Universitat d'Alacant
Universidad de Alicante

UA | UNIVERSIDAD DE ALICANTE
Vicerrectorado de Estudios, Formación y Calidad
ICE- Instituto de Ciencias de la Educación

# XI JORNADAS DE REDES DE INVESTIGACIÓN EN DOCENCIA UNIVERSITARIA

## Retos de futuro en la enseñanza superior: Docencia e investigación para alcanzar la excelencia académica

ISBN: 978-84-695-8104-9

Diseño: Gabinete de Imagen y Comunicación Gráfica de la Universidad de Alicante

# XI JORNADES DE XARXES D'INVESTIGACIÓ EN DOCÈNCIA UNIVERSITÀRIA

## Reptes de futur en l'ensenyament superior: Docència i investigació per a aconseguir l'excel·lència acadèmica

**Coordinadores**

**María Teresa Tortosa Ybáñez**
**José Daniel Álvarez Teruel**
**Neus Pellín Buades**


**Universidad de Alicante**
**Vicerrectorado de Estudios, Formación y Calidad**
**Instituto de Ciencias de la Educación (ICE)**

**ISBN: 978-84-695-8104-9**

**Revisión y maquetación: Neus Pellín Buades**

# Pensamiento meta-analítico: educación estadística

D. Frías-Navarro*, L. Badenes-Ribera*, M. Pascual-Soler**, & H. Monterde-i-Bort*

*Universidad de Valencia

**ESIC, Business & Marketing School, Valencia

**RESUMEN**

El procedimiento de significación de la hipótesis nula ha sido la estrategia de análisis dominante en el área de las ciencias sociales y de la salud. Sin embargo, también han sido muchas las críticas que han atacado el pensamiento dicotómico que implica dicho procedimiento. El movimiento de la reforma estadística plantea un cambio de perspectiva desde la significación estadística a la estimación de efectos. Desde esta nueva perspectiva destaca el desarrollo del pensamiento meta-analítico como una estrategia de análisis centrada en la estimación de efectos junto con sus intervalos de confianza y en la acumulación de evidencias con el objetivo de mejorar la precisión de las estimaciones. La mejora en las prácticas estadísticas requiere una educación estadística y un cambio en la conducta del investigador. Nuestro trabajo se centra en el estudio del impacto de la reforma estadística y las nuevas necesidades de formación entre los investigadores españoles de Psicología y Educación. Mediante una metodología de encuesta vía Internet se ha medido el cambio en la conducta vinculado a los elementos que la actual reforma estadística destaca: tamaño del efecto y sus intervalos de confianza y meta-análisis. (Investigación subvencionada por el Ministerio de Economía y Competitividad, MINECO, España. Proyecto EDU2011-22862).

**Palabras Clave:** metodología, pensamiento meta-analítico, educación estadística, reforma estadística

## 1. INTRODUCCIÓN

Las técnicas de inferencia estadística constituyen la herramienta fundamental del diseño estadístico en Psicología y Educación (Carver, 1993; Shaver, 1993; Steiger & Fouladi, 1997). Sin embargo, desde sus inicios han sido muchas las críticas que han recibido, sobre todo por su ejecución ritual o por la aplicación del mágico valor de 0.05. En 1989 Rosnow y Rosenthal destacaban cuatro cuestiones críticas dentro del área de la metodología y la conducta del investigador: la dependencia excesiva de la decisión estadística dicotómica (mantener la hipótesis nula/rechazar la hipótesis nula) sin valorar la utilidad práctica de los hallazgos, la tendencia a hacer estudios con baja potencia estadística, el hábito de definir los resultados de la investigación solamente en términos de niveles de significación y el énfasis sobre los estudios originales y únicos. Estas cuatro cuestiones han sido debatidas durante décadas y continúan siendo un tema destacado dentro de la denominada reforma estadística y la práctica basada en la evidencia (Cohen, 1994; Cortina & Dunlap, 1997; Frick, 1996; Frías-Navarro, 2011; Hagen, 1997; Kaufman, 1998; Nickerson, 2000; Schmidt, 1996).

Algunos de los errores más comunes de interpretación de las pruebas de significación estadística están relacionados con la misma decisión estadística, la importancia sustantiva que se le otorga a los resultados estadísticamente significativos o el alcance de la interpretación de los resultados estadísticamente no significativos. La decisión estadística dicotómica (mantener la hipótesis nula / rechazar la hipótesis nula) otorga importancia directa a los resultados cuyo valor de probabilidad, p, es menor o igual a 0.05 mientras que anula de forma directa la consideración de los resultados con p>0.05. Sin embargo, se olvida en muchas ocasiones la estimación de la potencia estadística a priori y la relación directa que hay entre el valor p de probabilidad y el tamaño de la muestra. Un valor p de probabilidad señala la probabilidad del resultado de la prueba o estadístico muestral, dado un tamaño de muestra y asumiendo que la hipótesis nula es cierta. Y no dice nada de la probabilidad de la hipótesis nula o de la hipótesis alternativa (ver Tabla 1).

Tabla 1. *Errores de interpretación más comunes vinculados al proceso de contraste estadístico*

| |
|---|
| 1. Un resultado estadísticamente significativo demuestra que el efecto encontrado es importante y cuanto menor sea el valor *p* de probabilidad (o más asteriscos tenga en el informe) más importante es el hallazgo. |
| 2. Un resultado estadísticamente no significativo demuestra que la hipótesis nula es cierta y por lo tanto no hay ningún efecto o relación entre las variables. |
| 3. Un valor $p > \alpha$ señala que la hipótesis nula es cierta. Por lo tanto, el valor *p* es la probabilidad de que los resultados se deban al azar. |
| 4. Un valor $p > \alpha$ señala que la hipótesis alternativa es falsa. |

| |
|---|
| 5. Un valor $p < \alpha$ señala que la hipótesis nula es falsa. Por lo tanto, el valor *p* es la probabilidad de que los resultados no se deban al azar. |
| 6. Un valor $p < \alpha$ señala que la hipótesis alternativa es cierta. |
| 7. El valor *p* es la probabilidad que tiene la hipótesis nula de ser verdadera. |
| 8. El valor *p* indica la probabilidad de que la hipótesis científica sea verdadera. |
| 9. El valor *p* indica la probabilidad de encontrar los mismo hallazgos si se replica la investigación. |

Usted lector piensa que ¿un resultado estadísticamente significativo con un valor *p* de 0.002 es bastante mayor que el resultado obtenido con un valor de $p = .045$? ¿Usted considera que la significación estadística otorga importancia al efecto detectado? Los valores *p* de probabilidad no son un índice del tamaño del efecto, ni tampoco indican la probabilidad de que la hipótesis nula sea verdadera o falsa. Del mismo modo, una falta de significación estadística no significa que la hipótesis nula sea verdadera ni que los efectos de los dos grupos sean equivalentes. Ausencia de evidencia no es evidencia de ausencia de efectos (Altman y Bland, 1995).

El valor *p* de probabilidad es un valor que limita el rechazo o no rechazo de la hipótesis nula dentro del procedimiento de significación de la hipótesis nula. De forma arbitraria Sir Ronald A. Fisher (1925) fijó el nivel de alfa en .05 (5%). El criterio de $p < .05$ se basa en la idea que tuvo de Fisher acerca de la razonable confianza que representaba para señalar que un efecto existe. No implica ningún valor mágico, ni detecta importancia del hallazgo. Por ello es totalmente incorrecto concluir que un resultado con $p = .04$ es importante y otro con $p = .06$ no es importante. Seguramente si los tamaños de la muestra fuesen iguales esos efectos serían muy similares. De ahí la importancia de informar de la probabilidad exacta del resultado (por ejemplo $p = .03$ en lugar de $p < .05$).

Un resultado estadísticamente significativo indica que la probabilidad del resultado observado (o más extremno) es menor a .05 ($p \leq .05$), si la hipótesis nula fuese cierta. Como consecuencia se rechaza la hipótesis nula y se concluye que probablemente existe un efecto. Pero este valor de probabilidad que implica el rechazo de la hipótesis nula si el alfa fijado a priori es de .05 no dice nada ni del tamaño del efecto ni de la significación clínica del efecto observado. Por ello, puede ocurrir que un tamaño del efecto pequeño detectado en un estudio con un tamaño muestral grande tenga el mismo valor de *p* que un tamaño del efecto grande de otro estudio cuyo tamaño muestral en cambio es pequeño.

Por lo tanto, una vez realizado el proceso de decisión estadística falta que el investigador interprete los hallazgos y aquí sus creencias y atribuciones sobre el significado de los resultados no siempre son correctas. El factor humano se convierte en tema de debate

sobre los usos y abusos por parte de los investigadores de las pruebas de significación estadística. Las interpretaciones incorrectas de las pruebas de contraste de hipótesis estadísticas han quedado reflejas en diferentes estudios con muestras de estudiantes y también con muestras de expertos e investigadores (Gordon, 2001; Mittag & Thompson, 2000; Monterde-i-Bort, Frías-Navarro & Pascual-Llobell, 2010; Monterde-i-Bort, Pascual-Llobell, & Frías-Navarro, 2006; Nelson, Rosenthal, & Rosnow, 1986; Oakes, 1986; Rosenthal & Gaito, 1963; Tryon, 1998; Zuckerman, Hodgins, Zuckerman, & Rosenthal, 1993; Wagenmakers, 2007). Y ese factor humano junto con la escasa información que el procedimiento facilita sobre el tamaño del efecto son dos argumentos utilizados por algunos investigadores para reemplazar el procedimiento tradicional de significación de la hipótesis nula por otras alternativas de análisis (Schmidt, 1996; Schmidt y Hunter, 1997).

La reforma estadística que se está llevando a cabo desde hace años y que es el origen de los cambios en las políticas editoriales de las revistas, enfatiza la estimación del tamaño del efecto y su intervalo de confianza junto a los valores *p* de probabilidad para evitar los sesgo de interpretación que durante años ha minado los trabajos de investigación confundiendo por ejemplo 'significación estadística' con 'importancia de los efectos hallados' (American Psychological Association, 2001, 2010; Cohen 1990, 1994; Cumming & Finch, 2001; Fidler, Thomason, Cumming, Finch, & Leeman, 2004, 2005; Frías-Navarro, 2011). El planteamiento de la reforma estadística destaca que junto a la clásica 'significación estadística' hay que considerar otros tipos de significación como el tamaño del efecto y la significación clínica o sustantiva. Por ejemplo, ¿qué es más interesante o más útil para el profesional, conocer que la intervención A es significativamente mejor que el placebo con una $p < 0.0000000001$ y que la Terapia B lo es con una $p < 0.0000001$? o ¿conocer que la intervención A reduce la sintomatología depresiva en un 32% mientras que la intervención B lo hace en un 20% respecto al grupo placebo? Desde luego, en la investigación aplicada es más importante conocer el cambio clínico o sustantivo que la significación estadística (Ogles, Lunnen y Bonesteel, 2001).

Nuestro trabajo de investigación tiene como objetivo conocer las opiniones de profesores de las universidades españolas de Psicología y Educación sobre los usos que hacen de la herramienta de la estadística. De este modo, se podrá analizar el alcance que la reforma estadística está teniendo entre dichos profesores así como analizar los posibles errores de interpretación que se puedan tener sobre la decisión estadística y el valor p de probabilidad.

## 2. METODOLOGÍA

### 2.1. Participantes

La muestra está constituida por 150 profesores de las universidades españolas de Psicología y Educación. El 46.67% son hombres y el 53.33% son mujeres, con una antigüedad media como profesores universitarios de 12.38 años. El 89.33% desarrollan su actividad en una universidad pública y el 10.67 lo hacen en una universidad privada.

### 2.2 Instrumentos

La encuesta fue enviada mediante un servicio a través de Internet con el programa SurveyMonkey. La encuesta incluye una serie de preguntas como por ejemplo "En su opinión, ¿qué cuestiones estadísticas o de diseño de investigación están sometidas a debate actualmente?", "¿Ha leído o utilizado algún trabajo de meta-análisis?", "En su opinión lo que el investigador o investigadora desea conocer principalmente cuando realiza una investigación empírica es…?, "En su opinión, ¿obtener un resultado estadísticamente significativo implica de forma indirecta que el efecto detectado es importante?", "¿Cuando ejecuta una prueba estadística ¿considera prioritario informar siempre de la significación estadística obtenida?" o "¿Podría señalar el nombre de algunos estadísticos del tamaño del efecto?

## 3. RESULTADOS

Los resultados señalan que un alto porcentaje de profesores universitarios desconocen que exista algún tipo de debate abierto dentro de sus disciplinas (60.67%). Un 1.33% opina que no hay ningún tipo de debate abierto y un 38% consideran que sí hay debates (ver Gráfica 1).

La mitad de los encuestados opina que lo que el investigador o investigadora desea conocer principalmente cuando realiza una investigación empírica es conocer si se ha producido algún efecto o diferencia entre los grupos que sea estadísticamente significativa (50.67%). El 22.67% de los participantes opina que lo que se busca es la importancia sustantiva o clínica del efecto y el 14% considera que lo relevante es conocer la magnitud del tamaño del efecto.

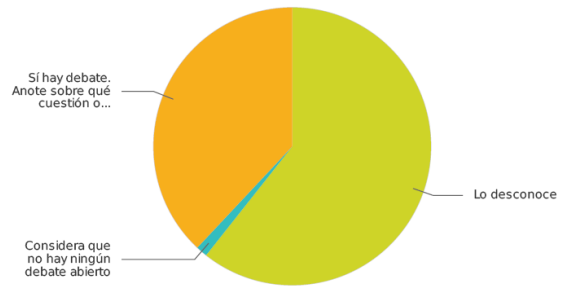
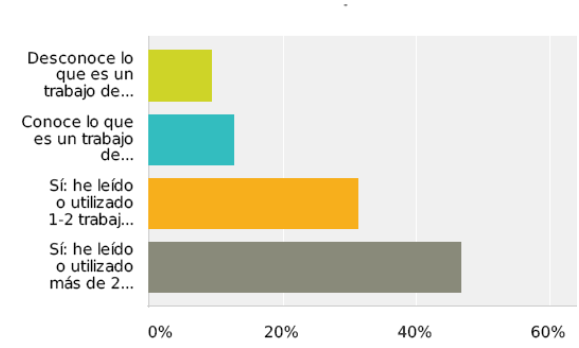
Cuando se trata de valorar si un resultado estadísticamente significativo implica de forma indirecta que el efecto detectado es importante el 50.67 da una respuesta negativa y el 26% opina que sí se puede equiparar significación estadística con importancia. Un 23.33% matiza su respuesta.

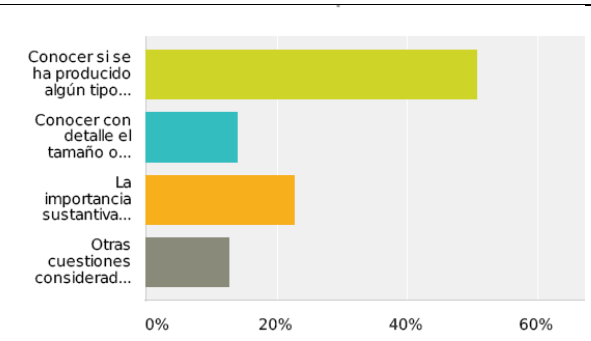
Una gran parte de los encuestados opinan que cuando ejecutan una prueba estadística consideran prioritario informar siempre de la significación estadística obtenida pero difieren en la forma de expresar el valor p de probabilidad. El 47.33 utiliza expresiones como $p< 0.05$ o $p>0.05$ y el 43.33 señala el valor exacto de probabilidad. Solamente el 9.33 no considera prioritario informar de la significación estadística.

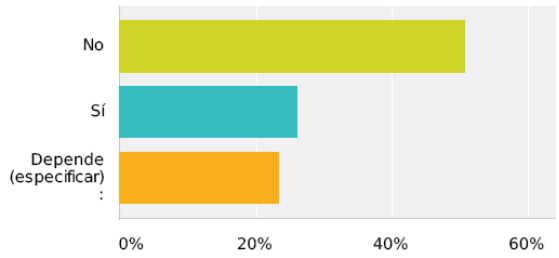
Respecto al conocimiento que tienen del tamaño del efecto se han planteado tres cuestiones. En primer lugar se ha analizado si conocen lo que es un trabajo de meta-análisis. Con esta pregunta se ha evaluado de manera indirecta el conocimiento del tamaño del efecto dado que es el estadístico que se utiliza en dichos estudios. El 9.33% de los participantes desconoce lo que es un trabajo de meta-análisis y el 12.97 lo conoce pero nunca lo ha leído o utilizado. Entre los dos tipos de respuesta aproximadamente el 22% no ha tenido entre sus manos la lectura de un trabajo de meta-análisis. El 31.33% opina que sí ha leído o utilizado un estudio de meta-análisis y el 46.67% han trabajado con más de dos trabajos de meta-análisis.

Cuando se trata de nombrar a algunos estadísticos del tamaño del efecto, el 38.67% desconoce el nombre de algún estadístico de tamaño del efecto y el 61.33% nombra al menos un estadístico, destacando la presencia de los estadísticos de la familia de la diferencia estandarizada de medias tipo d de Cohen o g de Hedges.

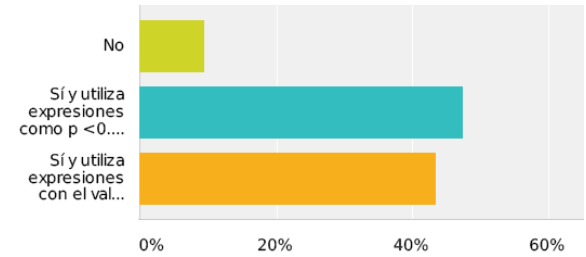
Gráfica 1. Porcentajes de respuestas

<table>
<tr>
<td>“En su opinión, ¿qué cuestiones estadísticas o de diseño de investigación están sometidas a debate actualmente?”<br> 
</td>
<td>“¿Ha leído o utilizado algún trabajo de meta-análisis?”<br> 
</td>
</tr>
<tr>
<td>“En su opinión lo que el investigador o investigadora desea conocer principalmente cuando realiza una investigación empírica es…?</td>
<td>“En su opinión, ¿obtener un resultado estadísticamente significativo implica de forma indirecta que el efecto detectado es importante?”</td>
</tr>
</table>

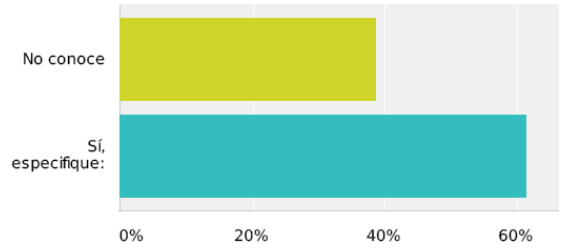



"¿Cuando ejecuta una prueba estadística ¿considera prioritario informar siempre de la significación estadística obtenida?"




"¿Podría señalar el nombre de algunos estadísticos del tamaño del efecto?"




## 4. CONCLUSIONES

Un primer resultado que llama la atención es el desconocimiento que los profesores universitarios tienen sobre los posibles debates abiertos en sus áreas de especialización. Tanto en Psicología como en Educación se abren y reabren debates de forma continua, impulsando el avance de la Ciencia. En concreto, en el área de la metodología el uso y abuso de las pruebas de significación estadística es un debate presente desde prácticamente principios del siglo XIX, activándose continuamente. Fruto de este debate el Manual de la American Psychological Association (2001, 2010) ha adoptado nuevas políticas editoriales que son seguidas por miles de revistas científicas y, por lo tanto, seguidas por los autores de los artículos. Otra consecuencia del mencionado debate metodológico ha repercutido en el ya clásico debate sobre significación clínica o social y significación estadística donde el cómputo del estadístico del tamaño del efecto trata de acapar la atención sobre la cuestión de magnitud del efecto, tratando de eliminar la asociación tantas veces errónea entre estadísticamente significativo e importancia del efecto detectado.

En segundo lugar, se observa que la mitad de los encuestados considera que la búsqueda de la significación estadística de los efectos (p<alfa) es el objetivo de los estudios, solamente un 14% considera que lo relevante es conocer la magnitud de dichos efectos y

aproximadamente el 23% destacan la búsqueda de la significación sustantiva. Este tipo de actuaciones impiden el desarrollo del denominado pensamiento meta-analítico y el pensamiento de estimación que favorecen la interpretación de la magnitud del efecto detectado en el propio estudio individual, contextualizándolo con el conjunto de efectos que se han estimado en un área concreta de investigación, favoreciendo con ello la acumulación de evidencia científica y de este modo incrementando la precisión de las estimaciones. Ya no se trata de buscar 'algún' efecto que sea estadísticamente diferente de cero (procedimiento tradicional de significación de la hipótesis nula) sino buscar un tamaño del efecto concreto que pueda tener relevancia sustancia, social o clínica dentro del área concreta de investigación en la que se encuadra el estudio. Un resultado estadísticamente significativo solamente indica que es probable que haya alguna relación entre las variables. Es decir, el valor p de probabilidad del resultado estadístico señala la probabilidad de que ese resultado (o un resultado más extremo) pueda ocurrir si la hipótesis nula es cierta. Pero, no proporciona ningún tipo de información sobre la fuerza o magnitud de la relación (tamaño del efecto) y tampoco informa de si la relación detectada es útil o práctica (significación sustantiva social o clínica). Y, llegados a este punto conviene reflexionar de nuevo si la búsqueda de algún efecto satisface realmente las necesidades de la investigación científica.

Conectados con el punto anterior, se observa que a pesar de las recomedaciones del Manual de la American Psychological Association, aproximadamente el 48% de los investigadores y profesores españoles siguen utilizando expresiones como $p<0.05$ y $p>0.05$. De este modo, se otorga validez e importancia al mágico valor de 0.05 olvidando que dado un tamaño del efecto concreto, obtener un valor de probabilidad de 0.49 o de 0.06 puede suponer incluir cuatro o cinco sujetos más en la investigación. El conocimiento de los valores concretos de probabilidad vinculados a un resultado estadístico es imprescindible para facilitar la lectura crítica de los hallazgos y potenciar la Práctica Basada en la Evidencia entre los profesionales y expertos de Psicología y Educación.

Los resultados de nuestro estudio corroborán de nuevo el escaso impacto de la reforma estadística entre los investigadores y profesores universitarios españoles ya que casi el 40% de los encuestados no conoce ningún estadístico del tamaño del efecto. Por supuesto, partiendo de este dato resulta totalmente inviable pensar que los participantes conocen y actúan en sus estudios siguiendo las recomendaciones de la reforma estadística. Conviene anotar que dichas recomendaciones se encuentran en el Manual de la American Psychological

Association y, por lo tanto, son de obligado cumplimiento de los autores de los artículos. Entonces, ¿hasta qué punto las instrucciones a los autores de las normas editoriales de las revistas científicas españolas son tenidas en cuenta por los revisores o por los editores? Esta es una cuestión que ahora estamos analizando vaciando las publicaciones más relevantes de la Psicología y Educación.

En definitiva, urge la educación estadística centrada en el desarrollo del pensamiento de estimación y el pensamiento meta-analítico, enfatizando el uso del tamaño del efecto, los intervalos de confianza y los trabajos de meta-análisis. Educación que se puede lograr con la docencia, la actualización de los programas estadísticos como SPSS, las políticas editoriales de las revistas y la propia conducta del investigador. Muy probablemente estos cambios permitirán el desarrollo de teorías más elaboradas, mejorando el progreso de la investigación científica.

## 5. REFERENCIAS BIBLIOGRÁFICAS


Altman, D.G., & Bland, J.M. (1995) Statistics notes: absence of evidence is not evidence of absence. *British Medical Journal*, *311*, 485.

American Psychological Association (2001). *Publication Manual of the American Psychological Association* (5th Ed.). Washington, DC: American Psychological Association.

American Psychological Association (2010). *Publication Manual of the American Psychological Association* (6th Ed.). Washington, DC: American Psychological Association.

Carver, R. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education*, *61*, 287-292.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.

Cortina, J. M., &Dunlap,W. P. (1997). Logic and purpose of significance testing. *Psychological Methods*, *2*, 161-172.

Cumming, G., & Finch, S., (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, *61*, 532-575.

Fidler, F., Thomason, N., Cumming, G., Finch, S., & Leeman, J. (2004). Editors can lead

researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine. *Psychological Science, 15,* 119-126.

Fidler, F., Thomason, N., Cumming, G., Finch, S., & Leeman, J. (2005). Confidence intervals, still much to learn: Reply to Rouder & Morey. *Psychological Science, 16*, 494-495.

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, *22*, 700-725.

Frías-Navarro, D. (2011). Reforma estadística: tamaño del efecto En D. Frías-Navarro, *Técnica estadística y diseño de investigación*. Valencia: Ediciones Palmero.

Frías-Navarro, D., Badenes-Ribera, L., Pascual-Soler, M., & Monterde-i-Bort, H. (2013). Pensamiento meta-analítico: educación estadística. *XI Jornadas de Redes de Investigación en Docencia Universitaria 2013. Retos de futuro en la enseñanza superior: Docencia e investigación para alcanzar la excelencia académica*. Universidad de Alicante, 4 y 5 de julio.

Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, *1*, 379-390.

Gordon, H. R. D. (2001). American Vocational Education Research Association members' perceptions of statistical significance tests and other statistical controversies. *Journal of Vocational Educational Research, 26*, 1-18.

Hagen, R. L. (1997). In praise of the null hypothesis statistical test. *American Psychologist, 52*, 15-24.

Kaufman, A. S. (1998). Introduction to the special issue on statistical significance testing. *Research in the Schools, 5*, 1.

Mittag, K. C., y Thompson, B. (2000). A national survey of AERA members' perceptions of statistical significance tests and other statistical issues. *Educational Researcher, 29*, 14-20.

Monterde i Bort, H., Pascual, J., & Frías-Navarro, D. (2006). Errores de interpretación de los métodos estadísticos: importancia y recomendaciones. *Psicothema*, *18*, 848-856.

Monterde-i-Bort, H., Frías-Navarro, D., & Pascual, J. (2010). Uses and abuses of statistical significance tests and other statistical resources: A comparative study. *European Journal of Psychology of Education, 25*, 429-447.

Nelson, N., Rosenthal, R., & Rosnow, R. L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *American Psychologist*, *41*, 1299-1301.

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods, 5*, 2, 241-301.

Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioural sciences*. Chichester, England: Wiley.

Ogles, B. M., Lunnen, K.M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review, 21*, 421-446.

Rosenthal, R., & Gaito, J. (1963). The interpretation of level of significance by psychological researchers. *Journal of Psychology, 55*, 33-39.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44*, 1276–1284.

Schmidt, F. L. & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. En L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (p. 37-64). Hillsdale, NJ: Erlbaum.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods, 1*, 115-129.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *The Journal of Experimental Education, 61*(4), 293-316.

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik,&J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Erlbaum.

Tryon, W. W. (1998). The inscrutable null hypothesis. *American Psychologist, 53*, 796.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*, 779-804.

Zuckerman, M., Hodgins, H. S., Zuckerman, A., & Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 531 psychologists. *Psychological Science, 4*, 49-53.