Ranking Methodology

Selection of Journals

To produce department rankings of replicability, I downloaded articles from 105 major (high impact journals) that cover a wide range of disciplines (cognitive, developmental, clinical, neuro, social, personality, applied). Nevertheless, department rakings depend on the selection of journals and can change when the number of journals increases (similarly rankings based on the H-Index depend on the corpus of journals that is used to compute the H-Index).

Selection of Departments

The first list of departments includes private universities in the US and UK (Cambridge), top ranked universities (top 50 in the QS World Rankings of Psychology Departments), a wide range of nationalities which is limited to the concentration of psychology in Anglo-Saxon and Northern European nations.

Estimation Method

For each department, I downloaded the 2015 list of faculty members. If faculty members moved from one university to another, their past publications count towards their current university. For each faculty member, a computer program searched the articles for the author name and created a list of articles. Another program then searched for reported test statistics (mostly t and F tests reported in the text). Not all research is covered by this method (e.g., modeling research), but t and F-tests are used to claim a discovery. The test statistics for each author were saved in a file. The data from faculty members of a department were then pooled. If articles were co-authored by faculty members of the same department, an article was used only once. The pooled data were then used to estimate replicability using Brunner and Schimmack's method. Researchers who publish more significance tests have a stronger influence on the department score, just like these researchers contribute more to the H-Index and reputation of a department.

Reliabilty of Estimates

The rankings are based on the years 2010-2015. A six year period ensures that a large number of results are used and the 95% confidence intervals around the estimates are reasonably small. The data were also analyzed separately for the periods 2010-2012 and 2013-2015. The correlation between these estimates was r = .3. This implies a reliability of r = .46. Thus, even with a rather long period of six years, small differences between estimates are not interpretable. The reason for the low reliability is that the range of estimates is rather narrow (45% to 75%), suggesting that most psychology department follow a common implicit norm about the strength of evidence that is used for published results.