Review of: "The Impact of Study Environment on Students' Academic Performance: An Experimental Research Study"

M. Shane Tutwiler[1]

1 University of Rhode Island

Potential competing interests: No potential competing interests to declare.

This study has progressed nicely since I first read it a few weeks ago. Keep up the good work!

This review is primarily methodological.

1) Hypotheses are not stated correctly. As currently framed, they are:

“Null Hypothesis (H0): There is no significant difference in academic performance between participants studying in high-noise environments and participants studying in low-noise environments.

Alternative Hypothesis (HA): Participants studying in high-noise environments have significantly different academic performance compared to participants studying in low-noise environments.”

There are a few things wrong here. First, the null hypothesis is about the POPULATION, not the sample. Differences in the population can’t be “(statistically) significant”, as that is a property of the sample that helps justify the inference to the population. This same error is replicated in the alternative hypothesis.

Also, your alternative hypothesis should be directional if the literature review suggests a direction for the effect under study.

So, your actual hypotheses are, perhaps:

H0: There is no relationship between the noise level of the study environment and student academic performance, on average, in the population.

HA: Students in high-noise study environments exhibit lower academic performance.

If your alternative hypothesis is *truly* non-directional, then perhaps you need to do a more comprehensive scoping review of the literature to understand the scientific consensus on the matter first. This will help to focus your analysis.

2) You’re likely underpowered.

Randomization is a critical component of experimental design, yes. But interpretation of effects and hypothesis tests only make sense in studies that are adequately powered (have a sufficiently large sample size). As Gelman and Carlin (2014) point out, when conducting null hypothesis significance testing, small (underpowered) samples can often result in overly

large estimates of effects due to the increased variance in the sample. This is a Type-M (magnitude) error.

How did you decide on your sample size? Was it based on a power analysis derived from the effects in previous studies? If not, what was your mechanism to determine the sample size for the study?

Why does your procedure say you selected 200 subjects, but only 30 are shown in the table?

3) Noise isn't the only potential environmental factor.

It's clear that other factors, such as temperature and lighting, might also be impacting the outcomes. They're confounders, introduced because they weren't all controlled to be the same between the high and low noise groups. As such, the way to properly answer your research question would be to *control for these conditions* in a multiple regression model:

Score = B0 + B1*Noise + B2*Light + B3* Temperature + e

You would evaluate the magnitude, direction, and precision (often via the p-value) of the B1 coefficient in this model, which is the "Effect of noise on score, controlling for light and temperature levels."

Moreover, it's quite likely that the impact of noise *statistically interacts* with other environmental factors, which your study is under-powered to detect.

The t-test alone is inadequate. Consider an expanded analysis.

4) Interpretation of results throughout.

Please revisit your understanding of what it means for something to be "Statistically significant". This phrase denotes the observation that your p-value was lower than a *pre-specified* threshold (usually a=.05), which allows you to then reject the null hypothesis. You frequently use it as a stand-in for "real effect" or "important effect", which it does not denote. Your use of the phrase "a significant difference between…" hints at this misapplication.

All you can say is that you've rejected the null hypothesis. It requires domain-level knowledge of the questionnaire or the construct under study to determine if this difference is truly meaningful.

5) Details of the assessment are missing.

You need to provide some key details about the assessment, such as evidence of construct validity, reliability, and descriptive statistics (mean and sd) between groups. Also, in the same table, you should show the mean and sd of ages between the two groups, as this will help you determine the degree to which randomization was successful.

Works Cited

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641-651.