An aberrant abundance of Cronbach’s alpha values at .70

*Supplementary Materials*

## Table 1S. List of journals included in the psychology dataset (all APA journals 1985-2013)

|  |
| --- |
| **Journal name** |
| American Journal of Orthopsychiatry |
| American Psychologist |
| Behavioral Neuroscience |
| Canadian Journal of Behavioural Science |
| Canadian Journal of Experimental Psychology |
| Canadian Journal of Psychology |
| Canadian Psychology |
| Clinical Practice in Pediatric Psychology |
| Consulting Psychology Journal: Practice and Research |
| Couple and Family Psychology: Research and Practice |
| Cultural Diversity and Ethnic Minority Psychology |
| Cultural Diversity and Mental Health |
| Decision |
| Developmental Psychology |
| Dreaming |
| Emotion |
| Evolutionary Behavioral Sciences |
| Experimental and Clinical Psychopharmacology |
| Group Dynamics: Theory, Research, and Practice |
| Health Psychology |
| Health: the Journal of Collaborative Family Healthcare |
| History of Psychology |
| International Journal of Play Therapy |
| International Journal of Stress Management |
| International Perspectives in Psychology: Research, Practice, Consultation |
| Journal of Abnormal Psychology\*\* |
| Journal of Applied Psychology\* |
| Journal of Comparative Psychology |
| Journal of Consulting and Clinical Psychology |
| Journal of Counseling Psychology |
| Journal of Diversity in Higher Education |
| Journal of Educational Psychology |
| Journal of Experimental Psychology: Applied |
| Journal of Experimental Psychology: General |
| Journal of Experimental Psychology: Human Perception and Performance |
| Journal of Experimental Psychology: Learning, Memory, and Cognition |
| Journal of Family Psychology |
| Journal of Neuroscience, Psychology, and Economics |
| Journal of Occupational Health Psychology\* |
| Journal of Personality and Social Psychology |
| Journal of Psychotherapy Integration |
| Journal of Rural Mental Health |
| Journal of Social, Evolutionary, and Cultural Psychology |
| Journal of Theoretical and Philosophical Psychology |
| Journal of Threat Assessment and Management |
| Law and Human Behavior |
| Military Psychology |
| Motivation Science |
| Neuropsychology |
| Peace and Conflict: Journal of Peace Psychology |
| Personality Disorders: Theory, Research, and Treatment |
| Psychiatric Rehabilitation Journal |
| Psychoanalytic Psychology |
| Psychological Assessment |
| Psychological Bulletin |
| Psychological Methods |
| Psychological Review |
| Psychological Services |
| Psychological Trauma: Theory, Research, Practice, and Policy |
| Psychology of Addictive Behaviors |
| Psychology of Aesthetics, Creativity, and the Arts |
| Psychology of Consciousness: Theory, Research, and Practice |
| Psychology of Men & Masculinities |
| Psychology of Popular Media Culture |
| Psychology of Religion and Spirituality |
| Psychology of Sexual Orientation and Gender Diversity |
| Psychology of Violence |
| Psychology, Public Policy, and Law |
| Psychomusicology: a Journal of Research in Music Cognition |
| Psychomusicology: Music, Mind and Brain |
| Psychomusicology: Music, Mind, and Brain |
| Psychosocial Rehabilitation Journal |
| Psychotherapy |
| Psychotherapy: Theory, Research, Practice, Training |
| Qualitative Psychology |
| Rehabilitation Psychology |
| Review of General Psychology |
| School Psychology |
| School Psychology Quarterly |
| Spirituality in Clinical Practice |
| Sport, Exercise, and Performance Psychology |
| Stigma and Health |
| Training and Education in Professional Psychology |
| \* *Journal was included in both psychology and I/O datasets* |
| \*\* *Subsequently renamed Journal of Psychopathology and Clinical Science* |

## Table 2S. List of journals included in the I/O dataset (metaBUS: 1980-2017)

|  |
| --- |
| **Journal name** |
| Australasian Journal of Organisational Psychology |
| Academy of Management Journal |
| British Journal of Management |
| Employee Responsibilities and Rights Journal |
| European Journal of Work and Organizational Psychology |
| Group & Organization Management |
| Human Performance |
| Human Relations |
| Human Resource Management |
| Human Resource Management Journal |
| International Journal of Human Resource Management |
| International Journal of Selection and Assessment |
| Journal of Applied Psychology\* |
| Journal of Applied Social Psychology |
| Journal of Business and Psychology |
| Journal of Management |
| Journal of Managerial Psychology |
| Journal of Organizational Behavior |
| Journal of Occupational Health Psychology\* |
| Journal of Occupational and Organizational Psychology |
| Journal of Personnel Psychology |
| Journal of Vocational Behavior |
| The Leadership Quarterly |
| Organizational Behavior and Human Decision Processes |
| Personality and Individual Differences |
| Personnel Psychology |
| Work & Stress |
| \* *Journal was included in both psychology and I/O datasets* |

## Note 1S. Analyses of the I/O dataset after removing DOIs already present in the psychology dataset

The test of the first hypothesis found a 14% excess of values of .70, *Z* = 4.84, *p* < .00001. The test of the second hypothesis found excesses across the three bins, *Z* = 4.68, *p* = .00007, with an excess at .80 = 3%, excess at .90 = 1%. We therefore rejected the null hypothesis that there was no evidence of no excesses of values at common rule-of-thumb thresholds.

## Figure 1S. Observed counts of values with kernel smoothing (upper panel) and residuals (lower panel) in the I/O dataset, when overlap with the psychology dataset was removed



## Figure 2S. Frequency of use of constructs in the I/O dataset.



## Note 2S. Caliper tests

We judged caliper tests to be less suitable for our current purposes than the kernel smoothing method on the basis that there are plausible distributional differences between adjacent bins (i.e., the distribution of α values is non-uniform, see Figures 1 and 2). For example, assuming the general distribution of values observed in Figure 1 is roughly approximate to their true distribution, there are likely to be more observations of = .78 than .77 even in the absence of any distortions. Nonetheless, given the conceptual analogy between excesses of barely-significant *p* values and barely reliable α values, it is useful to include caliper tests as a robustness test given that they were employed in studies examining the distribution of *p* values (Hartgerink et al., 2016; Masicampo & Lalande, 2012). Caliper tests were not preregistered and were therefore exploratory. We employed a caliper width of one bin (i.e., = .01) and therefore compare counts between each bin and the preceding bin (i.e., .69 vs. .70, .79 vs. .80, and .89 vs. .90). Caliper ratios were calculated for each comparison in each dataset (i.e., counts of .70 divided by counts of .69). See Figure 2S for a visual illustration of how the kernel smoothing and caliper tests differ.

In the psychology dataset, the caliper ratio for the = .70 threshold 1.84. That is, nearly twice as many values of .70 were observed as .69. This larger over abundance relative to the kernel density approach can be attributed to the fact that the caliper tests also take the under-abundance of values in the pre-threshold bins (e.g., .69) as well as the over-abundance of values in the threshold bins (e.g., .70). Ratios for the other two thresholds were less extreme ( = .80: ratio = 1.21; = .90: ratio = 1.04). See Figure 3S. Rather than compare these ratios against a null hypothesis of zero (due to the above distributional considerations), we instead then calculated a ratio for every bin and its preceding bin between = .50 and .99. In order to test whether the caliper ratios at the thresholds were larger than other ratios, we then applied permutation tests to compare the threshold ratios against all non-threshold ratios. Results demonstrated that the = .70 caliper ratio was larger than the other ratios, *Z* = 3.40, *p* = .02; and that the .70, .80, and .90 ratios were larger than the other ratios, *Z* = 2.40, *p* = .014.

This pattern of results also generalized to the I/O dataset for two of the thresholds ( = .70: ratio 1.64; = .80: ratio = 1.13; = .90: ratio = 0.96). See Figure 4S. The = .70 caliper ratio was again found to be larger than the other ratios, *Z* = 2.34, *p* = .04; however the .70, .80, and .90 ratios were not found to be larger than the other ratios, *Z* = 1.40, *p* = .092. The pattern of excesses at = .70 were therefore robust to the choice of analytic method. The collective excesses at all three thresholds were not robust in the I/O dataset. This may be because of large ratios observed at other round values of which may have also been used as thresholds, but which we did not attempt to analyze (i.e., at .50 and .60; see Figure 4S).

## Figure 3S. Comparison between the kernel smoothing and caliper test methods in the psychology dataset.



## Figure 4S. Caliper ratios for each bin and the preceding bin in the psychology dataset.

****

## Figure 5S. Caliper ratios for each bin and the preceding bin in the I/O dataset



## Figure 6S. Caliper ratios for each bin and the preceding bin in the I/O dataset, when overlap with the psychology dataset is removed



## Table 3S. Standard errors associated with Cronbach’s = .70 for different sample sizes and numbers of items.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Number of items (*k*) | | | | | |
| Sample size (*n*) | 3 | 5 | 10 | 25 | 50 | 100 |
| 50 | 0.076 | 0.069 | 0.064 | 0.062 | 0.061 | 0.061 |
| 100 | 0.053 | 0.048 | 0.045 | 0.043 | 0.043 | 0.043 |
| 250 | 0.033 | 0.030 | 0.028 | 0.027 | 0.027 | 0.027 |
| 500 | 0.023 | 0.021 | 0.020 | 0.019 | 0.019 | 0.019 |
| 1000 | 0.016 | 0.015 | 0.014 | 0.014 | 0.014 | 0.013 |
| 2000 | 0.012 | 0.011 | 0.010 | 0.010 | 0.010 | 0.010 |
| 5000 | 0.007 | 0.007 | 0.006 | 0.006 | 0.006 | 0.006 |

## Note 3S. Caliper tests

Cronbach’s ’s variance is a function of (a) the value of , (b) the sample size (*n*), and (c) the number of items or parcels (*k*) from which it was calculated (van Zyl et al., 2000). In order to assess whether is sufficiently precisely estimated in common study designs that researchers could plausibly discriminate between scales whose population values were .69 versus .70, we calculated the standard error of = .70 for a range of sample sizes and numbers of items. As can be seen in the table, standard errors 0.006 even when the number of participants and items are both very large (i.e., *n* = 5000, *k* = 100). Differences in population reliabilities of just = .01 would not be detectable in typical study designs, and therefore researchers could not genuinely calibrate their decisions on differences this small.

# References

Hartgerink, C. H. J., van Aert, R. C. M., Nuijten, M. B., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Distributions of p-values smaller than .05 in psychology: What is going on? *PeerJ*, *4*, e1935. https://doi.org/10.7717/peerj.1935

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2271–2279. https://doi.org/10.1080/17470218.2012.711335

van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach’s alpha. *Psychometrika*, *65*(3), 271–280. https://doi.org/10.1007/BF02296146