

Why replication studies are essential: learning from failure and success

Anne-Wil Harzing

Version July 2016

Accepted for Cross Cultural & Strategic Management, vol. 23, no 4,
2016

Copyright © 2016, Anne-Wil Harzing

All rights reserved.

Prof. Anne-Wil Harzing
Middlesex University
The Burroughs, Hendon
London NW4 4BT
Email: anne@harzing.com
Web: www.harzing.com

Why replication studies are essential: learning from failure and success

Introduction

Van Witteloostuijn's (2016) commentary "What happened to Popperian Falsification?" is an excellent summary of the many problems that plague research in the (Social) Sciences in general and (International) Business & Management in particular. As van Witteloostuijn (2016:pp) admits his "[...] *diagnosis is anything but new – quite the contrary*", nor is it applicable only to the Social Sciences. When preparing this note, I was reminded of Cargo Cult Science, a 1974 Caltech commencement address by Physicist Richard Feynman (Feynman, 1974), which – more than four decades ago – makes many of the same points, including the pervasive problem of a lack of replication studies, which will be the topic I will focus on in this short rejoinder.

Conducting replication studies is more difficult in International Business (IB) than it is in many other disciplines. For instance in Psychology – a discipline that favours experimental research – one might be able to replicate a particular study within weeks or, in some cases, even days. However, in IB data collection is typically very time-consuming and fraught with many problems not encountered in purely domestic research (for a summary see Harzing, Reiche & Pudelko, 2013). Moreover, most journals in our field only publish articles with novel research findings and a strong theoretical contribution, and are thus not open to replication studies. To date, most studies in IB are therefore unique and are never replicated. This is regrettable, because even though difficult, replication is even more essential in IB than it is in domestic studies, because differences in cultural and institutional environments might limit generalization from studies conducted in a single home or host country.

Somehow though, pleas for replication studies – however well articulated and however often repeated – seem to be falling on deaf ears. Academics are only human, and many humans learn best from personal stories and examples, especially if they evoke vivid emotions or associations. Hence, in this note, instead of providing yet another essayistic plea for replication, I will attempt to argue "by example". Below, I will present two short case studies from my own research: one in which the lack of replication resulted in the creation of myths, and another in which judicious replication strengthened arguments for a new – less biased – measure of research performance. Finally, I will provide a recommendation on how to move forward that can be implemented immediately without the need for a complete overhaul of our current system of research dissemination.

Learning from failure: when a lack of replication creates myths and undermines scholarship

My first example refers to the topic of my first-ever academic journal publication, entitled: "*The persistent myth of high expatriate failure rates*" (Harzing, 1995). Expatriate failure is normally defined as the percentage of expatriates returning home before their assignment contract expires. One would be hard-pressed to find many articles on expatriate management that do *not* make the case for their study by stating that expatri-

ate failure rates are (very) high. However, after a forensic examination of referencing patterns, I found that there was almost no empirical foundation for the existence of high expatriate failure rates. The persistent myth of high expatriate failure rates seemed to have been created by massive (mis)quotations of three articles, as well as careless copying of references. Only **one** of the three articles (Tung, 1981) contained solid empirical evidence on expatriate failure rates and in fact showed them to be rather low.

As my article didn't quite have the impact I had hoped for and academics kept making the same unjustified assertions, I updated my analysis in 2001, generalising it to provide twelve guidelines for good academic referencing (Harzing, 2002). All twelve guidelines were habitually violated in the citation network of expatriate failure rates. These violations led to the self-perpetuating myth of high expatriate failure rates, thus seriously undermining the field's academic credibility and hindering its progress. However, this self-perpetuating myth would not have been sustainable if – rather than relying on argumentation simply by repetition of inaccurate interpretations – academics would have instead replicated Tung's study. Although this case study teaches us that even clear-cut empirical evidence *can* be distorted – remember Tung's study did *not* show that expatriate failure rates were high – this outcome would have been much less likely if empirical evidence to the contrary had accumulated through replication studies.

Learning from success: when replication supports the adoption of less biased research metrics

My second story is a success story. Since 2005, I have had an interest in bibliometric research, and in particular in research that redresses the traditionally disadvantaged position of the Social Sciences and Humanities in the evaluation of research performance. In this context, I have developed a new research metric based on the h-index (Hirsch, 2005). The h-index has taken the research community by storm: Hirsch's article has drawn nearly 6,000 Google scholar citations and the field of bibliometrics has witnessed an almost bewildering explosion of publications proposing h-index variants. However, none of these variants corrected for both disciplinary and career stage differences, a shortcoming that is addressed by our proposed hI,annual index (or hIa-index for short) (Harzing, Alakangas & Adams, 2014).

The hIa-index represents the average annual increase in the individual h-index, which is an h-index corrected for the number of co-authors. As such, the hIa-index measures the average number of single-author equivalent h-index points that an academic has accumulated in each year of their academic career. A hIa-index of 1.0 means that an academic has consistently published one article per year that, when corrected for the number of co-authors, has accumulated enough citations to be included in the h-index.

We tested this metric in a sample of 146 associate and full professors at the University of Melbourne – an elite university ranked in the top-30 worldwide – with an average academic age of 24 years. We found that whereas the h-index privileged full over associate professors and academics in the Life Sciences and Natural Sciences over the three other disciplines, using the hIa-index leveled the playing field both between junior and senior academics, and between the disciplines. The disciplinary effect is shown in Figure 1.

Recently, Ryan (2016) conducted an exact replication of our study with a much larger, but substantially different, and arguably more representative, sample: academics at **all**

levels, with an average academic age of 12 years, in a leading Middle Eastern University. He found both the career stage and the disciplinary effect to be almost identical to our original study, the only difference being the relatively low performance of Science academics in his sample. The disciplinary effect in Ryan’s study is shown in Figure 2.

Figure 1: h-index compared with hIa index for different disciplines (Harzing et al. 2014)

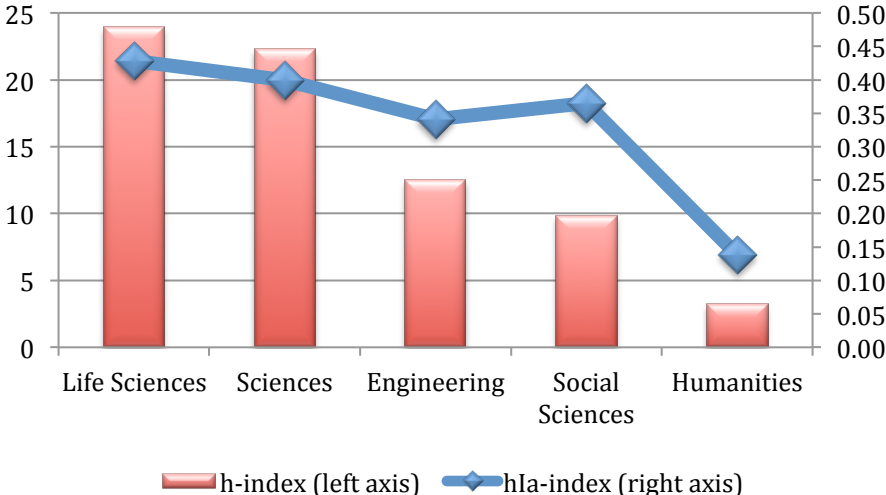
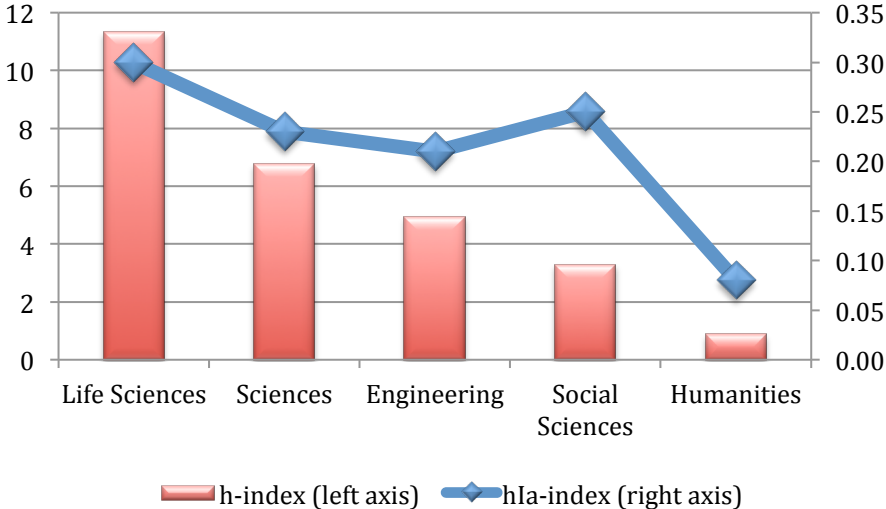


Figure 2: h-index compared with hIa index for different disciplines (Ryan, 2016)



Although the average h-index of the Middle Eastern sample is only just over a third of that of the Australian sample, the average hIa-index for the Middle Eastern is nearly two thirds of that of the Australian sample. This reflects the fact the hIa-index corrects for the very different academic age distribution in Ryan’s sample and thus provides us with further evidence that the hIa-index is more suitable than the h-index in comparing heterogeneous groups of researchers. There is still more work to do to in replicating these findings in other settings, such as countries in which research is mainly published in languages other than English, and with other databases, such as Google Scholar. However, I would argue that Ryan’s replication study contributed far more to our collective knowledge in bibliometrics than the publication of yet another new h-index variant.

How to promote replication studies? A call to journal editors

This short note has shown that a lack of replication lead to the creation of harmful myths, whereas judicious replication strengthened arguments for a new – less biased – measure of research performance. So why do we still not give replication studies a chance in our discipline? Van Witteloostuijn (2016) suggests the establishment of a new journal dedicated to replications to promote replication studies. However, I fear it is quite likely that academics will ignore such as journal. What might be a better solution is for **each** journal in our field to have a section called Replication Studies. That section would publish replication studies in the **exact** field of the journal. In that way, we can be assured that the studies are actually read by people who care about the topics in question.

I do realize that many journal editors will be reluctant to sacrifice journal pages to replication studies. First, they might argue that precious journal space should be reserved for novel research findingsⁱ. But isn't ultimately what should matter is whether, collectively, published research advances our knowledge and improves the quality and reliability of the research that is conducted? The positive effect of replication studies in this respect might be both direct, i.e. replications would allow us to separate chance findings from systematic results, and indirect, i.e. the knowledge that studies are likely to be replicated might counteract the temptation to compromise research integrity in the search for novel results. Moreover, replication studies would be particularly attractive to junior researchers, both as readers and as authors. As readers, junior researchers will benefit from demystification of the research process, as these articles would demonstrate that this process is rarely smooth and linear. As authors, junior academics will benefit from the opportunity to start their publishing career with more structured replication studies.

The second reason for journal editors' reluctance to publish replication studies might be that they expect these articles to be cited less frequently than original research articles. Thus they might be concerned about the adverse effect on the journal's impact factor and its associated standing.ⁱⁱ However, whether or not replication studies are cited less is an empirical question. Replication studies might be cited quite heavily in the literature review of an article, as authors would give preference to results that have been successfully replicated. In addition, the novelty of replication papers might draw curious readers to a journal and might hence lead to higher citation levels of its original research articles as well.

So what's stopping us? Editors can be powerful change agents. So let's stop waiting for things to happen and take the matter into our own hands. If we act now, in ten years time replication studies might be a standard fixture of our academic repertoire; I am confident our research will be all the better for it.

References

- Adler, N.; Harzing, A.W. (2009) When Knowledge Wins: Transcending the sense and nonsense of academic rankings, *The Academy of Management Learning & Education*, 8(1): 72-95.
- Feynman, R. P. (1974). Cargo Cult Science. *Engineering and Science*, 37(7), 10-13.
- Harzing, A.W. (1995) The persistent myth of high expatriate failure rates, *International Journal of Human Resource Management*, 6(May): 457-475.

- Harzing, A.W. (2002) Are our referencing errors undermining our scholarship and credibility? The case of expatriate failure rates, *Journal of Organizational Behavior*, 23(1): 127-148.
- Harzing, A.W.; Alakangas, S.; Adams, D. (2014) hIa: An individual annual h-index to accommodate disciplinary and career length differences, *Scientometrics*, 99(3): 811-821.
- Harzing, A.W.; Reiche B.S.; Pudelko, M. (2013) Challenges in international survey research: A review with illustrations and suggested solutions for best practice, *European Journal of International Management*, 7(1): 112-134.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.
- Ryan, J. C. (2016). A validation of the individual annual h-index (hIa): application of the hIa to a qualitatively and quantitatively different sample, in press for *Scientometrics*, DOI 10.1007/s11192-016-1972-8.
- Tung, R. L. (1981). Selection and training of personnel for overseas assignments. *Columbia journal of world business*, 16(1): 68-78.
- van Witteloostuijn, A. (2016). What happened to Popperian falsification? Publishing neutral and negative findings: moving away from biased publication practices. *Cross Cultural & Strategic Management*, 23(3): in press.

Endnotes

ⁱ This presumes that the number of journal pages is fixed. Given that the move to online-only publication is likely to be less than 10 years ahead, and hence the marginal cost of additional publications will decline, this should soon be a problem of the past. Moreover, given the astronomical profit margins of academic publishers, maybe the time has come to re-negotiate journal space with publishers?

ⁱⁱ I am not implying that this concern is healthy or justified. In fact, I have been very critical of journal rankings (see Adler & Harzing, 2009). However, I am just reflecting on the current reality.