

Referencesⁱ regarding null hypothesis significance testing and alternatives, including the importance of providing effect sizes and/or confidence intervals.

- Balluerka, N., Gómez, J., & Hidalgo, D. (2005). The controversy over null hypothesis significance testing revisited. [Journal Article]. *Methodology: European Journal of Research Methods for the Behavioral & Social Sciences*, 1(2), 55-70.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997-1003.
- Cowles, M., & Davis, C. (1982). On the origins of the .05 level of statistical significance. [Peer Reviewed]. *American Psychologist*, 37(5), 553-558.
- Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. [Peer Reviewed]. *Theory & Psychology*, 5(1), 75-98.
- Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)*, 292(6522), 746-750.
- Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What If There Were No Significance Tests?* New York, NY: Psychology Press, Taylor & Francis Group.
- Harrison, J., Thompson, B., & Vannest, K. J. (2009). Interpreting the Evidence for Effective Interventions to Increase the Academic Performance of Students With ADHD: Relevance of the Statistical Significance Controversy. [Review]. *Review of Educational Research*, 79(2), 740-775. doi: 10.3102/0034654309331516
- Hubbard, R., & Lindsay, R. M. (2008). Why p values are not a useful measure of evidence in statistical significance testing. [Peer Reviewed]. *Theory & Psychology*, 18(1), 69-88.
- Kline, R. B. (2004). *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. Washington, DC: American Psychological Association (2004). (2004) xii, 325 pp.
- Pollard, P., & Richardson, J. T. (1987). On the probability of making Type I errors. [Peer Reviewed]. *Psychological Bulletin*, 102(1), 159-163.
- Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. [Peer Reviewed]. *American Psychologist*, 65(1), 1-12. doi: 10.1037/a0018326
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1(2), 115-129.

- Thompson, B. (1989). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. *Measurement & Evaluation in Counseling & Development*, 22(1), 2-6.
- Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. *Journal of Experimental Education. Special Issue: Statistical significance testing in contemporary practice*, 61(4), 361-377.
- Thompson, B. (1997). Rejoinder: Editorial policies regarding statistical significance tests: Further comments. *Educational Researcher*, 26(5), 29-32.
- Thompson, B. (1999). If statistical significance tests are broken/misused, what practices should supplement or replace them? *Theory & Psychology*, 9(2), 165-181.
- Thompson, B. (2002). "Statistical," "practical", and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling & Development*, 80(1), 64-71.
- Vacha-Haase, T., & Thompson, B. (1998). Further comments on statistical significance tests. *Measurement & Evaluation in Counseling & Development*, 31(1), 63-67.
- Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T. S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology*, 10(3), 413-425.
- Vacha-Haase, T. (2001). Statistical significance should not be considered one of life's guarantees: Effect sizes are needed. [Peer Reviewed]. *Educational and Psychological Measurement*, 61(2), 219-224.
- Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473-481.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. [Peer Reviewed]. *American Psychologist*, 54(8), 594-604. doi: 10.1037/0003-066x.54.8.594

ⁱ This list is by no means an exhaustive list about hypothesis testing and the problems associated with it, not are these references "the best." However, some of these are "must read," and all together, they will get you into the literature about this issue.