PERGAMON

Personality and Individual Differences 28 (2000) 229–237

PERSONALITY AND INDIVIDUAL DIFFERENCES

www.elsevier.com/locate/paid

Coefficient alpha: a useful indicator of reliability?

M. Shevlin[a,*], J.N.V. Miles[b], M.N.O. Davies[c], S. Walker[c]

[a]*School of Behavioural and Communication Sciences, University of Ulster at Magee, Northern Ireland, BT48 7JL, UK*
[b]*Institute of Behavioural Sciences, Derby University, Mickleover, Derby, DE3 5GX, UK*
[c]*Division of Psychology, Quantitative Research Group, Department of Social Sciences, Nottingham Trent University, Nottingham, NG1 4BU, UK*

Received 21 August 1998

Abstract

A series of Monte Carlo simulations were carried out to examine the performance of Cronbach's alpha as an index of reliability. Data were generated to be consistent with a single factor measured with six items. The magnitude of the factor loadings, systematic error and sample size were manipulated and alpha calculated from random samples. The results showed that alpha is influenced by factors other than the reliability of the items that comprise a scale. In particular the amount of systematic error, or deviation from unidimensionality, increased the estimate of alpha. The results are discussed in terms of traditional interpretation of alpha. © 1999 Elsevier Science Ltd. All rights reserved.

1. Introduction

Cronbach's alpha (Cronbach, 1951) is extensively reported in the psychological literature, in particular personality research, as an index of reliability (Bollen, 1989; Cortina, 1993). Its popularity may be attributable to its relevance to traditional psychometric practice that relies on the use of multiple indicators to measure latent constructs and the importance placed on reliability of measurements. In addition, alpha has more desirable properties than other indices of reliability such as split-half correlations.

However, Cortina (1993) found that there was little consensus in the literature regarding the definition of alpha. Further, there is no universal agreement on the appropriate interpretation of alpha or what constitutes an acceptable level of alpha (Boyle, 1991). It has been noted that

* Corresponding author.

PII: S0191-8869(99)00093-8

a common interpretation of alpha is that it represents a lower-bound estimate of reliability when the assumption of tau-equivalence is violated (Cortina, 1993; Raykov, 1997). The implication of this is that a reported alpha could be considered a conservative estimate of true reliability given the likelihood of violating the assumptions of tau-equivalence.

This paper aims to demonstrate the factors that influence the magnitude of estimates of alpha. The purpose is to demonstrate sources of variation that are reflected by alpha, and to aid in the interpretation of a reported value of alpha in a particular context. In particular, the study will examine the effects of the reliability of the items (operationalised as factor loadings) that constitute a scale, the amount of systematic error and sample size. Population values of alpha were computed and used as baseline values to determine the additional contribution made to alpha by systematic error and sample size. These estimates of reliability differ from the strict psychometric definition of reliability in that they do not include systematic sources of variation that would increase the estimate of alpha. The population alpha represents the variance/covariance of the items *attributable only to a single common factor*. Therefore any differences from the population values indicate the change in alpha due to deviations from unidimensionality and sample size. The number of items that comprise a scale was not included, as the positive relationship between the number of items and the value of alpha is well documented and given by the Spearman–Brown prophecy formula (see Nunnally & Bernstein (1994) for a description and discussion).

It is expected that the magnitude of factor loadings would have a strong positive association with estimates of alpha. Factor loading estimates are derived from sample correlations and reflect the strength of the association between the factor and the item. The squared standardised factor loading can be interpreted as the reliability of that item, therefore the higher the factor loadings for items that comprise a scale, the higher the expected value of alpha. The variance in an item that is not explained by the factor is measurement error, or

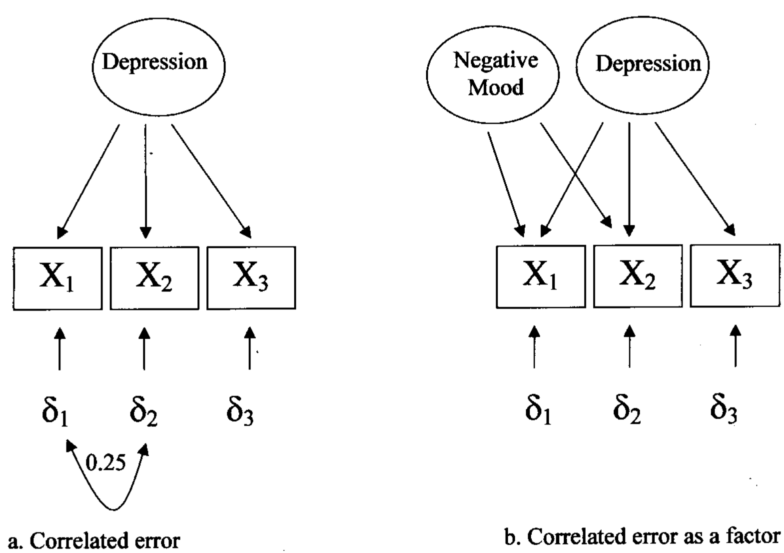

Fig. 1. Correlated error modelled as a factor.

unique variance. Unique variance is comprised of two components, random error and systematic error. Random error may be due to situational factors, the wording and response format of the item (Alwin & Krosnick, 1991) or administration errors (Saris & Andrews, 1991). However, random measurement errors associated with each item are assumed to be independent. Alternatively, systematic error represents that variance due to extraneous variables rather than variance due to the trait being measured and/or random measurement error. The presence of systematic error can be inferred from correlations between the unique variances of items and can be represented by additional factors. As an example, Fig. 1a shows three observed variables measuring the factor of depression. The observed variables could relate to items assessing (X_1) sadness, (X_2) loneliness and (X_3) insomnia.

The unique variance associated with the first two items (X_1 and X_2) is not random, therefore they are correlated (r = 0.25). Fig. 1b shows an alternative but equivalent representation of this correlated error. In this model the systematic error, or correlated error is represented as a factor with loadings of 0.5. This factor could be labelled 'negative mood' as it represents the association between the sadness and loneliness items with the effect of depression removed. It should be noted that a correlated error of 0.25 represents the presence of an additional common factor with non-trivial factor loadings.

The variability of the sampling distribution of correlations is negatively related to sample size. As the calculation of alpha is based on correlations, it is expected that the variability of the estimates of alpha will decrease as sample sizes increase.

2. Method

A series of Monte Carlo simulations were carried out using a general factor model consisting

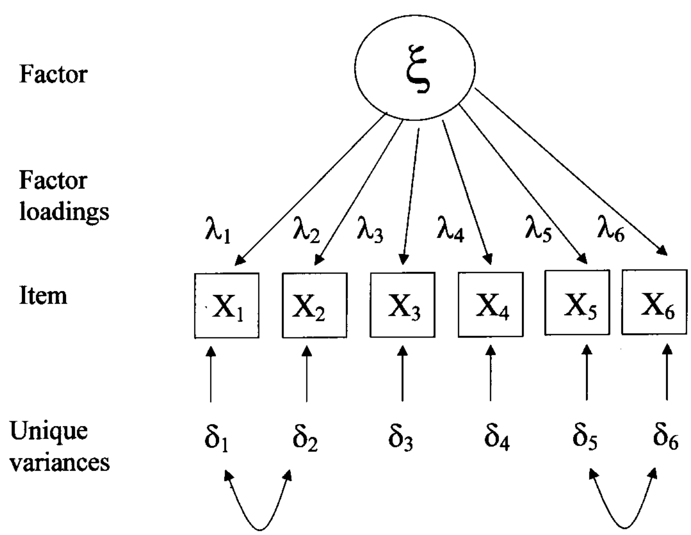

Fig. 2. Structure of the general factor model.

of six indicators, or items, specified to be measuring a single factor. The structure of this general model is shown in Fig. 2.

This general model was varied by specifying various population values for the factor loadings and correlated errors. The study employed a 3(loadings) × 4(systematic error) × 4(sample size) design. The three levels of the 'loadings' factor represented specified factor loadings (λ) of 0.3, 0.5, and 0.7. The three levels of systematic error related to correlated errors between two pairs of unique variances (δ_1, δ_2 and δ_5, δ_6). The values of the correlated errors were specified to be 0, 0.1, 0.2, and 0.3.

For each of these conditions, the calculation of the population covariance matrix was based on the specified values of the factor loadings and correlated errors. Fifty thousand cases were generated to be consistent with each population matrix (Jöreskog & Sörbom, 1993). The data were specified to be standardized random variables with a multivariate normal distribution. Using SPSS 7.5, 50 random samples were drawn from the generated data from each condition, and alpha calculated. The size of each sample was 50, 100, 200, and 400. These sample sizes represented the 'sample size' factor.

3. Results

The means and standard deviations of alpha for each condition are presented in Table 1.

Levene's test for homogeneity of variance was significant ($F(47,2352) = 23.34$, $P < 0.05$) indicating significant differences between the variances across the cells. The correlation between

Table 1
Mean and standard deviation of alpha estimates

Sample size	Correlated error	Factor loadings 0.3	0.5	0.7
50	0.0	0.405 (0.100)	0.652 (0.073)	0.789 (0.092)
	0.1	0.452 (0.096)	0.679 (0.078)	0.861 (0.029)
	0.2	0.445 (0.111)	0.694 (0.064)	0.871 (0.026)
	0.3	0.512 (0.093)	0.721 (0.054)	0.876 (0.024)
100	0.0	0.405 (0.080)	0.656 (0.045)	0.810 (0.030)
	0.1	0.455 (0.075)	0.682 (0.048)	0.863 (0.022)
	0.2	0.453 (0.067)	0.702 (0.047)	0.871 (0.022)
	0.3	0.527 (0.073)	0.728 (0.032)	0.881 (0.020)
200	0.0	0.386 (0.059)	0.665 (0.033)	0.809 (0.020)
	0.1	0.438 (0.056)	0.680 (0.035)	0.865 (0.015)
	0.2	0.467 (0.053)	0.699 (0.034)	0.871 (0.015)
	0.3	0.520 (0.052)	0.730 (0.026)	0.883 (0.014)
400	0.0	0.369 (0.044)	0.665 (0.029)	0.807 (0.017)
	0.1	0.405 (0.027)	0.672 (0.015)	0.860 (0.007)
	0.2	0.471 (0.017)	0.703 (0.007)	0.872 (0.002)
	0.3	0.528 (0.024)	0.735 (0.016)	0.886 (0.007)

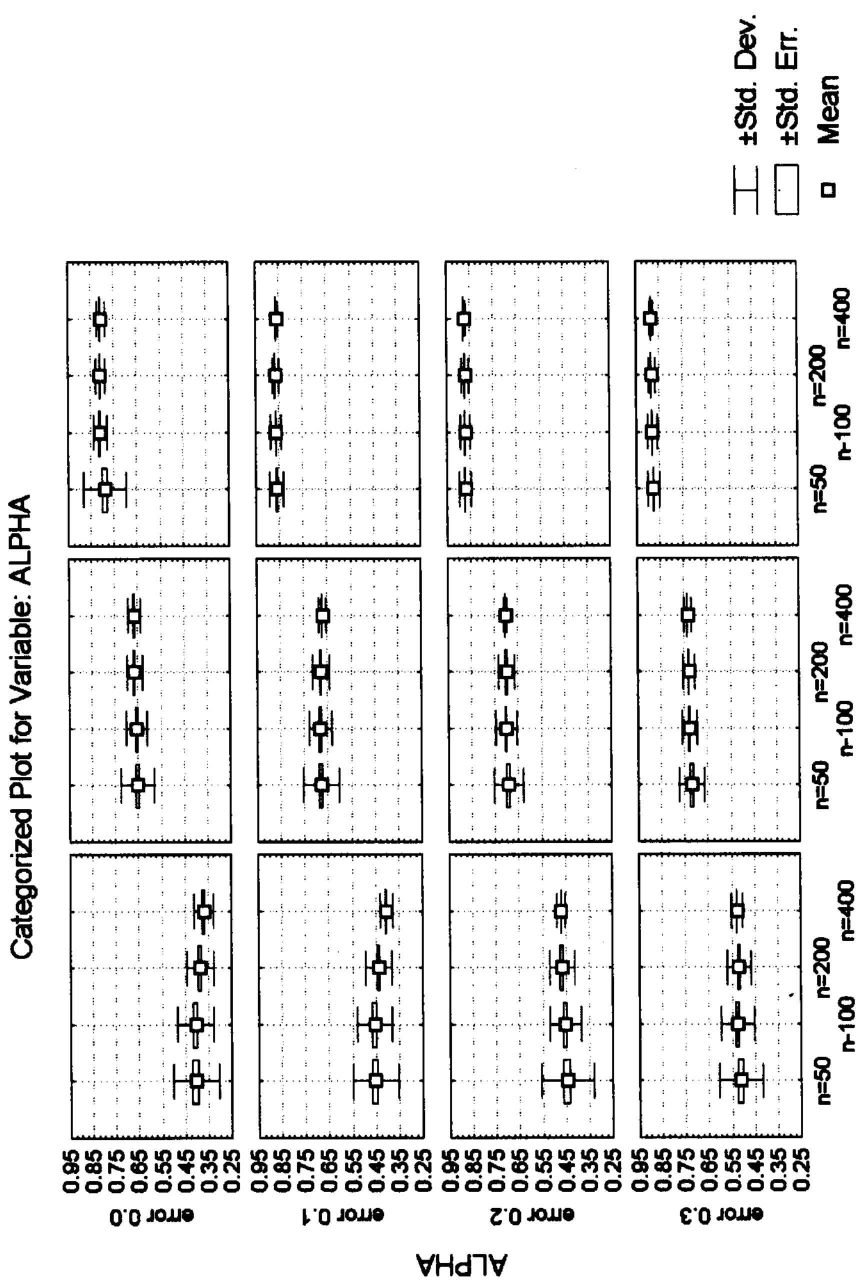

Fig. 3. Categorized plot for estimates of alpha.

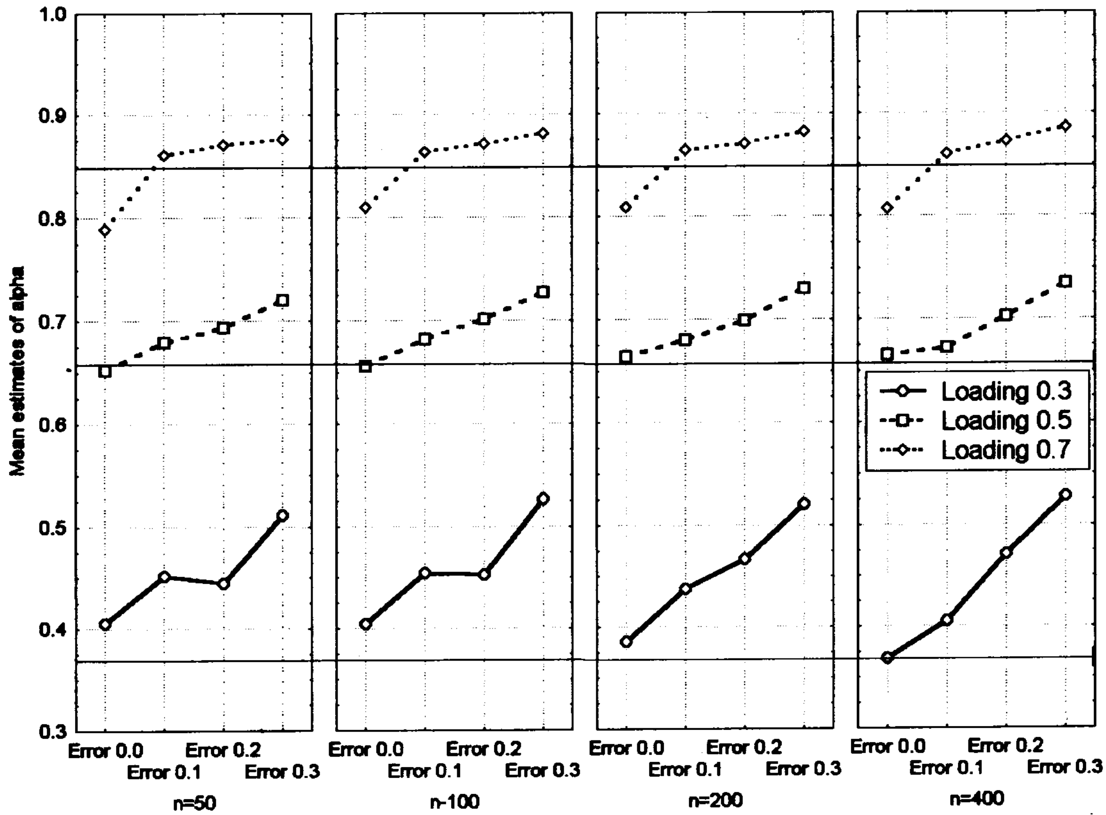

Fig. 4. Interaction plot showing mean estimates of alpha.

the means and standard deviations across all conditions was −0.635 ($P < 0.05$). A categorized box-plot showing the distribution of alpha within each condition is shown in Fig. 3.

The estimates of alpha were used in a 3(loadings) × 4(systematic error) × 4(sample size) ANOVA[1]. As expected there was a significant main effect for the loadings factor ($F(2,2353) = 12,745$, $P < 0.05$) and for the systematic error factor ($F(3,2353) = 344$, $P < 0.05$). The main effect for sample size was not significant ($F(2,2353) = 1.90$, $P > 0.05$). The sample size by loadings interaction was not significant, although the sample size by systematic error ($F(9,2353) = 2.57$, $p < 0.05$) and loadings by systematic error ($F(6,2353) = 23.29$, $P < 0.05$) interactions were statistically significant. The three-way interaction was also significant ($F(18,2353) = 1.851$, $P < 0.05$). The three-way interaction is plotted in Fig. 4. Reference lines are included to identify the population value of alpha for each level of the loadings factor

[1] Although Levene's test of homogeneity was significant, it was expected that the ANOVA results would be robust—given the large and equal number of cases in each cell.

(alpha was calculated using the population correlation matrix for each condition and the formula in Bollen, 1989).

4. Discussion

The results indicate that the performance of alpha is largely attributable to the reliabilities of the items that comprise a scale, but also to the amount of systematic error and to sample size. The effects of these factors can be identified from Fig. 3, and the relationship between the sampling distributions of alpha and sample size is shown in Table 1.

Fig. 3 clearly shows that across all levels of the systematic error and sample size factors, the effect of the factor loadings are strong. This indicates that alpha is strongly related to the reliability of the items that constitute a scale and therefore to the reliability of the scale as a whole. This effect is consistent with the findings of Cortina (1993) who found a strong positive association between average inter-item correlations and estimates of alpha. However, although the association between the reliability of the items and estimates of alpha is in the predicted direction, the mean values of alpha deviate markedly from the population value of alpha. The deviations from the population value indicate multiplicative effects between the three factors. The population values of alpha are the values of alpha calculated from the correlation matrix implied by the three models where the factor loadings were 0.3, 0.5 and 0.7 and the systematic error was zero. Therefore deviations from this value indicate the change in alpha that can be attributed to extraneous variables and to the influence of sample size. That is, the population alpha is the reliability of the common factor only.

Generally across all levels of reliability and sample size, there is a strong positive association between the level of systematic error and size of estimates of alpha. More specifically, alpha will tend to deviate most from the population value when the factor loadings are low. When systematic error is low (0 and 0.1), larger sample sizes will tend to lead to estimates of alpha that are closer to the population value. However, higher levels of systematic error result in increasing positive deviations from the population with larger sample sizes. As expected, the mean estimates of alpha best approximate the population value when there is no systematic error and when there are larger sample sizes. This shows that extraneous variables or systematic error will lead to higher levels of alpha even though the factor loadings associated with the common factor of interest remain constant. Put simply, as a scale deviates from unidimensionality, alpha increases.

In the conditions of medium factor loadings (0.3), the influence of systematic error is still evident although there is no visible significant sample size effect. The mean estimates of alpha are similar to the population value in the absence of systematic error but these estimates increase in relation to the population value in the presence of systematic error—across all levels of sample size. Again, deviations from unidimensionality result in an increase in alpha. A similar pattern is evident when the factor loadings are high (0.7) although it is interesting to note that the mean estimate of alpha underestimates the population value of alpha when there is no systematic error. This illustrates that alpha can underestimate the population value in the absence of systematic error (Novick & Lewis, 1967), although this attenuation is reduced with larger sample sizes.

In addition to information on the point estimate precision of alpha, Fig. 3 shows that the sampling distribution of alpha is related primarily to sample size and factor loadings, and systematic error has a relatively small effect. The categorized box-plots in Fig. 3 show that at any given level of factor loadings the variability of the estimates decrease as the sample size increases. In addition, when the factor loadings are 0.5 and 0.7 the variability of the estimates decreases as the amount of systematic error increases. In addition, the variability decreases as the factor loadings increase.

The reported results have implications for the interpretation of a reported alpha and suggest that additional information regarding the dimensionality of a scale is a necessary prerequisite for reporting sample estimates of alpha.

The difference between statistical and practical interpretations of alpha can lead to ambiguity. As Bollen (1989) states, systematic error is 'a consistent and *reliable* component of x' and, as has been demonstrated contributes to the value of alpha. This definition is statistically correct in that systematic error will be a stable element of the observed measures. However, the accepted meaning and interpretation of alpha in the psychological literature is not consistent with the strict psychometric definition. For example, authors reporting alpha for a self-esteem scale imply that the value for alpha represents the reliability of the 'self-esteem' factor rather than the reliability of self-esteem and all other unknown factors that are being measured. In the psychological literature measurements of a particular trait are referenced in terms of that particular trait, not in terms of the trait and all other stable components of the (e.g. 'Self-esteem was used as an IV' rather than 'Self-esteem, and all other stable components measured by the self-esteem scale, was used as an IV'). Further, high reliability is generally cited as evidence of good psychometric properties of a scale. However, a high estimate of alpha may be indicating the presence of systematic errors. These extraneous variables can make a substantial contribution to alpha rather than the actual trait that is being attempted to be measured. In particular it has been shown that when the factor loadings of the trait being measured are low, the presence of systematic errors can greatly inflate the estimate of alpha, especially with large sample sizes. Along with a smaller sampling distribution (associated with large sample sizes), suggesting a smaller standard error, the estimate of alpha can be misleading. An example: take a self-esteem scale with a reported alpha of 0.8. Three interpretations are possible:

1. The scale is measuring only self-esteem with a reliability of 0.8.
2. The scale is measuring self-esteem but the items that comprise the scale are not tau-equivalent. Therefore 0.8 is a conservative estimate of reliability.
3. The scale may be measuring self-esteem and a number of other variables. In this case a reliability of 0.8 represents the reliability of self-esteem and all these other variables; the reliability of the self-esteem factor will be less than 0.8.

Existing practices in scale construction can further confuse the interpretation of alpha. Using the Spearman–Brown prophecy formula, the additional number of items that need to be included in a scale in order to achieve a desired level of reliability can be easily calculated. However, it is not known whether these additional items increase the reliability of the scale by including items that are measuring the trait of interest, or items that are contributing to reliability through introducing systematic error.

Clearly in order to adequately interpret a value of alpha, additional information regarding the dimensionality of the scale is required. Exploratory factor analysis is not an appropriate method as it confounds random error with systematic error and does not allow a statistical test of dimensionality. However, within a structural equation modelling framework confirmatory factor models can be specified and tested. These models allow tests of dimensionality, identification of correlated errors and statistical tests of particular measurement models such as tau-equivalence. Only with such information can an unambiguous interpretation of alpha be made. In conclusion, it is suggested that interpretations of reported values of alpha are interpreted cautiously.

References

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement. *Sociological Methods and Research*, *20*, 139–181.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, *12*, 291–294.

Cortina, J. M. (1993). What is coefficient alpha: an examination of theory and applications. *Journal of Applied Psychology*, *78*, 98–104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL VIII user's reference guide*. Chicago: Scientific Software International.

Novick, M. R., & Lewis, C. L. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, *32*, 1–13.

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

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*, 173–184.

Saris, W. E., & Andrews, F. M. (1991). Evaluation of measurement instruments using a structural equation modeling approach. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. Mathiowetz, & S. Sudman, *Measurement errors in surveys* (pp. 575–597). New York: Wiley & Sons.