JOURNAL OF CONSUMER PSYCHOLOGY, *13*(4), 478–487

METHODOLOGICAL ARTICLE

Advancing Alpha: Measuring Reliability With Confidence

Dawn Iacobucci and Adam Duhachek
Department of Marketing
Northwestern University

In this research, we present the inferential statistics for Cronbach's coefficient alpha. This index of reliability is extremely important in consumer research. The estimation of alpha and its confidence intervals are described and an analytical demonstration illustrates the effects on these statistics of their components, including the number of items, the item intercorrelations, and sample size. SAS and SPSS programs are offered for easy implementation. We conclude with a prescription that every time a researcher reports a coefficient alpha, the confidence interval for alpha should accompany the alpha estimate.

Measurement development is an integral part of consumer research. Numerous articles have appeared in the consumer behavior literature dedicated to scale development and assessment. The extensive volumes by Bearden and Netemeyer (1999) and Bruner and Hensel (1994) together contain questionnaires intended to tap almost 500 marketing and consumer behavior constructs, which also speaks to the centrality of the pursuit of measurement in the consumer research enterprise.

There are also numerous texts and articles that guide the researcher in the construction of scales, from the early stages of theoretical demarcation and domain sampling, through to the statistical testing of the convergent and discriminant validity of the scale's factorial structure (e.g., Churchill, 1979; Gerbing & Anderson, 1988; Peter & Churchill, 1986). Establishing measurement reliability is of inarguable importance in both applied and theoretical research because reliability constitutes a necessary first step toward ensuring construct validity (e.g., Aiken, 2002; Allen & Yen, 1979; Anastasi & Urbina, 1996; Cronbach, 1951). Reliability is deemed so important that even when authors are not creating a scale but only using established scales, readers nevertheless expect a reliability index to be reported. By far the most frequently reported reliability index is Cronbach's coefficient alpha (Hogan, Benjamin & Brezinski, 2000; Peterson, 1994).

Thus reliability is important, and coefficient alpha in particular is important. In this article, we investigate the psychometric ramifications of recent statistical developments regarding alpha. There now exist methods to make inferential tests about the size of alpha. As of yet, the implications of these developments have not been reported in the literature.

Why does this matter? Inferential statistics are superior to simple descriptive statistics in the scientific pursuit of theory building and theory testing. In the *Journal of Consumer Psychology*, many experiments are reported. F statistics reported for main effects and interactions are tests—specifically testing hypotheses and inferences about the population. It would not be acceptable to a reader or reviewer for an author to make statements interpreting mean differences until those differences were established as statistically significant. We know enough about inferential statistics to know that a simple statement descriptive of a sample may or may not be true when projecting back to the population. The only way to truly test the hypothesis of no mean differences is to compute an inferential test statistic, such as the F test in analysis of variance to find out whether the seeming difference is likely to be replicable in subsequent research or is a haphazard finding in idiosyncratic data. Analogously, no one would simply report a beta weight or an R^2 in a regression model, nor a factor loading or path coefficient from Lisrel, and try to con-

Requests for reprints should be sent to Dawn Iacobucci, Department of Marketing, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. E-mail: d-iacobucci@kellogg.northwestern.edu

vince the reader of its size; rather, the proper inferential tests would be reported to support the magnitude argument. Once inferential statistics are available, it becomes no longer sufficient to subjectively judge a finding by a rule of thumb.

Consider an example. For a simple two-item scale with an item intercorrelation of .60 and a sample of 30, alpha is .75 and the standard error is .055. The 95% confidence interval is then $.64 < \hat{\alpha} < .86$. Researchers might couch their interpretations more conservatively if they realized that their findings were contaminated by up to 36% noise.

Or, consider the typical assessment by a researcher who achieves an alpha that exceeds .70 and concludes, based on Nunnally's rule-of-thumb, that the level of reliability is "sufficient." What Nunnally actually said is that "in the early stages of predictive or construct validation research," it may be "satisfactory" to "have only modest reliability, e.g., .70" (Nunnally & Bernstein, 1994, pp. 264–265). For other scenarios, Nunnally goes on to state that .80 or even .90 may be required. Further, if one computes an alpha equal to .70, with a confidence interval that ranges from .60 to .80, it is not so obvious that an "acceptable level of reliability" has been achieved. Thankfully, for the pursuit of social science, poorer reliability typically makes statistical tests more conservative, hence the strengths of the focal relations will be attenuated. However, the additional information included in a confidence interval allows readers and reviewers alike to more critically evaluate the reliability of the measures in question.

In this article, we explain the alpha test statistic and analytically illustrate its performance over a factorial design of component elements. The SAS and SPSS programming code is provided for the user to compute all the necessary ingredients to compose the statistical test and assess the size of the observed alpha. The sensitivity of the equations for the alpha and confidence intervals to the component factors of the number of items, the number of respondents, and the level of item intercorrelations is demonstrated. Let us begin by reviewing the equation for coefficient alpha.

CRONBACH'S COEFFICIENT ALPHA

Cronbach's coefficient alpha is widely known and defined as follows (Cronbach, 1951; Li, Rosenthal, & Rubin, 1996; Mendoza, Stafford, & Stauffer, 2000; Osburn, 2000; van Zyl, Neudecker, & Nel, 2000):

$$\alpha = \frac{p}{p-1}\left[1 - \frac{\sum_{i=1}^{p}\sigma_i^2}{\sigma_T^2}\right], \tag{1}$$

where p is the number of items in the scale (given the denominator of the first term, p must be 2 or greater); σ_i^2 is the variance of the i^{th} item, $i = 1,2, \ldots p$; and σ_T^2 is the variance of the entire test, hence it is the sum of the item variances and covariances: $\sigma_T^2 = \sum_{i=1}^{p}\sigma_i^2 + \sum_{i\neq}\sum_{j}\sigma_{ij}$. Equation 1 is the formula that is familiar to researchers employing reliability indexes.

Cortina (1993) is a nice treatment of the issues surrounding reliability. He presented an overview of assessments of internal consistency and described the effect of scale length and the relation among the scale items on reliability indexes like coefficient alpha. He laments that "there seems to be no real metric for judging the adequacy of the statistic" (p. 101), a shortcoming in the reliability and measurement literature that we alleviate with the confidence intervals we present in this article. With the benign assumption of a standard distribution and the computation of standard errors and confidence intervals, we offer exactly that metric that Cortina seeks.

Whereas Cortina (1993) looked at coefficient alpha via analytical comparisons, Peterson (1994) is a contribution particularly to the marketing and consumer behavior literatures in that he investigates the qualities of alpha as reported in published research articles. Collecting hundreds of articles and several thousand alphas allowed him to describe with clarity the 'typical' levels of alpha (e.g., its mean among these accounts of .77). Researchers may compare their obtained alpha to Nunnally's suggested .70, or this normative value, .77. Beyond these descriptive comparisons, our general, inferential statistic on alpha will allow for the assessment of alpha, for any literature, any scenario, any sample size, item intercorrelation, scale length, and so forth.

Kopalle and Lehmann (1997) investigated the well-known phenomenon of examining coefficient alpha, dropping poorly fitting items, and recomputing alpha. The items that get dropped in these measurement development studies will be presumably those items less correlated with the remainder. The result of their deletion will be the enhancement of alpha. Although this strategy is standard fare, Kopalle and Lehmann posed the question that perhaps the resulting alphas are biased over-estimates, at least when the subsequent alphas are recomputed on the same sample as that which initially determined the items to be deleted. (We have not explored the impact on our statistics of such two-phased scale development processes.)

In sum, measurement is extremely important and coefficient alpha has a long tradition of being the center of a great deal of research attention. The inferential test offered in this article addresses issues of concern to Cortina (1993) and Peterson (1994), and could easily be investigated for those conditions of Kopalle and Lehmann (1997). We return to our particular research question, the inferential statistics for alpha.

Recent research has affirmed that Equation 1 is the form of the maximum likelihood estimator of alpha based on an assumption of multivariate normality (van Zyl et al., 2000). With this assumption, they derive the distribution of alpha; as $n \to \infty$, then $\sqrt{n}(\hat{\alpha} - \alpha)$ has a Normal distribution with a mean of zero and a variance of:

$$Q = \left[\frac{2p^2}{(p-1)^2 (j'Vj)^3} \right] \left[(j'Vj)(trV^2 + tr^2V) - 2(trV)(j'V^2j) \right], \quad (2)$$

where n represents sample size, $\hat{\alpha}$ is the MLE of α and $\mathbf{V}$ is the population item covariance matrix.

Armed with a variance, we now derive a standard error, use that in conjunction with the distribution (Normal), calculate a z score and p value to assess the significance of alpha, and create confidence intervals to complement the knowledge about the size of alpha. The z score will test the null hypothesis, H_0: $\alpha = c$, where c is some constant, say, H_0: $\alpha = .50$, and it takes the form:

$$z = \frac{(\hat{\alpha} - \alpha)}{\sqrt{\frac{Q}{n}}}, \quad (3)$$

where $\hat{\alpha}$ is coefficient alpha computed on the sample at hand and α is the value in the null hypothesis. If z exceeds 1.96, the researcher would conclude that the sample alpha, $\hat{\alpha}$, is significantly greater than the hypothesized value of α at a 95% level of confidence. (It is not immediately obvious what hypothetical value for alpha should be tested, so we will return to this question later.)

Analogously, the 95% confidence interval is derived to be:[1]

$$\hat{\alpha} \pm (1.96)\left(\sqrt{\frac{Q}{n}}\right). \quad (4)$$

To cultivate an understanding of the implications of the z test and confidence intervals for consumer psychologists, we present an analytical illustration of these statistics for varied levels of n (sample size), p (number of items comprising the scale), and r_{ij} (the correlations among the items).

ANALYTICAL COMPARISONS BASED ON: *n, p,* AND r_{ij}: THE DESIGN

In this section, the confidence intervals for $\hat{\alpha}$ for each combination of n, p, and $\bar{r}_{ij}$ are examined. Sample size, n, ranges from a relatively small sample, $n = 30$ (and the point at which the Central Limit Theorem assists the behavior of test statistics) to $n = 200$ (reasoning that not many consumer research articles report larger samples). This upper bound should suffice: Alpha itself is insensitive to sample size and the results will indicate the negligible improvement in the accuracy of its estimation (i.e., the confidence intervals) with samples greater than 100 (or even 50 with sufficient numbers of items and strong item intercorrelations).

Similarly, we selected values for p, beginning with 2, adding the levels of 3, 5, 7, and 10. We did not explore p beyond 10 in part because literature reviews and meta-analytical studies (e.g., Peter & Churchill, 1986; Peterson, 1994) suggest that the use of longer scales is rare. The empirical performance of the confidence interval estimates indicates that little additional information is gained in terms of enhancing the consistency (i.e., precision) of the estimates with more items than five or certainly seven. One may also certainly argue that it becomes increasingly difficult to create unidimensional measures for constructs as p increases (Clark & Watson, 1995; Peterson, 1994).

Finally, the impact of $\bar{r}$, the average intercorrelation among the items, on the alpha and confidence interval estimates is considered. The full range of $\bar{r}$ is varied, from .0, .1, .2, …, .9, 1.0.

At this point, the design is fully presented. All factors that can impact coefficient alpha have been incorporated. As many researchers have pointed out, and as the reader can verify by examining Equation 1, alpha is a function of p (the number of items), and r_{ij}, the correlations among the items (Cortina, 1993; Cronbach, 1988; Gerbing & Anderson, 1988; Iacobucci, 2001; Kopalle & Lehmann, 1997). Although sample size does not enter into the computation of alpha, it *does* enter into the calculation of the confidence interval limits. Even intuitively, the reader can understand that the item intercorrelations themselves are more stably estimated as sample size increases.

RESULTS

Figure 1 contains the estimated alphas as a function of the factors, p, n, and $\bar{r}$. The lower and upper bounds of the confidence intervals are plotted in Figure 2. Here are the main observations.

In Figure 1, note that as p (the number of items) increases, coefficient alpha also increases. This effect could be anticipated by examining Equation 1 and has been propounded by researchers over the years (e.g., Allen & Yen, 1979; Cortina, 1993). However, analytical investigation allows us to examine precisely the strength of the impact of p as it increases from 2 to 10, as well as examine the effects on the confidence intervals.

When items are not correlated ($r = .0$), $\alpha = .0$, regardless of the number of items. This finding is sensible given that alpha is intended to reflect the internal consistency of the measurement of a set of p items; if those items are not cor-

[1]If a researcher computes the confidence interval and the lower bound is negative, (e.g., for small samples), the estimate should be truncated and the lower bound should be reported to be zero.

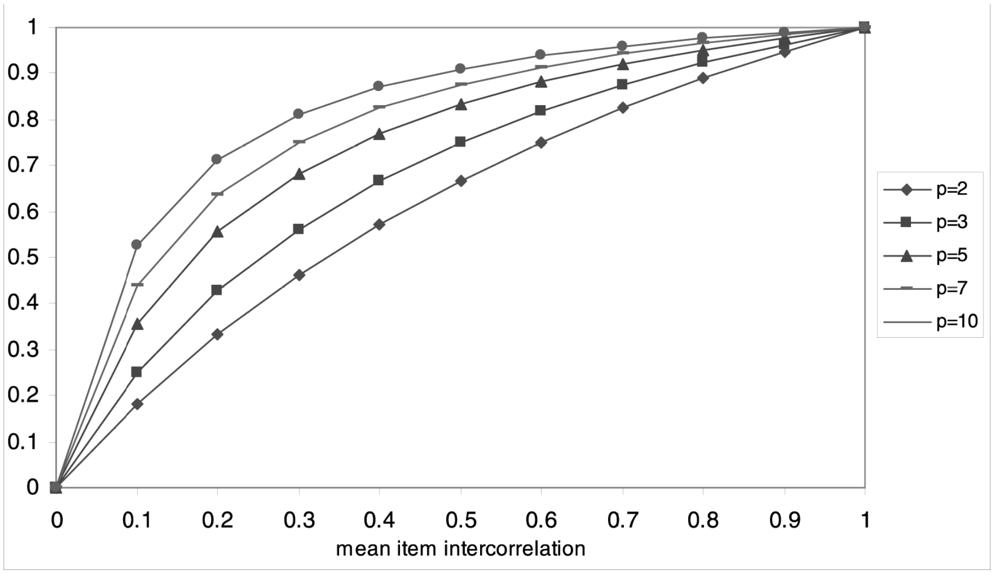

FIGURE 1 Coefficient Alpha.

related, it would be difficult to argue that they were tapping the same underlying construct. Conversely, when items are extremely highly correlated($\bar{r}$ = 1.0), α = 1.0 regardless of the number of items. This finding might be slightly surprising given that we all "know" that alpha is enhanced with the addition of items; however, in this special case (i.e., when items are this highly intercorrelated), alpha has already maxed out and adding items leaves little room for improvement in alpha.

Although instructive for highlighting the role of item intercorrelation on alpha, neither of the extremes, $\bar{r}$ = .0 or $\bar{r}$ = 1.0, are likely to occur in real data. The center of the *x* axis, where $\bar{r}$'s range between approximately .3 and .7, are likely to be more representative of what researchers encounter with their scales. Within this range, *p* and $\bar{r}$ function nearly linearly, in that longer scales and stronger item correlations both contribute to greater alphas. The item correlations are rarely under the control of the researcher, but *p* is. When $\bar{r}$ = .7, the improvement in alpha from $p = 2$ to $p = 10$ is slight, so a "cost-benefit analysis" might suggest that the additional items are not worth the likely boredom induced in the respondents, or the incremental cost accrued to the researcher. At the other extreme, when $\bar{r}$ = .3, the additional items improve alpha from just over .4 to around .8, an improvement of 100%. Researchers develop experience with the scales they use frequently (e.g., say, "need for cognition") and will be able to judge the likely level of $\bar{r}$ in their item correlations. If those $\bar{r}$'s tend to be high (.6 or .7), p = 3 or 5 will suffice to yield acceptable alphas. If those $\bar{r}$'s are expected to be low (.3 or .4), $p = 7$ may be minimally acceptable. (This summarization of course is based on these analytical findings. Should the researcher require additional items for domain and facet coverage, those theoretical issues would encourage the lengthening of the scale, apart from the fact that this empirical work indicates that statistically they are not necessary. However, in all likelihood, those additional facets would result in another factor, for which reliability should be assessed separately.)

Note also the compensatory effect between *p* and $\bar{r}$. Adding an item or two (e.g., going from $p = 2$ to $p = 3$ or from $p = 3$ to $p = 5$) has approximately the same effect on alpha as an increase in the item intercorrelations of .10.

Figure 2 contains the patterns of the confidence intervals. There are four striking insights. First, the confidence intervals are tighter (i.e., the estimation of alpha more precise) as the item correlations increase. Confidence intervals begin widely for $\bar{r}$ = .0 regardless of *p,* and they converge to very narrow intervals as $\bar{r}$ approaches 1.0. Second, confidence intervals are always wider for smaller sample sizes, as one might expect, though the differences between $n = 30$ and $n = 200$ are nominal for $\bar{r}$ = .6 or higher even when there are only two items, and when $\bar{r}$ = .4 or higher when *p* = 5 or more. Third, the impact of *p,* the number of items, is also clear; the enhancement to alpha of $\bar{r}$ when $p = 2$ is nearly linear, but when p = 10, 7, or even 5, alpha increases rapidly from $\bar{r}$ = .0 to $\bar{r}$ = .4. (This finding is consistent with Villani & Wind, 1975, who found negligible decreases in alphas for shortened scales.) Fourth, with a greater number of items, confidence intervals begin tighter, even for relatively small samples and relatively small item correlations.

Reflecting on these various trends, note that the effect of sample size is the case of gaining power as one has obtained more information. The curves asymptote out in that a sample of size 200 is not much more effective in obtaining useful, precise estimates than a smaller sample, even n = 30, if *p* and/or $\bar{r}$ are large. In many statistical applications, ∞ is not much beyond 30, and in our analysis also, we see that the statistic is generally well-behaved by the time *n* reaches 30, at least for $p \geq 5$ and $\bar{r} \geq .5$, or even when $p \geq 2$

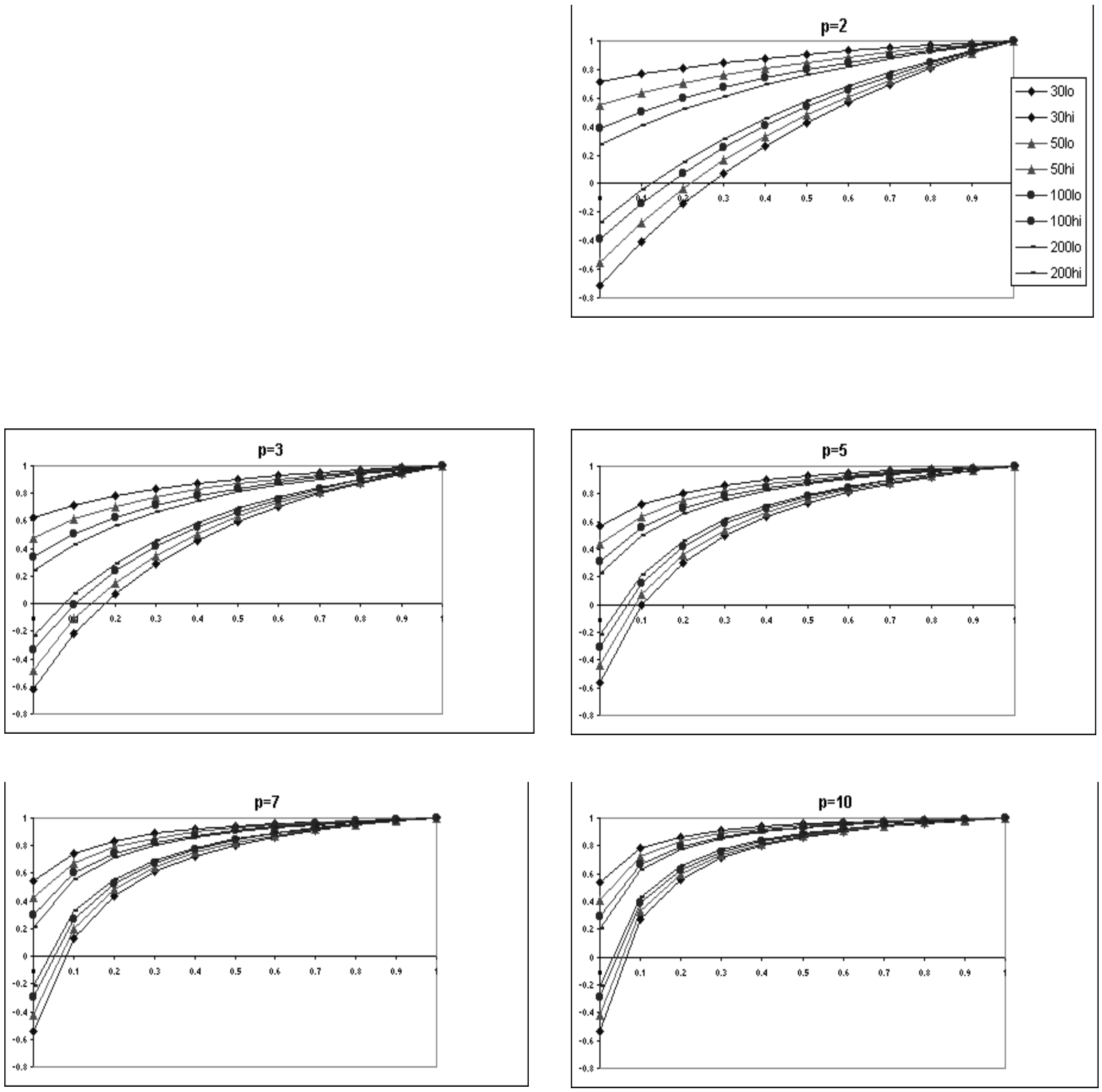

FIGURE 2 Confidence intervals.

if $\bar{r} \geq .7$. These results are encouraging and should provide assurance to consumer behavior researchers, given that many studies reported in the literature are based on small samples.

Note that the effect of $\bar{r}$ on confidence interval width is not a linear one, but rather the confidence intervals get narrower; that is, estimation is more precise, as the factor intercorrelations grow stronger in a manner that may be described as quadratic. To test this functional form directly, we fit a regression to our results. In this model, we predicted the values of coefficient alpha depicted in Figure 1. The predictors were the number of items (p), the mean item intercorrelation ($\bar{r}$), and the item intercorrelation term squared ($\bar{r}^2$). The results follow:

$$\text{alpha} = -.022 + .023\,(p) + 2.055\,(\bar{r}) - 1.207\,(\bar{r}^2)$$

The $R^2 = .9417$ and all the regression coefficients were significant ($p < .0001$). As p increases, alpha increases. As $\bar{r}$ increases, alpha increases. The final term is –1.207 for $\bar{r}^2$, reflecting precisely the effect depicted in the figures—a sharp rise and a tapering off. (Also note that counter to Peterson, 1994, who examined published reports and concluded there were no substantial relations between alpha and component terms such as those investigated here, these results clearly show that $\bar{r}$ has a substantial effect on alpha, though the effect of p is somewhat less. Perhaps published articles represent a restriction of range, and the current analytical investigation is factorially more complete.)

At this point, we have illustrated alpha and the confidence intervals across a large number of conditions for varying n, p, and $\bar{r}$. This analytical exercise has illustrated how each of these components affect measurement reliability. In the remainder of the article, SAS and SPSS programs available to the user to estimate these statistics are made, hypotheses surrounding alpha are discussed, and the logical bounds on these statistics are pushed by inverting the equations to obtain power estimates.

SAS and SPSS Codes

The SAS program to compute alpha, its standard error, and the confidence interval is presented in Appendix A. The SPSS code is in Appendix B. If a researcher simply relied on SAS's current Proc Corr Alpha option, the standard error and confidence intervals would not be available. SPSS produces analogously limited results. We are making the SAS and SPSS code publicly available, it is easy to implement, and it yields precise estimates for any combination of *p, n,* and *r*s.

To use the program, the user specifies the number of items in the scale and the sample size. The programs are versatile in allowing the inputs to be raw data, a correlation or covariance matrix produced (e.g., by SAS's Proc Corr in an output statement), or input by hand (e.g., computed via some other statistical computing package). The standardized and unstandardized alphas are produced (their distinction is described in Appendix C), along with the confidence intervals and a *z* test of a hypothesis about alpha, along with its corresponding *p* value to denote its significance.

The Null Hypothesis: H_0: $\alpha = ?$

In this section, we entertain hypotheses about alpha. We have the statistical tools to estimate the reliability index, its standard error, and the lower and upper bounds of confidence intervals. We may also use these values to compute a *z* statistic to test hypotheses about alpha. The question is: What hypothesis should we test? A test of H_0: $\alpha = .0$ would seem to be a straw man/person argument in that researchers never truly expect alpha to be so negligible, and the statistic would not be informative to the extent that every reported alpha would presumably be significant. On the other hand, a test of H_0: $\alpha = 1.0$ would require a standard that is probably unrealistically high. (Furthermore, this test leaves the researcher in the precarious position of not really wanting to reject the null.) We might compare our alpha against the familiar rule-of-thumb benchmark, H_0: $\alpha = .7$ (cf. Nunnally & Bernstein, 1994), or .77, or .79, the mean and median respectively from Peterson's (1994) meta-analytical study on alpha. However, note that doing so raises the standard—whereas an alpha that exceeded .7 would have been deemed acceptable, one's alpha now would need to be significantly greater than .7.

We recommend that the researcher report the alpha and its confidence interval, rather than a *z* test and its *p* value given that the choice of the null comparison is somewhat arbitrary. The confidence interval expresses to a reader the range of likely values for alpha, and like any confidence interval implicitly offers the tests of an infinite number of hypotheses using the logic that a value for alpha that falls outside the interval would be one for which a null hypothesis comparison would be rejected. Those values within the interval are all statistically plausible.[2]

POWER ESTIMATES FOR ALPHA CONFIDENCE INTERVALS: THE INVERTED EQUATION

The equations for the confidence intervals can be inverted to solve for *n.* Larger samples increase the precision of the estimation; that is, standard errors decrease as *n* increases, hence confidence intervals narrow. The right-hand term of Equation 4 is one half the width of the confidence interval, so for a given level of Q (from Equation 2, the variance estimate of alpha derived from a function of the sample covariance matrix among the *p* items), and a goal of say ±, .1, we derive *n* as:

$$n = \frac{1.96^2 Q}{w^2}, \tag{5}$$

where the new term, *w,* is one half the width of the confidence interval (i.e., the ± term), for example:

$$n = \frac{1.96^2 Q}{.1^2}. \tag{6}$$

If Q is small, fewer respondents will be necessary to obtain the level of precision we desire (± .1). If Q is large, a larger sample will be required to achieve the same level of precision. If one seeks a higher level of confidence (99% vs. 95%), a larger sample will be required. And if one seeks tighter precision (i.e., decrease *w* from .10 to .05) more respondents will be necessary. This power equation for *n* resembles analogous estimates for sample size (e.g., Churchill & Iacobucci, 2002, p. 504).

Consider several examples. For $Q = .20$, a researcher would need 307.328 (or 308) respondents to get as precise as ± .05. Smaller samples would be sufficient to achieve less precision, for example, for ± .10, $n = 76.832$ (i.e., 77); for ± .20, $n = 19.208$ (i.e., 20); for ± .30, $n = 8.537$ (i.e., 9), for ± .40; $n = 4.802$ (i.e., 5);

[2]We thank two of our reviewers for pointing out that alpha has been used as an (upper bound) estimate for communality in factor analysis (Nunnally & Bernstein 1994, p. 522). Squared multiple correlations are used more frequently because they offer the more conservative, lower bound estimates. It would be interesting to compare their performance with the lower end of the alpha confidence interval. Confirmatory factoring and structural equations models may start with either of these estimates, but iterate to the sum of squared factor loadings. In particular, alpha can be of assistance in releasing degrees of freedom when the number of parameters to be estimated is large. Nevertheless, subsequent even to structural equations modeling, researchers tend to report coefficient alpha estimates of reliability, rather than those based on Lisrel estimates.

and for ±.50, n = 3.073 (i.e., 4). For a larger Q, say .40, these confidence widths ±.05, .10, .20, .30, .40, .50 could be achieved only by larger sample sizes: 614.656 (615), 153.664 (154), 38.416 (39), 17.074 (18), 9.604 (10), and 6.147 (7), respectively.

CONCLUSION

The equations and analytical illustrations presented in this article should have great applicability to the consumer researcher. Early reports indicate that these statistics might prove to be fairly hardy; Yuan and Bentler (2002) showed in their exploration of skewness and kurtosis that these indexes are fairly robust to violations of the assumption of multivariate normality. In addition, previous attempts to derive inferential statistics for special reliability indexes (e.g., Charter, 2000, for Spearman–Brown; Feldt, 1965, for KR20;[3] Feldt, Woodruff, & Salih, 1987, with an analysis of variance approach; and Mendoza, Stafford, & Stauffer, 2000, using selected samples and validity coefficients) may now be subsumed into this more general, elegant approach (van Zyl et al., 2000).

The results also provide largely good news about the behavior of coefficient alpha. Unless one is working under extreme conditions (e.g., p = 2 and r = .0 or very small), the alphas function in a predictably robust manner, even for small samples (n = 30).

Given these results and the ready availability of our program, every alpha should be reported with its confidence interval to allow the reader to assess the size of the reliability index. Thus our prescriptions are these:

1. Compute alpha (via SAS, SPSS, or our program).
2. Compute the confidence interval for alpha (via our program).
3. Report both the alpha and the confidence interval wherever you would have previously reported alpha, for example, "our scale reliability was α = .75 (95% confidence interval: .68 to .82)."

Hopefully .70 or even .80 will be within (or below!) the confidence interval and researchers can proceed to use their reliable scales with greater confidence in their results.

ACKNOWLEDGMENTS

Dawn Iacobucci is a professor of marketing and Adam Duhachek is a doctoral student at Northwestern University.

We are grateful to Professor van Zyl for his thoughtful and encouraging correspondence and to Kent Grayson, Jennifer Escalas, Bob Wyer, and the three anonymous reviewers for their helpful feedback on this research.

REFERENCES

Aiken, L. R. (2002). *Psychological testing and assessment* (11th ed.). Boston: Allyn & Bacon.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA: Brooks/Cole.

Anastasi, A., & Urbina, S. (1996). *Psychological testing* (7th ed.). New York: Prentice Hall.

Bearden, W. O., & Netemeyer, R. G. (1999). *Handbook of marketing scales: Multi-item measures for marketing and consumer behavior research.* Thousand Oaks, CA: Sage.

Bruner, G. C., II, & Hensel, P. J. (1994). *Marketing scales handbook: A compilation of multi-item measures*. Chicago: American Marketing Association.

Charter, R. A. (2000). Confidence interval formulas for split-half reliability coefficients. *Psychological Reports, 86,* 1168–1170.

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16,* 64–73.

Churchill, G. A., Jr., & Iacobucci, D. (2002). *Marketing research: Methodological foundations* (8th ed.). Fort Worth, TX: Harcourt.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309–319.

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.

Cronbach, L. J. (1988). Internal consistency of tests: Analyses old and new. *Psychometrika, 53,* 63–70.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder–Richardson reliability coefficient twenty. *Psychometrika, 30,* 357–370.

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11,* 93–103.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 25,* 186–192.

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods. *Educational and Psychological Measurement, 60,* 523–531.

Iacobucci, D. (Ed.). (2001). Methodological and statistical concerns of the experimental behavioral researcher [Special issue]. *Journal of Consumer Psychology, 10,* 55–62.

Kopalle, P. K., & Lehmann, D. R. (1997). Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. *Organizational Behavior and Human Decision Processes, 70,* 189–197.

Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman–Brown to maximal reliability. *Psychological Methods, 1,* 98–107.

Mendoza, J. L., Stafford, K. L., & Stauffer, J. M. (2000). Large-sample confidence intervals for validity and reliability coefficients. *Psychological Methods, 5,* 356–369.

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

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5,* 343–355.

Peter, J. P., & Churchill, G. A. (1986). Relationships among research design choices and psychometric properties of rating scales. *Journal of Marketing Research, 23,* 1–10.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research, 21,* 381–391.

van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika, 65,* 271–280.

[3]KR20 is the Kuder–Richardson formula 20, applied to binary data (e.g., responses scored "right" or "wrong").

Villani, K. E. A., & Wind, Y. (1975). On the usage of "modified" personality trait measures in consumer research. *Journal of Consumer Research, 2,* 223–228.
Yuan, K. H., & Bentler, P. M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67,* 251–259.

Received, November 12, 2002
Accepted, February 8, 2003

APPENDIX A
The SAS Code to Compute Alpha, Standard Error, *Z,* Confidence Intervals[1]

Option 1: User Provides a Correlation or Covariance Matrix.

As the first four comments indicate, the user must provide four things: *p, n,* their matrix of item covariances or correlations, and a value to use as the hypothesized benchmark in the *z* test.

```
proc iml;
*TO USER: you need to fill in
the hypothesized value you want
your alpha tested against;          hypalpha = { 0.7 };
*TO USER: you need to fill in
the number of items you have in
your scale;                         numbitem = { 3 };
*TO USER: you need to fill in
your sample size;                   numbsubj = { 100 };
*TO USER: you need to cut and
paste your item covariance matrix
into the following form (i.e., begin
and end with braces, add commas
to delineate matrix rows);          itemcov = { 1 .5 .5 ,
                                    .5 1 .5 , .5 .5 1 };

*next are analyses;
one=j(numbitem,1); jtphij = (one`)*itemcov*one;
myalpha = 1 – ((trace(itemcov))/jtphij); myalpha =
   (numbitem / (numbitem -1)) *myalpha;
trphisq = trace(itemcov*itemcov); trsqphi =
   (trace(itemcov))**2;
jtphisqj = (one`)*(itemcov*itemcov)*one; omega =
   jtphij*(trphisq+trsqphi);
omega = omega-(2*(trace(itemcov))*jtphisqj); omega =
   (2/(jtphij**3))*omega;
s2 = (numbitem**2) / ((numbitem-1)**2); s2 =
   s2*omega;
se = sqrt(s2/numbsubj); z = (myalpha-hypalpha)/se; pv =
   1-probnorm(z);
cimin95 = myalpha – (1.96*se); cimax95 = myalpha +
   (1.96*se);
print 'Your Covariance Matrix was:'; print itemcov;
print 'Your number of items and sample size were:'
   numbitem numbsubj;
print 'Your unstandardized coefficient alpha is:'
   myalpha;
print 'The z score for alpha and its p-value are:' z pv;
print 'The lower and upper 95% confidence limits fol-
   low:' cimin95 cimax95;
if cimin95 < .00 then print 'You should report your
   confidence interval as: .0';
if cimin95 < .00 then print cimin95 cimax95;
*scale cov matrix to corr matrix; s=diag(itemcov);
   s=sqrt(s); s=s**(-1); itemcov=s`*itemcov*s;
jtphij = (one`)*itemcov*one;
myalpha = 1 – ((trace(itemcov))/jtphij); myalpha =
   (numbitem / (numbitem -1)) *myalpha;
trphisq = trace(itemcov*itemcov); trsqphi =
   (trace(itemcov))**2;
jtphisqj = (one`)*(itemcov*itemcov)*one; omega =
   jtphij*(trphisq+trsqphi);
omega = omega-(2*(trace(itemcov))*jtphisqj); omega =
   (2/(jtphij**3))*omega;
s2 = (numbitem**2) / ((numbitem–1)**2); s2 =
   s2*omega;
se = sqrt(s2/numbsubj); z = (myalpha-hypalpha)/se; pv =
   1-probnorm(z);
cimin95 = myalpha – (1.96*se); cima×95 = myalpha +
   (1.96*se);
print 'Your Correlation Matrix was:'; print itemcov;
print 'Standardized coefficient alpha (the one to report)
   equals:' myalpha;
print 'Your z score and its p-value are:' z pv;
print 'The lower and upper 95% confidence limits are:'
   cimin95 cimax95;
if cimin95 < .00 then print 'You should report your
   confidence interval as: 0.00 to' cima×95;
quit; run;
```

Option 2: User Reads in Raw Data.

1. Prior to the "proc iml;" statement in the program above, insert:

```
data myabc; input ×1 ×2 ×3; cards;
1 1 0
2 2 3
2 1 3
…
4 2 1
run;
```

2. After the "proc iml;" statement above, insert:

[1]An interested user need not type in this syntax. Our program will be available on our websites or via attachment upon e-mail request. The program uses SAS's IML module. IML stands for "interactive matrix language," but this program is not run interactively; just submit it the way you would any SAS job.

```
use myabc var {x1 x2 x3};
read all var {x1 x2 x3} into x;
```

3. Then delete the "itemcov = { … };" statement.
4. After the comment "*next are analyses;" insert:

```
bigone=j(numbsubj,1); means=((bigone`)*x)/numbsubj;
xd=x-(bigone*means); itemcov = (1/(numbsub;-1)) *
  ((xd`)*xd);
```

Option 3: User Reads Raw Data Into Proc Corr, Produces a Correlation or Covariance Matrix to be Used in Proc IML

1. Prior to the "proc iml;" statement in the program above, insert:

```
data myabc; input x1 x2 x3; cards;
1 1 0
2 2 3
2 1 3
…
4 2 1
proc corr cov outp=mycorrs; var x1 x2 x3; run;
```

2. After the "proc iml;" statement above, insert:

```
use mycorrs var {x1 x2 x3};
read point {1 2 3} var {x1 x2 x3} into itemcov;
```

APPENDIX B

The SPSS Code to Compute Alpha, Standard Error, Confidence Intervals[1B]

* USER: fill in #items in scale, sample size, correlation matrix.

```
matrix.
compute numbitem = 3.
compute numbsubj = 100.
compute itemcov = { 1.0, .5, .5; .5, 1.0, .5; .5, .5, 1.0}.
compute one=make(numbitem,1,1).
compute jtphij=transpos(one).
compute jtphij=jtphij*itemcov.
compute jtphij=jtphij*one.
compute trmy=trace(itemcov).
compute trmy=trmy/jtphij.
compute myalpha=1-trmy.
compute nn1=numbitem-1.
compute nn1=numbitem/nn1.
compute myalpha=nn1*myalpha.
compute trphisq=itemcov*itemcov.
compute trphisq=trace(trphisq).
compute trsqphi=trace(itemcov).
compute trsqphi=trsqphi**2.
compute ttp=itemcov*itemcov.
compute jtphisqj=transpos(one).
compute jtphisqj=jtphisqj*ttp.
compute jtphisqj=jtphisqj*one.
compute omega=trphisq+trsqphi.
compute omega=jtphij*omega.
compute omegab=trace(itemcov).
compute omegab=omegab*jtphisqj.
compute omega=omega-(2*omegab).
compute omega=(2/(jtphij**3))*omega.
compute s2=(numbitem**2) / ((numbitem-1)**2).
compute s2=s2*omega.
compute se=sqrt(s2/numbsubj).
compute cimin95=myalpha-(1.96*se).
compute cimax95=myalpha+(1.96*se).
print myalpha /format ="f8.3"/title= 'Your coefficient alpha is (standardized if you had a corr matrix, the
  one to report; but unstandardized if you had entered
  a covariance matrix, not the one to report):'.
print cimin95 /format ="f8.3"/title= 'The lower 95%
  confidence limit follows:'.
print cimax95 /format ="f8.3"/title= 'The upper 95%
  confidence limit follows:'.
end matrix.
```

[1B]An interested user need not type in this syntax. Our program will be available on our websites or via attachment upon e-mail request. The SPSS code must be entered in the syntax window.

APPENDIX C
Standardized Versus Unstandardized Coefficient Alpha

That SAS produces both a "standardized" alpha and an "unstandardized" alpha has perplexed more than one researcher. The difference is one of alpha being computed on the standardized versus raw data (or, equivalently, on the correlation vs. covariance matrix, cf. http://v8doc.sas.com/sashtml/). The resulting alphas are rarely equal, though they are often close in size. Researchers should report the standardized alpha for the philosophical reason that the scales of our measurements tend to be arbitrary. It is also usually the larger index.

Consider the small example that follows. The covariance matrix is reported first, followed by the coefficient alpha derived from this unstandardized matrix, $\alpha = .727$. The correlation matrix is presented next, followed by the resulting newly scaled, standardized alpha, $\alpha = .750$.

Covariance Matrix:

$$\begin{bmatrix} 25 & 10 & 7.5 \\ 10 & 16 & 6 \\ 7.5 & 6 & 9 \end{bmatrix}$$

Equation 1 worked out for the covariance matrix, yielding the "unstandardized" alpha:

$$\alpha = \frac{3}{3-1}\left[1-\frac{(25+16+9)}{(25+16+9)+2(10+7.5+6)}\right]$$
$$= \frac{3}{2}\left[1-\frac{50}{97}\right] = 1.50 \times .485 = .727.$$

the corresponding correlation matrix:

$$\begin{bmatrix} 1 & .5 & .5 \\ .5 & 1 & .5 \\ .5 & .5 & 1 \end{bmatrix}$$

and its respective "standardized" alpha estimate:

$$\alpha = \frac{3}{3-1}\left[1-\frac{(1+1+1)}{(1+1+1)\_\,2(.5+.5+.5)}\right]$$
$$= \frac{3}{2}\left[1-\frac{3}{6}\right] = 1.50 \times .50 = .750.$$