



It's Just an Observation

Daniel H. Robinson¹ · Howard Wainer²

Accepted: 4 August 2023 / Published online: 14 August 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

To date, there have been four responses (Dumas & Edelsbrunner, *Educational Psychology Review*, 35, 48, 2023; Grosz, *Educational Psychology Review*, 35, 57, 2023; Mayer, *Educational Psychology Review*, 35, 64, 2023; Zitzmann et al., *Educational Psychology Review*, 35, 65, 2023) to the Brady et al. (*Educational Psychology Review*, 35, 36–37, 2023) observational study showing that the field of educational psychology has experienced a continued (1) decline in the proportion of empirical articles that include intervention and experimental studies, an (2) increase in the proportion that include observational studies, and an (3) increase in the proportion of the latter that include recommendations for practice. Educational psychology is continuing its movement toward the dark side of the quality continuum. There are fewer intervention studies and more armchair research—“analyzing” large datasets from the comfort of offices. More alarming is the increase in the tendency for authors to infer causality from correlation and offer recommendations based on shaky ground. The statistical models that flood our journals are casual, NOT causal. They are connected only through a vowel movement that is stinking up our field.

Keywords Causality · Modeling · Experiments · Intervention · Observation

To date, there have been four responses (Dumas & Edelsbrunner, 2023; Grosz, 2023; Mayer, 2023; Zitzmann et al., 2023) to the Brady et al. (2023) observational study showing that the field of educational psychology has experienced a continued (1) decline in the proportion of empirical articles that include intervention and experimental studies, an (2) increase in the proportion that include observational studies, and an (3) increase in the proportion of the latter that include recommendations for practice. Why is there such interest concerning these findings?

✉ Daniel H. Robinson
daniel.robinson@uta.edu

¹ College of Education, The University of Texas at Arlington, 504 Hammond Hall, Arlington, TX 76019, USA

² Pennington, NJ, USA

The earth's temperature has risen about 2°F since 1880. However, the rate of warming since 1981 is more than twice as fast compared to previous decades (Lindsey & Dahlman, 2023). Is this a cause for concern? Similar to the data reported by Brady et al. (2023), it is just an observation. Whether we should be concerned with such trends depends on the possible consequences. The earth's rising temperature may have disastrous consequences, whereas the replacement of experiments with observational studies likely will not kill anyone (at least not directly).

In the book *Fossil Men*, Tim White, a paleoanthropologist, rants about his field's degeneration. At a conference where he preaches from the pulpit, he says the field is producing too many ill-prepared and poorly trained PhDs who do not want to go out into the field and find fossils. Instead, they want others to share imprints of their fossils so they can “analyze” the data from the comfort of their offices (Pattison, 2020). Our interpretation of the Brady et al. (2023) findings is similar. Educational psychology is continuing its movement toward the dark side of the quality continuum. There are fewer intervention studies and more armchair research—“analyzing” large datasets from the comfort of offices. More alarming is the increase in the tendency for authors to infer causality from correlation and offer recommendations based on shaky ground.

What most refer to as causal modeling (e.g., structural equation modeling or SEM and a direct generalization of traditional factor analysis) is in fact not causal at all. SEM gives us only a model that conveys causal assumptions. Leaning heavily on such models requires also leaning very heavily on largely untestable, substantial, heroic assumptions. A recent typical example is a study by Clarke et al. (2023) that was promoted as follows in *Science Daily*: “Study of 600 UK teenagers suggests that having stronger self-awareness and sense of purpose may raise GCSE maths scores ‘by a couple of grades.’” The authors analyzed student questionnaire data collected at one time point using SEM. This “causal” conclusion is bogus. Raising self-awareness and sense of purpose is not likely to raise math performance. It is just as likely that increased math performance raises self-awareness. This “correlation is not causation” error should have been corrected in any introductory statistics or research methods course. It is sadly reminiscent of the self-esteem movement from the 1970s. The thought back then was that because self-esteem was positively correlated with student achievement, we could simply raise achievement by raising self-esteem. Unfortunately, efforts to raise students' self-esteem not only did not increase student achievement, but in some cases, achievement decreased (Baumeister et al., 2003). The lessons learned from 50 years ago have apparently been forgotten.

In educational psychology, we have moved away from a very important component of research—rigor. Instead, we have been seduced by what is simple and easy. The fruits of this latter type of research are rarely useful in providing causal evidence. More rigorous methods, such as interventions, pay bigger dividends. Improving education takes rigor and sustained effort over a long period of time. Even then, the effects might be neither impressive nor replicable. But we keep trying. We roll up our sleeves and apply “elbow grease” by using the most rigorous methods we have available.

In finance, the Markowitz model shows the positive relationship between risks and rewards. Higher returns typically only come from higher risks. In educational

research, a similar relationship exists between the difficulty of data gathering and the credibility of causal inference.

The causal effect of some treatment relative to a control is the difference between the outcome observed with the treatment and the outcome that would have occurred without it. This latter aspect is what is usually referred to as Hume's counterfactual (Holland, 1986). Observational studies typically assume that outcomes observed under the control condition represent what would have occurred in the experimental condition had they not received the treatment. This assumption is often heroic (e.g., those students living with their grandmothers who were in over-crowded, overheated, poorly taught, inner-city schools were just the same in all important ways as those who were in new suburban schools with A/C, highly qualified teachers, and two parents at home who were college graduates). The counterfactual assumption is more credible in a randomized experiment because there is nothing different, a priori, between the treatment and control groups (at least on average) beyond that generated by chance—and those differences shrink as sample sizes increase. But randomized controlled experiments are more difficult to carry out.

Easy-to-do studies such as conducting complex analyses on a huge database with no interventions or manipulations, on the other hand, provide little supporting evidence of causal inference. Such observational studies require heroic assumptions. Whether researchers should offer recommendations for practice should depend on the credibility of the evidence supporting those recommendations. In educational psychology, credible evidence is disappearing.

The limits of any analytic procedure are most easily explained in terms of missing data. All statistical methods can be thought of in terms of missing data. For example, regression is $y=Bx$, in which we observe the y s and the x s and the B s are missing. The analytic task is to estimate the B s. This is covered in the first statistics course. Of course, the greater the missingness, the greater the uncertainty. What about factor analysis/SEM? The basic model is the same as regression: $y=Bx$, except now both the B s and the x s are missing (called factor loadings and factor scores). All we have are the y s. Obviously, this is a much tougher task and cannot be accomplished at all without relying on heroic assumptions.

This is why, when it is possible, we ought to (vastly) prefer (and believe in) studies that are closer to what we can observe. Unfortunately, educational psychology continues to reject randomized experiments in favor of observational studies (Cook, 2001). SEM applied to observational data can only disconfirm causal hypotheses, not confirm them. Such models are casual, NOT causal. They are connected only through a vowel movement that is stinking up our field.

References

- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, 4, 1–44.

- Brady, A. C., Griffin, M. M., Lewis, A. R., Fong, C. J., & Robinson, D. H. (2023). How scientific is educational psychology research? The increasing trend of squeezing causality and recommendations from non-intervention studies. *Educational Psychology Review*, 35, 36–37. <https://doi.org/10.1007/s10648-023-09759-9>
- Clarke, T., McLellan, R., & Harold, G. (2023). Beyond life satisfaction: Wellbeing correlates of adolescents' academic attainment. *School Psychology Review*. <https://doi.org/10.1080/2372966X.2023.2217980>
- Cook, T. D. (2001). Sciencephobia: Why education rejects randomized experiments. *Education Next*, 1(4), 62–68. Retrieved July 21, 2023, from <https://www.educationnext.org/sciencephobia/>
- Dumas, D., & Edelsbrunner, P. (2023). How to make recommendations for educational practice from correlational data using structural equation models. *Educational Psychology Review*, 35, 48. <https://doi.org/10.1007/s10648-023-09770-0>
- Grosz, M. P. (2023). Should researchers make causal inferences and recommendations for practice on the basis of nonexperimental studies? *Educational Psychology Review*, 35, 57. <https://doi.org/10.1007/s10648-023-09777-7>
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*., 81, 945–970.
- Lindsey, R., & Dahlman, L. (2023). Climate change: Global temperature. *Climate.gov*. Retrieved July 23, 2023, from <https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature>
- Mayer, R. E. (2023). How to assess whether an instructional intervention has an effect on learning. *Educational Psychology Review*, 35, 64. <https://doi.org/10.1007/s10648-023-09783-9>
- Pattison, K. (2020). *Fossil men: The quest for the oldest skeleton and the origins of humankind*. William Morrow.
- Zitzmann, S., Machts, N., Hübner, N., et al. (2023). The yet underestimated importance of communicating findings from educational trials to teachers, schools, school authorities, or policy makers (Comment on Brady et al. (2023)). *Educational Psychology Review*, 35, 65. <https://doi.org/10.1007/s10648-023-09776-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.