Four Common Pitfalls of Quantitative Analysis in Experimental Research

Jimmie Leppink, PhD, postdoctoral researcher, Ellen M. Kok, MSc, PhD student, Esther M. Bergman, PhD, assistant professor, Mariëtte H. van Loon, PhD, assistant professor, and Anique B.H. de Bruin, PhD, associate professor, Maastricht University

A recently published AM Last Page presents five common methodological pitfalls of experimental research in medical education.[1] In this Last Page, we present four statistical pitfalls and their more appropriate alternatives. Pitfalls are illustrated with a case of a fictitious researcher who conducts a study with elements that are common in many medical education experiments (Figure 1).

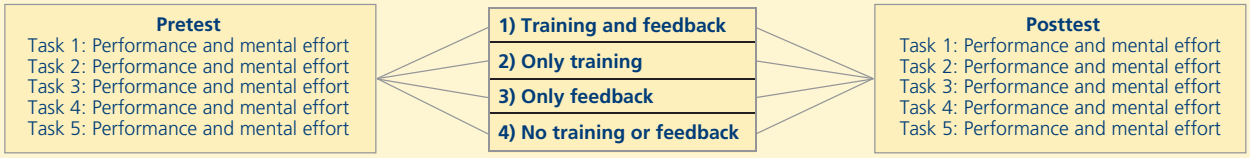

Figure 1: The researcher studies the effects of two factors, training and feedback, on student performance and mental effort measured during five different tasks. Students are randomly assigned to four treatment conditions: (1) training and feedback, (2) only training, (3) only feedback, and (4) no training or feedback.

Pitfall 1: Treating a two-factor design as a one-factor design

To test for differences between the four conditions in the experiment, the researcher uses a one-way analysis of variance (ANOVA) with four groups.

Problem with the one-way ANOVA approach

- It does not test for an interaction effect between training and feedback, but compares each combination of two groups (see Figure 2, below).
- It is less likely to detect main effects of training and/or feedback.

Alternative

- A two-way (2x2) ANOVA that tests for a main effect of feedback, a main effect of training, and an interaction effect between training and feedback (see Figure 2, below)[2]

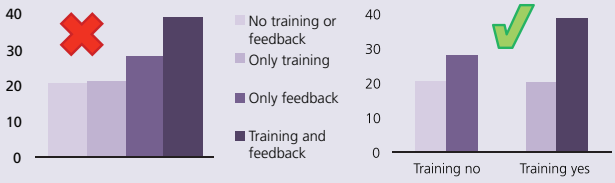

Figure 2: One-way ANOVA (left) versus 2x2 ANOVA (right).

Pitfall 2: Treating the pretest and posttest as repeated measures instead of treating the pretest as a covariate

In the experiment participants completed a pretest and posttest. The researcher treats the pre- and posttest as repeated measures in a within-subjects ANOVA.

Problem with the repeated-measures approach

- It is appropriate in quasi-experimental (nonrandomized group comparison) studies or when there are pretest differences between treatment conditions.
- When randomization of participants has resulted in no significant differences in pretest performance, there is no need to further test for these differences (Figure 3).

Alternative

- When successful randomization has resulted in no significant differences in pretest performance, treating pretest as a covariate in an analysis of covariance (ANCOVA) provides a more powerful test, because ANCOVA is more parsimonious than repeated-measures ANOVA.[3]

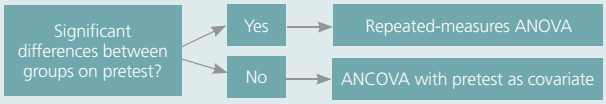

Figure 3: Statistical analysis depends on whether there is a significant difference between groups on the pretest.

Pitfall 3: Considering the time-on-task as a covariate versus as a mediator

The training took on average 1 hour and the feedback session on average ½ hour, resulting in differences in time-on-task between treatment conditions. In the analysis, the researcher includes time-on-task in an ANCOVA.

Problem with the ANCOVA approach

- The treatment influences time-on-task (or another variable of interest), and time-on-task may affect performance.[1]
- Including time-on-task in an ANCOVA will either underestimate or overestimate the treatment effect.

Alternative

- Path analysis enables for treating time-on-task as a mediator and allows estimating both direct and indirect effects of treatment (see Figure 4).[4]
- The total treatment effect is the sum of the treatment effect mediated by time-on-task (i.e., indirect effect) and the treatment effect not mediated by time-on-task (i.e., direct effect).

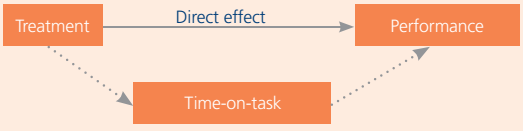

Figure 4: Direct and indirect effect of treatment condition on performance.

Pitfall 4: Ignoring a hierarchical structure of data by performing "ordinary" regression instead of multilevel regression

Posttest performance and mental effort are measured per task. To analyze their relation, the researcher wants to average individual scores across tasks or treat every individual-by-task combination as an independent observation.

Problem with these approaches

- The independent-observations approach ignores the intraindividual correlation between the five tasks. Averaging over five tasks means loss of information.
- Both approaches can result in an incorrect interpretation of the correlation of interest.

Alternative

- Using multilevel models constitutes a best practice for dealing with intra-individual and/or intraclass correlations (see Figure 5).[5]

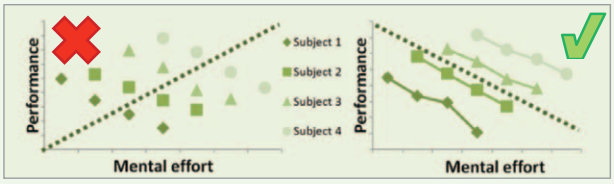

Figure 5: What can happen when using the averaging or independent observation approach (left), when the multilevel approach is more appropriate (right).

References

1. Van Loon MH, Kok EM, Kamp RJA, et al. AM Last Page: Avoiding five common pitfalls of experimental research in medical education. Acad Med. 2013;88:1588.
2. Leppink J, Paas F, Van Gog T, Van der Vleuten CPM, Van Merriënboer JJG. Effects of pairs of problems and examples on task performance and different types of cognitive load. Learn Instr. 2014;30:32–43.
3. Van Breukelen GJP. ANCOVA versus change from baseline has more power in randomized studies and more bias in nonrandomized studies. J Clin Epidemiol 2006;59:920–925.
4. Leppink J. On causality and mechanisms in medical education research: An example of path analysis. Perspect Med Educ. 2015;4:66–72.
5. Leppink J. Data analysis in medical education research: A multilevel perspective. Perspect Med Educ. 2015;4:14–24.

Author contact: jimmie.leppink@maastrichtuniversity.nl

First published online October 12, 2015