Educación XX1
ISSN: 1139-613X · e-ISSN: 2174-5374

The effect size in scientific publication

There are currently numerous style manuals that contain rules and recommendations for the preparation and presentation of scientific documents. Among the most widespread standards we can quote: the Council of Science Editors (CSE) style and the Harvard System of Referencing, widely used in biology, physics and chemistry; The Chicago Manual Style, widely used in history and law; the IEEE Standards Style Manual, used in engineering, computer science and technology; the MLA Handbook for Writers of Research Papers, common in literature, arts and humanities publications; the Publication Manual of the American Psychological Association (APA), frequently used in the social and behavioral sciences; and the Uniform Requirements for Manuscripts Submited to Biomedical Journals (URM) or Vancouver format, applied by the main biomedical journal publishers. Each of these manuals offer specific guidelines that logically conform to the conventions of the different scientific fields. Among other issues, they tend to address aspects such as the structure and organization of the manuscript, indications for presenting tables and figures, the format of citations and bibliographic references, and even offer ethical, legal and deontological recommendations.

Among scientific journals of educational research, the adoption of the APA publication manual is widespread. Proof of this is that the 10 Spanish educational journals with an international impact factor in the 2021 edition of the Journal Citation Reports (categories of *Education & Educational Research*, *Education, Scientific Disciplines*, Education, Special and Psychology, Educational) refer in their guidelines for the submission of manuscripts to the indications contained in this manual. In the specific case of Education XXI, it is expressly indicated that, in the preparation of the manuscript, authors should follow the APA publication guidelines in its 7th edition.

How to reference this editorial:
López-Martín, E., & Ardura-Martínez, D. (2023) The effect size in scientific publication. *Educación XX1*, *26*(1), 9-17. https://doi.org/10.5944/educxx1.36276

However, despite the acceptance of these recommendations, it is not infrequently observed that the aforementioned style guidelines are applied more to questions of form than of substance. In other words, while the requirement that all manuscripts should share a common structure and ensure that their elements are presented following the same format has been internalized, other guidelines established in the APA Publication Manual, such as the guidelines on the basic information that should be included in each section of the manuscript so that authors can communicate the results of their research in a clear, precise and transparent manner (Journal Article Reporting Standards, JARS), are not always taken into account. In this editorial we will reflect on one of the specific criteria for quantitative research articles (Quantitative Design Reporting Standards, JARS-Quant), referring to the presentation of the results, such as the need to accompany the statistical significance tests, whenever possible, with the estimated effect sizes and their corresponding confidence interval.

Many of the originals we receive in Education XXI do not take into account the recommendation of the APA (2020) to report the effect size so that readers can appreciate the magnitude or importance of the findings of the study, and this number increases notably if we consider providing a confidence interval for each effect size, which reports the precision of the estimate of the effect size. However, this does not seem to be a new issue, nor one that affects only educational research journals (Famus et al., 2022; Garcia et al., 2011; Sun et al., 2010; Wei et al., 2019). Although there is an increasing tendency to report effect size, there is ample room for improvement, especially in terms of its interpretation.

FROM STATISTICAL SIGNIFICANCE TO PRACTICAL SIGNIFICANCE. WHY IS IT IMPORTANT TO REPORT EFFECT SIZES AND HOW CAN THEY BE INTERPRETED?

Statistical inference seeks to draw conclusions about populations from information extracted from samples. Specifically, null hypothesis significance tests (NHSTs) allow researchers' prior expectations about the problem under study to be tested. In fact, as the name suggests, what is tested is the so-called null hypothesis, which implies the absence of effect or relationship between the variables under study. Thanks to the statistical test we can reject or not reject this hypothesis in the population from the data collected in the sample, assuming a previously established confidence level.

More than 100 years ago, the English chemist and mathematician William S. Gosset first proposed the use of the NHST (Student, 1908). Since then, the presence of this type of statistical test has become widespread as a research tool in many fields of knowledge, and the field of educational sciences has been no

exception. However, practically simultaneously, critical voices about its adequacy in making practical decisions related to validated hypotheses were emerging and have reached our days (Boring, 1919; Funder and Ozer, 2019). Indeed, although NHSTs are commonly used in quantitative research of all kinds, they have several limitations (Thomson, 1996). In particular, some aspects such as the arbitrary choice of the level of significance (α) or the fact that their results are dependent on the size of the sample used, have led to a debate about the appropriateness of these tests when making decisions based on the results they provide. For these reasons, among others, it is questioned whether finding a "statistically significant" result necessarily implies that it can be important or valuable in practice. In fact, one of the most common criticisms of statistical tests is that they relegate researchers' judgments to the background, leaving decisions in the hands of a mere mathematical calculation (Huberty and Morris, 1988).

To overcome these limitations, as indicated above, it is recommended to report evidence of the so-called practical significance indices, which are aimed at measuring the magnitude or size of the effects detected thanks to the NHST (Thomson, 2008). In other words, the study of practical significance makes it possible to estimate the extent to which the statistics deviate from what was assumed a priori in the statement of the null hypothesis.

Numerous procedures have been proposed to estimate the magnitude of the effects and, for the most part, they can be classified into measures of mean differences —d, g, Δ, etc. — and measures of the strength of association —r, r^2, h^2, e^2, w^2, etc.— (Kirk, 1996; Rosnow & Rosenthal, 2003). In the hope that it may be of help to those authors and readers of Education XXI less familiar with effect sizes, in Table 1 and Table 2 we synthesize the effect size measures that usually accompany the statistical tests that are mainly applied in educational research. Along with the effect size measures, we provide the reference values that tend to be used for their interpretation and that are based on the classification proposed by Cohen (1988, 1992).

Table 1

Effect sizes associated with the most widely applied NHSTs in educational research

Magnitude analyzed	Comparison type	Associated statistic	Effect size	Interpretation
Differences between two groups	Proportions	-	Cohen's H	< 0.20 very small 0.20-0.49 small 0.50-0.79 moderate ≥ 0.80 large
	Independent samples	Student's t[1]	Cohen's D (*d*), Hedges' G (*g*), Glass' Delta (Δ)	< 0.20 very small 0.20-0.49 small 0.50-0.79 moderate ≥ 0.80 large
		Z (Mann-Whitney U test)[2]	Rank-biserial correlation coefficient (r_b)	< 0.10 very small 0.10-0.29 small 0.30-0.49 moderate ≥ 0.50 large
	Paired samples	Student's t[1]	Cohen's D (*d*), Hedges' G (*g*), Glass' Delta (Δ)	< 0.20 very small 0.20-0.49 small 0.50-0.79 moderate ≥ 0.80 large
		Z (Wilcoxon test)[2]	Rank-biserial correlation coefficient (r_b)	< 0.10 very small 0.10-0.29 small 0.30-0.49 moderate ≥ 0.5 large
Differences between more than two groups	Independent samples	F (ANOVA)[1]	Eta-squared (η^2), Epsilon-squared (ε^2), Omega-squared (ω^2)	< 0.01 very small 0.01-0.05 small 0.06-0.13 moderate ≥ 0.14 large
		Kruskal-Wallis H[2]	Epsilon-squared (ε_R^2), Eta-squared (η_H^2)	< 0.01 very small 0.01-0.05 small 0.06-0.13 moderate ≥ 0.14 large
	Paired samples	F (Repeated measures ANOVA)[1]	Partial Eta-squared (η_p^2), Generalized Eta-squared (η_G^2), Omega-squared (ω^2)	< 0.01 very small 0.01-0.05 pequeño 0.06-0.13 moderate ≥ 0.14 grande
		χ^2 (Friedman test)[2]	Kendall's W	< 0.10 very small 0.10-0.29 small 0.30-0.49 moderate ≥ 0.50 large

<table>
<tr><th>Magnitude analyzed</th><th>Comparison type</th><th>Associated statistic</th><th>Effect size</th><th>Interpretation</th></tr>
<tr><td rowspan="3">Relationship between two variables</td><td>Quantitative variables</td><td>Pearson correlation coefficient</td><td>R</td><td>< 0.09 very small
0.10-0.29 small
0.30-0.49 moderate
≥ 0.5 large</td></tr>
<tr><td>Ordinal variables</td><td>Spearman correlation coefficient</td><td>r_s</td><td>Specified cut-off points proposed for psychological research:
< 0.09 very small
0.10-0.19 small
0.20-0.29 moderate
≥ 0.3 large</td></tr>
<tr><td>Contingency tables</td><td>Chi -Squared</td><td>Cramer’s V</td><td><table><tr><th>df_{min}</th><th>Small</th><th>Moderate</th><th>Large</th></tr><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></table></td></tr>
</table>

Note. Prepared by the authors based on Cohen (1988, 1992), Funder and Ozer (2019), Gignac and Szodorai (2016), Kirk (1996), Morse (1999), Tomczak and Tomczak (2014), and Volker (2006).
[1] Parametric hypothesis testing; [2] Non-parametric hypothesis testing; df_{min} = degrees of freedom of the rows or columns —the smaller of the two values—.

It should be noted that, faced with the use of traditional cut-off points to interpret effect sizes, some critical voices have emerged in recent years that recommend contextualizing their interpretation, i.e., discussing the results with the findings of other studies with similar characteristics and assessing them in terms of their scope (Bakker et al., 2019; Pek & Flora, 2018). However, neither does this perspective seem to be without limitations (Panzarella et al., 2021), nor without criticism (Simpson, 2018). In our opinion, beyond classifying effect sizes according to traditional or discipline-specific cut-off points, we encourage authors to compare effect sizes with those reported by other comparable studies in terms of research design, favoring the development of meta-analytic thinking when interpreting and contextualizing the empirical evidence derived from their research.

Table 2
Effect sizes associated with regression models

Magnitude analyzed	Variable type	Analysis type	Effect size	Interpretation
Explanatory power (regression models)	Quantitative dependent variable	Linear regression model	*Model*: R^2 / R^2 Ajusted	< 0.02 very small 0.02-0.12 small 0.13-0.25 moderate ≥ 0.26 large
			Predictors: Cohen's *f* (f^2)	< 0.02 very small 0.02-0.14 small 0.15-0.34 moderate ≥ 0.35 large
	Qualitative dependent variable	Logistic regression model	*Model*: Pseudo-R-squared	McFadden R^2: 0.2-0.4 excellent fit
			Predictors: Odds Ratio (OR)	< 1.44 very small 1.44-2.47 small 2.48-4.27 moderate ≥ 4.28 large

Note. Prepared by the authors based on Cohen (1988, 1992), and McFadden (1977).

IN CONCLUSION

Reporting effect sizes is a good scientific practice that consists, simply, in doing the right thing and we all —editors, authors and reviewers— must contribute to guarantee this minimum requirement (Hyde, 2001). According to professor Blanco-Blanco (2018), it is necessary to definitively adopt the habit of reporting the effect size and its corresponding confidence interval, to counteract a questionable use of classical statistical inference such as increasing type I error, i.e., claiming that a difference, an effect or a relationship is significant, when in fact it is not. This author calls on the editors of scientific journals to make explicit statistical-methodological standards that should be assumed by the authors.

This consideration, which is already present in the editorial policy of journals affiliated with the APA (Wilkinson & Task Force on Statistical Inference, 1999) or the American Educational Research Association (2006), among others, should be extended to all scientific publications if the aim is to ensure the quality of scientific research. The Editorial Team of Education XXI has updated the publication guidelines so that in the manuscripts we receive in future calls for papers, the authors, when

reporting the results of their research, should conveniently report the effect sizes. We hope this will help to promote more rigorous scientific practices.

Esther López Martín
Editor-in-chief of Educación XX1

Diego Ardura Martínez
Associated editor of Educación XX1

REFERENCES

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* (2st 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