Espiral. Cuadernos del Profesorado
Vol. 18, Issue 38
September 2025
ISSN 1988-7701

How to reference this article:

Rodríguez-Sabiote, C., Vázquez, L.M., López-López, J.A., & Sánchez-Martín, M. (2025). Effect size as an alternative to statistical significance testing. A practical example with JASP. *Espiral. Cuadernos del Profesorado, 18*(38), 129-141. https://doi.org/10.25115/ecp.v18i38.10234

Published on-line: 25 of September of 2025

Effect size as an alternative to statistical significance testing. A practical example with JASP[1]

El tamaño del efecto como alternativa a las pruebas de significación estadística. Un ejemplo práctico con JASP

Clemente Rodríguez Sabiote [1], Lindsay Michelle Vázquez [1], José Antonio López-López [2] y Micaela Sánchez-Martín [2]

[1] University of Granada, Granada, Spain; [2] University of Murcia, Murcia, Spain

"Statistical significance is not the same as practical importance" (David A. Friedman)

Abstract

The work addresses the criticism of the exclusive reliance on statistical significance (SS) in scientific research and proposes the use of effect size (ES) as a more robust and explanatory alternative. Two main weaknesses of SS are highlighted: its dependence on sample size and the arbitrariness of the p-value threshold (generally 0.05). These limitations can lead to misinterpretations and questionable practices such as "p-hacking" or data dredging. Effect size is presented as a quantitative measure of the magnitude of a phenomenon, independent of sample size, facilitating comparison between studies and diverse contexts. It is classified into various typologies: mean differences (Cohen's d, Hedges' g, Glass' Δ), correlations (Pearson's r, r^2), analysis of variance (η^2, ω^2, Cohen's f), and odds ratios (Odds Ratio, Risk Ratio). The interpretation of ES varies according to context, but general guidelines suggest that values such as 0.2, 0.5, and 0.8 in Cohen's d represent small, medium, and large effects, respectively. Finally, a practical example is included to illustrate the application of these measures and how SS and ES can lead to contrasting conclusions.

Keywords: Effect Sizes; Statistical Significance; JASP.

Resumen

El texto aborda la crítica a la dependencia exclusiva de la significación estadística en la investigación científica y propone el uso del tamaño del efecto (TE) como una alternativa más robusta y explicativa. Se destacan dos principales debilidades de la significación estadística: su dependencia del tamaño muestral y la arbitrariedad del umbral del valor p (generalmente 0.05). Estas limitaciones pueden llevar a interpretaciones erróneas y prácticas cuestionables como el "p-hacking" o dragado de datos. El tamaño del efecto se presenta como una medida cuantitativa de la magnitud de un fenómeno, independiente del tamaño de la muestra, que facilita la comparación entre estudios y contextos diversos. Se clasifica en varias tipologías: diferencias de medias (d de Cohen, g de Hedges, Δ de Glass), correlaciones (r de Pearson, r^2), análisis de varianza (η^2, ω^2, f de Cohen) y razones de probabilidades o cuotas (Odds Ratio, Risk Ratio). La interpretación del TE varía según el contexto, pero guías generales sugieren que valores como 0.2, 0.5 y 0.8 en d de Cohen representan efectos pequeños, medianos y grandes, respectivamente. Finalmente, se incluye un ejemplo práctico para ilustrar la aplicación de estas medidas y como la significación estadística y los tamaños del efecto pueden llegar a conclusiones contrarias.

Palabras clave: Tamaños del efecto; significación estadística; JASP.

Received: November 6, 2024
Accepted: January 30, 2025

Correspondence: Lindsay Michelle Vázquez, University of Granada, Spain

[1] JASP was chosen over JAMOVI (the software commonly used in methodological capsules), as it offers a wider range of effect size measures.

Email: lindsay@ugr.es

Key points

What is known

- Using statistical significance tests or hypothesis testing may not always be the most suitable method to identify differences in effects between groups.

What this work provides

- Fundamental concepts to assess potential differential effects between two groups (using a statistical significance test and calculating effect size) in a hypothetical situation, utilizing the *open-source software JASP.*

Practical scenario

A scale measuring anxiety levels was administered to 20 university students before the exam period of the first semester. Only the last item of the scale, or criterion item, was taken into consideration. The possible answers were: 1: no anxiety, 2: low anxiety, 3: moderate anxiety, 4: high anxiety and 5: very high anxiety. In addition, it should be noted that the firs 10 students do not practise any form of relaxation strategy, whereas the last 10 attend a yoga course organised by their university. Against this background, the following results were obtained:

a) No relaxation strategy practised: 1,2,2,3,3,3,3,4,4,5.
b) Attends the university-organised yoga course: 1,1,1,1,2,2,2,3,3,4.

The objective is to determine whether differential effects occur between the two groups based on anxiety levels. In order to meet this objective, it was determined that significance tests and effect sizes would be developed and the results compared.

Introduction

In scientific research, statistical significance has long served as a cornerstone for determining the relevance of results across various research methodologies. Traditionally, hypothesis testing and p-values have dominated decision-making processes, establishing an arbitrary threshold (typically 0.05) to decide whether a result is statistically significant, leading to the rejection of the null hypothesis when the value is equal to or below 0.05 ($p \leq 0.05$). However, this approach has faced increasing criticism for its inability to adequately represent effect size and the intrinsic uncertainty in the data, leading researchers to seek more robust and exploratory alternatives. Effect size can provide valuable complementary information alongside statistical significance. Additionally, Bayesian statistics may be emphasised, as its relevance warrants a dedicated discussion that could be incorporated in a future training module similar to the present one.

Main Limitations of Statistical Significance

Numerous studies have raised concerns about the reliability of statistical significance due to its limitations. In this context, the work of Fernández-Cano and Fernández-Guerrero (2009) stands out as particularly scholarly, comprehensive, and exhaustive, especially within the context of educational research. Without delving too deeply into this work, we highlight, along with Barriopedro (2015), Fernández-Cano and Fernández-Guerrero (2009), Kirk (2002), Onwuegbuzie and Levin (2003), Pascual Llobel et al. (2004), and Silva-Acayguer (2016), among others, the two main reasons why we should reconsider the exclusive use of statistical significance.

The primary reason is the dependence of statistical significance on the sample size under analysis. This is due to the fact that *p-value* is influenced by both the effect size and the sample size. A very small effect can be statistically significant with a sufficiently large sample size, while conversely, even a large effect may not be statiscally significant with a small sample size. In contrast, effect size

provides a direct measure of the magnitude of a difference or relationship, independent of sample size, providing a clearer understanding of the practical importance of the results.

The second major reason is the arbitrariness of the *p-value* threshold (0.05). In this respect, Rosnow and Rosenthal (1989) remarked that "surely God loves 0.06 almost as much as 0.05". The dichotomous decision that arises from establishing a specific arbitrary value increases the pressure on researchers to surpass this threshold. This often leads to "p-hacking" or data dredging, wherein researchers may manipulate data or study conditions to achieve statistical significance. This does not even account for the potential discord between statistical significance and substantive significance (Rodríguez-Sabiote et al, 2001).

Effect Size as a Substitute for Statistical Significance

Concept

The concept of effect size has its roots in the development of statistics and psychometrics in the 20th century. Jacob Cohen, a psychologist and statistician, was instrumental in popularising the concept of effect size, particularly through his 1969 book, Statistical Power Analysis for the Behavioral Sciences. Before Cohen's contributions, the primary focus in applied statistics was on statistical significance. Nevertheless, Cohen advocated for the importance of assessing the magnitude of observed effects rather than merely determining whether these effects were statistically significant.

In this context, effect size is a quantitative measure of the magnitude of a phenomenon. It indicates the scope or intensity of a relationship between variables, a difference between groups, or a change in a variable over time. Unlike the p-value, which only suggests whether an effect is likely to exist, effect size provides a measure of practical, substantive, or clinical impact. We use these different terms (practical, substantive, or clinical) to refer to the measure of impact conveyed by effect sizes, as they indicate relevant impacts beyond statistical significance, depending on the field in which the effect is applied. Regardless of the discipline in which we apply the tool of effect size, certain distinctive features should be highlighted, namely:

Magnitude: It provides a measure of how large an effect is, independent of sample size.

Therefore, and as a consequence of the above, we can identify the Independence of the Sample Size as a distinguishing feature: unlike the p-value, effect size does not depend on sample size.

Comparability: Effect size enables comparisons across different studies and contexts, facilitating meta-analysis, given that an effect size is a standardised z-score, at least in the case of the comparison of means.

Interpretability: effect size offers a more intuitive and practical interpretation of statistical results.

Typologies of Effect Size

There is a relatively broad consensus within the scientific community, which has focused on the application and interpretation of effect sizes, with Coe and Merino-Soto (2002), Cohen (2008) and Morales-Vallejo (2008), among others, stating that these effect sizes can be classified as outlined below. However, readers are advised that this is not intended to be a unique or exhaustive taxonomy.

Mean Difference-Based Effect Sizes

a) Cohen's *d*: measures the standardised difference between two means using the pooled standard deviation of the groups. It is useful for moderate to large sample sizes.

b) Hedges' *g*: similar to Cohen's d, but includes a correction for bias in small sample sizes.

c) Glass's Δ: uses the standard deviation of the control group, and is appropriate when the standard deviations of the groups differ significantly. This approach assumes that the standard deviation of the control group is considered to better reflect the measurement scale (as it is not affected by treatment effects).

d) Unstandardised mean difference: presents results on the original scale, which can facilitate interpretation. This option is advisable in cases where standardisation is not deemed necessary or recommended.

Correlation-based effect sizes

a) Pearson's *r*: measures the strength and direction of the linear relationship between two quantitative variables, with an interpretative range from -1 to +1. Other indices adapted for non-quantitative variables have similar interpretative meanings.

b) R^2 (Coefficient of Determination): the proportion of variance in the dependent variable explained by the independent variable.

ANOVA-based side effects

a) η^2 (Eta Squared): measures the proportion of total variance attributable to a factor, indicating effect size in terms of percentage of variance explained.

b) ω^2 (Omega Squared): similar to η^2 but adjusts for bias in small sample sizes, providing a more realistic estimate of effect size.

c) Cohen's *f*: used in ANOVA to measure the magnitude of differences between means, adjusted for within-group variance.

Probability or ratio-based effect sizes

a) *Odds Ratio*: calculated as the ratio of the odds of an event occurring in one group compared to another group. Commonly used in case-control studies and other designs.

b) *Risk Ratio*: the ratio of the risk (i.e., relative frequency or probability) of an event between two groups, often used in cohort studies and similar designs.

c) *Risk Difference*: the difference in the risk of an event between two compared groups. The inverse value of this index is known as the Number Needed to Treat (NNT), which is a useful indicator of the impact of an intervention. Used in cohort studies and related designs.

Interpreting Effect Sizes

The interpretation of effect size depends on the context and discipline in which it is applied. However, there are general guidelines that can serve as a starting point for interpretation, though they should not be considered absolute. According to Cohen (1992), it can be stated that Cohen's *d* and other similar indices (such as Hedges' *g* and Glass's Δ) may be classified as follows: Small (0.2): indicates a small yet significant effect; Medium (0.5): represents a moderate effect; and Large (0.8): indicates a large, easily observable effect. Nevertheless, at the end of this work, a table with more clearly labelled effect size intervals will be proposed for improved interpretative accuracy.

For Pearson's *r* (absolute value): Small (0.1 to 0.25): Weak correlation; Moderate (0.25 to 0.40): Moderate correlation; and Strong (>0.40): Strong correlation.

Finally, regarding effect sizes based on percentages of explained variance, a value around 0.10 is typically considered indicative of a low-to-moderate effect, while a value around 0.25 denotes a large effect.

Developing the Proposed Practical Example with JASP

Accessing the Previously Created Data Matrix

Once the data template or matrix has been created based on the practical scenario considered, open the JASP software and click on the three horizontal lines in the upper left corner.

Figure 1

Home screen of the JASP program

Once there, you need to navigate to the location of the file of interest, which in this case is prepared in SPSS format (ESPIRAL.sav), fully compatible with JASP. After locating the file, proceed to open it.

Figure 2

File opening screen

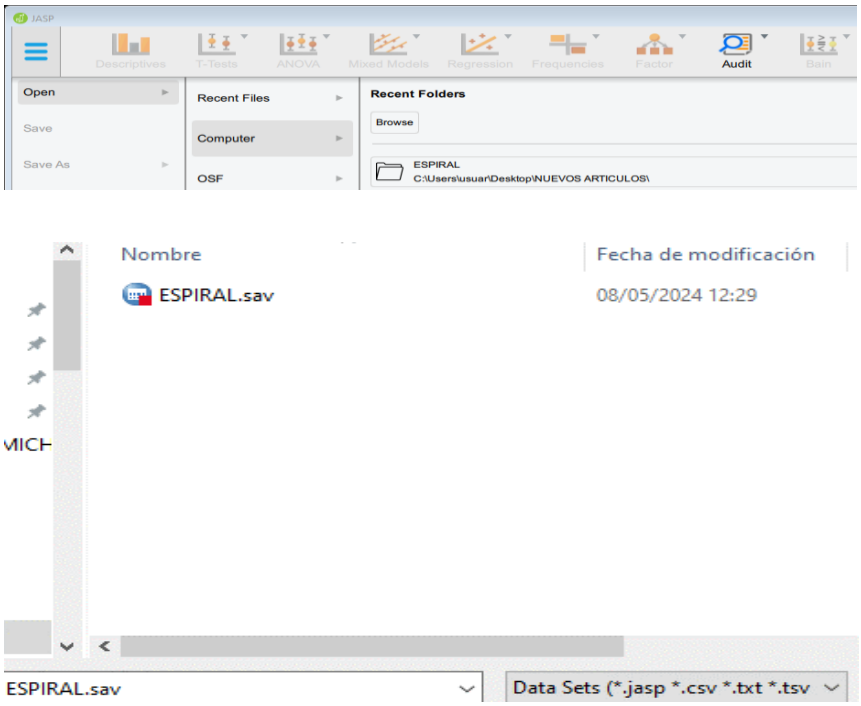
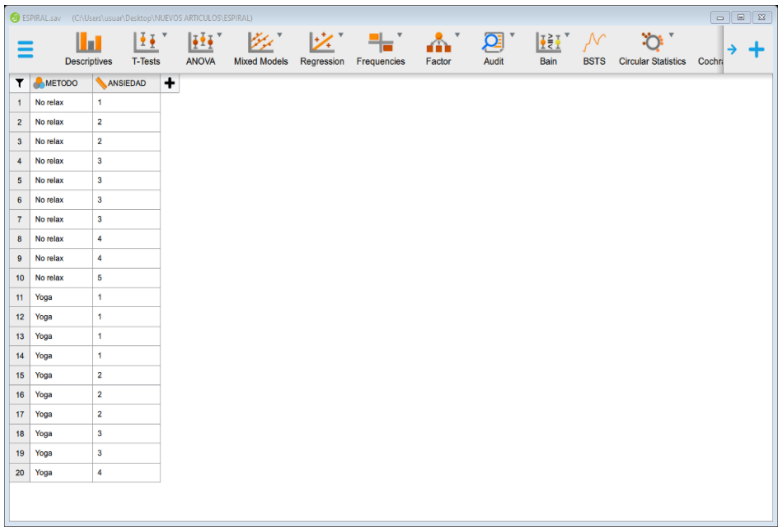


Figure 3

Open file screen



How to Access the Analysis Procedure to Address the Objective in the Practical Scenario

As an introduction, the arithmetic means of each group will be calculated. This data will assiste in interpreting the results obtained. To do this, navigate to → *Descriptives* → *Descriptive Statistics* → Place *ANSIEDAD* in the Variables window and *METODO* in the *Split* window. Finally, select only *Mean* to obtain Table 1.

Figure 4

Descriptive Statistics screen

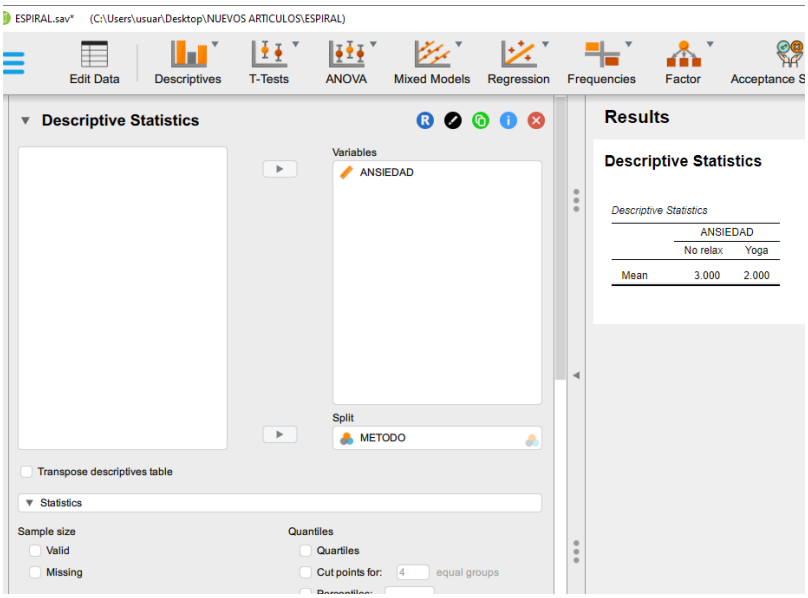


Table 1

Arithmetic means of the two groups (no relax vs. yoga) in the anxiety variable.

	Group	N	Mean
ANXIETY	No relax	10	3.000
	Yoga	10	2.000

Note. *N* = sample size; *Mean* = average

Furthermore, to conduct the statistical significance tests, as well as to calculate effect sizes, navigate to:

→*T-Test*→*Independent Sample t-test*

Figure 5

Access to the Independent Samples t-test screen

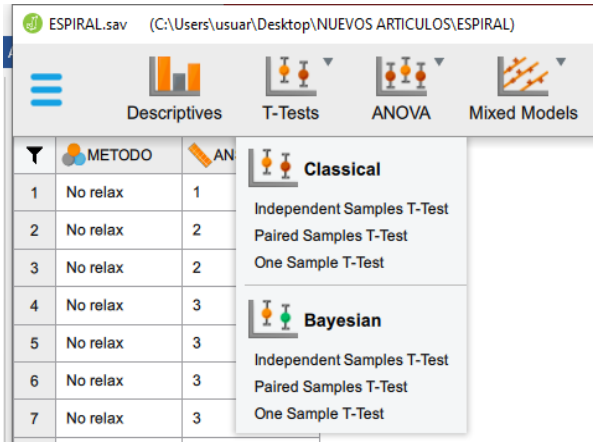
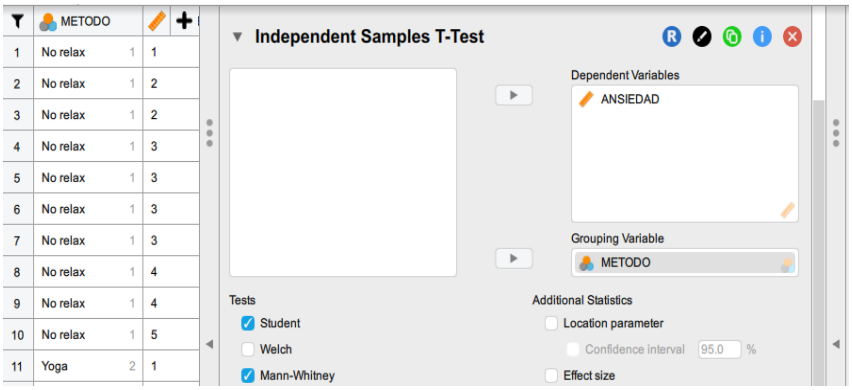


Figure 6

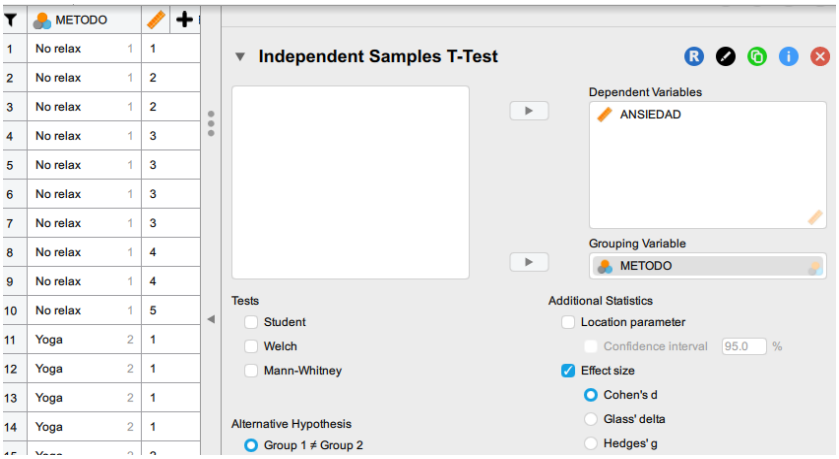
Independent Samples t-test screen



After this action, you should select the Student's t-tests (parametric in nature[2]) and their non-parametric alternative (Mann-Whitney U). Additionally, as an initial sample, although this will be developed in more detail later, you may also select *Effect Size* with the *Cohen's d* statistic.

Figure 7

Statistical Significance Tests screen



After these actions, the results will be displayed in Table 2.

Table 2

Results of the statistical significance tests (student's t- tests and Mann-Whitney U)

	Test	Statistic	df	*P*
ANXIETY	Student	2.023	18	0.058
	Mann-Whitney	74.000		0.067

Note. *df* = degrees of freedom; *p*= significance value

If the focus is on implementing alternative measures to Cohen's d-size, you could additionally select the statistics *Glass's delta* and *Hedges'g*. These statistics differ in the way the denominator of the above expression (Cohen's d) is calculated, but conceptually they also represent a standardisation of the observed difference. The results after this action would be presented in Tables 3, 4, and 5.

Table 3

Description of table content

	Independent Samples T-Test				
	T	df	*p*	Cohen's d	SE Cohen's d
ANXIETY	2.023	18	0.058	0.905	0.491

Note. Student's t-test; *df* = degrees of freedom; *p*= significance value; *SE*= standard error

Table 4

Results of the Effect Size Test Based on Glass's Δ

	Independent Samples T-Test				
	t	df	p	Glass' delta	SE Glass' delta
ANSIEDAD	2.023	18	0.058	0.949	0.495

Note. Glass' delta uses the standard deviation of group Yoga of variable METODO.

Note. Student's t-test; *df* = degrees of freedom; *p*= significance value; *SE*= standard error

[2] Tests where the assumptions of normality, independence, and, where applicable, homoscedasticity (or sphericity, in the case of a repeated measures ANOVA F-test) must be met.

Table 5

Results of the Effect Size Test Based on Hedges' g

	Independent Samples T-Test				
	t	df	p	Hedges' g	SE Hedges' g
ANSIEDAD	2.023	18	0.058	0.866	0.467

Note. Student's t-test. *df* = degrees of freedom; *p*= significance value; *SE*= standard error

Interpretation of Results

The interpretation of the results will be structured around four tables, beginning with Tables 1 and 2. These tables show that both the Student's t-statistic = 2.02 and Mann-Whitney U-statistic = 74, produced statistical probabilities of p > 0.05 with a two-tailed 95% confidence level and 18 degrees of freedom (df, 20 participants - 2). Specifically, p = 0.058 for the t-test and p = 0.064 for the Mann-Whitney U test.

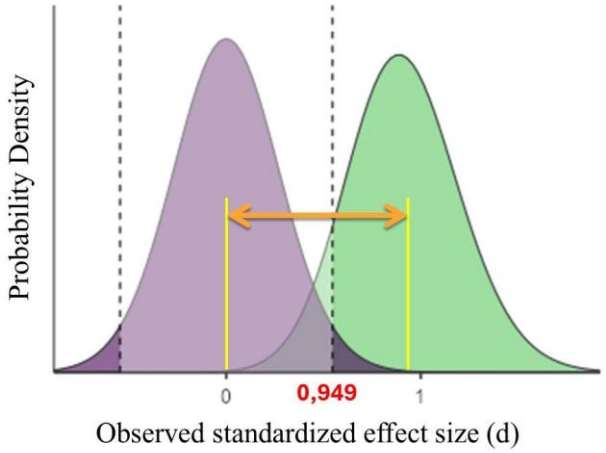
Given these p-values, there is no statistical evidence to asset that the means obtained (yoga group = 2 vs no-relax group = 3) are statistically different, as both probabilities exceed the 0.05 threshold. It should be noted that the values of the dependent variable range from 1 to 5, and the difference between the two groups—considering the minimum and maximum score range—is actually substantial, as it amounts to 1 point. Even so, the traditional approach, that is, statistical significance, through both parametric and non-parametric tests, has concluded that the 1-point difference is not substantial enough to conclude that the yoga training was effective in reducing anxiety.

However, consideration of effect sizes (as shown in Tables 3 to 5) suggests an alternative perspective. The three effect size measures calculated—Hedges' *g* (0.866), Cohen's *d* (0.905), and Glass's Δ (0.949)—indicate large and easily observable effects. Based on these results, it may be concluded, in contrast to earlier findings, that the yoga group may indeed exhibit a substantially lower mean anxiety level (mean = 2) than the no-relax group (mean = 3).

For further clarity, given that an effect size is essentially a standardised z-score, it can be stated that the means of the no-relax group and the yoga group differ by almost z = 1, or one standard deviation (taking, for instance, Glass's Δ = 0.949 as a reference). If this difference were visualized graphically with two normal distributions for each group, the distance between the curves would be evident, as shown in Figure 8.

Figure 8

Graphical representation of the standardised distance between two means using normal distributions, based on the implemented example and the value of Glass's Δ = 0.949.



It should be noted that the more overlapped the curves are, the smaller the standardised difference between the groups represented by these curves (smaller effect size). Conversely, the greater the separation between the curves, the larger the standardised difference between the groups (larger effect size).

At this point, we conclude that the obtained effect sizes are large, indicating a standardised distance between the group means (yoga = 2 and no-relax = 3) of nearly one standard deviation. However, it is still not entirely clear how substantial this distance is and why it might be considered large. To clarify this, the relationship between effect sizes, in terms of percentiles and the percentage of non-overlap between distributions, will be examined as shown in Table 6.

Table 6

Interpretative labels for effect sizes in terms of percentiles and percentage of non-overlap between distributions

Interpretative label	Effect size	Percentile	Non overlapping %
Small	0.0	50	0
	0.1	54	7.7
	0.2	58	14.7
	0.3	62	21.3
Medium	0.4	66	27.4
	0.5	69	33
	0.6	73	38.2
	0.7	76	43.0
Large	0.8	79	47.4
	0.9	82	51.6
	1.1	84	55.4
	1.2	86	58.9
	1.3	88	62.2
	1.4	90	65.3
	1.5	91.9	68.1
	1.6	93.3	73.1
	1.7	94.5	75.4
	1.8	95.5	77.4
	1.9	96.4	79.4
	2	97.1	81.1
	2.5	99.0	87.7
	3	99.9	92.6

Source: Cohen (1992).

In this case, let's assume we take Cohen's d value of 0.905 as a reference due to its proximity to 0.9. With this result, we can assert that the resulting effect size is large; 82% of participants in the no-relax group are above the mean anxiety level of the yoga group. Finally, the percentage probability (non-overlapping) that a participant from the yoga group has a lower anxiety level than someone from the no-relax group rises to 51.6%.

Solution to the Practical Scenario

A careful reading of this paper may reveal insights you had not previously considered. Understanding data analysis procedures for comparing groups beyond statistical significance likely opens up new avenues that complement, if not entirely replace, traditional approaches in complex or controversial situations. The limitations inherent in statistical significance highlight the importance of supplementing it with other tools, such as effect sizes (discussed in this paper), along with Bayesian statistics and power analysis for tests, which are undoubtedly valuable and intriguing areas for future exploration. In this context, it is essential to remember that when your scientific interest is focused on identifying potential differential effects between groups, as illustrated in this case, classical statistical methods should be employed. However, do not overlook the importance of complementing these with effect size measurements. Our recommendation stems from the potential for discrepancies between conclusions drawn from analyses based on statistical significance and those based on effect size magnitude. In this study, having reported such discrepancies, we have chosen to prioritise the effect size approach, concluding that students who participated in the yoga intervention exhibit substantially lower anxiety levels, which makes this an important consideration.

Conclusions

Hypothesis testing or significance testing, whether parametric or non-parametric, remains the primary reference option when attempting to determine differential effects between groups. While this approach is viable, legitimate, and statistically sound, it must be applied with caution, particularly in cases like those presented in this study, where small sample sizes can prevent large effects from reaching statistical significance. Conversely, with larger sample sizes, even small effects may appear statistically significant.

The potential divergence in conclusions drawn from classical significance tests versus effect size analysis suggests that researchers comparing groups should critically examine significance test results and assess whether effect size measures confirm the presence of substantive differential effects.

This is precisely the situation observed in the present case: the presence of a large effect that fails to reach statistical significance due to a small sample size. This finding is substantiated by effect sizes measures, which have demonstrated the existence of a substantial effect worth considering. In cases of such divergence, the researcher should be free to choose the approach that best reflects the data. However, in light of the limitations of statistical significance, it is generally more reasonable to prioritise effect size analysis or, alternatively, to employ another robust option, such as Bayesian statistics—a topic that will be discussed in future research.

Contributions of each author: Conceptualisation: C.R.S., J.A.L.L., and M.S.M.; Manuscript Writing: C.R.S., L.M.V., and J.A.L.L.; Writing, Review, and Editing: C.R.S., L.M.V., J.A.L.L., and M.S.M.; Supervision: C.R.S., L.M.V., J.A.L.L., and M.S.M.

Funding: This research has received funding from the Ministry of Science, Innovation and Universities, University Teacher Training (FPU22/01938).

Conflict of interests: The authors declare that they have no conflicts of interest.

References

Barriopedro, M. (2015). La significación estadística no es suficiente. *RICYDE. Revista Internacional de Ciencias del Deporte, 11,* (40) 101-103. DOI: http://dx.doi.org/10.5232/ricyde2015.040ed

Coe, R. y Merino Soto, C. (2003). Magnitud del Efecto: Una guía para investigadores y usuarios. *Revista De Psicología*, *21*(1), 145-177. https://doi.org/10.18800/psico.200301.006

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, B.H. (2008). *Explaining psychological statistics*. John Wiley & Sons.

Fernández-Cano, A. y Fernández-Guerrero, I. (2009). *Crítica y alternativas a la significación estadística en el contraste de hipótesis*. La Muralla.

JASP Team (2024). *JASP (Version 0.18.3)* [Computer software].

Kirk, R.E. (2001). Promoting Good Statistical Practices: Some Suggestions. *Educational and Psychological Measurement, 61* (2), 213-218.

Morales Vallejo, P. (2008). *Estadística Aplicada a las Ciencias Sociales.* Universidad Pontificia Comillas.

Onwuegbuzie, A.J. y Levin, J. R. (2003). Without supporting statistical evidence, where would reported measures of substantive importance lead? To no good effect. *Journal of Modern Applied Statistical Methods*, *2*, (1), 133-151.

Pascual Llobell, J., Frías Navarro, M.D. y García Pérez, F. (2004). Usos y abusos de la significación estadística: propuestas de futuro (¿Necesidad de nuevas normativas editoriales?). *Metodología de las Ciencias del Comportamiento, Volumen Especial,* 465-469.

Rodríguez-Sabiote, C., Lorenzo-Quiles, O. y Navarro-Hernández, J. J. (2001). Un ejemplo práctico sobre la discordancia entre la significación estadística y la significación sustantiva con relación a la decisión de la hipótesis nula. *Publicaciones*, *31*, 173–186.

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

Silva-Acayguer, L. (2016). Las pruebas de significación estadística. Seis décadas de fuegos artificiales. *Revista Nacional de Salud Pública*, *34* (3), 372-379.

Effect Size as an Alternative to Statistical Significance Tests. A Practical Example with JASP

Why look for alternatives to Statistical Significance (SS)?

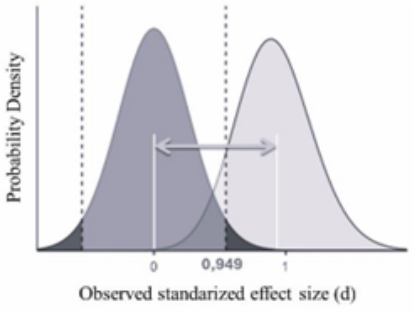
Dependence on Sample Size: It may lead to erroneously rejecting or accepting hypotheses if the sample size is too large or too small.

Arbitrariness of the p-value Threshold: The common threshold of 0.05 is arbitrary and may not reflect the practical importance of the results.

Alternative Proposal

Effect Size (ES)

What is it?



What use it?

A quantitative measure that evaluates the magnitude of a phenomenon, offering a perspective independent of sample size and facilitating the comparison of results across different studies and contexts.

Robustness

ES provides a more stable and representative measure of the results, without depending so much on sample size.

Explanatory

Helps to understand the actual magnitude of the observed effect, providing a richer context than a simple "yes/no" from SS.

Effect Size (ES)

How is it classified?

Mean Differences
Cohen's d
Hedges' g
Glass' Δ
Correlations
Pearson's r
Coefficient of Determination

ANOVA
Eta Squared (η^2)
Omega Squared (ω^2)
Cohen's f
Odds and Risk
Odds Ratio
Risk Ratio

How is it interpreted?

General Guidelines
Example – Cohen's d:
0.2 = small
0.5 = medium
0.8 = large

Practical Application
Examples of how ES and SS may lead to opposite conclusions.