

El tamaño del efecto en la publicación científica

En la actualidad existen numerosos manuales de estilo que recogen normas y recomendaciones para la elaboración y presentación de documentos científicos. Entre las normas más extendidas podemos citar: el estilo *Council of Science Editors* (CSE) y el *Harvard System of Referencing*, muy utilizados en biología, física y química; *The Chicago Manual Style*, de gran aplicación en historia y derecho; el *IEEE Standards Style Manual*, empleado en el ámbito de la ingeniería, la informática y la tecnología; el *MLA Handbook for Writers of Research Papers* habitual en las publicaciones de literatura, arte y humanidades; el *Publication Manual of the American Psychological Association* (APA), frecuente en el campo de las ciencias sociales y del comportamiento; y el *Uniform Requirements for Manuscripts Submitted to Biomedical Journals* (URM) o formato Vancouver que aplican las principales editoriales de revistas biomédicas. Cada uno de estos manuales ofrece directrices específicas que lógicamente se ajustan a las convenciones de los distintos campos científicos. Entre otras cuestiones, tienden a abordar aspectos como la estructura y organización del manuscrito, las indicaciones para presentar tablas y figuras, el formato de las citas y las referencias bibliográficas e, incluso, ofrecen recomendaciones éticas, legales y deontológicas.

Entre las revistas científicas de investigación educativa está muy generalizada la adopción del manual de publicación de la APA. Prueba de ello es que las 10 revistas españolas de temática educativa con factor de impacto internacional en la edición del año 2021 del *Journal Citation Reports* (categorías *Education & Educational Research*, *Education*, *Scientific Disciplines*, *Education*, *Special* y *Psychology*, *Educational*) hacen referencia en sus directrices para el envío de originales a las indicaciones recogidas en este manual. En el caso concreto de Educación XX1 se indica expresamente que, en la preparación del manuscrito, los autores deben seguir las normas de publicación de la APA en su 7ª edición.

Cómo referenciar este editorial:

López-Martín, E., & Ardura-Martínez, D. (2023). El tamaño del efecto en la publicación científica [The effect size in scientific publication]. *Educación XX1*, 26(1), 9-17. <https://doi.org/10.5944/educxx1.36276>

Sin embargo, y a pesar de la aceptación de estas recomendaciones, no en pocas ocasiones se observa cómo las citadas normas de estilo se aplican más a cuestiones de forma que de fondo. Es decir, mientras que se ha interiorizado la exigencia de que todos los manuscritos deben compartir una estructura y garantizar que sus elementos se presentan siguiendo un mismo formato, otras de las orientaciones establecidas en el Manual de publicaciones de la APA, como son las directrices sobre la información básica que debe incluirse en cada una de las secciones del manuscrito para que los autores puedan comunicar los resultados de sus investigaciones de manera clara, precisa y transparente (*Journal Article Reporting Standards, JARS*), no siempre son tenidas en cuenta. En este editorial reflexionaremos sobre uno de los criterios específicos para artículos de investigación cuantitativa (*Quantitative Design Reporting Standards, JARS-Quant*), referido a la presentación de los resultados, como es la necesidad de acompañar las pruebas de significación estadística, siempre que sea posible, de los tamaños del efecto estimados y su correspondiente intervalo de confianza.

No son pocos los originales que recibimos en Educación XX1 que no tienen en cuenta la recomendación de la APA (2020) de reportar el tamaño del efecto para que los lectores puedan apreciar la magnitud o importancia de los hallazgos del estudio y, este número se incrementa notablemente, si atendemos a la consideración de proporcionar un intervalo de confianza para cada tamaño del efecto, que informe de la precisión de la estimación de dicho tamaño. Sin embargo, esta no parece ser una cuestión nueva, ni que afecte solo a las revistas de investigación educativa (Famus et al., 2022; García et al., 2011; Sun et al., 2010; Wei et al., 2019). Aunque cada vez más se tiende a reportar el tamaño del efecto, existe un amplio margen de mejora, especialmente en lo que se refiere a su interpretación.

DE LA SIGNIFICACIÓN ESTADÍSTICA A LA SIGNIFICACIÓN PRÁCTICA. ¿POR QUÉ ES IMPORTANTE REPORTAR LOS TAMAÑOS DEL EFECTO Y CÓMO SE PUEDEN INTERPRETAR?

La inferencia estadística busca establecer conclusiones sobre las poblaciones a partir de la información extraída de las muestras. En concreto, las pruebas de significación de hipótesis nula (PSHN) permiten someter a evaluación las expectativas previas de los investigadores sobre el problema que está estudiando. En realidad, como su nombre indica, lo que se somete a prueba es la llamada hipótesis nula, que implica la ausencia de efecto o de relación entre las variables objeto de estudio. Gracias a la prueba estadística podremos rechazar o no rechazar esta hipótesis en la población a partir de los datos recogidos en la muestra, asumiendo un nivel de confianza previamente establecido.

Hace ya más de 100 años, el químico y matemático inglés Willian S. Gosset propuso por primera vez la utilización de las PSHN (Student, 1908). Desde entonces, la presencia de este tipo de pruebas estadísticas se ha generalizado como una herramienta de investigación en multitud de campos de conocimiento, y el de las ciencias de la educación no ha sido una excepción. Sin embargo, de manera prácticamente simultánea, fueron surgiendo voces críticas sobre su adecuación a la hora de tomar decisiones prácticas relacionadas con las hipótesis validadas, que han llegado hasta nuestros días (Boring, 1919; Funder y Ozer, 2019). Efectivamente, aunque las PSHN se utilizan de forma habitual en investigaciones cuantitativas de todo tipo, llevan apareadas diversas limitaciones (Thomson, 1996). En particular, algunos aspectos como la arbitrariedad en la elección del nivel de significación (α) o el hecho de que sus resultados sean dependientes del tamaño de la muestra empleada, han propiciado un debate en torno a la adecuación de estas pruebas a la hora de tomar decisiones a partir de los resultados que aportan. Por estas razones, entre otras, se cuestiona el hecho de que encontrar un resultado “estadísticamente significativo” implique necesariamente que pueda ser importante o valioso en la práctica. De hecho, una de las críticas que más habitualmente reciben las pruebas estadísticas es que relegan los juicios de los investigadores a un segundo plano, dejando las decisiones en manos de un mero cálculo matemático (Huberty y Morris, 1988).

Para superar estas limitaciones, como se ha indicado anteriormente, se recomienda reportar evidencias de los llamados índices de significación práctica, que están orientados a medir la magnitud o el tamaño de los efectos detectados gracias a la PSHN (Thomson, 2008). En otras palabras, el estudio de la significación práctica permite estimar en qué medida los estadísticos se desvían de lo supuesto a priori en el enunciado de la hipótesis nula.

Son numerosos los procedimientos propuestos para calcular la magnitud de los efectos y , en su mayor parte, se pueden clasificar en medidas de diferencias de medias — d , g , Δ , etc.— y en medidas de la fuerza de asociación — r , r^2 , h^2 , e^2 , w^2 , etc.— (Kirk, 1996; Rosnow & Rosenthal, 2003). Con el ánimo de que pueda servir de ayuda a aquellos autores y lectores de Educación XX1 menos familiarizados con los tamaños del efecto, en la Tabla 1 y en la Tabla 2 sintetizamos las medidas de tamaño del efecto que suelen acompañar a las pruebas estadísticas que principalmente se aplican en investigación educativa. Junto con las medidas del tamaño del efecto, se ofrecen los valores de referencia que se tienden a utilizar para su interpretación y que están basados en la clasificación propuesta por Cohen (1988, 1992).

Tabla 1
Tamaños del efecto asociados a las PSHN de mayor aplicación en investigación educativa

Magnitud analizada	Tipo de comparación	Estadístico asociado	Tamaño del efecto	Interpretación
Diferencias entre dos grupos	Proporciones	-	H de Cohen	< 0.20 muy pequeño
				0.20-0.49 pequeño
	Grupos independientes	T de Student ¹	D de Cohen (<i>d</i>), G de Hedges (<i>g</i>), Delta de Glass (Δ)	< 0.20 muy pequeño
				0.20-0.49 pequeño
	Grupos relacionados	T de Student ¹	D de Cohen (<i>d</i>), G de Hedges (<i>g</i>), Delta de Glass (Δ)	0.50-0.79 moderado
				≥ 0.80 grande
	Grupos independientes	Z (U de Mann-Whitney) ²	Correlación biserial por rangos (r_b)	< 0.10 muy pequeño
				0.10-0.29 pequeño
Grupos relacionados	Z (Prueba de Wilcoxon) ²	Correlación biserial por rangos (r_b)	0.30-0.49 moderado	
			≥ 0.50 grande	
Diferencias entre más de dos grupos	Grupos independientes	F (ANOVA) ¹	Eta al cuadrado (η^2), Epsilon al cuadrado (ϵ^2), Omega al cuadrado (ω^2)	< 0.01 muy pequeño
			0.01-0.05 pequeño	
	Grupos relacionados	F (ANOVA de medidas repetidas) ¹	Eta parcial al cuadrado (η_p^2), Eta cuadrado generalizado (η_G^2), Omega al cuadrado (ω^2)	0.06-0.13 moderado
				≥ 0.14 grande
	Grupos independientes	H de Kruskal-Wallis ²	Epsilon al cuadrado (ϵ_R^2), Eta al cuadrado (η_H^2)	< 0.01 muy pequeño
				0.01-0.05 pequeño
Grupos relacionados	χ^2 (Prueba de Friedman) ²	W de Kendall	0.06-0.13 moderado	
			≥ 0.14 grande	
Grupos relacionados	χ^2 (Prueba de Friedman) ²	W de Kendall	< 0.10 muy pequeño	
			0.10-0.29 pequeño	
Grupos relacionados	χ^2 (Prueba de Friedman) ²	W de Kendall	0.30-0.49 moderado	
			≥ 0.50 grande	

El tamaño del efecto en la publicación científica

Magnitud analizada	Tipo de comparación	Estadístico asociado	Tamaño del efecto	Interpretación																							
Relación entre dos variables	Variables cuantitativas	Coficiente de correlación de Pearson	R	< 0.09 muy pequeño 0.10-0.29 pequeño 0.30-0.49 moderado ≥ 0.5 grande																							
	Variables ordinales	Coficiente de correlación de Spearman	r _s	Puntos de corte propuestos para investigación psicológica: < 0.09 muy pequeño 0.10-0.19 pequeño 0.20-0.29 moderado ≥ 0.3 grande																							
	Tablas de contingencia	Chi-Cuadrado	V de Cramer	<table border="1"> <thead> <tr> <th>gl_{min}</th> <th>Pequeño</th> <th>Moderado</th> <th>Grande</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0.10</td> <td>0.30</td> <td>0.50</td> </tr> <tr> <td>2</td> <td>0.07</td> <td>0.21</td> <td>0.35</td> </tr> <tr> <td>3</td> <td>0.06</td> <td>0.17</td> <td>0.29</td> </tr> <tr> <td>4</td> <td>0.05</td> <td>0.15</td> <td>0.25</td> </tr> <tr> <td>5</td> <td>0.04</td> <td>0.13</td> <td>0.22</td> </tr> </tbody> </table>	gl _{min}	Pequeño	Moderado	Grande	1	0.10	0.30	0.50	2	0.07	0.21	0.35	3	0.06	0.17	0.29	4	0.05	0.15	0.25	5	0.04	0.13
gl _{min}	Pequeño	Moderado	Grande																								
1	0.10	0.30	0.50																								
2	0.07	0.21	0.35																								
3	0.06	0.17	0.29																								
4	0.05	0.15	0.25																								
5	0.04	0.13	0.22																								

Nota. Elaborado a partir de Cohen (1988, 1992), Funder y Ozer (2019), Gignac y Szodorai (2016), Kirk (1996), Morse (1999), Tomczak y Tomczak (2014) y Volker (2006).

¹ Contraste de hipótesis paramétrico; ² Contraste de hipótesis no paramétrico; gl_{min} = grados de libertad de las filas o de las columnas —el más pequeño de los dos—.

Conviene advertir que, frente al uso de los tradicionales puntos de corte para interpretar los tamaños del efecto, han surgido algunas voces críticas durante los últimos años que recomiendan contextualizar su interpretación, es decir, discutir los resultados con los hallazgos de otros estudios de características similares y valorarlos en función de su alcance (Bakker et al., 2019; Pek & Flora, 2018). Sin embargo, tampoco esta perspectiva parece estar exenta de limitaciones (Panzarella et al., 2021), ni de críticas (Simpson, 2018). En nuestra opinión, más allá de clasificar los tamaños del efecto en función de los puntos de corte tradicionales o de los específicos de una disciplina, animamos a los autores a comparar los tamaños del efecto con los reportados por otros estudios equiparables en lo que al diseño de investigación se refiere, favoreciendo el desarrollo de un pensamiento meta-analítico a la hora de interpretar y contextualizar la evidencia empírica derivada de sus investigaciones.

Tabla 2
Tamaños del efecto asociados a los modelos regresión

Magnitud analizada	Tipo de variable	Tipo de análisis	Tamaño del efecto	Interpretación
Capacidad explicativa (modelos de regresión)	Variable dependiente cuantitativa	Modelo de regresión lineal	<i>Modelo:</i> R ² / R ² ajustado	< 0.02 muy pequeño 0.02-0.12 pequeño 0.13-0.25 moderado ≥ 0.26 substancial
			<i>Predictores:</i> f de Cohen (f ²)	< 0.02 muy pequeño 0.02-0.14 pequeño 0.15-0.34 moderado ≥ 0.35 grande
	Variable dependiente cualitativa	Modelo de regresión logística	<i>Modelo:</i> Pseudo R cuadrado	R ² McFadden: 0.2-0.4 excelente ajuste
			<i>Predictores:</i> Odds Ratio (OR)	< 1.44 muy pequeño 1.44-2.47 pequeño 2.48-4.27 moderado ≥ 4.28 grande

Nota. Elaborado a partir de Cohen (1988, 1992) y McFadden (1977).

A MODO DE CONCLUSIÓN

Reportar los tamaños del efecto es una buena práctica científica que consiste, simplemente, en hacer lo correcto, y todos —editores, autores y revisores— debemos contribuir a garantizar este requisito mínimo (Hyde, 2001). De acuerdo con la profesora Blanco-Blanco (2018), resulta necesario adoptar definitivamente el hábito de reportar el tamaño del efecto y de su correspondiente intervalo de confianza, para contrarrestar un uso cuestionable de la inferencia estadística clásica como es la de aumentar el error tipo I, es decir, afirmar que una diferencia, un efecto o una relación es significativa, cuando en realidad no lo es. Esta autora hace un llamamiento a los editores de revistas científicas para que hagan explícitos unos estándares estadístico-metodológicos que deban ser asumidos por los autores.

Esta consideración que ya está presente en la política editorial de las revistas afiliadas a la APA (Wilkinson & Task Force on Statistical Inference, 1999) o a la *American Educational Research Association* (2006), entre otras, debe extenderse al conjunto de publicaciones científicas si lo que se pretende es asegurar la calidad de la investigación científica. El Equipo Editorial de Educación XX1 ha actualizado las normas de publicación para que en los manuscritos que recibamos en próximas

convocatorias los autores, a la hora de informar de los resultados de su investigación, reporten convenientemente los tamaños del efecto. Esperamos con ello contribuir a promover prácticas científicas más rigurosas.

Esther López Martín
Editora jefe de Educación XX1

Diego Ardua Martínez
Editor asociado de Educación XX1

REFERENCIAS BIBLIOGRÁFICAS

- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33-40. <https://doi.org/10.3102/0013189X03500603>
- Bakker, A., Cai, J., English, L., Kaiser, G. y Mesa, V. Beyond small, medium, or large: points of consideration when interpreting effect sizes. *Educational Studies in Mathematics*, 102, 1–8 (2019). <https://doi.org/10.1007/s10649-019-09908-4>
- Blanco Blanco, A. (2018). Estado de las prácticas científicas e investigación educativa. Posibles retos para la próxima década. *Revista de Educación*, (381), 207-232. <https://doi.org/10.4438/1988-592X-RE-2017-381-386>
- Boring, E. (1919). Mathematical vs. scientific importance. *Psychological Bulletin*, 16(10), 335-338. <https://doi.org/10.1037/h0074554>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155-159. <https://doi.org/10.1037//0033-2909.112.1.155>
- Farmus, L., Beribisky, N., Martinez Gutierrez, N., Alter, U., Panzarella, E., & Cribbie, R. A. (2022). Effect size reporting and interpretation in social personality research. *Current Psychology*, 1-11. <https://doi.org/10.1007/s12144-021-02621-7>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- García, J. G., Campos, E. O., & De la Fuente Sánchez, 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. http://doi.org/10.5209/rev_SJOP.2011.v14.n2.49

- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Hyde, J. S. (2001). Reporting effect sizes: The roles of editors, textbook authors, and publication manuals. *Educational and Psychological Measurement, 61*(2), 225–228. <https://doi.org/10.1177/0013164401612005>
- Huberty, C., & Morris, J.D. (1988). A single contrast test procedure. *Educational and Psychological Measurement, 48*(3), 567–578. <https://doi.org/10.1177/0013164488483001>
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746–759. <https://doi.org/10.1177/0013164496056005002>
- McFadden, D. (1977). *Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments*. Cowles Foundation Discussion Papers 474, Cowles Foundation for Research in Economics, Yale University. <https://cowles.yale.edu/sites/default/files/files/pub/d04/d0474.pdf>
- Morse, D. T. (1999). MINSIZE2: A computer program for determining effect size and minimum sample size for statistical significance for univariate, multivariate, and nonparametric tests. *Educational and Psychological Measurement, 59*(3), 518–531. <https://doi.org/10.1177/00131649921969901>
- Panzarella, E., Beribisky, N., & Cribbie, R. A. (2021). Denouncing the use of field-specific effect size distributions to inform magnitude. *PeerJ, 9*, e11383. <https://doi.org/10.7717/peerj.11383>
- Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology, 57*(3), 221–237. <https://doi.org/10.1037/h0087427>
- Student (1908). The probable error of a mean. *Biometrika, 6*(1), 1–25. <https://doi.org/10.2307/2331554>
- Sun, S., Pan, W., & Wang, L. 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. <https://doi.org/10.1037/a0019507>
- Thompson, B. (1996). Research news and comment: AERA editorial policies regarding statistical significance testing: three suggested reforms. *Educational Researcher, 25*(2), 26–30. <https://doi.org/10.3102/0013189X025002026>
- Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. Guilford Publications.
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in sport sciences, 1*(21), 19–25.

- Volker, M. A. (2006). Reporting effect size estimates in school psychology research. *Psychology in the Schools, 43*(6), 653-672. <https://doi.org/10.1002/pits.20176>
- Wei, R., Hu, Y., & Xiong, J. (2019). Effect size reporting practices in applied linguistics research: A study of one major journal. *SAGE Open, 9*(2). <https://doi.org/10.1177/2158244019850035>
- 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>