setenta a Carver (1978) y Meehl (1978); en los años ochenta Dar (1987), Rosnow y Rosenthal (1989), etc. En los años noventa la crítica se intensifica (e.g., Cohen, 1990, 1994; Schmidt, 1996; Thompson, 1999a). Los anteriores y otros numerosos autores no solo criticaron la inadecuada interpretación del método en mención, sino, también, su imposibilidad de brindar resultados acordes a los objetivos del investigador en psicología.

La aplicación del NHST en la investigación en psicología también motivó severas expresiones por parte de reconocidos psicólogos especialistas en investigación cuantitativa, los que fueron publicando a lo largo de varias décadas en las más importantes revistas internacionales de psicología. Así se tiene, por ejemplo, que con el título de “La falacia del test significancia de la hipótesis nula”, Rozeboom (1960) se expresa sobre el uso del método en mención como: “... costumbres estadísticas de un pasado primitivo continúan dominando la escena local” (p. 417). Cronbach (1975), preocupado por la práctica de investigadores de preferir los resultados “significativos”, escribe: “Ha llegado el momento para exorcizar la hipótesis nula... la dicoto-

mía significativo/no significativo implica solo una desesperada inconsistencia” (p. 124). Rosnow y Rosenthal (1989) se expresarán con una famosa ironía sobre el nivel de significancia estadística: “... seguramente, Dios ama al .06 tanto como al .05” (p. 1277). En su artículo de alta difusión titulado “The Earth is round ($p < .05$)”, Cohen (1994) expresa: “El NHST no solo ha fallado en darle soporte al avance de la Psicología, sino, seriamente, la ha impedido” (p. 997). Schmidt (1996), en la reconocida revista *Psychological Methods*, expresa:

la confianza en el test de significancia estadística en el análisis e interpretación de resultados en la investigación ha retardado sistemáticamente la acumulación del conocimiento en Psicología. (p. 115)

Es interesante notar que a pesar de las duras críticas que ha recibido el método del NHST a lo largo de las décadas este haya continuado en aplicación constante. Una importante explicación a esto es que los investigadores conservan falsas creencias sobre este procedimiento, creencias acerca de que este método ofrece un conjunto de beneficios para la investigación, pero que, sin embargo, en la realidad no lo hace (Schmidt, 1996). Como expresa Cohen (1994) sobre el NHST: “... eso no nos dice lo que nosotros queremos saber, y nosotros queremos saber tanto aquello que queremos saber que, por desesperación, sin embargo, creemos que lo hace” (p. 997).

Schmidt (1996) expresa tres comunes creencias sobre la aplicación del NHST. La primera consiste en que un resultado estadísticamente significativo haría referencia a un resultado replicable; la segunda, de que el test de significancia estadística provee un índice de la importancia o tamaño de una diferencia o relación, cuando en verdad esto solo puede ser provisto por las medidas del tamaño del efecto; y la tercera, que un resultado no significativo implica una diferencia o relación igual a cero, o tan pequeños que puede ser considerado como nulo. Estos y otros tipos de creencias y usos e interpretaciones erróneos del NHST pueden encontrarse en Schmidt (1996), Schmidt y Hunter (1997), Kline (2004), entre otros. En la práctica profesional, las creencias descritas se observan cuando muchos profesionales, investigadores y docentes raramente distinguen entre las implicancias de un resultado con significancia estadística y otro con significancia práctica, o incluso peor, sus resultados, siendo solo estadísticamente significativos, son interpretados como si ellos tuvieran significancia práctica. Esto ocurre cuando un investigador interpreta una significancia estadística como “significante” o “altamente significativo” (Ellis, 2010).

Sin embargo, y a pesar de las coherentes y constantes objeciones, el uso del NHST, paradójicamente, se ha difundido. Esto, incluso, a pesar de que es ampliamente aceptado que el NHST es comúnmente mal usado y mal inter-

pretado (Fidler, 2002). Esta situación se desarrolla entre expresiones de reforma que plantean la eliminación del uso del NHST (Schmidt, 1996), así como de hacer incorporación complementaria y obligatoria de los tamaños del efecto e intervalos de confianza (Kirk, 2001; Thompson, 1999a), esto para dar un mejor sentido e interpretación a los resultados.

EL TAMAÑO DEL EFECTO

La necesidad del reporte del tamaño del efecto

En el contexto de crecimiento de las críticas al método de NHST, y en particular por el artículo de Cohen (1994), el Comité de Asuntos Científicos de la American Psychological Association (APA) convino formar un comité denominado Task Force on Statistical Inference (TFSI), cuyo objetivo era “elucidar algunos de los aspectos controvertidos que rodean la aplicación de la estadística incluyendo el test de significancia y sus alternativas” (Wilkinson y TFSI, 1999, p. 594). Este comité reunió a expertos en investigación cuantitativa y análisis estadístico en Psicología, además de editores de revistas, autores de libros y renombradas personalidades en estos campos, los que harán expresión de un conjunto de lineamientos y requerimientos para la elaboración de artículos científicos. Entre estos se encuentra el de considerar, en los reportes de investigación, los tamaños del efecto.

Esta expresión se concretiza en el artículo titulado “Statistical methods in psychology journals: Guidelines and explanations”, publicado en la revista *American Psychologist* (Wilkinson y TFSI, 1999).

Es notoria, en cuanto a las recomendaciones del TFSI, la insistencia en la inclusión de las medidas del tamaño del efecto en las publicaciones: “Siempre presentar tamaños del efecto en los resultados... Nosotros enfatizamos que reportar e interpretar los tamaños del efecto en el contexto de reportes previos del tamaño del efecto es esencial para una buena investigación” (Wilkinson y TFSI, 1999, p. 599). Aunque el TFSI no recomendó la suspensión del uso del NHST, sus recomendaciones han sido bien recibidas y frecuentemente citadas por las publicaciones posteriores (Fidler, 2002).

Previo a las recomendaciones de Wilcoxon y TFSI (1999), el *Manual de publicaciones de la APA*, en su cuarta edición, por primera vez ya alentaba, pero no requería, a los autores a reportar los tamaños del efecto en sus resultados estadísticos (Kline, 2004). En la quinta edición del *Manual de publicaciones*, con emisión posterior a la publicación del TFSI, se realiza un mayor énfasis en el reporte de los tamaños del efecto, haciendo referencia a que el no reporte de estos es uno de los tipos de “defectos en el diseño y reporte de investigación” (APA, 2001, p. 25). En su más reciente edición del *Manual de*

publicaciones, la APA hace una expresión más explícita para la presentación de los tamaños del efecto:

Los tamaños del efecto, los intervalos de confianza y una extensa descripción son necesarios para transmitir el significado más completo de los resultados... Uno de los requerimientos mínimos de todas las publicaciones periódicas de la APA es una presentación completa de todas las hipótesis examinadas y las estimaciones de tamaños del efecto e intervalos de confianza adecuados. (APA, 2010, p. 33)

Definición y cálculo del tamaño del efecto

Cohen (1988) define el tamaño del efecto como “el grado en que el fenómeno se presenta en la población” (p. 9), o “el grado en el que la hipótesis nula es falsa” (p. 10). Esta definición, en la que se considera al tamaño del efecto a partir de la relación con el valor de la hipótesis nula que corresponde al NHST, también es considerada por otros autores, como Grissom y Kim (2005), para los que “un tamaño del efecto mide el grado en que la hipótesis nula es falsa” (p. 4). También Vacha-Hasse y Thompson (2004) expresan que el término tamaño del efecto lo usarán para “referirse a cualquier estadístico que cuantifica el grado en que los resultados de la muestra divergen de la hipótesis nula” (p. 473). Sin embargo, y siguiendo a Kelley y Preacher (2012),

el relacionar el tamaño del efecto con el término de hipótesis nula, y por ende con el método del NHST, debería ser evitado ya que ambos representan diferentes maneras de analizar los datos cuando se hacen inferencias.

De acuerdo con Kelley y Preacher (2012), “el tamaño del efecto es una reflexión cuantitativa de la magnitud de algún fenómeno, el que se usa para abordar alguna cuestión de interés” (p. 140). Es así que una medida del tamaño del efecto evalúa, ya sea en el análisis de una diferencia de grupos, en una correlación, etc., el tamaño, grado o magnitud del fenómeno presente, entendiéndose como fenómeno cualquiera de los análisis anteriores. Una expresión simplificada del tamaño del efecto correspondería a la medida del grado en que el fenómeno de interés está presente en la muestra.

Para el cálculo del tamaño del efecto se debe considerar una formulación específica según el fenómeno estudiado. Así, para una comparación de dos grupos independientes se tiene como expresión más común a la diferencia de medias de los grupos en función de su desviación estándar: $d = (M_1 - M_2) / DE^*$, donde d es el tamaño del efecto, M_1 y M_2 son las medias de cada uno de los grupos, respectivamente, y DE^* es la desviación estándar. Se tienen tres formas de estimar d , las que básicamente difieren por la consideración particular que hacen para el cálculo de DE^* . De esta manera se tienen las fórmulas de

Cohen, Glass y Hedges (d_{Cohen} , d_{Glass} y g_{Hedges} , respectivamente).

En el caso de la correlación de dos variables, es necesario considerar la utilidad de r^2 para una mejor comprensión de r como una medida del tamaño del efecto.

Al cuadrado del coeficiente de correlación, r^2 , se le denomina coeficiente de determinación, el cual es la proporción de varianza que comparten dos variables que se correlacionan, lo que también puede expresarse como el porcentaje de variabilidad con la que una variable explica a la otra. Con esa consideración, r^2 y r son importantes alternativas para la expresión del tamaño del efecto en el análisis de correlación bivariada.

El tamaño del efecto para los análisis de varianza, o diseños de ANOVA, como análisis en la comparación de más de dos grupos independientes es el eta cuadrado ($\hat{\eta}^2$). Sin embargo, una mejor alternativa para estimar el tamaño del efecto como fuerza de asociación de dos variables en estos diseños es el omega cuadrado ($\hat{\omega}^2$).

Por otro lado, en el caso de asociación de dos variables categóricas y análisis de bondad de ajuste se realiza con la medida de asociación V de Cramer. Una presentación y discusión detallada de los estadísticos anteriores, así como la distinción y el cálculo de tamaños del efecto paramétricos y no paramétricos, rebasan los objetivos del presente ensayo.

Significancia práctica de los resultados

Sobre el análisis de resultados y la interpretación que se hace de estos, Kirk (2001) plantea que a través de sus estudios los investigadores deben responderse a las siguientes tres preguntas: a) ¿existe un efecto real o los resultados deberían atribuirse al azar?, b) si el efecto es real, ¿qué tan grande es?, y c) ¿es el efecto lo suficientemente grande para considerarse importante o útil? De estas preguntas, el método del NHST solo puede responder la primera de ellas, e incluso esta respuesta sería muy limitada y no consistente con esta interpretación. La segunda, que corresponde al interés del investigador, se responde de manera coherente con el cálculo y la presentación del tamaño del efecto correspondiente. Finalmente, la tercera pregunta, que hace referencia a la significancia práctica de los resultados (Kirk, 1996), se puede responder a partir del desarrollo de guías, que hagan interpretación de los tamaños del efecto calculados, basando estos en los antecedentes y la experiencia sobre el campo particular de estudio.

El tamaño del efecto y el metaanálisis

Además, los tamaños del efecto son importantes porque con ellos es posible sistematizar varios estudios de un determinado tema de interés o fenómeno, esto se realiza mediante los me-

taanálisis, que es el análisis estadístico a un número amplio de resultados de estudios individuales con el propósito de integrar los hallazgos (Glass, 1976), siendo, desde su introducción en los años setenta, una importante herramienta para la investigación de síntesis de resultados (Kline, 2004). Los metaanálisis no solo contribuyen para sistematizar evidencia de los efectos de las intervenciones, sino, también, tienen implicaciones sobre el modelo que debe usarse para analizar los datos, qué análisis considerar y cómo los resultados deben ser interpretados (Borenstein, Hedges, Higgins y Rothstein, 2009). Debe tenerse en cuenta que para realizar un metaanálisis no se requiere el resultado de algún método del NHST, sino los tamaños del efecto de cada estudio individual. Los metaanálisis, a diferencia del método tradicional de análisis, sí permiten llegar a conclusiones certeras y, por ende, contribuyen a la acumulación de conocimiento (Schmidt, 1996).

Actualidad en el reporte del tamaño del efecto

A pesar de la importancia de las medidas del tamaño del efecto para la expresión, interpretación, valoración y sistematización de un determinado fenómeno estudiado, sea diferencia o correlación, por ejemplo, y cuya importancia se hace innegable a partir de las limitaciones lógicas y prácticas del método del NHST, la presentación de los tamaños del efecto no se suelen aplicar en todos los repor-

tes de investigación (Thompson, 1999b). Esto, incluso, además de los requerimientos explícitos que la APA y otras importantes asociaciones de investigación psicológica han solicitado en las últimas dos décadas.

Como evidencia empírica sobre la aplicación de las medidas del tamaño del efecto, mediante una metodología de revisión metodológica de investigación (Keselman et al., 1998), hay estudios que hacen revisión y análisis sobre la aplicación del tamaño del efecto en los resultados de investigaciones en importantes y reconocidas revistas internacionales de Psicología. Entre estos estudios se tienen los desarrollados por Sun, Pan y Wang (2010), García, Ortega y De la Fuente (2008, 2011), Vacha-Haase y Ness (1999), Keselman et al. (1998), Snyder y Thompson (1998), entre otros. Estos autores encontraron un porcentaje de reporte del tamaño del efecto de 57 %, 20 %, 9 % y 54 %, respectivamente.

En nuestro contexto no se han realizado publicaciones previas en las que se haga el análisis con una metodología de revisión sistemática de investigación sobre la frecuencia de publicaciones que hacen reporte del efecto en las revistas de investigación en psicología. Sin embargo, se tiene una primera contribución de Coe y Merino (2003), quienes elaboran una guía introductoria sobre el cálculo del tamaño del efecto y discuten la importancia de diferenciar la significancia de la significancia estadística.

Sobre el tamaño del efecto no habría publicaciones posteriores que abran la discusión acerca de esta importante medida, más aún, no existen publicaciones que hagan una revisión de la aplicación del tamaño del efecto en la investigación psicológica peruana.

A lo anterior, y para menoscabo de la investigación psicológica, los textos más referenciados en nuestro contexto sobre metodología cuantitativa no presentan indicación explícita a los investigadores sobre la importancia y obligatoriedad del reporte de los tamaños del efecto. Esto a diferencia de textos de metodología de la investigación de editoriales extranjeras, donde se dedican secciones de los libros para la explicación de los tamaños del efecto y su importancia para la discusión de resultados; por ejemplo: Bordens y Abbott (2011), Goodwin (2010), Myers y Well (2003), Shaughnessy, Zechmeister y Zechmeister (2012), entre otros. De la misma forma, en los cursos de pregrado de la formación en psicología no se suelen tener como referencia libros de estadística coherentes con la naturaleza de la medición y de las variables de investigación, como bien podrían ser los de Aron, Coups y Aron (2013), Kirk (2008) y Pagano (2013).

La evaluación sobre el reporte del tamaño del efecto en publicaciones de psicología es una práctica realizada en importantes revistas de investigación. Estas revisiones permiten visualizar de manera empírica el ajuste del análisis

de resultados a estándares internacionales y observar la práctica que van teniendo los autores. Sin embargo, y a pesar de la importancia de este tipo de estudios, estos no se han realizado en el contexto de publicaciones peruanas.

CONCLUSIONES Y RECOMENDACIONES

El método del NHST es el método que desde la década de 1940 se utiliza con mayor frecuencia para el análisis de datos cuantitativos en psicología, pero su uso y la interpretación de sus resultados han sido constantemente cuestionados, siendo las medidas del tamaño del efecto los indicados para dar las respuestas que el NHST no puede por su lógica misma. Los tamaños del efecto son medidas que desde su conceptualización están orientados a dar información sobre la magnitud de un determinado fenómeno o análisis en psicología; es decir, dan información y permiten el análisis de la significancia práctica de un resultado. Además, son el insumo para posteriores estudios de sistematización objetiva de los resultados sobre un tema de interés, los metaanálisis. Sin embargo, no obstante la trascendencia de esta medida, los tamaños del efecto no se reportan con regularidad, a pesar de las directivas de importantes instituciones académicas, las que hacen demanda explícita de su presentación e interpretación.

Es necesario sensibilizar o capacitar a los editores de las revistas de investigación en psicología, así como a los docentes y alumnos de formación superior,

sobre la importancia de la presentación e interpretación de los tamaños del efecto en toda investigación de metodología cuantitativa que se haga en psicología.

No solo para alinear el nivel de dicha investigación a los estándares internacionales, sino, principalmente, para hacer coherente la interpretación de los resultados en la producción científica. También es conveniente la realización de estudios que desde revisiones metodológicas de investigación permitan obtener evidencia empírica del grado de aplicación del tamaño del efecto en las publicaciones.

REFERENCIAS

- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5.^a ed.). Washington DC: Autor.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6.^a ed.). Washington, DC: Autor.
- Aron, A., Coups, E., y Aron, E. (2013). *Statistics for psychology* (6.^a ed.). Buenos Aires: Pearson Education.
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66(6), 423-437.
- Bordens, K., y Abbott, B. (2011). *Research design and methods* (8.^a ed.). New York: McGraw-Hill.
- Borenstein, M., Hedges, L., Higgins, J., y Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons.
- Carver, R. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48(3), 378-399.
- Coe, R., y Merino, C. (2003). Magnitud del efecto: Una guía para investigadores y usuarios. *Revista de Psicología [PUCP]*, 21(1), 146-177.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2.^a ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312.
- Cohen, J. (1994). The Earth is round ($p < .05$). *American Psychologist*, 49(12), 997-1003.
- Cowles, M. (2001). *Statistics in Psychology: An historical perspective* (2.^a ed.). Mahwah, New Jersey: Lawrence Erlbaum.
- Cronbach, L. (1975). Beyond the two disciplines of psychology. *American Psychologist*, 30(2), 116-127.
- Dar, R. (1987). Another look at Meehl, Lakatos, and the scientific practices of psychologists. *American Psychologist*, 42(2), 145-151.
- Ellis, P. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, UK: Cambridge University Press.
- Fidler, F. (2002). The fifth edition of the APA publication manual: Why

its statistics recommendations are so controversial. *Educational and Psychological Measurement*, 62(5), 749-770.

García, J., Ortega, E., y De la Fuente, L. (2008). Tamaño del efecto en las revistas de Psicología indizadas en Redalyc. *Informes Psicológicos*, 10(11), 173-188.

García, J., Ortega, E., y De la Fuente, L. (2011). The use of the effect size in JCR spanish journals of Psychology: from theory to fact. *The Spanish Journal of Psychology*, 14(2), 1050-1055.

Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.

Goodwin, J. (2010). *Research in Psychology: Methods and Design* (6.^a ed.). Toronto: John Wiley and Sons.

Grissom, R., y Kim, J. (2005). *Effect sizes for research*. Mahwah, New Jersey: Lawrence Erlbaum.

Hernández, R., Fernández, C., y Baptista, P. (2010). *Metodología de la investigación* (5.^a ed.). México, D. F.: McGraw-Hill.

Hubbard, R., y Ryan, P. (2000). The historical growth of statistical significance testing in psychology and its future prospects. *Educational and Psychological Methods*, 60(5), 661-681.

Kelley, K., y Preacher, K. (2012). On effect size. *Psychological Methods*, 17(2), 137-152.

Keselman, H., Huberty, C., Lix, L., Olejnik, S., Cribbie, R., Donahue, B., ... Levin, J. (1998). Statistical practices of educational researchers: An analysis of their Anova, Manova, and Ancova analyses. *Review of Educational Research*, 68(3), 350-386.

Kirk, R. (1996). Practical significance: a concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759.

Kirk, R. (2001). Promoting good statistical practices: some suggestions. *Educational and Psychological Measurement*, 61(2), 213-218.

Kirk, R. (2008). *Statistics: an introduction* (5.^a ed.). Belmont, CA: Thomsom Wadsworth.

Kline, R. (2004). *Beyond significance testing: reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Lehmann, E. (2011). *Fisher, Neyman, and the creation of classical statistics*. New York: Springer.

Meehl, P. (1978). Theoretical risks and tabular asterisk: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.

Myers, J., y Well, A. (2003). *Research design and statistical analysis* (2.^a ed.). Mahway, New Jersey: Lawrence Erlbaum Associates.

Nickerson, R. (2000). Null hypothesis significance testing: A review of an old

- and continuing controversy. *Psychological Methods*, 5(2), 241-301.
- Pagano, R. (2013). *Understanding statistics in the behavioral sciences* (10.^a ed.). México, D. F.: Cengage Learning.
- Rosnow, R., y Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44(10), 1276-1284.
- Rozeboom, W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, 57, 416-428.
- Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1(2), 115-129.
- Schmidt, F., y Hunter, J. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. En A. Harlow, S. Mulaik, y J. Steiger, *What if there were no significance tests?* (pp. 37-64). New York: Lawrence Erlbaum Associates.
- Shaughnessy, J., Zechmeister, E., y Zechmeister, J. (2012). *Research methods in psychology* (9.^a ed.). New York: McGraw-Hill.
- Snyder, P., y Thompson, B. (1998). Use of tests of statistical significance and other analytic choices in a school psychology journal: review of practices and suggested alternatives. *School Psychology Quarterly*, 13(4), 335-348.
- Sun, S., Pan, W., y Wang, L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology*, 102(4), 989-1004.
- Thompson, B. (1999). If statistical significance tests are broken/misused, what practice should supplement or replace them? *Theory & Psychology*, 9(2), 165-181.
- Thompson, B. (1999). Statistical significance tests, effect size reporting and the vain pursuit of pseudo-objectivity. *Theory & Psychology*, 9(2), 191-196.
- Vacha-Haase, T., y Ness, C. (1999). Statistical significance testing as it relates to practice: use within Professional Psychology: Research and Practice. *Professional Psychology: Research and Practice*, 31, 104-105.
- Vacha-Haase, T., y Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473-481.
- Wilkinson, L., y TFSI - Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.