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

Supplementary Materials

Ian Hussey, Taym Alsalti, Frank Bosco, Malte Elson[*] & Ruben Arslan[*]

[*] Shared last author

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: 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)

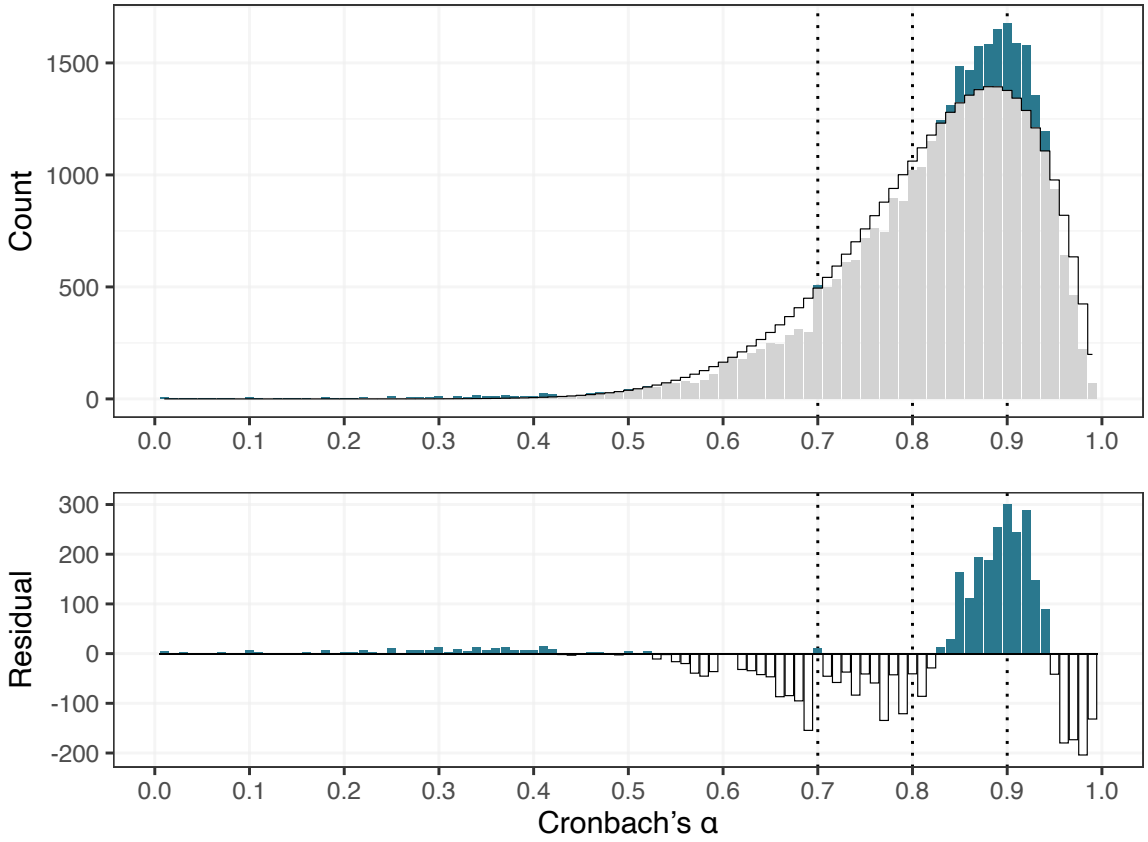
Journal name
Australasian Journal of Organisational Psychology
Academy of Management Journal
British Journal of Management
Employee Responsibilities and Rights Journal
European Journal of Work and Organizational Psychology
Group & Organization Management
Human Performance
Human Relations
Human Resource 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. Exploratory Beta regression fit to the psychology dataset

An exploratory Beta regression was fit to the observed distribution of binned α values, with binned αs as the dependent variable and intercepts for the location (μ) and precision (ϕ) parameters. Model fit parameters were μ = 1.545, SE = 0.004, p < .0001 and ϕ = 2.541, SE = 0.008, p < .0001. Inspection of the model fit (see Table 1S) indicated worse fit than kernel smoothing with much unmodelled variance.

Figure 1S. Observed counts of α values with fitted Beta regression (upper panel) and residuals (lower panel) in the psychology dataset



Note 2S. 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 α 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 α values at common rule-of-thumb thresholds.

Figure 2S. Observed counts of α values with kernel smoothing (upper panel) and residuals (lower panel) in the I/O dataset, when overlap with the psychology dataset was removed

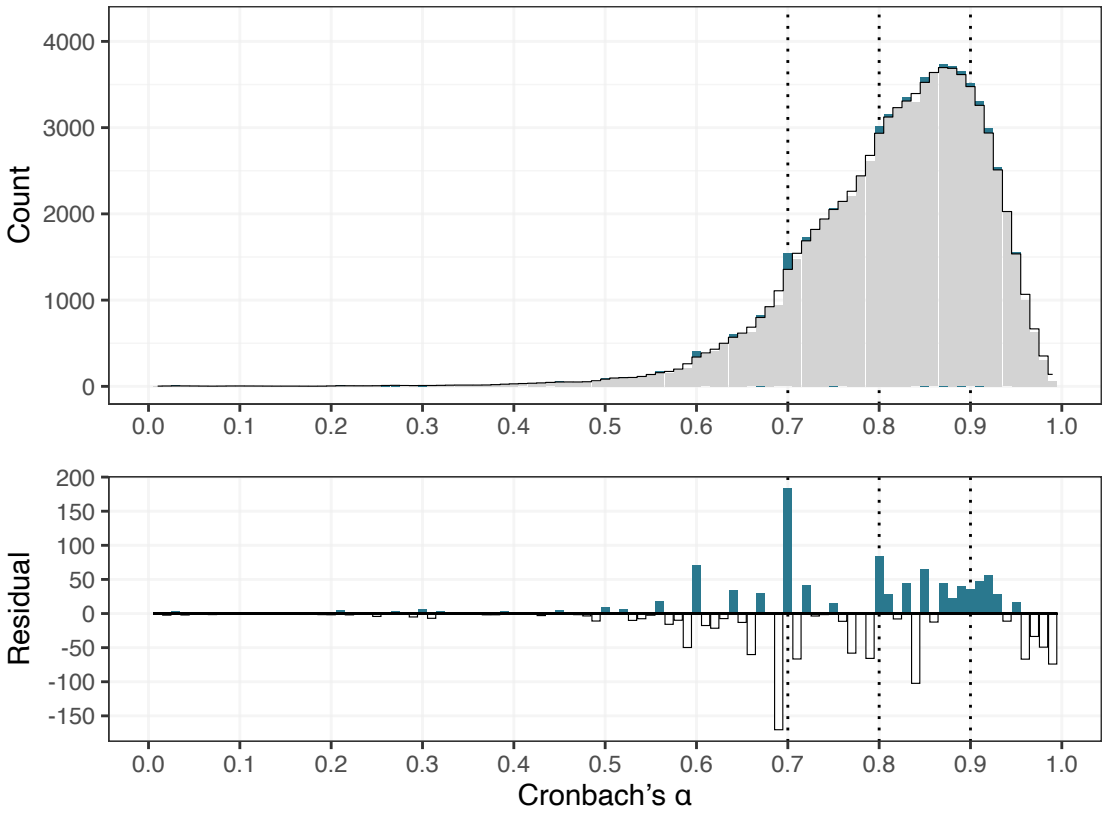


Figure 3S. Frequency of use of constructs in the I/O dataset

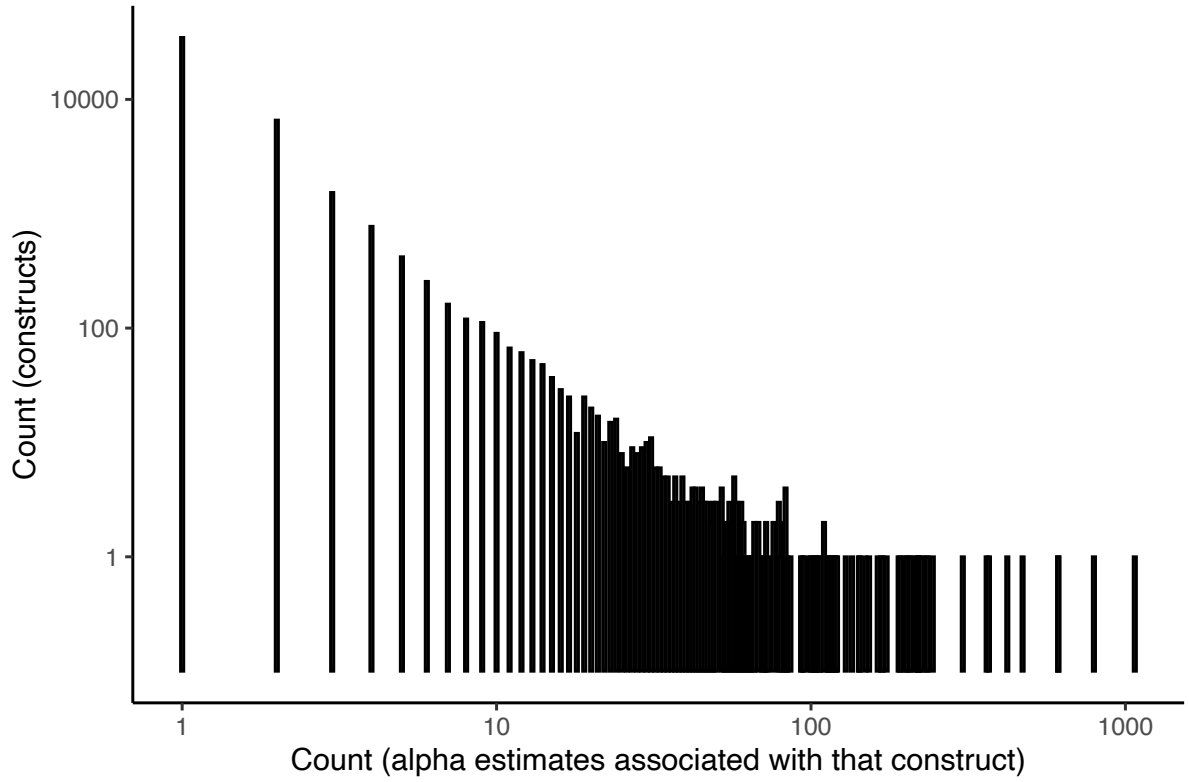


Figure 4S. Observed counts of α values with kernel smoothing (upper panel) and residuals (lower panel) in original measures in the PsycTests dataset

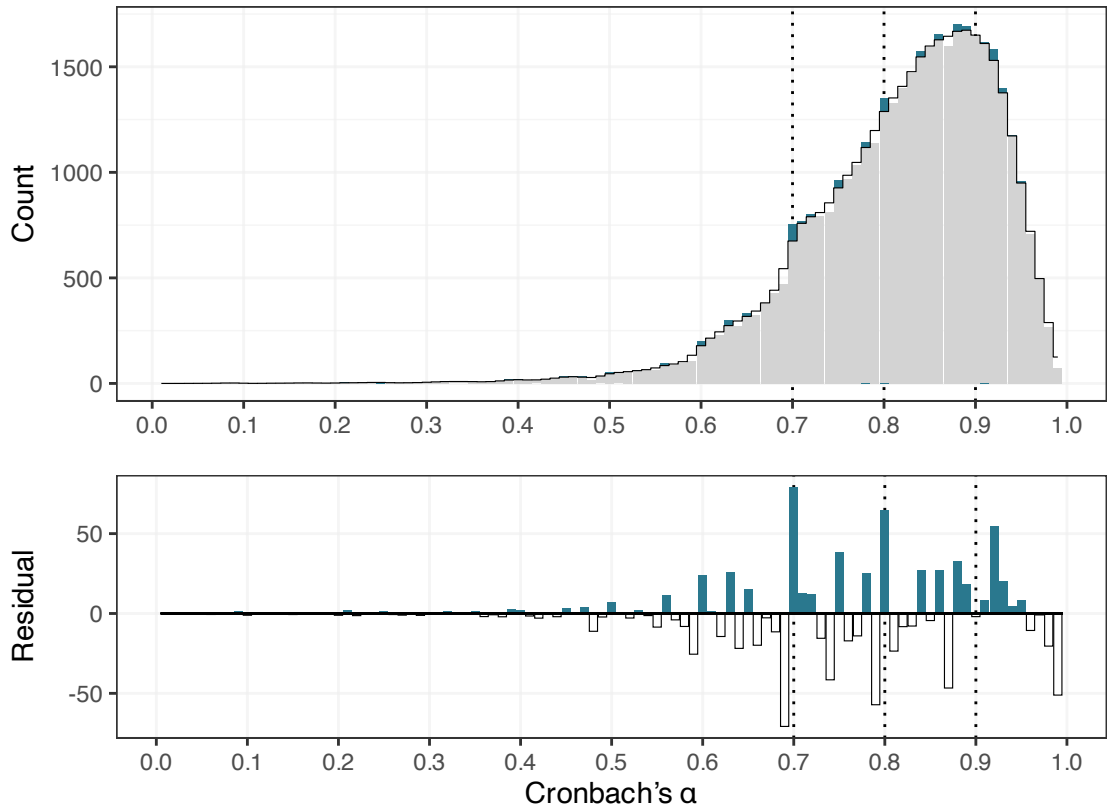
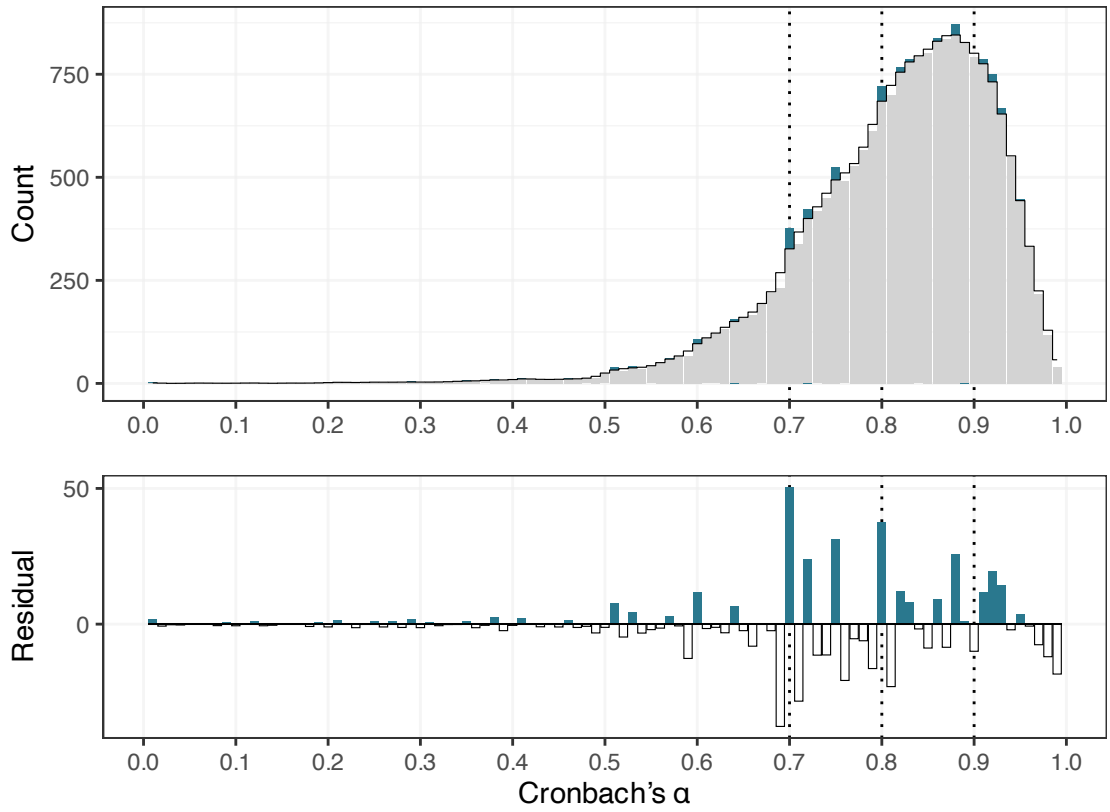


Figure 5S. Observed counts of α values with kernel smoothing (upper panel) and residuals (lower panel) in revised and translated measures in the PsycTests dataset



Note 3S. 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 to 3). For example, assuming the general distribution of α values observed in Figure 1 is roughly approximate to their true distribution, there are likely to be more observations of α = .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., α = .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 6S for a visual illustration of how the kernel smoothing and caliper tests differ.

In the psychology dataset, the caliper ratio for the α = .70 threshold was 1.71. That is, 71% more α values of .70 were observed than .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 α 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 (α = .80: ratio = 1.16; α = .90: ratio = 1.02). See Figure 7S. 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 α = .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 α = .70 caliper ratio was larger than the other ratios, $Z = 2.96$, $p = .020$; and that the .70, .80, and .90 ratios were larger than the other ratios, $Z = 1.97$, $p = .026$.

This pattern of results also generalized to the I/O dataset for two of the thresholds (α = .70: ratio 1.64; α = .80: ratio = 1.13; α = .90: ratio = 0.96). See Figure 4S. The α = .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, $Z = 1.40$, $p = .092$.

This pattern of excesses at .70 were found to generalize to the PsycTests dataset (α = .70: ratio 1.60), but not at the other two thresholds (α = .80: ratio = 1.19; α = .90: ratio = 0.98). See Figure 4S. The α = .70 caliper ratio was again found to be larger than the other ratios, $Z = 2.46$, $p = .04$; however the .70, .80, and .90 ratios were not found to be larger than the other ratios, $Z = 1.644$, $p = 0.065$.

The pattern of excesses at α = .70 were therefore robust to the choice of analytic method. The collective excesses at all three thresholds were not robust in the I/O or PsycTests datasets. This may be because of large ratios observed at other round values of α which are also sometimes used as thresholds, but which we did not attempt to analyze (i.e., at .50 and .60; see Figures 7S to 12S). Visual inspection of the excesses at .50 and .60 in the caliper plots (Figures 7S to 12S) may suggest that there were also excesses at those values. Future research may wish to examine excesses at these values.

Figure 6S. Comparison between the kernel smoothing and caliper test methods in the psychology dataset

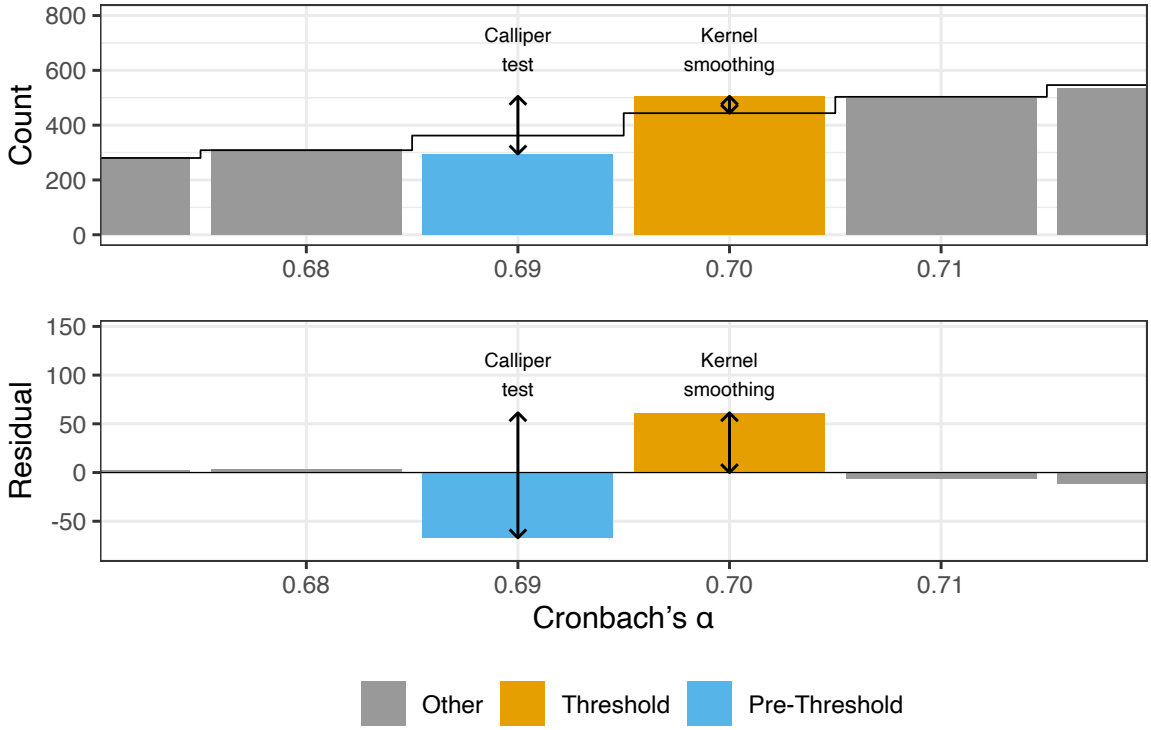


Figure 7S. Caliper ratios for each bin and the preceding bin in the psychology dataset

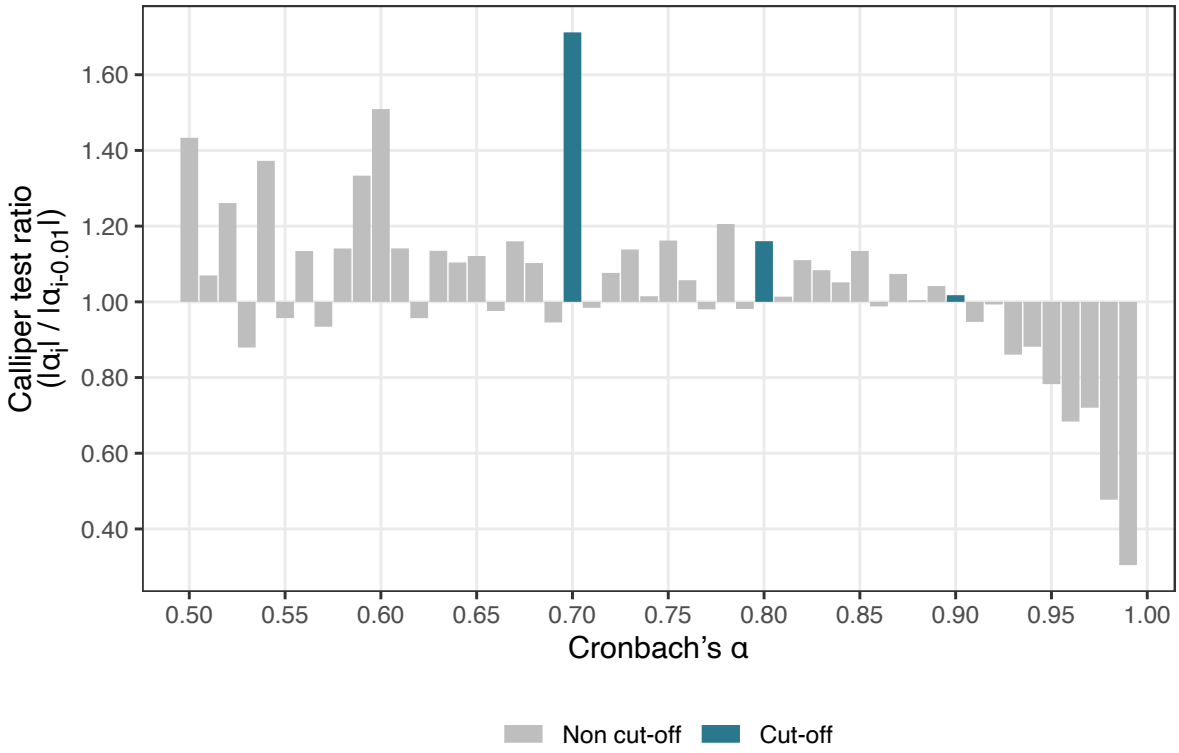


Figure 8S. Caliper ratios for each bin and the preceding bin in the I/O dataset

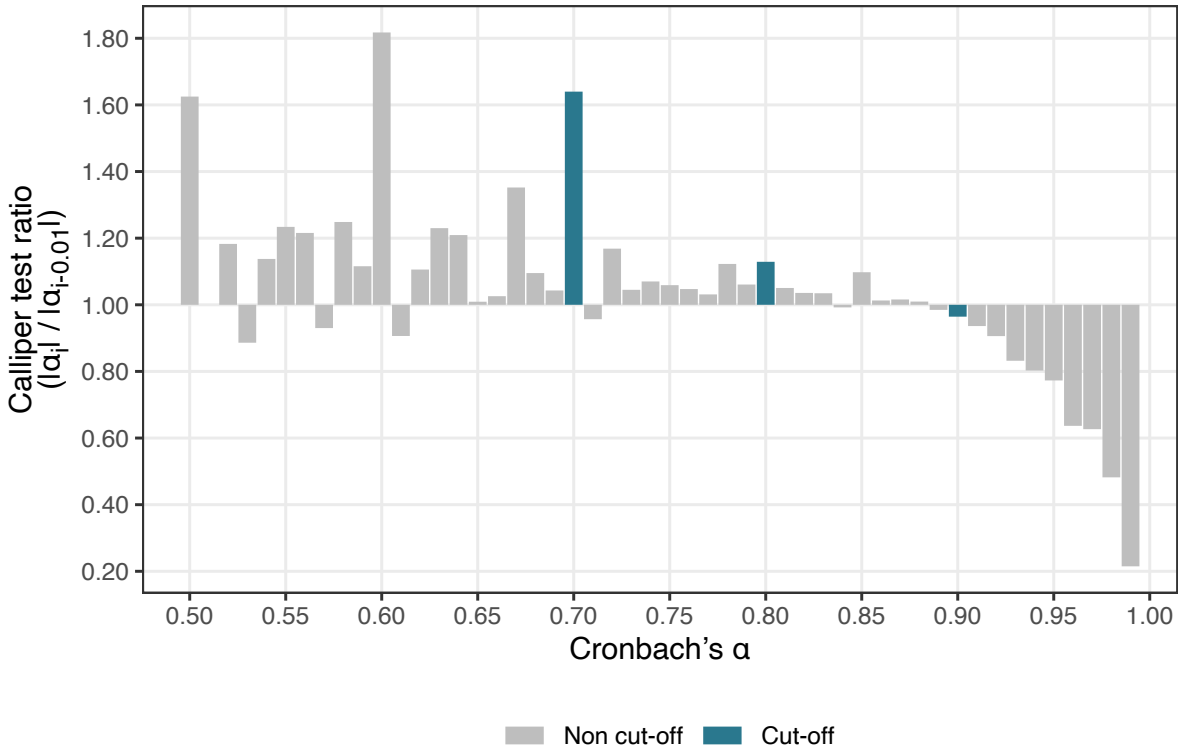


Figure 9S. Caliper ratios for each bin and the preceding bin in the I/O dataset, when overlap with the psychology dataset is removed

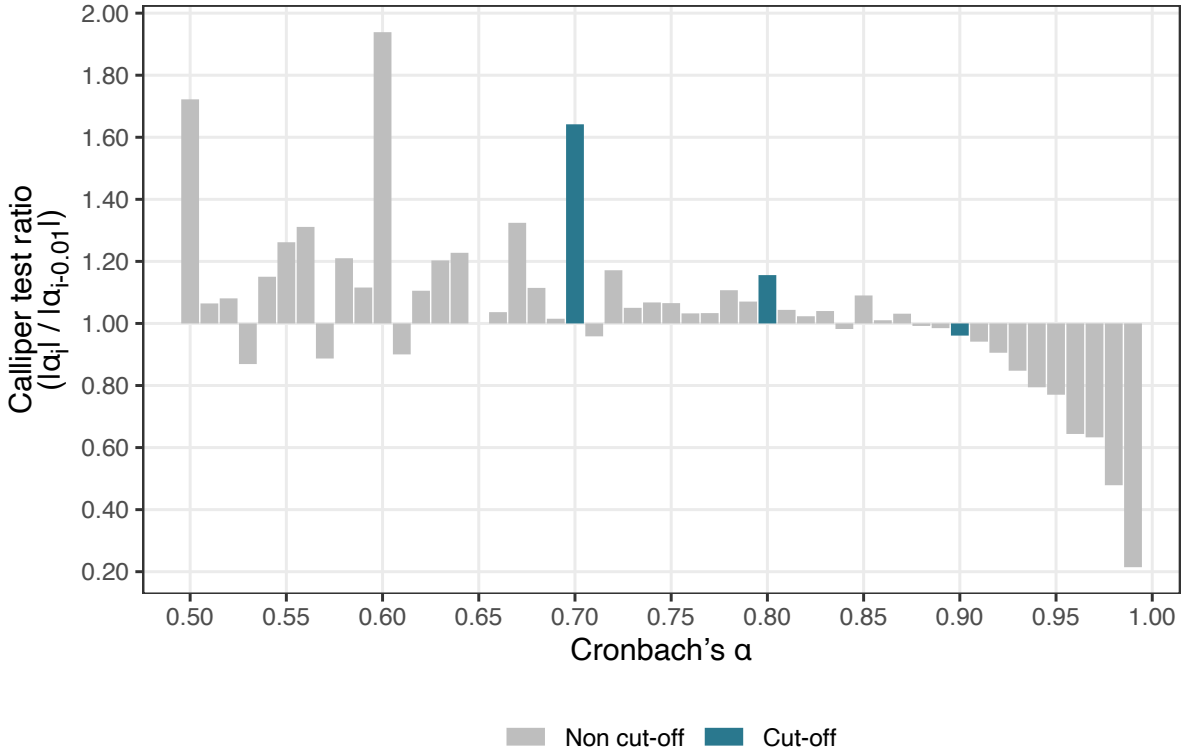


Figure 10S. Caliper ratios for each bin and the preceding bin in the full PsycTests dataset

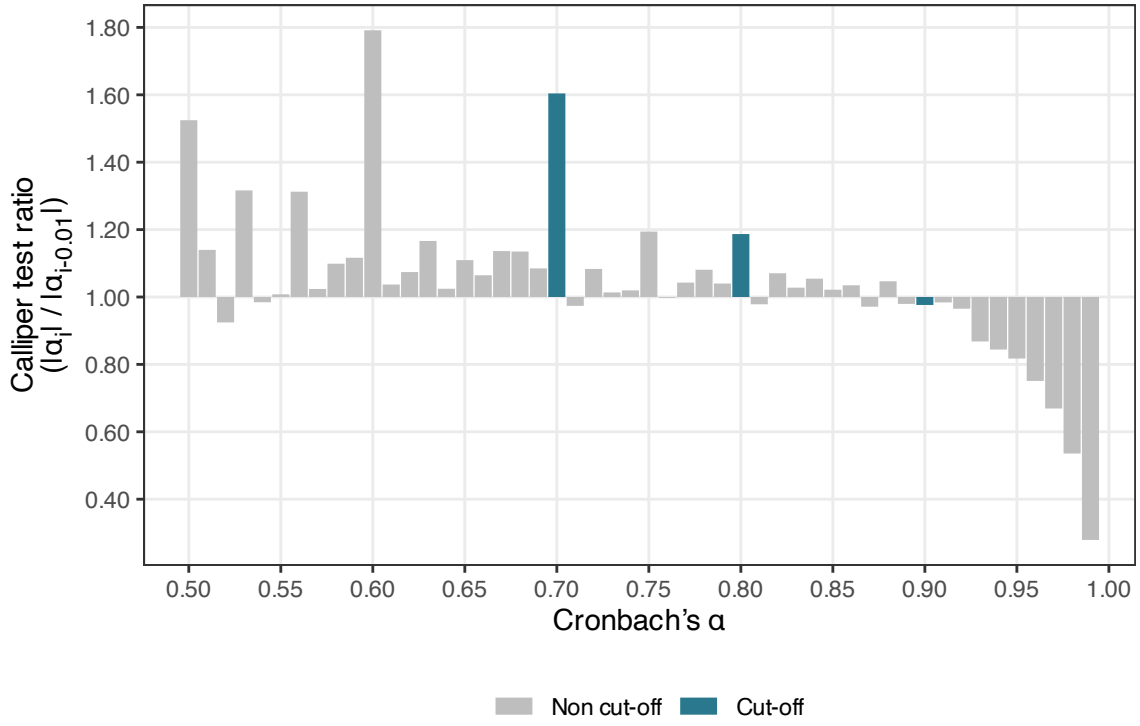


Figure 11S. Caliper ratios for each bin and the preceding bin in original measures included in the PsycTests dataset

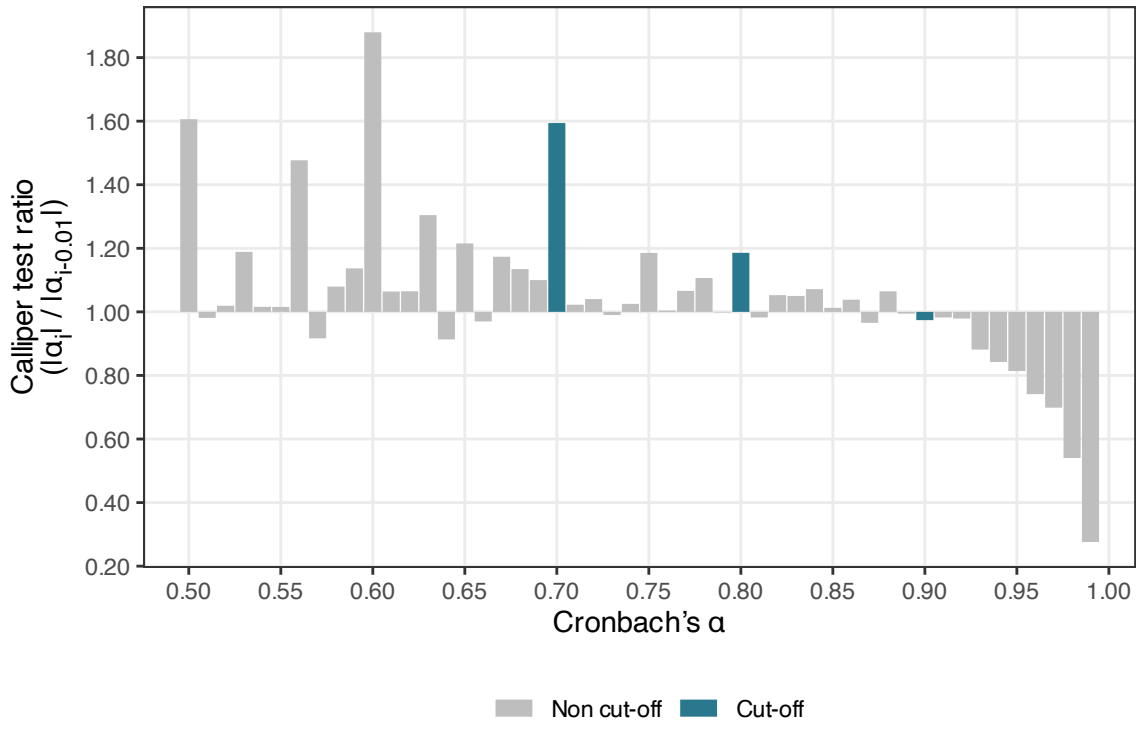


Figure 12S. Caliper ratios for each bin and the preceding bin in revised and translated measures included in the PsycTests dataset

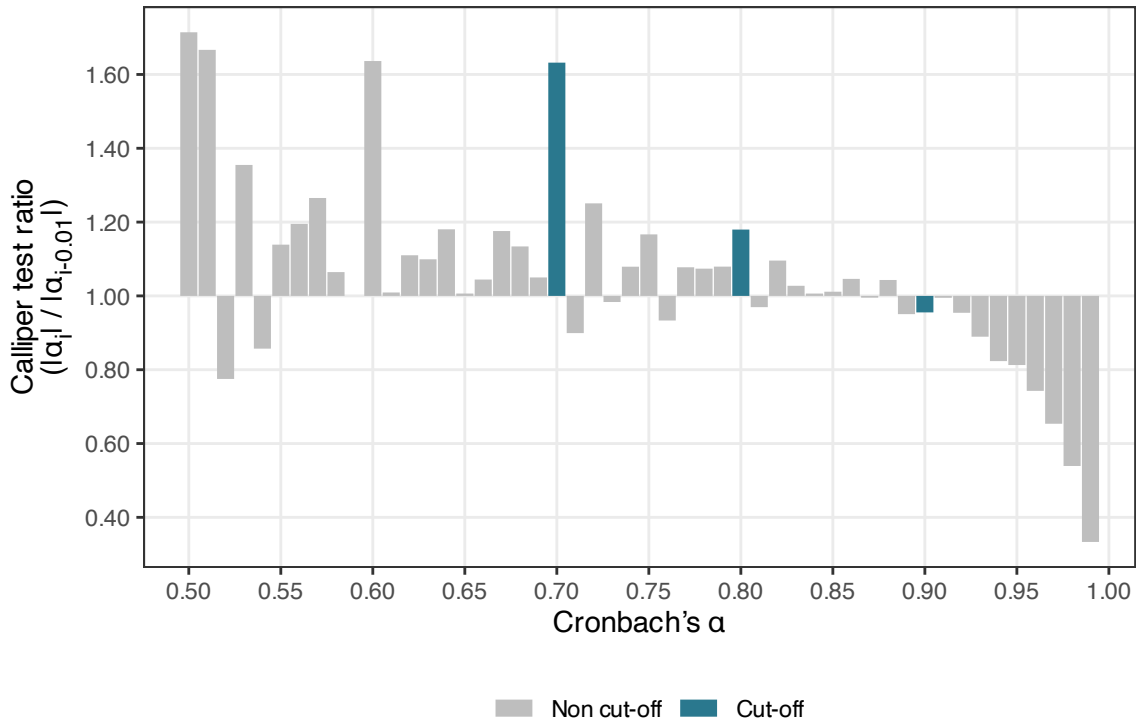


Table 3S. Standard errors associated with Cronbach's α = .70 for different sample sizes and numbers of items.

Number of items (*k*)	Cumulative percent of all psychological measures with ≤ *k* items	Standard Error of Cronbach's α = .70						
		n = 50	*n* = 100	*n* = 250	*n* = 500	*n* = 1000	*n* = 2000	*n* = 5000
3	5.8 %	0.076	0.053	0.033	0.023	0.016	0.012	0.007
5	14.8 %	0.069	0.048	0.030	0.021	0.015	0.011	0.007
10	34.7 %	0.064	0.045	0.028	0.020	0.014	0.010	0.006
15	50.1 %	0.063	0.044	0.028	0.020	0.014	0.010	0.006
20	63.5 %	0.062	0.044	0.028	0.019	0.014	0.010	0.006
25	73.1 %	0.062	0.043	0.027	0.019	0.014	0.010	0.006
50	92.5 %	0.061	0.043	0.027	0.019	0.014	0.010	0.006
100	98.0 %	0.061	0.043	0.027	0.019	0.013	0.010	0.006

Note that data for the cumulative percent of all psychological measures with up to *k* number of items was derived from the APA's PsycTests database.

Note 4S. The standard error associated with Cronbach's α = .70 across a range of sample sizes (*n*) and numbers of items (*k*)

Cronbach's α's variance is a function of (a) the value of α, (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 α is sufficiently precisely estimated in common study designs that researchers could plausibly discriminate between scales whose population α values were .69 versus .70, we calculated the standard error of α = .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 α = .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