Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide

Anne M. Gadermann, *Harvard Medical School*
Martin Guhn & Bruno D. Zumbo, *University of British Columbia*

This paper provides a conceptual, empirical, and practical guide for estimating ordinal reliability coefficients for ordinal item response data (also referred to as Likert, Likert-type, ordered categorical, or rating scale item responses). Conventionally, reliability coefficients, such as Cronbach's alpha, are calculated using a Pearson correlation matrix. Ordinal reliability coefficients, such as ordinal alpha, use the polychoric correlation matrix (Zumbo, Gadermann, & Zeisser, 2007). This paper presents (i) the theoretical-psychometric rationale for using an ordinal version of coefficient alpha for ordinal data; (ii) a summary of findings from a simulation study indicating that ordinal alpha more accurately estimates reliability than Cronbach's alpha when data come from items with few response options and/or show skewness; (iii) an empirical example from real data; and (iv) the procedure for calculating polychoric correlation matrices and ordinal alpha in the freely available software program R. We use ordinal alpha as a case study, but also provide the syntax for alternative reliability coefficients (such as beta or omega).

Reliability is an important source of evidence when establishing the validity of the inferences one makes based on scores from tests and measures (e.g., Kane, 2006; Zumbo, 2007). Throughout, we use the term 'test' to refer to any type of quantitative, multi-item measurement, such as tests, scales, and surveys in the social sciences. There are, of course, multiple definitions and types of reliability (e.g., internal consistency, retest, inter-rater), and multiple ways of obtaining reliability coefficients or indices (e.g., via different estimation methods and by using correlation or covariance matrices). Given the importance and the complexities of the concept(s) of reliability, the field has witnessed recurring debates on the interpretations and purposes of different types of reliability, on the advantages and disadvantages of different reliability indices, and on the methods for obtaining them (e.g., Bentler, 2009; Cortina, 1993; Revelle, & Zinbarg, 2009; Schmitt, 1996; Sijtsma, 2009).

A topic that has attracted particular attention in the psychometric literature is Cronbach's alpha (Cronbach, 1951), which remains the most widely and frequently used reliability index (Sijtsma, 2009). Some of this attention has been motivated by the fact that Cronbach's alpha has repeatedly been misinterpreted and misused (cf. Cortina, 1993; Schmitt, 1996; Sijtsma, 2009)—as noted by Cronbach himself (2004). In addition, some of the attention has centered on the question of whether there are some alternative reliability coefficients, such as omega, that may be more appropriate under certain circumstances (Revelle & Zinbarg, 2009; Zinbarg, Revelle, Yovel, & Li, 2005).

Some of the debates on reliability indices and on Cronbach's alpha have been fairly technical, including Cronbach's original paper (1951; see also, for example, Bentler, 2009; Green & Yang, 2009a, 2009b; Lord & Novick, 1968; Sijtsma, 2009). The implications of those technical debates are, however, not just of interest to a technical audience, but are critically relevant to practitioners and researchers in the social sciences in general. In fact, using Cronbach's alpha—or any other reliability coefficient—under circumstances that violate its assumptions and/or prerequisites might lead to substantively deflated reliability estimates (e.g., Gelin, Beasley, & Zumbo, 2003; Maydeu-Olivares, Coffman, & Hartmann, 2007; Osburn, 2000). A substantively deflated

estimation of a test's reliability, in turn, might potentially entail some misinformed inferences, such as discarding a test due to its seemingly low reliability.

Purpose of the paper

In this paper, we provide a tutorial for when and how to calculate ordinal reliability coefficients—rather than non-ordinal coefficients, such as Cronbach's alpha—for the very common scenario that one's data come from measurements based on ordinal response scales (e.g., rating scales or Likert-type response formats).

We focus on *ordinal alpha*, which was introduced by Zumbo, Gadermann, and Zeisser (2007), and which was shown to estimate reliability more accurately than Cronbach's alpha for binary[1] and ordinal response scales. We focus on alpha because it is the most widely used reliability coefficient, and because it is useful to use a familiar scenario as a concrete example. We note, however, that our discussion applies to other reliability coefficients as well. In other words, the rationale for using an ordinal version of a reliability coefficient is not restricted to alpha, but is equally valid for other reliability coefficients, such as McDonald's omega or Revelle's beta (please see Zinbarg et al. (2005) for a description of the omega and beta reliability coefficients).

The main purpose of this paper is to (i) provide researchers and practitioners with the psychometric and conceptual rationale for when, why, and how to use ordinal reliability coefficients, (ii) present an empirical example that illustrates the degree to which Cronbach's alpha and its ordinal equivalent can differ, and (iii) present step-by-step practical instructions and an example for how to compute ordinal alpha and other reliability coefficients, such as ordinal beta and ordinal omega, in the freely available software package R (http://www.R-project.org).

Coefficient alpha for ordinal data: psychometric rationale

Ordinal alpha is conceptually equivalent to Cronbach's alpha. The critical difference between the two is that ordinal alpha is based on the polychoric correlation matrix, described in detail below, rather than the Pearson covariance matrix, and thus more accurately estimates alpha for measurements involving ordinal data.

---

[1] A special case of coefficient alpha is KR-20, which is computed from binary data.

In general, the computation of coefficient alpha involves the matrix of correlations or covariances among all items of a scale. For Cronbach's alpha, the Pearson covariance matrix is routinely used; for example, as a default in statistical software programs, such as SPSS and SAS. An important assumption for the use of Pearson covariances is that data are continuous, and if this assumption is violated, the Pearson covariance matrix can be substantively distorted (e.g., Flora & Curran, 2004). In social science measurement, it is very common to use the Likert-type item response format. (For example, respondents are asked to indicate their level of agreement with an item by choosing one of a given number of ordered response categories, e.g., with five categories ranging from 'strongly agree' to 'strongly disagree'.) The data arising from such items are not continuous, but ordinal; however, they are often treated as if they were continuous; that is, they are treated "as if the data had been measured on an interval scale with desired distributional properties" (p. 443, Olsson, 1979).

It has been shown that the Pearson correlation coefficient severely underestimates the true relationship between two continuous variables when the two variables manifest themselves in a skewed distribution of observed responses. A tetrachoric/polychoric correlation, on the other hand, more accurately estimates the relationship of the underlying variables (Carroll, 1961). Accordingly, for ordinal data, the method of choice is to use the polychoric correlation matrix. Based on this argument, Zumbo et al. (2007) introduced a coefficient alpha for ordinal data—ordinal alpha—that is derived from the polychoric correlation matrix.

It needs to be noted that ordinal alpha, in line with the longstanding psychometric tradition of interpreting ordinal responses as manifestations of an underlying variable, is focusing on the reliability of the unobserved continuous variables underlying the observed item responses. Using a polychoric matrix for calculating alpha thus represents an underlying variable approach to covariance modeling of ordinal data. That is, when using a polychoric correlation matrix, an item's observed responses are considered manifestations of respondents exceeding a certain number of thresholds on that item's underlying continuum. Conceptually, the idea is to estimate the thresholds and model the observed cross-classification of response categories via the underlying continuous item response variables. Formally, the observed ordinal response for item *j* with *C* response categories, where the response option $c = 0, 1, 2, …, C\text{-}1$, is defined by the underlying variable y* such that

$$y_j = c \quad \text{if} \quad \tau_c < y_j^* < \tau_{c+1},$$

where τ_c, τ_{c+1} are the thresholds on the underlying continuum, which are typically spaced at non-equal intervals and satisfy the constraint $-\infty = \tau_0 < \tau_1 < \cdots < \tau_{c-1} < \tau_c = \infty$. The underlying distribution does not necessarily have to be normally distributed, although it is commonly assumed so due to its well understood nature and beneficial mathematical properties (cf. Liu, Wu, & Zumbo, 2010).

In summary, ordinal reliability coefficients may differ from their non-ordinal counterparts because of their scaling assumptions. The non-ordinal coefficients focus on the reliability of the observed scores by treating the observed item responses as if they were continuous, whereas the ordinal coefficients focus on the reliability of the unobserved continuous variables underlying the observed item responses. In this way, the ordinal coefficients are nonparametric reliability coefficients in a nonlinear classical test theory sense (Lewis, 2007).

Review of findings from a simulation study

Zumbo et al. (2007) present findings from a simulation study that compared ordinal alpha and Cronbach's alpha for all combinations of the following conditions:

  (i)   The theoretical reliability of a test was simulated so that it was .4, .6, .8, or .9. As Zumbo et al. (2007) note, the theoretical reliability was determined in reference to the underlying continuum of variation.
  (ii) The number of response options of the items was set to 2, 3, 4, 5, 6, or 7.
  (iii) The amount of skewness of the data was set to 0 or –2.

For all conditions, the number of items (*p*) was set to 14. We reanalyzed the data from Zumbo et al.'s (2007) simulation study, which invoked a paradigm introduced by Zumbo and Zimmerman (1993), specifying the underlying continuous scale as the reference for the theoretical reliability. Figure 1 illustrates the degree of underestimation of the theoretical reliability by

Cronbach's alpha and ordinal alpha, for those different conditions, respectively. In Figure 1, the degree to which ordinal alpha as well as Cronbach's alpha accurately estimate or underestimate the theoretical reliability of the underlying variable is illustrated in terms of an *attenuation index*, which is calculated by the following formula (Equation 1):

Percent attenuation   =   [100 * (alpha – theoretical reliability) / theoretical reliability)]

In equation 1, alpha denotes either ordinal alpha or Cronbach's alpha. When alpha is equal to the theoretical reliability, the term becomes zero, indicating no attenuation. The more alpha diverges from the theoretical reliability, the closer the term gets to (-100), which would indicate the highest possible degree of attenuation.

The graphs in Figure 1 indicate that ordinal alpha provides a suitable estimate of the theoretical reliability, regardless of (i) the magnitude of the theoretical reliability, (ii) the number of scale points, and (iii) the skewness of the scale point distributions. The accuracy of Cronbach's alpha, on the other hand, decreases (i) as the skewness of the scale items increases, (ii) as the number of response options becomes smaller, and (iii) as the theoretical reliability of the scale is lower.

The findings from the simulation study thus corroborate the general psychometric recommendation to use a polychoric correlation matrix for ordinal data, and they indicate that ordinal alpha is an unbiased estimator of the theoretical reliability for ordinal data (at least for scenarios like or similar to those tested in the simulation study). If one assumes that the observed item responses are manifestations of a continuous underlying item response variable, particular care should be taken in the interpretation of Cronbach's alpha, especially when one has very few item response options and/or highly skewed observed item responses. In those cases, Cronbach's alpha is a substantially attenuated estimate of the lower bound of the reliability of the underlying item response variables, whereas ordinal alpha tends to estimate this reliability more accurately—as the polychoric correlations correct for the attenuation caused by the scaling of the items (cf. Carroll, 1961).
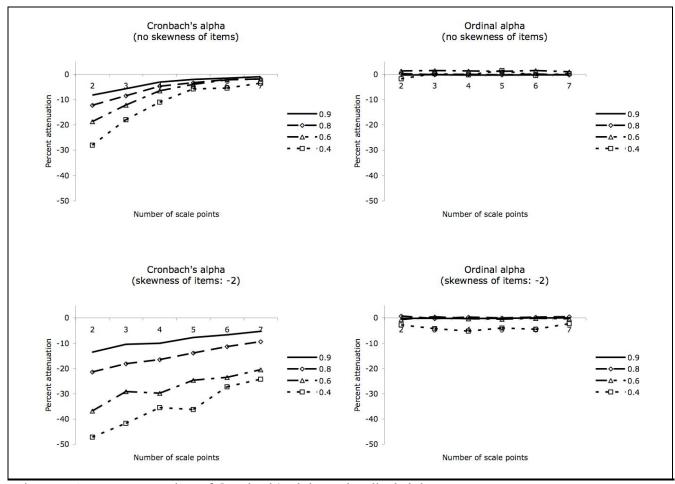
Figure 1: Percent attenuation of Cronbach's alpha and ordinal alpha

Illustration of an example

We now provide an example that illustrates the potential for a large discrepancy between ordinal alpha and Cronbach's alpha, using real data. The data are from a sample of 43,644 kindergarten children (48% girls; M_{age} = 5.7 years), and were collected with the *Early Development Instrument* (EDI; Janus & Offord, 2007). The EDI is a teacher-administered rating scale with 103 ordinal/binary items on children's developmental status in Kindergarten (see also Guhn, Janus, Hertzman, 2007; Guhn, Zumbo, Janus, & Hertzman, 2011). For our example, we used data from the *physical independence* subscale of the EDI. This subscale is composed of 3 binary items, which are scored as 0 (no) and 1 (yes). From a statistical point of view, the binary case can be thought of as a special case of ordinal data, and therefore the same methods apply here. Table 1 shows the means,

standard deviations, skewness, and kurtosis for the 3 items. Table 2 shows the Pearson correlations, Pearson covariances, and polychoric/tetrachoric correlations for the three items, the average correlations/covariance, and the respective alphas. Table 3 shows the factor loadings, communalities, and uniquenesses for the 3 items, and the alphas, calculated based on a factor analysis model, and using the matrix of the (i) Pearson correlations, (ii) Pearson covariances, or (ii) polychoric correlations[2],

---

[2] We note that the term *polychoric correlation* refers to all correlations based on ordinal variables that measure an (assumably) continuous underlying variable. In the special case that this involves two dichotomous/binary variables, the term *tetrachoric correlation* is used. Because the tetrachoric correlation is a special case of the polychoric correlation, calculating a polychoric correlation for binary variables is, in fact, equivalent to calculating a tetrachroic correlation; see Uebersax, 2006.

respectively. As can be seen in Table 1, the skewness and kurtosis of these three items are very high.

Table 1: Means, standard deviations, skewness and kurtosis of the three items of the physical independence subscale of the EDI

Item Would you say this child …	Mean *(SD)*	Skew *(SE)*	Kurtosis *(SE)*
… is independent in washroom habits most of the time (item 1)	*.98* *(.12)*	*-7.9* *(.03)*	*59.9* *(.07)*
… shows an established hand preference (item 2)	*.98* *(.16)*	*-6.1* *(.03)*	*34.9* *(.07)*
… is well coordinated (moves without running into things) (item3)	*.92* *(.28)*	*-3.0* *(.03)*	*7.2* *(.07)*

The results in Table 2 illustrate the degree to which the polychoric/tetrachoric correlations differ from the Pearson correlations in this case. The differences are all statistically significant (as tested via a Fisher z-Transformation; cf., Cohen, Cohen, West, & Aiken, 2003), and also substantial in their magnitude. If one applies the guidelines provided by Cohen (1988), the Pearson correlation coefficient for items 1 and 2 ($r_{Pearson}$=.23) is considered a small-medium effect, whereas the size of the polychoric correlation coefficient for the same two items ($r_{Polychoric}$=.63) is considered to indicate a large effect.

Table 2: Polychoric and Pearson correlations/covariances, and alphas for the physical independence subscale items

	Items 1,2	Items 1,3	Items 2,3	*Average*
Pearson correlation (*r*)	*.23*	*.25*	*.30*	*.26*
Covariances (*cov*)	*.004*	*.008*	*.013*	*.0085*
Polychoric/tetrachoric correlation[3] (*r*)	*.63*	*.65*	*.68*	*.65*
	Item 1	Item 2	Item 3	
Variances (*var*)	*.015*	*.024*	*.074*	*.038*
Standardized alpha	$\alpha = (k* r_{average})/(1+(k-1) * r_{average}) = (3*.26)/(1+(3-1)*.26) = .51$			
Cronbach's (raw) alpha	$\alpha = (k* cov_{average})/(var+(k-1) * cov_{average}) = (3*.0085)/(.038+(3-1)*.0085) = .46$			
Polychoric ordinal alpha	$\alpha = (k* r_{average})/(1+(k-1) * r_{average}) = (3*.65)/(1+(3-1)*.65) = .85$			

k: number of items in the scale

---

[3] Polychoric covariances are equal to polychoric correlations, as they are based on standardized variables.

Table 3 shows that, for this example, the factor loadings (λ) obtained from a factor analysis of the Pearson correlation or covariance matrices are substantially lower than those obtained from a polychoric correlation matrix. Correspondingly, these differences are reflected in the communalities (h^2) and uniquenesses (u^2). In some cases, differences in Pearson-based loadings versus polychoric correlation-based loadings might lead to different decisions about which items to include or not to include in one's factor model—for example, if item loadings, respectively, fall below or above a commonly used convention to only consider items with factor loadings greater than .40 (Ford, MacCallum, & Tait, 1986).

In addition, ordinal alpha and Pearson covariance-based (Cronbach's/raw) alpha are substantially different (.46 versus .85, respectively), with a percent attenuation of (-46). (The table provides the factor model-based formula for the 1-factor model for calculating alpha[4]; cf. McDonald, 1985). Typically, the psychometric literature (e.g., Nunnally, 1978) recommends that alpha for a scale should not be smaller than .70 when used for research purposes, at least .80 for applied settings, and greater than .90 or even .95 for high-stake, individual-based educational, diagnostic, or clinical purposes. In our example, interpreting the reliability of the physical independence scale by using ordinal versus Cronbach's alpha would make a difference with regard to these conventional recommendations.

General procedure for computing ordinal reliability coefficients in R

To calculate ordinal reliability coefficients, one needs to estimate a polychoric correlation matrix, and then calculate the reliability coefficient from the polychoric correlation matrix. In this paper, we show how to calculate these steps in the statistical software package *R* (R Development Core Team, 2011). There are alternative options for obtaining ordinal reliability coefficients, but the procedure in R has the following advantages: (1) Recent advancements and newly installed applications in R (Fox, 2005, 2006, 2011; Revelle, 2011) allow one to obtain polychoric correlations, ordinal reliability coefficients, and corresponding (ordinal) item statistics in a few simple steps; (2) R can be downloaded for free, and can be installed for Windows, Unix (Macs),

---

[4] By using this formula, readers who obtain the polychoric correlation matrix in MPlus or PRELIS/LISREL (via a 1-factor EFA with categorical data, see syntax in Elosua Oliden & Zumbo, 2008) may calculate ordinal alpha.

Table 3: Item characteristics, ordinal alpha, Cronbach's alpha, and discrepancy

	Pearson correlation matrix			Pearson covariance matrix			Polychoric matrix		
	λ	h^2	u^2	λ	h^2	u^2	λ	h^2	u^2
Item 1	.44	.19	.81	.46	.21	.79	.78	.60	.40
Item 2	.53	.29	.71	.49	.24	.76	.81	.66	.34
Item 3	.57	.32	.68	.48	.23	.77	.84	.71	.29
Average	**.51**	**.27**	**.73**	**.475**	**.23**	**.77**	**.81**	**.66**	**.34**

Formula for alpha based on a 1-factor model (cf. McDonald, 1985)

$$\alpha = [\ k\ /\ (k\text{-}1)\ ] * [\ (\ k * (\lambda_{average})^2 - h^2_{average}\ )\ /\ (\ k * (\lambda_{average})^2 + u^2_{average}\ )\ ]$$

Standardized $\alpha_{Pearson\ correlations}$ $[\ 3/(3\text{-}1)\ ] * [\ (\ 3*(.51)^2 - .27\ )\ /\ (\ 3*(.51)^2 + .73\ )\ ]\ = .51$

Cronbach's (raw) $\alpha_{Pearson\ covariances}$ $[\ 3/(3\text{-}1)\ ] * [\ (3*(.475)^2 - .23\ )\ /\ (3*(.475)^2 + .77\ )\ ]\ = .46$

Ordinal $\alpha_{polychoric\ correlations}$ $[\ 3/(3\text{-}1)\ ] * [\ (\ 3*(.81)^2 - .66\ )\ /\ (\ 3*(.81)^2 + .34\ )\ ]\ = .85$

Percent attenuation: $(100 * (\text{Cronbach's alpha} - \text{Ordinal alpha})\ /\ \text{ordinal alpha}) = (100 * (.46 - .85)/.85) = -\ 46$

λ: Factor loading; h^2: Communality (for a 1-factor model $h^2 = \lambda^2$); u^2: Uniqueness ($u^2 = 1 - h^2$); k: number of items on the scale

and Linux operating systems; (3) A graphic user interface developed for R—*R Commander* (Fox, 2005, 2006)—allows one to easily import data files in textfile, URL, clipboard, Minitab, SPSS, or Stata format into R; (4) The procedures to obtain polychoric correlation matrices in (proprietary) software programs—such as MPlus, SAS, Stata, and PRELIS/LISREL—involve more elaborate syntax (or additional macros) [5], and/or one would have to calculate ordinal reliability "by hand" once the polychoric correlation matrix is obtained; (5) Some widely used statistical software packages (e.g., SPSS) do not produce a polychoric correlation matrix.

We note that one limitation of using ordinal reliability coefficients may arise from the procedure with which the polychoric correlation matrix is calculated: Depending on the nature of the data as well as on the estimation method that is employed (i.e., whether correlations are calculated pair-wise for variables or simultaneously for the entire matrix), the correlation matrix may be *non-positive definite* which is not an issue when one calculates the Pearson correlation matrix from the same data, in the case of no missing data (Rigdon, 1997; Wothke, 1993). We note that R simultaneously estimates polychoric correlations from the entire data matrix, and we did not encounter this problem with the data used for this study. However, this limitation may arise in other software environments (i.e., with other estimation methods) or with other data. On that note, it should also be mentioned that the estimation of polychoric correlation matrices for scales containing a relatively large number of items may require substantial time/computer processing power.

In the remainder of the paper, we provide an example for how to calculate ordinal reliability coefficients with a data set that is included in the R software. We also provide instructions for preparing and importing data from other sources into R, so that researchers can easily calculate ordinal reliability coefficients in R for data files that already exist in the databases of their respective software programs.

R can be downloaded at the R website, http://www.R-project.org. Appendix A provides a description for downloading and installing R, and lists a number of useful online resources. Once R is downloaded, starting R will open the R menu and console, and one can install and load so-called R packages—installing specific packages will allow one to conduct specialized analyses, such as, in our case, calculating polychoric correlation matrices, ordinal reliability coefficients, and ordinal item statistics. The packages that need to be installed for our purpose are *psych* (Revelle, 2011), and *GPArotation* (Bernaards &

---

[5] For example, for SAS, see
http://www.ats.ucla.edu/stat/sas/faq/tetrac.htm, and for STATA, see
http://www.ats.ucla.edu/stat/stata/faq/tetrac.htm.

Jennrich, 2005). Once R and the required packages within R have been installed and loaded, we suggest closing R, and opening a new session, so that the following example can be replicated by readers in a manner that represents a typical new session in R.

In Appendix B, we present the syntax that needs to be *typed* (not copied and pasted[6]) into the R console (which opens when R is started) to run the example. Please note that (i) the syntax is entered after the *greater-than* sign (>) that always appears on the last line of the R console (i.e., the syntax file), (ii) R is case sensitive, and (iii) the return key needs to be pressed at the end of each syntax command to run the command. Please also note that the #-signs in our example are not part of the syntax, but simply indicate that we inserted an explanatory comment. In the syntax, **bolded font** indicates necessary steps, and regular font indicates steps that are optional, but will help to obtain commonly requested information in the context of calculating ordinal reliability coefficients.

Importing data into R via the graphic user interface R Commander

Once the R package *Rcmdr* has been installed and loaded, the syntax command *library (Rcmdr)*, entered into the R console, will open the graphic user interface, R Commander (see Fox, 2005 for a tutorial on R Commander). R Commander lets you import data in the following file formats: url, textfile (e.g., ASCII), clipboard, SPSS/PASW, Stata, and Minitab. Here, we briefly delineate the procedure for importing an SPSS file: In the menu of the graphic user interface R Commander, which is located on the top, click on 'Data', then on 'Import data', and then on 'from SPSS data set…'. In the window that opens ('Import SPSS Data Set'), highlight the word 'Dataset' (in the box 'Enter name for data set'), choose and type a name under which you wish to save the dataset to be imported (e.g., *mydata*), and click 'OK'. This opens your computer's directory. Choose the dataset to be imported (e.g., *spssdatafile.sav*), and click on 'Open'. Please note that unless one wishes to create a new data frame[7] in R, (i) the imported dataset

should only contain the items of the scale for which the reliability coefficient(s) is to be computed, (ii) the items should have ordinal data with *consecutive numbers* (e.g., 1, 2, 3; or 0, 1, 2, 3, 4)—so if, for example, an ordinal variable is coded as 10, 20, 30, 40, 50, it should be recoded to 1, 2, 3, 4, 5; otherwise, R will produce erroneous results, (iii) item labels should be a maximum of 8 characters long, (iv) the SPSS column 'Values' should be set to 'None' for all items (so that all values are displayed as numbers, not as text/response categories), (v) the SPSS column 'Missing' should be set to 'None' for all items, (vi) missing data should be empty cells (in R, cells with missing data will then appear as *NA*), not numerical values, such as 88 or 999, and (vii) the SPSS column 'Measure' should be set to 'Ordinal'. Once the data file is imported, clicking on 'View data set' in the menu will display the imported data file. As a last step, entering the syntax command *attach (mydata)* into the R console will attach the imported data set (in this case, *mydata*) to the current R session. Then, the syntax for calculating ordinal reliability coefficients, as described above, can be used for the imported dataset.

Conclusions

We recommend considering ordinal, polychoric correlation-based versions of reliability coefficients, such as alpha or omega, when one's data are binary and/or ordinal—that is, from Likert-type or mixed items, with 2 to 7 response options. In so doing, one invokes an underlying continuous variable for each item and the reliability coefficient is then defined by the covariation among these underlying variables. In this light, it is useful to think of the tetrachoric and polychoric strategy as akin to a data transformation, so that one is quantifying the reliability of the item response data in this transformed metric.

This recommendation is in line with general current thinking in the psychometric literature about using polychoric correlations for ordinal data (cf., Flora & Curran, 2004). Also, since the introduction of ordinal alpha by Zumbo et al. (2007), the use of a polychoric correlation-based version of alpha for ordinal or binary data has been applied elsewhere (e.g., Bentler, 2009, Green & Yang, 2009a). We reiterate that the strategy of using the polychoric correlation could be applied to any reliability estimate that can be computed from a correlation matrix. We provided the R syntax for alpha and alternative reliability coefficients, such as omega, but it needs to be noted that one could also compute an ordinal version of generalizability theory (e.g., G

---

[6] Please note that copying syntax into the R console sometimes leads to error messages.

[7] It is possible to import data sets, and to then create subsets—so-called data frames (using the syntax command data.frame—in R, in order to conduct analyses, such as calculating ordinal alpha, for a subset of the imported items/data (see, e.g., http://cran.r-project.org/doc/manuals/R-intro.html)).

coefficients) or test-retest reliability by using the polychoric correlation with the respective equations.

For future research, it will be of particular interest to better understand the interdependent, interacting effects that a scale's number of items, number of item response options, skewness, kurtosis, and factor structure have on ordinal alpha and Cronbach's alpha, and the discrepancy between them. Our data suggest a diminishing return model with regard to the number of items and number of response options, and also indicate that item skewness is associated with the attenuation of Cronbach's alpha. However, the exact nature of the multivariate relationship between these factors remains to be determined.

We would like to conclude with a note of caution and with an endorsement of a unitary, holistic approach to validation. In a recent review (Cizek, Rosenberg, & Koons, 2008), it was found that a majority of articles in the social sciences that report on the 'validity' of tests rely on none (7%), one (29%), or two (33%) sources of evidence for 'validity'. Cronbach's alpha is the most commonly reported piece of validity evidence for tests (reported in 76% of the cases). This practice is not in line with current recommendations provided by the large scientific and professional associations in the psychological and educational fields (e.g., American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Furthermore, such practice is not in line with current scholarly thinking in the areas of reliability analysis, generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; see also Brennan, 2001; Shavelson & Webb, 1991; Shavelson, Webb, & Rowley, 1989), and holistic perspectives on validity theory (e.g., Lissitz, 2009; Zumbo, 2007). Rather, a unitary, holistic perspective on validity emphasizes the importance (i) of uncovering and understanding multiple sources of measurement variance, and (ii) of validating the *interpretations, meanings, inferences, and social consequences* that are attributed to or based on measurement scores. In line with this thinking, we recommend using ordinal reliability coefficients for binary and Likert-type and mixed response data as one of several sources of information on a scale's reliability and validity.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Bentler, P. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74,* 137-143. doi: 10.1007/s11336-008-9100-1

Bernaards, C. A., & Jennrich, R. I. (2005) Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement, 65,* 676–696. doi: 10.1177/0013164404272507

Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlag.

Carroll, J. B. (1961). The nature of data, or how to choose a correlation coefficient. *Psychometrika, 26*, 347-372. doi: 10.1007/BF02289768

Cizek, G. J., Rosenberg, S., & Koons, H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68,* 397-412. doi: 10.1177/0013164407310130

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Maywah, NJ: Erlbaum.

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

Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64,* 391-418. doi: 10.1177/0013164404266386

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.

Elosua Oliden, P., & Zumbo, B. D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema, 20,* 896-901.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9,* 466-491. doi: 10.1037/1082-989X.9.4.466

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor-analysis in applied psychology—a critical review and analysis. *Personnel Psychology, 39*, 291-314. doi: 10.1111/j.1744-6570.1986.tb00583.x

Fox, J. (2005). The R Commander: A basic-statistics graphical user interface to R. *Journal of Statistical Software, 14,* 1-42. Retrieved from http://www.jstatsoft.org

Fox, J. (2006). *Getting started with the R Commander.* Retrieved from http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/Getting-Started-with-the-Rcmdr.pdf

Fox, J. (2011). *R Commander. Reference manual.* Retrieved from http://cran.r-project.org/web/packages/Rcmdr/Rcmdr.pdf

Gelin, M. N., Beasley, T. M., & Zumbo, B. D. (2003, April). *What is the impact on scale reliability and exploratory factor analysis of a Pearson correlation matrix when some respondents are not able to follow the rating scale?* Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Chicago, Il.

Green, S. B., & Yang, Y. (2009a). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika, 74,* 155-167. doi: 10.1007/s11336-008-9099-3

Green, S. B., & Yang, Y. (2009b). Commentary on coefficient alpha: A cautionary tale. *Psychometrika, 74,* 121-135. doi: 10.1007/s11336-008-9098-4

Guhn, M., Janus, M. & Hertzman, C. (Eds.) (2007). The Early Development Instrument [Special issue]. *Early Education and Development, 18*(3).

Guhn, M., Zumbo, B. D., Janus, M., & Hertzman, C. (Eds.) (2011). Validation theory and research for a population-level measure of children's development, wellbeing, and school readiness [Special issue]. *Social Indicators Research: An International Interdisciplinary Journal for Quality of Life Measurement, 103*(2).

Janus, M., & Offord, D. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioral Science*, 39, 1-22. doi: 10.1037/cjbs2007001

Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Washington, DC: American Council on Education and National Council on Measurement in Education.

Lewis, C. (2007). Classical test theory. In C. R. Rao and S. Sinharay (Eds.), *Handbook of Statistics, Vol. 26: Psychometrics,* (pp. 29-43). Amsterdam, The Netherlands: Elsevier Science B.V.

Lissitz, R. W. (Ed.) (2009). *The concept of validity: Revisions, new directions and applications.* Charlotte, NC: Information Age Publishing.

Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/rating scale item responses. *Educational & Psychological Measurement, 70*, 5-21. doi: 10.1177/0013164409344548

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Welsley Publishing Company.

Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution free (ADF) interval estimation of coefficient alpha. *Psychological Methods, 12,* 157-176. doi: 10.1037/1082-989X.12.2.157

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale NJ: Erlbaum.

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

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44,* 443-460. doi: 10.1007/BF02296207

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5,* 343-355. doi: 10.1037/1082-989X.5.3.343

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Revelle, W. (2009a). *Appendix A. R: Getting started.* Retrieved from http://personality-project.org/r/book/AppendixA.pdf

Revelle, W. (2009b). *An introduction to R.* Retrieved from http://personality-project.org/r/book/R_short_course.pdf

Revelle, W. (2011). *An overview of the psych package.* Retrieved from http://www.personalitytheory.org/r/book/overview.pdf

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74,* 145-154. doi: 10.1007/s11336-008-9102-z

Rigdon, E. (1997). *Not positive definite matrices—causes and cures.* Retrieved from http://www2.gsu.edu/~mkteer/npdmatri.html

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

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: A Primer.* Newbury Park, CA: Sage Publications.

Shavelson, R. J., & Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist, 44*, 922-932. doi: 10.1037/0003-066X.44.6.922

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74,* 107-120. doi: 10.1007/s11336-008-9101-0

Uebersax, J. S. (2006). *The tetrachoric and polychoric correlation coefficients. Statistical Methods for Rater Agreement web site.* Retrieved from http://john-uebersax.com/stat/tetra.htm

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing*

structural equation models (pp. 256-93). Newbury Park, CA: Sage.

Zinbarg, R.E., Revelle, W., Yovel, I., & Li. W. (2005). Cronbach's Alpha, Revelle's Beta, McDonald's Omega: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika.* 70, 123-133. doi: 10.1007/s11336-003-0974-7

Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao and S. Sinharay (Eds.) *Handbook of Statistics, Vol. 26: Psychometrics,* (pp. 45-79). Amsterdam, The Netherlands: Elsevier Science B.V.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6,* 21-29.

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology, 34*, 390-400. doi: 10.1037/h0078865

Appendix A

Downloading and installing R and loading the R packages pscyh and Rcmdr

R can be downloaded from the R website, at http://www.r-project.org/. The website provides background information on the R project, manuals, a FAQ page, the open access journal *The R Journal,* and links to multiple additional resources (e.g., R search; R conferences; related projects; see also Revelle, 2009a, and 2009b). The software can be downloaded by clicking on *CRAN mirror,* in the box *Getting started.* This opens the site *CRAN Mirrors,* on which users can choose an URL that is close to one's (geographic) location. Clicking on the link for your location will open a page containing a text box entitled *Download and Install R.* Here, users may choose the "precompiled binary distributions of the base system and contributed packages" for Linux, Mac OS X, and Windows. Mac users, after clicking on Mac OS X, can download R under *Files,* by clicking on *R-2.##.#.pkg (latest version).* Please note that you can choose between a 32- and a 64-version, and that this choice depends on the settings—under *Applications* → *Utilities* → *Terminal*—of your computer. Please refer to the frequently asked questions section, under the hyperlink *R for Mac OS X FAQ,* at http://cran.stat.sfu.ca/bin/macosx/RMacOSX-FAQ.html. Windows users can download R, by clicking on *Windows*, and then on *base*.

Once R is installed, starting R opens the R menu and console, and the R packages and their *dependencies* that one needs for specific calculations—in our case, *Rcmdr, psych, and GPArotation*, and their dependencies—can be installed by clicking on the menu option *Packages & Data,* choosing *Package Installer*, and then clicking on *Get List.* In the list, highlight the needed package, and install it by checking the boxes *At System Level* (or *At User Level*) and *Install dependencies*, and by clicking *Install Selected.* Once R packages are installed and loaded, they become part of the R environment. However, every time R is started for a new session, and one wishes to use one of the packages, one needs to type in the syntax *library*, and specify the name of the package in parentheses—e.g., *library (psych).* For each package, users may open and/or download a pdf-format user manual (www.personality-project.org/R/psych.manual.pdf for the *psych* package; and http://cran.r-project.org/web/packages/Rcmdr/Rcmdr.pdf for the *R Commander*,)), and the R help function allows one to open package-specific information in the R help window (by entering *??psych* or *??Rcmdr* into the R console, and then clicking on the respective package name in the list of help topics).

Please note that R is an open source software program environment that develops quickly. Our syntax was developed and tested for R 2.14.0 (32 bit version) on a computer with a Mac OS X 10.6.8 and on a computer with a Windows 7 operating system.

Appendix B

Syntax for calculating ordinal alpha and other ordinal reliability coefficients in R

library(psych)	# This activates the R package[8] '*psych*' (Revelle, 2011) for the current session in R.
data(bfi)	# This loads the dataset *bfi* contained in the R package *psych*.
attach(bfi)	# This attaches the dataset *bfi* to the current session in R.
bfi5items<-data.frame(N1,N2,N3,N4,N5)	# This creates a new dataset, labeled *bfi5items*, containing only five (ordinal) variables, N1 to N5, of the original 15-item dataset *bfi*.
describe(bfi5items)	# This describes the dataset *bfi5items*, providing descriptives, such as *n*, *mean*, *sd*, *min*, *max*, *range*, *skew*, and *kurtosis*.
bfi5items	# This displays the object/dataset called *bfi5items*.
polychoric(bfi5items)	# This provides the polychoric correlation matrix for the dataset *bfi5items*.
cor(bfi5items, y=NULL, use="complete.obs", method=c("pearson"))	# This calculates the Pearson correlation matrix for the dataset *bfi5items*, only taking into account cases with complete data ("complete.obs").
cov(bfi5items, y=NULL, use="complete.obs", method=c("pearson"))	# This calculates the Pearson method covariance matrix for the dataset *bfi5items*, only taking into account cases with complete data ("complete.obs").
skew(bfi5items)	# This provides the skewness for all items in the *bfi5items* dataset.
kurtosi(bfi5items)	# This provides the kurtosis[9] for all items in the *bfi5items* dataset.
scree(bfi5items)	# This provides the scree plots of the eigenvalues for a factor analysis and a principal component analysis for the dataset *bfi5items*.
examplename<-polychoric(bfi5items)	# This saves the polychoric correlation matrix, and corresponding *tau* values, under the name *examplename*. You may choose any name to save the matrix. (Note: R will not produce any output for this step.)
alpha(examplename$rho)	# This provides (raw and standardized) alpha, and corresponding item statistics, based on the data set or matrix that is specified in brackets. (The $rho command specifies that only the correlation matrix is used for the calculation, disregarding the tau values that are saved in conjunction with the matrix.) In the output of this calculation, alpha represents ordinal alpha, because it is based on the polychoric correlation matrix for the *bfi5items* dataset saved under the name *examplename*. One should obtain the following results as part of the R output: *raw_alpha* = .84; *std.alpha* = .84; *average_r* = .51. (Please note that raw alpha and standardized alpha are the same when they are calculated from a correlation matrix.)
alpha(bfi5items)	# This provides raw/Cronbach's and standardized alpha of the object specified in brackets. In this case, the object is a data

---

[8] As stated in the previous section, a package in R has to be installed once first, before it can be loaded for a current session in R. Please refer to Appendix A.

[9] Please note that, in R, the command for kurtosis is spelled without the final 's' (i.e.: kurtosi).

	matrix (*bfi5items*), and R therefore calculates raw/Cronbach's and standardized alpha, respectively, from the Pearson covariance and the Pearson correlation matrices of the data set. This step, in combination with the previous one, will allow one to compare ordinal alpha with raw/Cronbach's alpha. One should obtain the following results as part of the R output: *raw_alpha* = .81; *std.alpha* = .81; *average_r* = .47.
fa(bfi5items)	# This provides the factor loadings (*MR1*), communalities (h^2), and uniquenesses (u^2) for a 1-factor solution of the *bfi5items* data matrix.
fa(examplename$rho)	# This provides the factor loadings (MR1), communalities (h^2), and uniquenesses (u^2) for a 1-factor solution of the polychoric correlation matrix that was saved under the name *examplename*.
guttman(examplename$rho)	# This provides alternative estimates of reliability for the data matrix that is specified in brackets (i.e., examplename$rho). In the R output, these estimates are labeled as beta, Guttman bounds L1, L2, L3 (alpha), L4 (max), L5, L6 (smc), TenBerge bounds mu0, mu1, mu2, mu3, alpha of the first PC (=principal component), and the "estimated greatest lower bound based upon communalities". Since the specified data matrix is, in this case, a polychoric correlation matrix, all the reliability estimates represent ordinal versions. (We note that the guttman syntax command includes alpha (=L3) as one of the reliability estimates—however, the alpha syntax command provides additional item characteristics, such as the item-total correlations, that may be of interest to the user.)
	[Further details and references with regard to the different reliability coefficients featured in the guttman command can be found in Revelle, 2011.]
guttman(bfi5items)	# Equivalent to the command above, this provides a list of alternative estimates of reliability for the data matrix specified in brackets. Since *bfi5items* is a raw data matrix, the reliability estimates represent, in this case, Pearson correlation based reliability estimates.
omega(examplename$rho)	# This provides the ordinal versions of the reliability coefficients omega (hierarchical, asymptotic, and total), because their calculation is based on the polychoric correlation matrix 'examplename'.
omega(bfi5items)	# This provides omega coefficients for the data matrix *bfi5items*. (For details, see Revelle, 2011.)

Bolded font indicates necessary steps, and regular font indicates steps that are optional, but will help to obtain commonly requested information in the context of calculating ordinal reliability coefficients.

Citation:

Acknowledgement

Authors:

Anne M. Gadermann
Harvard Medical School
180 Longwood Avenue
Boston, MA 02115
AnneGadermann [at] googlemail.com

Martin Guhn
The Human Early Learning Partnership
University of British Columbia
Suite 440, 2206 East Mall
Vancouver, BC, V6T 1Z3, Canada
Martin.Guhn [at] ubc.ca

Bruno D. Zumbo, Corresponding Author
University of British Columbia
Scarfe Building, 2125 Main Mall
Vancouver, B.C. CANADA V6T 1Z
Bruno.Zumbo [at] ubc.ca