educade
Revista de Educación en Contabilidad, Finanzas y Administración de Empresas
The Spanish Journal of Accounting, Finance and Management Education

Nº 14 · 2023 ·
ISSN 2173-478X
pp. 1-4

The extreme relevance of avoiding the file drawer effect in educational research (Editorial)

José L. Arquero (Editor)

Universidad de Sevilla (España)

https://dx.doi.org/10.12795/EDUCADE.2023.i14.01

Scheel et al. (2021) note that if what is published as scientific literature were an accurate representation of the results obtained by researchers, it could be possible to infer what is true about the topics studied. However, those authors highlight that what is finally published is just a biased view of the reality investigated. This bias, as is systematic, causes aggregation not to cancel errors but even worsens the situation drifting away from a faithful representation of reality.

There is a strong tendency to publish only those outputs that obtain *positive results*; that is, where the hypotheses are supported, and the experiments generate the expected results. Meanwhile, a giant corpus of research, in which the results are not conclusive, experiments fail to produce the expected results, or the hypotheses are not supported, are simply forgotten and remain forever unpublished and unknown for any other researcher forming what Lemaire et al. (2021) label as dark science; maybe, better, *hidden science*. This problem is known as the *file drawer effect*, or *publication bias,* and could be defined (Dickersin, 1990, p.1385) as '*the tendency on the parts of investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings*.' This bias and its consequences have been attracting the attention of researchers several decades ago. Dickersin points to Sterling (1959) as one of the first authors to highlight the problem that results from rejecting negative findings while accepting only positive ones. Sterling analysed the published papers in four major psychology journals; his results were overwhelmingly clear: only 8 out of 286 papers reported negative results. None of them was a replication of a previously published experiment. On the same line, Smart (1964) reported that for psychological journals, studies with negative results accounted for less than 10% of the total volume of published papers, while the real proportion of 'unsuccessful' research was much higher. Recently, Allen and Melher (2019) indicated a wider range 5%-20% of null findings published, but this proportion is still much lower than the distribution observed in preregistered reports and Scheel et al. (2021) found 96% of positive results in standard Psychology literature, while this percentage dropped to 44% in registered reports (which is also a –lesser- biased representation of actual research). Consequently, published science in traditional journals is not a faithful representation of research results, but a misleading and biased depiction that systematically ignores an enormous corpus of research.

But, what kind of pieces of research are buried in those file drawers? Lemaire et al., (2021) indicate that the hidden science is mainly conformed by:

(I) findings that are not statistically significant, null findings, which could be categorize into:
   a. inconclusive findings
   b. conclusive findings
(II) replications
(III) research with weak designs, or flawed experiments

Usually, research with inconclusive findings (Ia) and those with weak designs (III) do not pass editorial or reviewer filters and consequently are less likely to be published in ranked journals. This is logical, given that their relevance for an average reader is lower.

Replications (II) are also less likely to be published, at least for those replications which confirm reported results; maybe replications that obtain unsuspected results, contradicting published results, have better chances (Lemaire et al., 2021). However, replication is extremely relevant given that in educational contexts orthodox experimental designs are almost impossible to implement, even if we try, and -as studies cannot employ samples that fully reflect diversity- generalisation is limited (Taber, 2019).

For us, in educational research, as well as in other critical areas such as medicine or pharmacology, the problem of keeping so many research unpublished is related **to null conclusive findings** (Ib). This kind of research is often kept by the authors unpublished, who even do not try to send it to journals, who prefer positive results given to the higher possibility of receiving more citations (Fanelly, 2013). However, the consequences of keeping this research unpublished are tremendous and extremely costly.

It is true that as editors and readers of educational papers, we are initially more interested in papers that report innovations, ideas, or materials that could be helpful in improving the teaching-learning process, helping our students better acquire the desired competencies. This kind of research is able to provide us with insights about what could be done to improve; they inform us how to obtain good similar results to those reached by the authors.

However, those papers that report negative results, with well-designed experiments, relevant data, and proper samples, are equally important. They tell us what ***DO NOT WORK***, which paths are not worthwhile to follow, and which innovations are not valuable to be implemented, acting as a danger signal. If the readers, teachers, are aware of positive **and** negative results, they could better choose which innovations are more likely to be successful for their students. Let us remember that the implementation of innovations is costly, maybe not in financial terms, but certainly in personal terms. Bad previous experiences with weakly designed innovations have a strong influence on the –negative- attitude of students, teaching staff, etc. against new 'experiments' (Arquero, 2016), even if the new one is better designed and grounded.

Image by Wepik on Freepik

Recently, we highlighted (Arquero, 2022) that any changes in educational processes must follow a utilitarian objective and must respond to a process of rigorous analysis and reflection that, at least, contemplates the following questions.

1. Which is the problem that I want to solve?
2. What are the causes of this problem, and can they be solved by our field of action?
3. What proven solutions or alternatives are there?
4. Are they applicable (efficiently) in the context of our problem?
5. Are they sustainable over time and capable of surviving a generalised application?

The answers to questions 3 to 5 require a process of critical analysis of the literature to consider the best alternatives. If the literature only covers positive results, ignoring negative ones, we could never be aware of which innovations could fail, and why!

Consequently, we are doomed to try once and once again the same inadequate "solutions" because only the successful cases were reported.

The literature on innovations is useful as far as the results are transferable, in other words, the results are replicable! This is why we insist in EDUCADE that articles reporting the results of educational innovations should always include an analysis of the key aspects that have influenced the positive results, especially if there are contextual factors that cannot be reproduced in other areas, and a reflection on the limitations and factors that prevent their generalisation. This allows a proper assessment of the usability by other teachers in other subjects and contexts (transferability-replication). But it is not possible to properly know if a certain treatment has (or not) chances of success if all negative results are hidden!

In educational research, the bias towards positive results is extremely costly because it impedes the proper balance of the pros and cons of any educational intervention. Therefore, negative results in education, if those results are conclusive, are as relevant as positive ones. The reflection on the key factors that could motivate the negative results is essential, so the readers could have a more complete view of things that could be done and (extremely important) things that should be avoided or done with extreme caution to prevent failure.

We should also remember here that subjects are different, students enrolling in different degrees have substantially different personal characteristics, motivations, attitudes, previous knowledge, etc. (e.g., Arquero et al., 2015; 2017; 2024), so we could not expect that innovations that are working perfectly well for certain students at a certain degree could give the same results with other students in a different degree. The objectives of such programmes, the competences to be developed, the characteristics of the contents to be taught, and the personal characteristics and profiles of the students are completely different (Arquero, 2022), with some of the acting as barriers (e.g. Arquero et al., 2022; 2023). Consequently, replication should be contextualised paying attention to differences (Elias et al., 2003); and we insist, negative-conclusive results should be reported, and positive results need to be published explaining clearly the limitations that future users could find to replicate, transfer, or generalise their findings (e.g. Gonzalez et al., 2014). This information could act as a clear warning in an area in which 'Trial and error' approaches are something that cannot be afforded.

We all, editors, reviewers, and also authors, must commit to present a complete view of relevant research results. Papers should be presented and accepted on the basis of their methodological rigour and potential relevance, taking into account that reporting no substantial effect in a certain innovation is as relevant as if the results were positive. If reviewers and editors are more committed to accepting such papers, the authors could find a motivation to send these results to journals instead of keeping them in a drawer forever. If not lucky enough to be published, research registers and repositories could also play a role to disseminate such results (Dickersin, 1990) and are often used to conduct meta-analyses. If meta-analyses and systematic literature reviews are conducted over a biased corpus of research, their results are going to be biased (Olson et al., 2002).

References

Allen C, and Mehler D.M.A. (2019). Correction: Open science challenges, benefits and tips in early career and beyond, *PLOS Biology,* 17:e3000587

Arquero, J. L. (2016). Innovación, indicadores y la confusión entre medidas y realidad. *Educade: Revista de Educación en Contabilidad, Finanzas y Administración de Empresas, (7), 1-2*. http://dx.doi.org/10.12795/EDUCADE.2016.i07.01

Arquero, J. L. (2022). Innovations, passing fads and serious questions to consider. *Educade: revista de educación en contabilidad, finanzas y administración de empresas,* (13), 1-3. https://dx.doi.org/10.12795/EDUCADE.2022.i13.01

Arquero, J. L., Fernández Polvillo, C., & Valladares García, D. (2017). Communication apprehension and students' educational choices: An exploratory analysis of Spanish secondary education students. *Education and Training, 59 (9), 946-956.* https://doi.org/10.1108/ET-04-2017-0049

Arquero, J. L., Fernández-Polvillo, C., & Hassall, T. (2022). Non-technical skills and students' overconfidence in accounting. *Education and Training*, *64*(5), 716-733. https://doi.org/10.1108/ET-08-2021-0309

Arquero, J. L., Fernández-Polvillo, C., & Jiménez-Cardoso, S. M. (2024). Financial literacy in tourism and management & business administration entry-level students: A comparative view. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *34*, 100474. https://doi.org/10.1016/j.jhlste.2023.100474

Arquero, J. L., Fernández-Polvillo, C., Hassall, T., & Joyce, J. (2023). Developing teamwork skills in accounting students: is communication apprehension a potential barrier? *Revista de Contabilidad-Spanish Accounting Review*, *26*(1), 97-110. https://doi.org/10.6018/rcsar.451151

Arquero, J. L., Fernández-Polvillo, C., Hassall, T., & Joyce, J. (2015). Vocation, motivation and approaches to learning: a comparative study. *Education and Training*, *57*(1), 13-30. http://dx.doi.org/10.1108/ET-02-2013-0014

Arquero, J. L., Polvillo, C. F., Hassall, T., & Joyce, J. (2017). Relationships between communication apprehension, ambiguity tolerance and learning styles in accounting students. *Revista de Contabilidad-Spanish Accounting Review*, *20*(1), 13-24. https://doi.org/10.1016/j.rcsar.2015.10.002

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama*, *263*(10), 1385-1389.

Elias, M. J., Zins, J. E., Graczyk, P. A., & Weissberg, R. P. (2003). Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School psychology review*, *32*(3), 303-319.

Fanelli, D. (2013). Positive results receive more citations, but only in some disciplines. *Scientometrics,* 94, 701–709. https://doi.org/10.1007/s11192-012-0757-y

González-González, J. M., Arquero, J. L., & Hassall, T. (2014). The change towards a teaching methodology based on competences: a case study in a Spanish university. *Research Papers in Education*, *29*(1), 111-130. http://dx.doi.org/10.1080/02671522.2012.745895

Lemaire, B. S., Lenoble, R., Zanon, M., Jacquel, T., & Vallortigara, G. (2021). The file drawer effect–a long-lasting issue in the sciences. *Peer reviewed In&Vertebrates, https://doi.org/10.52732/KUWG1044*.

Olson, C. M., Rennie, D., Cook, D., Dickersin, K., Flanagin, A., Hogan, J. W., ... & Pace, B. (2002). Publication bias in editorial decision making. *Jama*, *287*(21), 2825-2828.

Scheel, A. M., Schijen, M. R., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science*, *4*(2), 25152459211007467.

Smart, R. G. (1964). The importance of negative results in psychological research. *Canadian Psychologist / Psychologie canadienne, 5a*(4), 225–232. https://doi.org/10.1037/h0083036

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American statistical association*, *54*(285), 30-34.

Taber, K. S. (2019). Experimental research into teaching innovations: responding to methodological and ethical challenges. *Studies in Science Education*, *55*(1), 69-119. https://doi.org/10.1080/03057267.2019.1658058