



cocron: A Web Interface and R Package for the Statistical Comparison of Cronbach's Alpha Coefficients

Birk Diedenhofen, Jochen Musch

University of Düsseldorf, Germany

Abstract: To compare two or more measures with regard to their internal consistency reliability, Cronbach's alpha coefficients are usually computed for each measure. For a valid comparison between Cronbach's alpha coefficients, substantive differences need to be distinguished from circumstantial differences that occur due to chance. Unfortunately, however, popular statistics packages lack the ability to test for statistically significant differences between alpha coefficients. To remedy this problem, we present cocron, a platform-independent R package that provides functions to conduct significance tests for the comparison of two or more Cronbach's alpha coefficients for both dependent and independent groups of participants. The cocron package is part of the R statistical computing environment and is thus available for scripting. In addition, two graphical user interfaces – a web interface and a plugin for Rkward – are offered to provide comfortable access to the functionality of cocron.

Keywords: Cronbach's alpha, reliability, statistical test, significance, web interface, R

Introduction

Coefficient alpha (Cronbach, 1951) is by far the most widely used measure of reliability in the social sciences, and the computation of coefficient alpha has become routine practice whenever multiple-item scales are used to measure a single construct. As of 2016, Cronbach's (1951) seminal article on the computation of the internal consistency of a test has received 28,496 citations according to Google Scholar (Google, n.d.). A main reason for this popularity is that determining coefficient alpha has several advantages over alternative procedures for computing an estimate of the reliability of a scale. Unlike test-retest reliability, Cronbach's alpha can be determined in a single session; and unlike parallel-forms reliability, the computation of Cronbach's alpha does not require a parallel test that has to conform to strict requirements. The computation of Cronbach's alpha as a general measure of internal consistency is also superior to the computation of a split-half reliability coefficient because there are many ways to split a test into two halves, and the arbitrary choice of any particular splitting criterion offers an unwanted researcher degree of freedom that calls into question the generalizability of split-half reliability coefficients (Lord, 1956; Simmons, Nelson, & Simonsohn, 2011). Unlike the computation of split-half reliabilities, the computation of the more general coefficient alpha does not hinge on arbitrary decisions. For a test consisting of k items (or parts) that measures a quantity y , Cronbach's alpha is defined as

$$\hat{\zeta} = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \hat{\sigma}_{x_j}^2}{\hat{\sigma}_y^2} \right) \quad (1)$$

(Feldt, Woodruff, & Salih, 1987, formula 1) with $y = x_1 + x_2 + \dots + x_k$. $\hat{\sigma}_{x_j}^2$ denotes the variance of item j , and $\hat{\sigma}_y^2$ is the variance of the test score for a sample of individuals that completed the test. Thus, Cronbach's alpha is based on the ratio of the sum of item variances to the test score variance, employing a correction factor to take the number of items into account. (To avoid confusion with the statistical significance level α , the population alpha coefficient and the sample alpha coefficient are denoted by ζ and $\hat{\zeta}$, respectively.) The upper and lower bounds of a $100(1 - \alpha)$ percent confidence interval for ζ are calculated as

$$\zeta_U = 1 - \left[(1 - \hat{\zeta}) F \left(\frac{\alpha}{2} \right) \right] \quad (2)$$

and

$$\zeta_L = 1 - \left[(1 - \hat{\zeta}) F \left(1 - \frac{\alpha}{2} \right) \right] \quad (3)$$

(Feldt et al., 1987, formulae 6 and 7), respectively. $F \left(\frac{\alpha}{2} \right)$ and $F \left(1 - \frac{\alpha}{2} \right)$ denote the $100 \left(\frac{\alpha}{2} \right)$ and $100 \left(1 - \frac{\alpha}{2} \right)$ percentiles of an F -distribution with $n - 1$ and $(k - 1)(n - 1)$ as its *df*. α specifies the desired alpha level of the confidence interval and n denotes the size of the sample. A major drawback of Cronbach's alpha is that it cannot be computed if data are missing for one or more items or testees, and we are not aware of any validated approaches to impute missing values in the context of computing Cronbach's alpha. Furthermore, the computation of Cronbach's alpha is not appropriate to estimate intercoder or interrater reliabilities.

Statistical comparisons of alpha coefficients are frequently needed in online research settings. However, even though they are commonly required for substantive reasons, such statistical comparisons are often missing. For example, researchers often want to compare the reliability of competing measures administered to the same sample or the reliability of one measure that was administered to different samples. In these cases, a statistical test of two or more alpha coefficients is necessary to obtain an accurate assessment of measurement precision.

For example, with the advent of online testing, numerous studies (e.g., Ihme et al., 2009; Templer & Lange, 2008; Schreiner, Reiss, & Schweizer, 2014) have investigated whether online tests provide the same results as equivalent offline versions of the same test. To answer this question, it is necessary to calculate Cronbach's alpha and to compare online and offline tests with regard to their internal consistency. Even though descriptive differences in coefficient alpha have frequently been reported, they have rarely been tested for statistical significance (e.g., Barke, Nyenhuis, & Kröner-Herwig, 2012; Hewson & Charlton, 2005; Riva, Teruzzi, & Anolli, 2003). For instance, Barke et al. (2012) validated a German translation of the Internet Addiction Test. Even though the comparison of the online and offline versions of this test was central to their investigation, they neglected to report that (according to our own computation) their online version achieved a significantly higher reliability than their offline version. Similarly, Hewson and Charlton (2005) compared responses to the Multidimensional Health Locus of Control Scale that were collected online with responses that were collected offline. On the basis of a descriptive comparison of Cronbach's alpha coefficients, they concluded that the internal consistency of this test was higher for the online version than for the offline version. However, an (unreported) statistical test of the difference between the alpha coefficients reveals that the supremacy of the online test holds only for one of the three subscales of the test. For the two remaining subscales, the difference is not statistically reliable. Riva et al. (2003) also administered surveys to both an online and an offline sample. For the Computer Use Survey and the Internet Attitudes Survey, they reported that the online version had "slightly lower values" (p. 77) of Cronbach's alpha. Even though this may potentially cast doubt on the comparability of the two modes of administration, they did not report whether this difference was statistically significant.

Statistical tests for differences in reliability are also important in another area of online research. A common finding is that drop-out rates are higher in online studies than in offline studies (Reips, 2000; Nestler, Thielsch, Vasilev, & Back, 2015). Because drop-out rates increase with test length (Galesic, 2006; Yan, Conrad, Tourangeau, & Couper, 2011), online surveys should be kept short to hold drop-out rates at a low level. For this reason, online studies often employ short scales or short versions of existing measures. However, to shorten an existing scale, a reduced set of items that still shows convincing psychometric properties needs to be identified. Ideally, the short scale's alpha coefficient should not differ significantly from the long version of the scale. Failing to properly test the reliability of short forms runs the risk of using defective measures and has been referred to as one of the "sins of short form development" (Smith, McCarthy, & Anderson, 2000, p. 104). Comparisons between the coefficient alphas of different item sets are necessary for constructing a short scale with as little loss in reliability as possible. Unfortunately, however, many studies that used Cronbach's alpha to develop short scales did not apply tests of statistical significance to compare the difference in reliability between tests of varying length

(e.g., D’Ath, Katona, Mullan, Evans, & Katona, 1994; Putnam, Helbig, Gartstein, Rothbart, & Leerkes, 2014; Richins, 2004).

The above-mentioned examples illustrate that statistical comparisons of Cronbach’s alphas are rarely reported, although such tests have been around for many years. Feldt et al. (1987) give an overview of the development of a sampling error theory for Cronbach’s alpha. In their seminal article, these authors outline proofs, derivations, and numerical illustrations for significance tests comparing two or more dependent or independent alpha coefficients. The test statistic for the comparison of m independent alpha coefficients is based on the transformation $(1 - \hat{\zeta}_i)^{-\frac{1}{3}}$ with $\hat{\zeta}_i$ being the sample alpha coefficient (Feldt et al., 1987). The variance of this transformation can be estimated by

$$S_i^2 = \frac{2}{9(\tilde{k}_i - 1)(1 - \hat{\zeta}_i)^{\frac{2}{3}}} \quad (4)$$

(Feldt et al., 1987, formula 19), where $\tilde{k}_i = (k_i - 1)n_i/(k_i + 1)$. k_i are the number of items in the test and n_i the number of testees. If all $\hat{\zeta}_i$ are based on independent samples, the test statistic is

$$UX_{indep} = \sum_{i=1}^m \frac{\left((1 - \hat{\zeta}_i)^{-\frac{1}{3}} - \bar{\mu} \right)^2}{S_i^2} \quad (5)$$

(Feldt et al., 1987, formula 21) with

$$\bar{\mu} = \sum_{i=1}^m \frac{(1 - \hat{\zeta}_i)^{-\frac{1}{3}}}{m} \quad (6)$$

(Feldt et al., 1987, formula 20) and \bar{S}^2 being the arithmetic mean across all S_i^2 . The test statistic is approximately distributed as χ^2 with $df = m - 1$ when the null hypothesis is true. The comparison of m independent alpha coefficients can be considered as a special case of the comparison of m dependent coefficients. Hence, the test statistic for the comparison of m dependent alpha coefficients also builds upon the $(1 - \hat{\zeta})^{-\frac{1}{3}}$ transformation (Feldt et al., 1987). However, in addition to the variance S_i^2 , an approximation of the covariance S_{ij} between $(1 - \hat{\zeta}_i)^{-\frac{1}{3}}$ and $(1 - \hat{\zeta}_j)^{-\frac{1}{3}}$ is required that is based on the sample correlation $\hat{\rho}_{ij}$ between the scores on the tests i and j :

$$S_{ij} = \frac{2\hat{\rho}_{ij}^2}{9(\tilde{k} - 1)(1 - \hat{\zeta}_i)^{\frac{1}{3}}(1 - \hat{\zeta}_j)^{\frac{1}{3}}} \quad (7)$$

(Feldt et al., 1987, formula 25). The test statistic for the comparison of m dependent alpha coefficients is

$$UX_{dep} = \sum_{i=1}^m \frac{\left((1 - \hat{\zeta}_i)^{-\frac{1}{3}} - \bar{\mu} \right)^2}{(\bar{S}^2 - \bar{C})} \quad (8)$$

(Feldt et al., 1987, formula 26), where \bar{C} is the average of the covariances S_{ij} . Under the null hypothesis, the test statistic is approximately distributed as χ^2 with $df = m - 1$.

An obvious reason why statistical tests for the comparison of Cronbach’s alphas are often missing is the lack of readily available software. Popular statistics packages such as SPSS or SAS do not include the necessary tests. Even though there are several packages for the R statistical computing environment (R Core Team, 2016) that implement functions to determine Cronbach’s alpha (e.g., Falissard, 2012; Revelle, 2015), none of these packages allow users to test differences between alpha coefficients for statistical significance. Thus, there is currently no R package that allows users to determine whether two or more alpha coefficients differ to a larger extent than can be expected by chance.

There are only two stand-alone programs that are capable of comparing alpha coefficients: the FORTRAN-77 based *ALPHATST* written for DOS (Lautenschlager, 1989) and *AlphaTest* written for Windows (Lautenschlager & Meade, 2008). Both programs, however, are closed source and have not been integrated into a statistical computing environment. They are therefore not available for scripting. Abd-El-Fattah and Hassan (2011)

introduced the *Dependent-Alpha Calculator*, a Microsoft Excel spreadsheet that allows for comparison of alpha coefficients. However, this tool is restricted to a comparison of two dependent alpha coefficients and it requires a Microsoft Office installation to run. The platform-independent *cocron* package that is introduced here provides free and open software to compare coefficient alpha in the R statistical computing environment and can be used for scripting. Moreover, *cocron* is available as a web service similar to other websites offering interfaces to conduct statistical analyses (e.g., Diedenhofen & Musch, 2015; Freelon, 2010, 2013).

Web interfaces provide worldwide and platform-independent access to statistical methods and allow users to conduct analyses in a web browser without installing additional software. Thus, web interfaces lower the hurdles for users who are unacquainted with a particular method, and thereby, web interfaces have the potential to expand the number of people using this method. Most websites that provide interfaces for statistical analyses are free of charge (e.g., Diedenhofen & Musch, 2015; Freelon, 2010, 2013). However, not all websites grant access to the underlying source code (e.g., Uitenbroek, n.d.). Similar to most proprietary software, their source code is often not published and is therefore not available for users to review. Another drawback of web interfaces that offer statistical analyses is that their users are dependent on the availability of the service that is provided. Also, such analyses might not be reproducible if the service changes or is discontinued. Therefore, not only should web sites that offer statistical analyses share the source code behind the software, but they should also support users in conducting the same analyses offline on their own computers. The *cocron* package meets all of these requirements.

Comparing Cronbach's alpha coefficients with *cocron*

The software package *cocron* (an abbreviation for "COMparing CRONbach's alphas"; version 1.0-1) provides functions for comparing Cronbach's alpha coefficients in the R statistical computing environment, which is freely available for Windows, Mac, and Linux systems. It can be used to compare two or more alpha values that are based on either dependent or independent groups. The implementation relies on methods described by Feldt et al. (1987). The *cocron* R package can be downloaded from the Comprehensive R Archive Network (CRAN) at <http://cran.r-project.org/package=cocron>. To install *cocron* from the R console, it is sufficient to run the command `install.packages("cocron")`. The functionality of the package is made available by typing `library("cocron")`. The function `cocron()` calculates and compares Cronbach's alpha coefficients from raw data. If raw data are not available, the function `cocron.n.coefficients()` allows users to compare dependent or independent alpha coefficients that have already been determined (formulae 5 and 8). Internally, `cocron()` passes the calculated alpha values to `cocron.n.coefficients()`. The `cronbach.alpha.CI()` function calculates confidence intervals for Cronbach's alpha coefficients (formulae 2 and 3) that can be used to test if the alpha coefficient of a given test exceeds a certain benchmark value. The *cocron* R package includes test routines that are run whenever the package is built. They verify the accuracy of all calculations using the results of AlphaTest (Lautenschlager & Meade, 2008) and published examples (Feld et al., 1987).

Graphical user interfaces

A web interface for *cocron* is available at <http://comparingcronbachalphas.org>. It includes a wizard dialogue that will guide the user through the analysis. Figure 1 shows a screenshot of the *cocron* web interface. On the first page of the wizard, the user has to indicate whether the alpha coefficients that are to be compared are dependent or independent. Alpha coefficients are dependent if the responses to the measures that are to be compared were provided by the same participants. Alpha coefficients are independent if they rely on the responses of different groups of test takers. On the second page, the user is asked to specify the alpha coefficients, the sample sizes, and the number of items for each test. If the alpha coefficients are dependent, the user is asked to indicate the correlation between the sets of test scores from which the alpha coefficients were calculated. This is because the power of the test depends on this correlation such that differences in reliability can be detected more easily if the correlation between two tests is high. On the next page, the user has to specify the level of significance for the test (e.g., .05) and the level of confidence for the confidence intervals that are calculated (e.g., .90 or .95). On the final page, the test results and the R code that is required to attain the results are presented and may be copied, printed, or downloaded.

cocron comparing Cronbach alphas

Please provide the Cronbach alphas you want to compare and the sample sizes and number of items they are based on:

	Cronbach's alpha	Sample size	Item count	
1	<input type="text" value="0.52"/>	<input type="text" value="420"/>	<input type="text" value="30"/>	<input type="button" value="Remove"/>
2	<input type="text" value="0.43"/>	<input type="text" value="342"/>	<input type="text" value="30"/>	<input type="button" value="Remove"/>
3	<input type="text" value="0.48"/>	<input type="text" value="243"/>	<input type="text" value="30"/>	<input type="button" value="Remove"/>

Figure 1. Screenshot of the cocron web interface on <http://comparingcronbachalphas.org>.

Comparing Cronbach alphas

Please provide the Cronbach alphas you want to compare and the sample sizes and number of items they are based on:

Number of alpha coefficients to be compared:

Alpha	Sample size	Item count
1 .73	1 109	1 20
2 .76	2 120	2 20
3 .77	3 124	3 20

Figure 2. Screenshot of the cocron GUI plugin for the R front-end RKWard.

As an additional graphical interface, the cocron package also includes a plugin for the platform-independent R front-end RKWard (Rödiger, Friedrichsmeier, Kapat, & Michalke, 2012). After installing the R package, the cocron interface becomes directly available in the RKWard menu. A screenshot of the dialogue is shown in Figure 2. Similar to the web interface, users are guided through the analysis step by step. The RKWard plugin allows users to either enter alpha coefficients manually or to compute them from raw data. At the end of the wizard dialogue, users may either run the analysis directly or add the R code of the analysis to their R script.

Code examples

In the following, code examples are provided to demonstrate how the cocron package is used in the R console. All data used in the examples are simulated data created for the purpose of illustrating the use of cocron. The first example shows how to conduct a statistical comparison between two measures that both consist of 20 items. We assumed that for the two measures, consistencies of $\alpha_1 = .78$ and $\alpha_2 = .72$ were calculated from the responses of two independent groups of $n_1 = 120$ and $n_2 = 200$ participants, respectively:

```
library(cocron) # load cocron package

cocron.n.coefficients(alpha=c(.78,.72), items=c(20,20), n=c(120,200),
dep=FALSE)
```

The above command creates the following console output, which shows that the difference between these two independent alphas is not statistically significant:

```
Compare n alpha coefficients

Comparison between: a1 = 0.78, a2 = 0.72
The coefficients are based on independent groups
95% confidence intervals: CI1 = 0.7186 0.8333,
CI2 = 0.6605 0.7733
Group sizes: n1 = 120, n2 = 200
Item count: i1 = 20, i2 = 20
Null hypothesis: a1 and a2 are equal
Alternative hypothesis: a1 and a2 are not equal
Level of significance: 0.05

chisq = 1.9139, df = 1, p-value = 0.1665
Null hypothesis retained
```

For the comparison of dependent alpha coefficients, a matrix containing the correlation coefficients in the upper triangle has to be passed as parameter r . In the next example, three tests consisting of 50, 40, and 35 items were administered to a sample of $n = 100$ participants. We assumed that for the three tests, consistencies of $\alpha_1 = .85$, $\alpha_2 = .87$, and $\alpha_3 = .80$ were determined. Moreover, we supposed that the correlation between test 1 and test 2 was $r_1 = .8$, the correlation between test 1 and test 3 was $r_2 = .6$, and the correlation between test 2 and test 3 was $r_3 = .7$. The following code compares the three alpha coefficients and provides the intercorrelations between the test scores in matrix r :

```
r <- rbind(
  c( 1, .8, .6),
  c(NA, 1, .7),
  c(NA, NA, 1)
)

cocron.n.coefficients(alpha=c(.85,.87,.80), items=c(50,40,35), n=100,
dep=TRUE, r=r)
```

The output for the above code shows that the null hypothesis that the three dependent alpha coefficients do not differ has to be rejected:

```
Compare n alpha coefficients

Comparison between: a1 = 0.85, a2 = 0.87, a3 = 0.8
The coefficients are based on dependent groups
```

```
95% confidence intervals: CI1 = 0.8048 0.8891,  
  CI2 = 0.8307 0.9040, CI3 = 0.7394 0.8523  
Group sizes: n1 = 100, n2 = 100, n3 = 100  
Item count: i1 = 50, i2 = 40, i3 = 35  
Null hypothesis: a1, a2, and a3 are equal  
Alternative hypothesis: a1, a2, and a3 are not equal  
Level of significance: 0.05
```

```
chisq = 8.8349, df = 2, p-value = 0.0121  
Null hypothesis rejected
```

To calculate and compare alpha coefficients from raw data, the function `cocron()` is available. In the following example, two tests with 30 items each were administered to a sample of $n = 312$ participants. First, the data set (`knowledge`) is loaded – a list that contains two matrices holding the responses to both tests. Next, the function `cocron()` determines Cronbach's alpha for the two tests and conducts a significance test for dependent alpha coefficients that takes into account the correlation between the underlying test scores:

```
data("knowledge") # load data  
cocron(knowledge, dep=TRUE)
```

The following output of the `cocron()` function shows that in this example, the alpha coefficients of the two tests differ significantly:

```
Compare n alpha coefficients
```

```
Comparison between: a1 = 0.7639, a2 = 0.8042  
The coefficients are based on dependent groups  
95% confidence intervals: CI1 = 0.7246 0.8000,  
  CI2 = 0.7716 0.8342  
Group sizes: n1 = 312, n2 = 312  
Item count: i1 = 30, i2 = 30  
Null hypothesis: a1 and a2 are equal  
Alternative hypothesis: a1 and a2 are not equal  
Level of significance: 0.05
```

```
chisq = 5.6173, df = 1, p-value = 0.0178  
Null hypothesis rejected
```

Discussion

In this article, we introduced `cocron`, a free software package for the R programming language. The `cocron` package allows users to test for differences in the reliability of tests measured via Cronbach's alpha coefficients. It is available for scripting within the R environment and can be accessed conveniently either via a web interface or as a plugin for the RKWard front-end. Thus, `cocron` allows users to conduct tests that are necessary to compare the reliability of competing measures in a convenient and user-friendly way. The `cocron` software package may be applied in Internet-based research to test online and offline versions of a measure for equivalence, or to create and assess the reliabilities of short test forms for online assessment. Beyond that, a broad application of `cocron` in different psychometric contexts is possible and recommended. Comparisons of Cronbach's alpha coefficients are of interest whenever competing test instruments, test construction techniques, item selection strategies, test administration modes, item formats, test scoring procedures, or subpopulations are investigated (Feldt et al., 1987). A common strategy for comparing multiple Cronbach's alpha coefficients is to first compare all coefficients in an omnibus test, followed by pairwise comparisons of the coefficients. If multiple tests are conducted, an error-rate control may be required. A significant difference between alpha coefficients of two tests signifies that the two tests differ in internal consistency reliability.

One limitation to `cocron` is that the program currently does not provide measures of effect size. The variance of alpha estimates depends on the population value of alpha; the difference between two alpha estimates is therefore not a good measure of effect size. Only one preliminary first step has been taken toward the difficult goal of establishing effect sizes for Cronbach's alpha. Liu and Weng (2009) have proposed an index that mirrors the formulation of Cohen's d , that is, the distance between two values divided by the common population standard deviation. However, their proposal is applicable only to the comparison of independent alphas, and only a single

table for a test with 30 components has yet been generated using computer simulations (Liu & Weng, 2009, p. 388). Moreover, the proposed index is based on the strong assumptions of equal test lengths, a minimum sample size of 100, normally distributed component scores, uncorrelated errors, and compound symmetry of the covariance matrix of the components. For the comparison of dependent Cronbach's alphas, there currently is no effect size index available (Liu & Weng, 2009).

A Cronbach's alpha of $\zeta = .70$ has sometimes been considered a minimum requirement for sound psychometric testing (e.g., Nunnally, 1978; Nunnally & Bernstein, 1994). However, the use of a fixed benchmark has been criticized (e.g., Cortina, 1993; Schmitt, 1996; Streiner, 2003). Several aspects have to be considered, when interpreting alpha coefficients. The magnitude of alpha coefficients may depend on the number of items in the test: A greater number of items on the test usually inflates alpha (Cortina, 1993). With the exception of very narrow traits, alpha values $\zeta > .90$ most likely indicate unnecessary redundancy rather than a high internal consistency (Streiner, 2003). Cronbach's alpha may also be inflated when a test measures multiple constructs (Cortina, 1993). Alpha coefficients should therefore not be used to test if a scale is unidimensional. Instead of relying on a fixed benchmark, the evaluation of Cronbach's alpha should rather depend on the requirements of the application context. For example, in basic research, a short scale with moderate reliability may be sufficient to establish two groups differing with regard to some construct. In individual assessments, however, the reliability of a test cannot be high enough to capture the individual characteristics of a person as precisely as possible. The cocron software package provides a convenient tool to establish what tests are best suited for this important purpose.

References

- Abd-El-Fattah, S. M. & Hassan, H. K. (2011). Dependent-Alpha Calculator: Testing the differences between dependent coefficients alpha. *Journal of Applied Quantitative Methods*, 6, 59–61. Retrieved December 5, 2016 from http://www.jaqm.ro/issues/volume-6,issue-2/pdfs/8_fattah_hassan.pdf
- Barke, A., Nyenhuis, N., & Kröner-Herwig, B. (2012). The German version of the Internet Addiction Test: A validation study. *Cyberpsychology, Behavior, and Social Networking*, 15, 534–542. doi:10.1089/cyber.2011.0616
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98–104. doi:10.1037/0021-9010.78.1.98
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. doi:10.1007/BF02310555
- D'Ath, P., Katona, P., Mullan, E., Evans, S., & Katona, C. (1994). Screening, detection and management of depression in elderly primary care attenders. I: The acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. *Family Practice*, 11, 260–266. doi:10.1093/fampra/11.3.260
- Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, 10, e0121945. doi:10.1371/journal.pone.0121945
- Falissard, B. (2012). psy: Various procedures used in psychometry. R package version 1.1. Retrieved April 28, 2016 from <http://cran.R-project.org/package=psy>
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93–103. doi:10.1177/014662168701100107
- Freelon, D. G. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5, 20–33. Retrieved April 28, 2016 from http://www.ijis.net/ijis5_1/ijis5_1_freelon.html
- Freelon, D. G. (2013). ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service. *International Journal of Internet Science*, 8, 10–16. Retrieved April 28, 2016 from http://www.ijis.net/ijis8_1/ijis8_1_freelon.html
- Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, 22, 313–328. Retrieved April 28, 2016 from <http://www.jos.nu/Articles/abstract.asp?article=222313>

Google (n.d.). Google Scholar Search. Retrieved April 28, 2016 from <https://scholar.google.com>

Hewson, C., & Charlton, J. P. (2005). Measuring health beliefs on the Internet: A comparison of paper and Internet administrations of the Multidimensional Health Locus of Control Scale. *Behavior Research Methods*, 37, 691–702. doi:10.3758/BF03192742

Ihme, J. M., Lemke, F., Lieder, K., Martin, F., Müller, J. C., & Schmidt, S. (2009). Comparison of ability tests administered online and in the laboratory. *Behavior Research Methods*, 41, 1183–1189. doi:10.3758/BRM.41.4.1183

Lautenschlager, G. J. (1989). ALPHATST: Testing for differences in values of coefficient alpha. *Applied Psychological Measurement*, 13, 284. doi:10.1177/014662168901300308

Lautenschlager, G. J., & Meade, A. W. (2008). AlphaTest: A windows program for tests of hypotheses about coefficient alpha. *Applied Psychological Measurement*, 32, 502–503. doi:10.1177/0146621607312307

Liu, H.-Y., & Weng, L.-J. (2009). An effect size index for comparing two independent alpha coefficients. *British Journal of Mathematical and Statistical Psychology*, 62, 85–400. doi:10.1348/000711008X315518

Lord, F. M. (1956). Sampling error due to choice of split in split-half reliability coefficients. *The Journal of Experimental Education*, 24, 245–249. Retrieved April 28, 2016 from <http://www.jstor.org/stable/20154008>

Nestler, S., Thielsch, M., Vasilev, E., & Back, M. D. (2015). Will they stay or will they go? Personality predictors of dropout in an online study. *International Journal of Internet Science*, 10, 37–48. Retrieved April 28, 2016 from http://www.ijis.net/ijis10_1/ijis10_1_nestler_et_al.html

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Putnam, S. P., Helbig, A. L., Gartstein, M. A., Rothbart, M. K., & Leerkes, E. (2014). Development and assessment of short and very short forms of the Infant Behavior Questionnaire–Revised. *Journal of Personality Assessment*, 96, 445–458. doi:10.1080/00223891.2013.841171

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved April 28, 2016 from <http://www.R-project.org/>

Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 89–117). San Diego, CA: Academic Press.

Revelle, W. (2015). psych: Procedures for psychological, psychometric, and personality research. R package version 1.5.8. Retrieved April 28, 2016 from <http://cran.R-project.org/package=psych>

Richins, M. L. (2004). The material values scale: Measurement properties and development of a short form. *Journal of Consumer Research*, 31, 209–219. doi:10.1086/383436

Riva, G., Teruzzi, T., & Anolli, L. (2003). The use of the Internet in psychological research: Comparison of online and offline questionnaires. *CyberPsychology & Behavior*, 6, 73–80. doi:10.1089/109493103321167983

Rödiger, S., Friedrichsmeier, T., Kapat, P., & Michalke, M. (2012). RKWard: A comprehensive graphical user interface and integrated development environment for statistical analysis with R. *Journal of Statistical Software*, 49, 1–34. doi:10.18637/jss.v049.i09

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350–353. doi:10.1037/1040-3590.8.4.350

Schreiner, M., Reiss, S., & Schweizer, K. (2014). Method effects on assessing equivalence of online and offline administration of a cognitive measure: The Exchange Test. *International Journal of Internet Science*, 9, 52–63. Retrieved April 28, 2016 from http://www.ijis.net/ijis9_1/ijis9_1_schreiner_et_al.html

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366. doi:10.1177/0956797611417632

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12, 102–111. doi:10.1037/1040-3590.12.1.102

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99–103. doi:10.1207/S15327752JPA8001_18

Templer, K. J., & Lange, S. R. (2008). Internet testing: Equivalence between proctored lab and unproctored field conditions. *Computers in Human Behavior*, 24, 1216–1228. doi:10.1016/j.chb.2007.04.006

Uitenbroek, D. G. (n.d.) SISA – Simple Interactive Statistical Analysis. Retrieved on December 5, 2016 from <http://www.quantitativeskills.com/sisa/>

Yan, T., Conrad, F. G., Tourangeau, R., & Couper, M. P. (2011). Should I stay or should I go: The effects of progress feedback, promised task duration, and length of questionnaire on completing web surveys. *International Journal of Public Opinion Research*, 23, 131–147. doi:10.1093/ijpor/edq046