DOI: https://doi.org/10.3126/jiaas.v37i1.56976

VITALITY AND APPLICATION OF EFFECT SIZE FOR QUALITY RESEARCH

Chuda Dhakal*, PhD.
Institute of Agriculture and Animal Science, Kirtipur, Kathmandu
* Corresponding author email: chuda.studies@gmail.com
https://orcid.org/0000-0001-5810-9103

ABSTRACT

This review article highlights the importance of effect size in research. P-value alone is not sufficient to determine the practical significance of a study, making effect size an essential component in hypothesis testing. Choosing the appropriate effect size for a specific study design can be challenging, and its interpretation may require modifications and personal judgement. Therefore, researchers should exercise caution when reporting and interpreting effect size, as it provides valuable information about the practical significance of their study, complementing hypothesis testing results. In conclusion, effect size should not be overlooked and should be carefully chosen and interpreted to ensure the validity and reliability of research results.

Keywords: P-value, practical significance, Cohen's d benchmarks, confidence interval.

INTRODUCTION

Answering research questions relying solely on p-values derived from statistical tests can be misleading, according to Dunkler et al. (2020). Hypothesis testing results in either being significant (p-value<0.05), suggesting that the effect is likely due to a factor of interest, or not significant (P-value >0.05), suggesting that the effect is likely due to chance. This dichotomy provides an explanation for the cause of the effect but not for its strength or weakness. This does not give readers enough information to determine if the study results are practically relevant or not. Sun *et al.* (2010) note that the limitation of hypothesis testing is its inability to describe the size of an effect in an experiment. Effect size, a quantitative measure of the strength of a phenomenon, emphasizes the size of the difference between groups or the relationship between two variables (Becker, 2000; Huberty, 2002). It provides insight into the practical importance of the results by quantifying the magnitude of the effect.

Durlak (2009) highlights the need to describe the magnitude of research results, which is not emphasized by hypothesis testing. Several studies have discussed the shortcomings, controversies, insufficiencies, and misconceptions of hypothesis testing, including Levine *et al.* (2008), Szucs and Ioannidis (2017), Hypothesis testing (2011), Null hypothesis significance testing (n.d.), Limitations of significance testing (2015), and Castillo & Torquato (2018). These studies emphasize that the P-value in hypothesis testing does not provide information on the magnitude of differences found in research and only indicates if findings are due to chance or sampling error. Carpenter (2020) also notes that there is no straightforward relationship between p-values and the magnitude of effect, as significant p-values can have little practical importance while insignificant p-values can have significant practical significance. Therefore, decisions based solely on hypothesis testing may not be adequate.

Multiple sources advocate combining hypothesis testing with techniques that focus on the magnitude of the difference between groups or the relationship between variables, referred to as effect size, rather than relying solely on hypothesis testing. These sources include Nickerson (2000), Hypothesis Testing: Methodology and Limitations (2001), Hypothesis testing (2011), Limitations of significance testing (2015), Null hypothesis significance testing (n.d.), and Pernet (2016). Szucs and Ioannidis (2017) and Huberty (2002) emphasize that hypothesis testing should no longer be the sole method for testing treatment effects in experiments.

This article emphasizes the importance of effect size and provides a clear understanding of standard measures of effect size, such as the difference between two groups and the relationship between two variables. It also familiarizes the reader with various options for calculating effect size and highlights important considerations for choosing the appropriate measure. Finally, it serves as a comprehensive introduction for readers interested in incorporating effect size into their research, providing guidelines for estimating and interpreting effect size with caution.

WHY REPORT EFFECT SIZE

Effect size is important in research because it helps indicate the practical significance of the results, not just the statistical significance (Sullivan & Feinn, 2012). Nakagawa and Cuthill (2007) explain that effect size quantifies the size of experimental effects and helps researchers understand the practical importance of their findings. To enhance the quality of a study, adequate estimates of effect size should be reported, as emphasized by Lakens (2013).

LeCroy and Krysik (2007) argue that effect size measures provide a more comprehensive understanding of a study's results, as they take into account both the statistical and practical significance of the findings. In experiments, the size of the effect helps researchers determine the practical significance of the treatment (Sun *et al.,* 2010). It's crucial to report effect size because the consequences of not doing so can be detrimental (Sun *et al.,* 2010). A small P-value does not necessarily indicate a significant practical effect, and a large P-value may be due to limited statistical power (Sullivan, 2012). Therefore, effect size should be reported in all studies, regardless of statistical significance (Thompson, 2007, as cited in Lakens, 2013).

Including effect size in a study has several benefits, such as enabling researchers to project adequate sample sizes, inform judgement about practical significance, and compare results with other studies (Sun *et al.,* 2010). The fifth edition of the Publication Manual of the American Psychological Association (2001) states that including an index of effect size or strength of the relationship in the results section is almost always necessary for the reader to fully understand the importance of the findings. Daniel (2017) also emphasizes the importance of including an index of effect size in the results section, and Sullivan & Feinn (2012) recommend reporting both the statistical significance (p-value) and the practical significance (effect size) for a complete understanding of the study's impact.

MEASURES OF EFFECT SIZE

Effect size is a way to quantify the magnitude of an experimental effect. It shows how big or small the difference is between two groups or the relationship between dependent and independent variables. Effect size is important because it indicates the practical significance of a research finding. A large effect size means the finding is significant, while a small effect size indicates limited practical applications (Bhandari, 2022). It is important to know the magnitude of effect size because a smaller effect on one outcome can be more important than a larger effect on another outcome (Durlak, 2009).

Effect size and statistical significance are not the same. Statistical significance is about the likelihood that a result is due to chance, while effect size is about the importance of the result. Significance tests depend on sample size, with larger sample sizes increasing the likelihood of a significant treatment effect. On the other hand, effect size indices are not dependent on sample size and can act as population parameters (Dunker *et al.,* 2020). It is important to calculate effect size statistics regardless of P-value and to report it with an estimate of precision, such as a 95% confidence interval.

Effect size can be either absolute or standardized. Absolute effect size is the difference between the means of two groups, while standardized effect size is a dimensionless statistic useful for comparing different studies (Sullivan & Feinn, 2012). The type of effect size used depends on the study design. For example, if the unit of measurement is meaningful, absolute effect size is used. However, if the study is based on population mean and standard deviation, then standardized mean difference method is used to determine effect size.

There are many measures of effect size, divided into two broad groups: mean difference between groups and strength of relation between variables (Sun *et al.,* 2010). Mean difference effect sizes, such as *Cohen's d*, Glass's Δ, and Hedges's g, are based on the standardized group mean difference. The strength of relations is based on the proportion of variance accounted for (r^2) in the correlation between two variables. When conducting a study, it is important to choose the appropriate effect size measure (Sullivan & Feinn, 2012). The method used to calculate effect size will vary depending on the study design.

Effect size of the Difference Between Two Groups

The effect size in research is a way to quantify the difference between two group means. It is calculated by dividing the difference between the two means by the standard deviation. The larger the effect size, the more impactful the study results are, meaning a large effect size suggests an important difference, while a small effect size suggests a difference that may be considered unimportant.

Comparing group means can often serve as an effective way to calculate effect size, but in many cases, choosing the appropriate statistical measure is not straightforward. In these situations, researchers must exercise judgement and consider their specific study conditions, as outlined by Hill and Thompson (2004).

Two commonly used effect size statistics for small sample sizes are *Cohen's d* and *Hedges' g*. According to Hill and Thompson (2004), these statistics are well-suited for small sample sizes. *Cohen's d* is appropriate when the two groups have similar standard deviations and are of the same size, while Glass' delta may be a more suitable choice if the sample sizes are large and the intervention is expected to affect the standard deviation. On the other hand, if each group has a different standard deviation, Glass's delta, which uses only the standard deviation of the control group, is the appropriate effect size measure (Glen, 2022).

Hedges' g, which takes into account the relative size of each sample, is a useful alternative when the sample sizes of the two groups are different (Effect Size Calculator for T-Test, 2022). It is important to note that if the sample sizes are not the same, *Hedges' g* should be used."

Cohen's d

Cohen's d measures the effect size of the difference between two means (*Cohen's d:* Definition, examples, formulas, 2022). Bhandari (2022) explains that it is designed for comparing two groups and considers the difference between two means in standard deviation units. The size of the effect is expressed by the number of standard deviations that fall between the two means. For example, a d of 1 indicates a difference of 1 standard deviation between the group means, while a d of 2 indicates a difference of 2 standard deviations. Cohen (1988) provides benchmark values of small (d = 0.2), medium (d = 0.5), and large (d = 0.8), but it is important to note that these values are arbitrary and should not be interpreted strictly.

To calculate *Cohen's d*, the mean difference between the two groups is divided by the pooled standard deviation. This is an effect size in which the mean difference is standardized by an average of the standard deviations of both groups.

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(s_1^2 + s_2^2)/2}} \text{.................................. [Lalongo, 2016]}$$

Cohen's d [t-test with equal sample size and variance]

Cohen's d can range from 0 to infinity (Bhandari, 2022). Generally, a greater *Cohen's d* indicates a larger effect size. This measure is unit-free and can be used to compare results across studies. For example, if two conditions have mean lengths of 2.3 cm and 1 cm, the simple effect size would be the difference in mean length, or 1.3 cm. But when standardized, the effect size would be 1.3 (assuming the values for the pooled standard deviation are arbitrarily designated as 1). Therefore, *Cohen's d* can be understood in terms of standard deviations.

According to Hill and Thompson (2004), if the group sizes are small and the intervention is not expected to affect the standard deviation, *Cohen's d* might be the most suitable choice. Similarly, if two groups of the same size have similar standard deviations, *Cohen's d* is recommended as the appropriate effect size measure, as stated by Effect Size Calculator for T-Test (2022).

When there are more than two group means, the *Cohen's d* effect size measure would be the difference between the largest and smallest means divided by the square root of the mean square error (Thompson, 2007, cited in Lakens, 2013).

Hedge's g

Hedges' g and *Cohen's d* are equal if the two sample sizes are equal (Zach, 2021). The difference between the two lies in the calculation of the overall effect size. While *Cohen's d* uses the pooled standard deviations, *Hedges' g* uses the pooled weighted standard deviations, taking into account the sample sizes. Therefore, it is advised to use *Hedges' g* when the two sample sizes are unequal.

The formula for *Hedges' g* is:

$$g = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \text{.................................. [Lalongo, 2016]}$$

Hedge's *g* [t-test on small samples/unequal size]

Hedges' g is a measure of effect size that shows the difference between two groups, typically an experimental group and a control group. The g value indicates the difference in standard deviations between the groups, with a g of 1 meaning a difference of 1 standard deviation and a g of 2 meaning a difference of 2 standard deviations.

Hedges' g and *Cohen's d* are similar measures of effect size, however, *Hedges' g* has been shown to be more effective when sample sizes are below 20 (Glen, 2022). There is a rule of thumb for interpreting *Hedges' g*, as proposed by Zach (2021), which states that if g = 0.2, it represents a small effect size; if g = 0.5, it represents a medium effect size; and if g = 0.8, it represents a large effect size. It is important to note that these terms may not have the same meanings in different contexts. For instance, a "small" change in sleep habits might be considered positive, while a "small" weight loss might not be considered significant.

Glasse's Delta

Glass's delta is another measure of effect size between two group means. When standard deviations are significantly different between the groups, Glass's delta is used. This uses only the control group's standard deviation. It addresses the issue of unequal group variances by using the control group standard deviation as the denominator.

A mathematical formula for Glasse's delta is:

$$\Delta = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s^2_{\text{control}}}}$$ .... [Lalongo, 2016]

Glass's Δ [t-test with unequal variances/control group]

Glass's delta is defined as the mean difference between the experimental and control group divided by the standard deviation of the control group (Becker, 2000). Martin (2020) and Effect Size Calculator for T-Test (2022) explain, if standard deviations are significantly different between groups, we should choose Glass's delta which uses only the standard deviation of the control group. If group sizes were both large and the intervention was expected to impact outcome score standard deviation, Glass' delta might be the most reasonable choice (Hill and Thompson, 2004).

Hedges' g effect size is an alternative where there are different sample sizes but when each group has a different standard deviation and only the standard deviation of the control group is used, Glass's delta is the alternative, explains the effect size calculator for t-test (2022).

Effect Size of the Relationship Between Two Variables

The Pearson product-moment correlation coefficient (r) is a measure that indicates the strength of the relationship between two variables. The formula for computing the correlation coefficient is as follows:

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2}\sqrt{\Sigma(x - \bar{x})^2}}$$ .... [Lalongo, 2016]

Pearson's r: Linear correlation

The Pearson correlation coefficient, represented by the symbol "r", measures the strength and direction of the relationship between two variables. It ranges from -1, which represents a perfect negative correlation, to +1, which represents a perfect positive correlation.

The value of “r” serves as an indicator of the effect size between two variables and conveys information about both the magnitude and direction of the relationship (Rosenthal, 1991).

When the value of “r” is closer to 0, it means the effect size is small. On the other hand, values closer to -1 or +1 indicate a higher effect size (Bhandari, 2022). According to Cohen’s rule of thumb, an effect size is considered low if “r” is around 0.1, medium if “r” is around 0.3, and large if “r” is greater than 0.5.

In multiple regression analysis, the magnitude of the total effect for the regression equation is represented by the multiple R.

INTERPRETATION OF EFFECT SIZE

Effect sizes are a crucial aspect of research and should always be reported along with p-values. An example of how an effect size is reported and interpreted could be: for a test comparing two groups, group A (n = 45, M = 20.92, SD = 7.05) performed better than group B (n = 43, M = 18.59, SD = 7.36) in solving algebra problems, “F (1, 84) = 5.96, p < .05”, effect size = 0.52, which represents a medium effect.

According to Sun *et al.* (2010), simply reporting an effect size is not enough and researchers must interpret and evaluate its practical significance. It is important to understand the magnitude of effect size and its context in a specific research scenario. Kirk (1996) (as cited in LeCroy & Krysik, 2007) and Nakagawa & Cuthill (2007) also argue that there is no point in presenting effect sizes if they are not correctly interpreted and discussed.

The interpretation of effect sizes depends on various factors such as the context and design of the study, type of effect size, benchmarks for interpretation, and confidence interval. Henson (2006) points out that the importance and meaning of effect sizes depend on factors such as the context of the study, the significance of the outcomes, and prior research findings. Durlak (2009) explains that any judgment about effect size should be based on the characteristics of the study.

The context of a study and the way data is analyzed can greatly affect effect size interpretation. Reporting and interpreting effect sizes in the context of previous results allows readers to evaluate the stability of results across samples, designs, and analyses. According to Thompson (2007), a proper interpretation of effect size should focus on directly comparing new results with prior effect sizes in the related literature. However, effect sizes from different studies may not be directly comparable as the context of each study can impact the resulting effects.

The type of effect size also affects interpretation. Simple effect sizes describe the size of the effect in the original units of the variables, while standardized effect sizes are unit-free. The Publication Manual of the American Psychological Association (2001) recommends using effects on the original measurement scale when feasible. Nakagawa & Cuthill (2007) explain that understanding the study system well and interpreting effect sizes in the original units is more interpretable than using standardized effect statistics.

Another common strategy for interpreting effect sizes is using Cohen’s benchmarks, where a d of 0.2 represents a small effect size, 0.5 represents a medium effect size, and 0.8 represents a large effect size. However, Cohen himself states that these terms are relative and can vary based on the content and method of the study. Laken (2013) explains that Cohen’s benchmarks should only be used in extremely novel studies with no comparison to existing literature. Glass et al. (1981) and Thompson (2007) argue that in established areas of research, Cohen’s guidelines should not be applied blindly.

Interpreting effect sizes is a challenging task and calculating, reporting, and discussing effect sizes should be highly valued to produce more conclusive evidence in research. Confidence intervals offer a range of values and are more informative than point estimate hypothesis testing. Stukas & Cumming (2014) stress the importance of acknowledging the uncertainty in an effect size estimate and the need for a more comprehensive interpretation.

Téllez *et al.* (2015) highlight the benefits of using confidence intervals in data interpretation and detecting non-significant or trivial effects. The authors stress the importance of reporting the confidence interval of effect size, as it provides practical context for understanding the magnitude and meaning of the effect size in relation to the data.

SUMMARY AND CONCLUSION

Hypothesis testing is a statistical tool that is used to determine whether the results of an experiment are likely due to chance or not. While this is an important aspect of research, it is limited in its scope and provides no information about the practical significance of the results. To address this, researchers turn to effect size measures.

Effect size measures provide a way to describe the magnitude of the relationship between variables in a study. There are several common measures of effect size, each with specific advantages and limitations. Choosing the appropriate measure depends on the design of the study and the type of data being analyzed.

This article serves as a guide for researchers who want to use effect size measures in their work. It covers the various types of effect sizes, the conditions for choosing a particular measure, and strategies for interpreting effect sizes. By providing this information, the article aims to help researchers communicate the practical significance of their results in a clear and meaningful way.

In conclusion, the use of effect size measures is crucial for ensuring the quality of research and providing valuable insights into the practical implications of the results. This article provides a comprehensive overview of the various types of effect sizes, their appropriate use, and how to interpret them correctly. It is a valuable resource for researchers looking to enhance the quality of their work.

ACKNOWLEDGEMENT

The authors of this review article are grateful to all the authors of the papers and books cited in this work. We also extend our sincere appreciation to any other authors whose works, including books, papers, notes, videos, etc., have helped us in the preparation of this review article.

CONFLICT OF INTEREST

The author declare that there is no conflict of interest to disclose.

REFERENCES

Becker, L. A. (2000). *Effect size (ES)*. Retrieved from https://www.uv.es/~friasnav/EffectSizeBecker.pdf. Accessed September 19, 2019.

Bhandari, P. (2022). *What is effect size and why does it matter?* (examples). SCRIBBR. Retrieved from https://www.scribbr.com/statistics/effect-size/#:~:text=Cohen's%20d%20can%20take%20on,indicates%20a%20higher%20effect%20size. Accessed October 13, 2022.

Glen, S. (2022). *Hedges' g*: *Definition, Formula.* StatisticsHowTo.com. Retrieved from https://www.statisticshowto.com/hedges-g/. Accessed October 14, 2022.

Carpenter, A. (2020). *Effect size: Moving beyond the p-value.* Towards Data Science. Retrieved from https://towardsdatascience.com/effect-size-d132b0cc8669. Accessed December 19, 2020.

Castillo, E.O., &Torquato, M.G. (2018). Bayesian Statistics in Archaeology. *Annual Review of Anthropology*, *47*(1), 435-453. https://doi.org/10.1146/annurev-anthro-102317-045834

Cohen's d : *Definition, Examples, Formulas*. (2022). Statistics How To. Retrieved from https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/cohens-d/. Accessed October 13, 2022.

Daniel. (2017, June 28). What Statistical Significance Really Means (10-6) [Video]. Research by Design. Retrieved from https://www.youtube.com/watch?v=wuB7CC9DrIE&t=20s.

Dunkler, D., Haller, M., Oberbauer, R., &Heinze, G. (2020). To test or to estimate? P-values versus effect sizes. *Transplant International: Official Journal of the European Society for Organ Transplantation*, 33(1), 50–55. https://doi.org/10.1111/tri.13535

Durlak, J.A. (2009). How to Select, Calculate, and Interpret Effect Sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928. https://doi.org/10.1093/jpepsy/jsp004

Effect Size Calculator for T-Test. (2022). *Social Science Statistics*. Retrieved from https://www.socscistatistics.com. Accessed on October 13, 2022.

Glass, G. V., McGraw, B., & Smith, M.L. (1981). *Meta-Analysis for Social Research. Sage Publications.* Retrieved from https://books.google.com.np/books?id=si1-AAAAIAAJ&q=Glass,+McGaw+and+smith+1981&dq=Glass,+McGaw+and+smith+1981&hl=en&sa=X&redir_esc=y. Accessed on October 28, 2022.

Glen, S. (2022). *Glass's Delta.* StatisticsHowTo.com. Retrieved from https://www.statisticshowto.com/glasss-delta/. Accessed on October 15, 2022.

Martin, K.G. (2020). *Two Types of Effect Size Statistics: Standardized and Unstandardized.* The Analysis Factor. Retrieved from https://www.theanalysisfactor.com/two-types-effect-size-statistic/. Accessed on June 25, 2020.

Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences*, *10*(2), 64–69. https://doi.org/10.1016/j.tics.2005.12.005

Hill, C. R., & Thompson, B. (2004). Computing and Interpreting Effect Sizes. *Higher Education: Handbook of Theory and Research*, *19*(1), 175–196. https://doi.org/10.1007/1-4020-2456-8_5

Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, *62*(2), 227-240. https://doi.org/10.1177/0013164402062002002

Hypothesis Testing: Methodology and Limitations. (2001). Elsevier Science Ltd. Retrieved from https://www.stats.ox.ac.uk/~snijders/Encycl_isb203057.pdf. Accessed on December 15, 2019.

Hypothesis Testing. (2011). WikiVet. Retrieved from https://en.wikivet.net/Hypothesis_testing#Limitations_of_null_hypothesis_tests. Accessed on December 15, 2019.

Lalongo, C. (2016). Understanding the effect size and its measures. *BiochemiaMedica*, *26*(2), 150–163. https://doi.org/10.11613/BM.2016.015

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

LeCroy, C. W., & Krysik, J. (2007). Understanding and interpreting effect size measures. *Social Work Research*, *31*(4), 243-248. https://doi.org/10.1093/swr/31.4.243

Levine, T. R., Weber, R., Hullett, C., Park, H. S., & Massi Lindsey, L. L. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research*, *34*(2), 171-187. https://doi.org/10.1111/j.1468-2958.2008.00317.x

Limitations of significance testing. (2015). 24 X 7 Editing.com. Retrieved from https://www.24x7editing.com/limitations-of-the-tests-of-hypotheses/ Accessed on December 13, 2019

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval, and statistical significance: A practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society*, *82*(4), 591-605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241-301. https://doi.org/10.1037/1082-989X.5.2.241

Null hypothesis significance testing. (n.d.). InfluentialPoints.com. Biology, images, analysis, design. Retrieved from https://influentialpoints.com/Training/null_hypothesis_significance_testing-principles-properties-assumptions.htm Accessed on December 15, 2019

Pernet, C. (2016). Null hypothesis significance testing: A short tutorial. F1000Research, 4, Article 621. https://doi.org/10.12688/f1000research.6963.5

Publication Manual of the American Psychological Association. (7th ed.). (2020). American Psychological Association. https://doi.org/10.1037/0000165-000

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Sage Publications, Inc. https://doi.org/10.4135/9781412984997

Stukas, A. A., & Cumming, G. (2014). Interpreting effect sizes: Toward a quantitative cumulative social psychology. *European Journal of Social Psychology*, *44*(7), 711-722. https://doi.org/10.1002/ejsp.2019

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P-value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279-282. https://doi.org/10.4300/JGME-D-12-00156.1

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*, Advanced online publication. https://doi.org/10.1037/a0019507

Szucs, D., & Ioannidis, J. P. A. (2017). When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience*, *11*, https://doi.org/10.3389/fnhum.2017.00390

Téllez, A., García, C. H., & Corral-Verdugo, V. (2015). Effect size, confidence intervals, and statistical power in psychological research. *Psychology in Russia: State of the Art*, *8*(3), 27-46. https://doi.org/10.11621/pir.2015.0303

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, *44*, 423–432. https://doi.org/10.1002/pits.20234

Zach, Statology. (2021). *What is Hedges' g? (Definition & Example)*. Statology. Retrieved from https://www.statology.org/hedges-g. Accessed October 15, 2022.