

*Paper presented at the meeting of the Midwestern Association
for Public Opinion Research, Chicago, IL, November 2005*

A Computational Tool for Survey Shortening Applicable to Composite
Attitude, Opinion, and Personality Measurement Scales

Andrew F. Hayes

The Ohio State University

School of Communication

Andrew F. Hayes
School of Communication
The Ohio State University
3016 Derby Hall
Columbus, OH 43210
hayes.338@osu.edu
614-688-3027

Abstract

Most of the existing and validated composite measures of individual differences (including personality, attitudes, opinions, etc.) are too lengthy to be included in public opinion polls and representative surveys of local or national populations that are more costly to conduct and require the participation of people who are not captive audiences of the researcher. I argue that public opinion research could benefit by developing “short forms” of existing measures of individual differences and including them in surveys of locally or nationally representative samples. To facilitate the development of such short forms, I describe a macro written for SPSS and SAS that constructs all possible short forms of a k -item composite measure. I illustrate the macro through the development of the Willingness to Self-Censor Scale Short Form.

A Computational Tool for Survey Shortening Applicable to Composite
Attitude, Opinion, and Personality Measurement Scales

Social and personality psychologists, political scientists, and others who study individual differences have produced an extraordinary number of measures of personality, attitudes, and opinions—measures that run the gamut from attitudes about specific topics or groups to broad measures of global personality (e.g., Robinson, Shaver, & Wrightsman, 1991, 1999; Rubin, Palmgreen, & Sypher, 2004). Such research tools are typically “composite measures,” (Hayes, 2005) meaning that scores on the measure are derived as an aggregation of responses to multiple self-report questions, typically using the same response scale, each of which is presumed to be an indicator of the construct presumably being measured. An example composite scale well known to public opinion researchers is the Need for Cognition scale (Cacioppo & Petty, 1982), a composite measure containing 34 statements (18 in short form, Cacioppo, Petty, & Kao, 1984) that respondents provide an answer to on an ordinal scale with respect to how self-descriptive the statement is. A person’s Need for Cognition score is defined as the sum of their responses to the questions on the scale after each answer is converted to a quantitative metric.

Research on the development and validation of such measures is typically conducted using captive research participants, often students from the investigators’ own university, who are asked to fill out pencil-and-paper questionnaires with few if any limits on time. As a result, the researchers can easily capitalize on the principle of aggregation to produce measurement instruments that yield highly reliable data. The principle of aggregation tells us that the mean or sum of quantitative responses to a set of

questions each of which is correlated with some criterion will be more strongly correlated with that criterion than will the individual items. In classical test theory and in its application to the construction of measurement instruments, the criterion of interest is typically the unmeasured “true score,” and so the principle of aggregation means that sums or averages of several items, each of which is correlated with the true score, will be a more reliable measure of the true score than any of the individual items, with reliability increasing as the number of items increases all else held constant (e.g., Rushton, Brainerd, & Pressley, 1983; Nunnally, 1978).

Research that relies on such methodology is fairly low in cost, as pencil and paper measurement procedures using samples of convenience are inexpensive to administer and implement. Moreover, there are few disadvantages to increasing instrument length when your participant pool is conveniently available, and there are many psychometric advantages to erring on the side of too many questions rather than too few to exploit the aggregation principle. Unfortunately, as a result, many of the resulting measures of personality, attitudes, and opinions are so lengthy that they are impractical to employ in public opinion research that requires the researcher to go beyond the halls of the university to obtain research participants. Much public opinion research is done over the phone or face-to-face, sometimes with people in their homes who are not captive, can hang up the phone, and who have other things to do with their time. It is unreasonable to expect the typical respondent to spend more than 20 to 30 minutes with an interviewer before his or her patience, generosity, and attention become taxed. Even when administering a survey through the mail or over the internet, there are limits on the amount of time that anyone is willing to give to answering questions when there is relatively little personal benefit to doing so.

Indeed, there is reason to be concerned that lengthier questionnaires and interviews can reduce response rates and lower data quality, although the research evidence on the effects of questionnaire and interview length is mixed. Some have found reduced response rates to longer surveys, and others find no effects of length (Heberlein & Baumgartner (1978; Kanuk & Berenson, 1975; Yammarino, Skinner, & Childers, 1991). Studies of positional effects suggest various response sets may become more likely near the end of longer surveys or interviews, perhaps because motivation and interest are lower and fatigue is higher near the end than compared to the beginning of a questionnaire or interview. Such reduced motivation to expend the necessary cognitive resources to give thoughtful responses can induce “satisficing,” a mind set which can manifest itself in a variety of different ways (Krosnick, 1991, 1999). However, the evidence supporting the existence of positional effects on data quality is not consistent (compare, e.g., Herzog & Bachman, 1981; Clancy & Wachler, 1971; Johnson, Sieveking, & Clanton, 1974; Kraut, Wolfson, & Rothenberg, 1975). The ambiguous nature of the evidence on the effects of length is likely the result of context and individual differences interacting with length in determining its effects.

Although the effects of questionnaire length on response rates and response quality may be in dispute to some extent, there is no disputing that lengthier questionnaires and interviews, whether conducted by mail, telephone, or face-to-face, are more costly administratively than shorter ones. Longer mail surveys require more postage, reproduction costs are higher, and data entry takes longer and is therefore more expensive in terms of human resources. Longer telephone interviews yield larger phone bills and reduce the number of interviews that can be completed in a given unit of time, thereby increasing costs. Similarly, longer face-to-face interviews increase the amount of

time it takes to achieve a given sample size or require more trained interviewers to yield the desired sample size in a fixed unit of time. Perhaps the only mode of administration in which longer surveys may be only negligibly more expensive is over the internet. Regardless, given that the potential for reduced data quality and response rates and the certainly higher costs of administering a longer compared to a shorter survey in large scale public opinion studies, it makes sense to try to reduce interview and questionnaire length as much as feasible.

There is ample reason to believe that the study of public opinion could be enhanced by capitalizing on the work of scholars of individual differences by including measures of personality, specific traits, or well validated measures of attitudes in public opinion polls. Berensky (2004), for instance, measured survey respondents' "impression management" and "self-monitoring" dispositions, two individual difference constructs stemming from social psychology, to test the hypothesis that people who differ in how concerned they are about the impressions they convey to others would be more likely to express socially desirable opinions. Although the evidence supporting this hypothesis was mixed, there was clearly evidence that personality was related to the opinions people were willing to openly express. Similarly, Marcus, Sullivan, Theiss-Morse, and Wood (1995) describe evidence that political tolerance is related to stable individual differences such as the "Big 5" personality dimensions. They found that people who were more tolerant of the speech of disagreeable groups tended to score higher on Openness and lower in Neuroticism and Extraversion as measured by the NEO.

For reasons described earlier, it is often impractical to use full length measures of individual difference in telephone or face-to-face interviews. Fortunately, there is a lot of redundancy built into many existing composite measures, meaning that

psychometrically satisfying short forms of existing measurement instruments can be produced that could be included in large scale surveys. Good examples are the 5-item short form of Delli Carpini & Keeter's (1993) general political knowledge scale, the 2-item short form of the Need for Cognition scale and the 3-item short form of Need to Evaluate (from the original 16 item version, Jarvis & Petty, 1996). The political knowledge scale was included in the American National Election Study in the 1990s and research has since been published using it (e.g., Brians & Watterberg, 1996). The need for cognition and need to evaluate short forms appeared on the 1998 American National Election pilot study and later in the 2000 study (Bizer, Krosnick, Petty, Rucker, & Wheeler, 2000). Research is now starting to appear based on these latter two short form measures (e.g., Bizer, Krosnick, Holbrook, Wheeler, Rucker, & Petty, 2004)—research which has helped us to better understand the role of individual differences and cognitive processes in answering questions and expressing opinions on topics that interest public opinion researchers, such as political behavior and candidate evaluations.

In the remainder of this paper, I describe a computational tool that can aid public opinion researchers interested in developing substantially shorter forms of existing self-report composite measurement instruments for use in public opinion research. This tool can also be used as a means for researchers developing their own composite measures or ad hoc measurement instruments to determine, through pilot testing, the shortest form that can be used without substantially reducing measurement quality. Using data from a pilot sample of participants who responded to all k items in a composite instrument, the procedure I describe here yields a data spreadsheet that contains each possible short form and its psychometric properties based on data from the pilot sample. As the number of items in the full composite scale increases, the number of possible shorter forms increases

exponentially (as does the size of the spreadsheet), so the procedure also produces summary statistics that make it easy for the researcher to find those the subscale that produces the most psychometrically appealing short form.

ALPHAMAX: An SPSS and SAS Macro for Composite Measure Shortening

ALPHAMAX is macro written for SPSS in the MATRIX language and SAS in PROC IML. A *macro* is a set of commands that is called with a single keyword and arguments that are used to control functioning of the program, much in the same way SPSS and SAS command syntax work. This section gives a broad overview of what the macro does. The use of the macro is illustrated with a small-scale example in which the macro was used to construct a 4-item short form of the 8-item Willingness to Self-Censor scale (Hayes, Glynn, & Shanahan, 2005).

ALPHAMAX takes as input the responses of n research participants to a set of k items used to measure a unidimensional construct, set up as an $n \times k$ data matrix. The macro produces as output a spreadsheet containing all possible short forms or “subscales”—combinations of the k items containing at least 2 items from the original pool of k . Along with information about which of the k items are in each subscale, the macro produces Cronbach’s α for each subscale (α_{sub}) and the correlation between the sum of the items in the subscale and the sum of the k items in the full scale (r_{sub}). The macro also produces summary tables containing information about the highest α_{sub} and highest r_{sub} achieved for a subscale of length m items, $m = 2$ to k .

Letting X_{ij} be respondent i ’s response to item j ,

$$T_i = \sum_{j=1}^k X_{ij} \quad (1)$$

is respondent i ’s *full scale score* using all k items, and

$$T_{sub(i)} = \sum_{j=1}^k w_j X_{ij} \quad (2)$$

is respondent i 's *subscale score* for a scale containing a subset of the k items, with that subset defined as the m items from the set of k items with $w_j = 1$ if item j is in the subscale and 0 if not. Notice that T_i is a special case of $T_{sub(i)}$ where all $w_j = 1$. As discussed below, there are $2^k - k - 2$ possible combinations of the k values of w_j , excluding the combinations with either all $w_j = 1$ or only one $w_j = 1$, corresponding to the possible subscales with at least 2 items. For example, for the subscale consisting of items 2, 4, and 5 from a 5-item full scale, $w_1 = 0$, $w_2 = 1$, $w_3 = 0$, $w_4 = 1$, and $w_5 = 1$.

With this notation, define the *subscale internal consistency reliability* as

$$\alpha_{sub} = \frac{\sum_{j=1}^k w_j}{\left(\sum_{j=1}^k w_j\right) - 1} \left(1 - \frac{\sum_{j=1}^k w_j V(X_j)}{V\left(\sum_{j=1}^k w_j X_j\right)} \right) \quad (3)$$

where $V(X_j)$ is the variance of the n responses to item j and $V(\sum w_j X_j) = V(T_{sub(i)})$ is the variance of the sum of the items in the subset of m items defined by the k values of w_j . If all $w_j = 1$, then α_{sub} is simply Cronbach's α for the scale composed of all k items, which I will denote α_{full} (Cronbach, 1951). When some $w_j = 0$, α_{sub} is Cronbach's α for a subset of the k items defined as those items with $w_j = 1$. Finally, define r_{sub} as

$$r_{sub} = \frac{n \left(\sum_{i=1}^n T_{sub(i)} T_i \right) - \left(\sum_{i=1}^n T_{sub(i)} \right) \left(\sum_{i=1}^n T_i \right)}{\sqrt{\left[n \sum_{i=1}^n T_{sub(i)}^2 - \left(\sum_{i=1}^n T_{sub(i)} \right)^2 \right] \left[n \sum_{i=1}^n T_i^2 - \left(\sum_{i=1}^n T_i \right)^2 \right]}} \quad (4)$$

which is the familiar Pearson correlation between T_i and $T_{sub(i)}$ and quantifies the strength of the association between the full scale and subscale scores using all n respondents.

The user of the ALPHAMAX macro would be interested in finding a small subset of m items, $2 \leq m < k$, for which α_{sub} is not much smaller than α_{full} and r_{sub} is close to 1. That is, the user desires a “shortened” version of the k -item full scale with internal consistency reliability not much worse (and perhaps better than) the full scale that correlates very highly with the full scale total and that retains at least the subjective validity qualities of the full scale (i.e., face and content). If such a subscale exists, and typically several do, then the $k - m$ remaining items in the full scale add little to the psychometric qualities of the scale and can potentially be abandoned for the sake of efficiency. That is, why spend the effort and cost administering a k -item measure when a small subset of those same items will suffice from a psychometric standpoint?

However, finding a psychometrically satisfying subset of the k items can be computationally tedious because the number of subsets is typically large. Specifically, if a scale contains k items, then there are $2^k - k - 2$ possible subscales with at least 2 items. For example, if $k = 8$, then there are 246 subscales, and if $k = 15$, the number of subscales is 32,751. It would be practically impossible to manually construct all possible subscales as well as α_{sub} and r_{sub} . Although some existing statistical packages can help to identify certain subscales, such programs are of limited use. For example, SPSS will generate all possible “leave-one-out” subscales in its RELIABILITY procedure, showing α_{sub} for k of the possible subscales constructed by throwing out one item from the set of k items. This procedure could be used iteratively by running the RELIABILITY procedure repeatedly, the first time using all k items, the second time throwing out the item that lowers α the most, etc., until one has found a subscale that is satisfying. But such “sampling” of the possible subscales is unsatisfying. First, it will produce only a small fraction of the possible subscales. Second, the values of α_{sub} produced will be “hierarchically

constrained” in that the values of α_{sub} at iteration t will depend on which of the k items was discarded at iteration $t - 1$. For example, if $k = 4$, SPSS would yield α_{sub} for subscales (1,2,3), (1,2,4), (1,3,4), and (2,3,4) at the first iteration. Suppose alpha is reduced most when item 1 is discarded. At the second iteration, SPSS would produce α_{sub} for subscales (2,3), (3,4), and (2,4). Alpha cannot be computed with fewer than two items, so the user would stop having identified α_{sub} for only 7 of the 10 possible subscales. More generally, assuming one iterates until only two items remain, this procedure would generate only $50 (k^2 + k - 6) / (2^k - k - 2)$ percent of the possible subscales. As just illustrated, if $k = 4$, a full 70% of the values of α_{sub} are produced using this method. But if $k = 10$, leave-one-out iteration yields only about 5% of the possible subscales and values of α_{sub} . If the analyst stops earlier, this number is even smaller. It is possible that the most psychometrically satisfying subscale is among the very many subscales this iterative procedure did not identify.

Another problem with a leave-one-out procedure based on coefficient α is that it cannot be used to identify subscales with large r_{sub} . Although Cronbach’s α is an important psychometric index of measurement quality, it could be argued that as important if not more important when constructing a short form is r_{sub} , the correlation between T and T_{sub} . Given that the goal is to construct a substitute for T using only a subset of the k items used to generate T , one would want a subscale score (i.e., T_{sub}) that is maximally correlated with the total using all k of the items. It is possible for α_{sub} to be relatively high but r_{sub} to be still be relatively far from 1. Thus, searching for a small subset of items with large α_{sub} may not yield a subscale that correlates highly with the full scale.

The ALPHAMAX macro avoids the problems associated with the iterative leave-one-out method by generating all $2^k - k - 2$ of the possible subscales, along with their corresponding values of α_{sub} and r_{sub} . The user can scan the resulting data matrix containing the possible subscales as well as some summary statistics the macro produces and select the subscale that is most psychometrically satisfying, using either α_{sub} , r_{sub} , or some combination or compromise of the two depending on the user's goal, along with subjective judgments about face and content validity of the subscale. Typically, the macro would be used with data from a pilot study in which all k items are administered. Once a satisfying subscale is constructed, the full study can be conducted using only the subset of the items identified as necessary to produce a psychometrically satisfying proxy of the original full scale.

Example Application: Constructing the Willingness to Self-Censor Scale Short Form

To illustrate the use of the ALPHAMAX macro, I next describe its application to the construction of a short form of the Willingness to Self-Censor (WTSC) scale. WTSC is an individual difference introduced by Hayes, Glynn, & Shanahan (2005) to explain the weak relationship observed in a meta-analysis of survey studies (Glynn, Hayes, & Shanahan, 1997) between a person's belief as to how widely shared his or her opinion on an issue is amongst the public and willingness to express that opinion in public situations. Hayes et al. (2005) argued that some people use cues about the climate of opinion more than do others when deciding whether or not to voice an opinion, and they developed the WTSC scale to measure this individual difference. The scale is designed to measure the extent to which a person is willing to withhold his or her true opinion around an audience perceived to disagree with it. The eight items on the full scale can be found in Table 1. A person's willingness to self-censor is defined as the sum (or average) of their responses

to these 8 statements after reverse scoring items four and eight. There is a growing body of evidence that the WTSC scale is a valid measure of this individual difference, in that it correlates with other individual differences and behavior that theory and intuition predict it should correlate with (Hayes, Glynn, & Shanahan, 2005; Hayes, Scheufele, & Huges, in review). Furthermore, experimental evidence indicates that it successfully distinguishes between people who are more versus less susceptible to a manipulation of the climate of opinion when asked, hypothetically, how willing they are to express their opinion in public (Hayes, Glynn, & Shanahan, in press). There is also evidence that self-censors (i.e., those who score relatively high on the measure) are slower to provide their opinions on domains in which they are in objectively in the minority compared to the majority (Huges, 2004).

Nevertheless, the scale could be improved. Although eight items is not a prohibitive number to include in a public opinion survey to measure a single construct, in the context of a lengthy survey, any means of reducing the number of items is helpful. There is also some redundancy in the items, suggesting some shortening might be worthwhile. For instance, items 1, 4, and 8 are quite similar, differing only slightly in wording. So it might be possible to reduce the number of items on the scale by eliminating one or more of these items and thereby increase the efficiency of administration and decrease costs without substantially lowering measurement reliability and validity. The ultimate goal in the construction of the short form was to reduce the number of items to between three and five without substantially sacrificing measurement quality. Any fewer than three items would likely yield an instrument low in content and face validity and destroy the benefits of statistical aggregation to reduce error, and producing a short form with 6 or more items seemed hardly an improvement given the

original scale contains only eight items. Additionally, given the redundancy in wording of statements 1, 4, and 8, it was deemed desirable that the short form include no more than one of these items.

Developing the short form progressed in several steps, and it is recommended that researchers using the macro follow this model. At the first step, responses to the willingness to self-censor scale from a group of conveniently available college student participants were obtained. For reasons described below, this student sample was randomly split in half. Respondents in the first half of the sample were treated as the *development* sample and their responses submitted to the ALPHAMAX macro and the macro output examined in search of one or more potential short forms that met certain criteria. The second step was crossvalidation. It is well known that exploratory, post-hoc “data snooping” can produce findings that do not replicate well. This is a problem when composite scales are constructed by examining potential items for a scale that do and do not perform psychometrically by examining statistics such as alpha-when-deleted (Kopalle & Lehman, 1997). It is also problematic in other correlational procedures in which a person seeks the best fitting model by attempting to maximize a measure of association. Validity shrinkage is nearly inevitable when the model is applied to a new sample (e.g., Darlington, 1990, p. 159-163). Applied to the use of the ALPHAMAX macro, the potential is high for the “best” short form to have satisfying α_{sub} and r_{sub} only in the sample used to develop it. To avoid this problem, α_{sub} and r_{sub} were calculated in the second half of the student sample, the *crossvalidation* sample, but only for the potential short forms identified using the development sample. As will be seen, a second crossvalidation sample from a national sample of U.S. residents was also used in a third step to examine which short form better replicated in a sample different in its

characteristics from both the development and first crossvalidation sample. This was possible to do because data from this second sample were available. In practice, researchers may not have a second crossvalidation sample, in which case this additional step can be eliminated.

The development and validation samples were derived from a convenience sample of 759 college students taking courses in communication at a large Midwestern university ($M_{\text{age}} = 20.58$, $SD = 2.64$, 54% female) who responded to all eight items on the Willingness to Self-Censor scale during class. The order of the cases in the data file was randomized and the first 359 cases treated as the development sample, with the remaining 359 cases held out for crossvalidation. Once the macro was activated (see Appendix A for instructions), the responses to the eight items from the development sample were submitted to the macro. The command syntax is

In SPSS: `alphamax wtsc1 wtsc2 wtsc3 wtsc4r wtsc5 wtsc6 wtsc7 wtsc8r.`

In SAS: `%alphamax (data = fname, vars = wtsc1 wtsc2 wtsc3 wtsc4r wtsc5 wtsc6 wtsc7 wtsc8r);`

With $k = 8$ items, there are $2^8 - 8 - 2 = 246$ possible short forms with at least two items.

Figure 1 displays an abbreviated version of the resulting spreadsheet the macro automatically creates in the SPSS data window. (In SAS the user must print the data file that results using PROC PRINT or open the file using the data matrix editor). Each row in the spreadsheet corresponds to a subscale or potential short form. Because the macro also produces information for the scale using all k items (near the top of the spreadsheet in this case), the spreadsheet has 247 rows rather than 246. The majority of the cells in the data matrix contain the k values of w_j for the subscales, set to 1 if the item in the column heading is in the subscale and 0 if not. So, for example, row 10 in the spreadsheet corresponds to the subscale with items 1, 3, 4, 5, and 8 (as numbered above).

The remaining columns contain α_{sub} and r_{sub} for the subscale (a_sub and r_sub, respectively), and the number of items in the subscale.

The summary statistics in Figure 2, also provided by the macro, allow for a quick determination of the maximum α_{sub} and r_{sub} obtained for a subscale of various lengths. For instance, in the table below “Maximum Alpha (a_sub) as a Function of Number of Items” the largest α_{sub} in the possible three item subscales is .7424. As can be seen in Figure 1 (subscale number 61), this subscale contains items 1, 4, and 5, and Figure 1 and Figure 2 both tell us this subscale has $r_{sub} = .8770$. And the largest α_{sub} in a 4-item subscale is .7655, with an r_{sub} value of .9114. Figure 1 shows that this corresponds to the subscale with items 1, 4, 5, and 8 (subscale number 31). The table below “Maximum Subscale-Full Scale Correlation as a Function of Number of Items” provides the largest r_{sub} for various subscales of different lengths. For example, the maximum subscale-full scale correlation in a three item subscale is .9019, attained for the subscale with items 1, 3, and 4 (number 93 in Figure 1). Figure 1 and 2 tell us that this subscale has $\alpha_{sub} = .7178$. Similarly, the largest r_{sub} for a 4 item scale is .9337, which is subscale number 152 in Figure 1 with items 1, 2, 3, and 6. This subscale has $r_{sub} = .6688$.

Using the summary statistics provided by the macro (Figure 2) it is easy to hone in on good candidates for a shortened form with good psychometric qualities, depending on whether interest is maximizing internal consistency reliability of the subscale or the subscale-full scale correlation. But in this particular example, all but one of the “best” potential subscales listed in Figure 2 do not satisfy the criterion that the subscale contains only one of the items from the (1, 4, 8) redundant set. The exception is a four item subscale containing items 1, 2, 3, and 6. However, this subscale’s internal consistency is a bit lower than we typically like to see, and this would likely decrease with

crossvalidation. The solution is to hunt through the spreadsheet looking for good alternatives that meet this criterion and also has decent α_{sub} and good r_{sub} . Although a perusal of the data yielded many possibilities, two seemed most promising. These are displayed in Figure 1. Subscale 81 containing items 1, 2, 3, and 7 has α_{sub} of .7273 and r_{sub} of .9168. And subscale 117, with items 1, 2, 3, and 5, yields α_{sub} of .6998 and r_{sub} of .9118. Both of these scales meet minimum acceptable internal consistency, have decent subscale-full scale correlations, and, importantly, also have desirable face and content validity in so far as possible with a short form. The items are phrased in such a way that they capture the essence of the construct as defined in Hayes et al. (2005) with relatively little redundancy.

The second step was to crossvalidate to see if the good psychometric qualities of these two subscales generalize when α_{sub} and r_{sub} are calculated in a sample that was not used to develop them. This was accomplished by calculating these two statistics using the second half of the data held out from the development phase for crossvalidation purposes. For the first candidate short form (items 1, 2, 3, and 7), $\alpha_{sub} = .6927$ and $r_{sub} = .9140$. Not surprisingly, there was a little slippage in these statistics at crossvalidation, but they are still decent. For the second short form (items 1, 2, 3, and 6), $\alpha_{sub} = .7090$ and $r_{sub} = .9213$, which are actually somewhat higher than obtained during the development phase. Although either of these subscales perform well at crossvalidation, it would be sensible to prefer the second option.

Both the development and crossvalidation data sets were derived from a student sample of convenience from a single university. It would be desirable to know which, if any, of these two subscales crossvalidate in a sample from a different population. If one clearly outperforms the other when applied to a different population, that subscale would

be preferred as the final short form. Although in practice this may not be feasible, in this case, the data were available to carry out a second crossvalidation. The data obtained for this third phase came from a national probability sample of 723 U.S. residents who participated in a telephone poll in November 2003 ($M_{\text{age}} = 49.63$, $SD_{\text{age}} = 17.01$, 55% female). Each participant responded to all eight of the willingness to self-censor items, making it possible to calculate not only α_{sub} but also r_{sub} in this second crossvalidation sample. In this national sample of U.S. residents, the first short form had $\alpha_{\text{sub}} = .7106$, $r_{\text{sub}} = .9126$, whereas the second short form had $\alpha_{\text{sub}} = .6914$, $r_{\text{sub}} = .9054$. Given that the first short form crossvalidates with higher internal consistency and larger subscale-full scale correlation than short form 2, short form 1 is preferred. Thus, the final Willingness to Self-Censor Scale Short Form contains items 1, 2, 3, and 7 from the full scale.¹

Extensions and Limitations

Using ALPHAMAX for Item Analysis

The example above illustrates how the ALPHAMAX macro can be used to derive a shortened version of an existing multiple item measurement instrument. The macro can also be used to help in the construction of new multiple item instruments from a pool of preliminary items. When constructing a new multi-item instrument of this sort, it is common to do an item analysis once data are available, examining how each item in the scale contributes to the psychometric qualities of the scale. For example, the investigator might calculate whether and by how much alpha increases or decreases when an item is removed from the scale, or the item-corrected correlation between the scores on an item and the total scale score excluding that item. This approach to item analysis has problems, however. The change in alpha when an item is removed is dependent on the presence of the other items in the scale, some of which may themselves be worthy of

exclusion. Similarly, the item-corrected correlation includes items in the corrected scale total that may be detrimental to the reliability and validity of the scale. It seems pointless to define an item as good if it correlates well with a total score that itself potentially includes faulty items.

Information obtained from the ALPHAMAX helps to overcome the shortcomings of such leave-one-out methods of item analysis. If an item is faulty and lowers the reliability of the composite scale, then it would be expected that a subscale that includes that item would tend to have a lower alpha, on average, than would a subscale that excludes it when considering all possible subscales that can be created from the k items. The ALPHAMAX macro provides the “average gain in alpha” when an item is included in the scale. This statistic, displayed in Figure 2 in the table labeled “Mean Scale Alpha for All Possible Scales With and Without Item,” is derived as the difference between the average α_{sub} for all possible subscales that include the item and the average α_{su} for all possible subscales that exclude it. If the average gain in alpha is positive, this means that alpha tends to be higher when the item is included compared to when it is excluded—a desirable feature of an item in a composite scale. In Figure 2, it can be seen that all items except number 6 have positive alpha gains. This is not surprising given that the Willingness to Self-Censor scale went through a rigorous development procedure that excluded undesirable items as they became known. The small decrement in reliability attributable to item 6 is likely a product of this specific data set (i.e., sampling error) rather than evidence that this is a bad item.

This average gain in alpha can be used during the development of a new multi-item measurement scale by identifying preliminary items that tend to lower the internal consistency of the scale when the item is included. Importantly, because this statistic is

derived using all possible subscales, it more accurately gauges the performance of the item. That is, it is not reliant exclusively on a single comparison scale that includes every preliminary item other than the one under scrutiny.

The spreadsheet the macro produces can also be used to conduct a similar form of item analysis to what the average alpha gain table provides. By regressing either α_{sub} or r_{sub} on the k columns of values of w_j , the result is k regression weights each quantifying the average amount by which α_{sub} or r_{sub} increases (or decreases) when the item is included in the scale, partialing out the effects of the other items. Poor items will have regression weights near zero or negative.

Shortening Multidimensional Scales

In order to use the ALPHAMAX procedure, one must either make the assumption that the k items tap a unidimensional construct or be willing ignore the possibility that the full scale or potential subscales are multidimensional. Coefficient alpha is not a measure of dimensionality, so knowing that a subscale has high internal consistency does not rule out multidimensionality (Cortina, 1993; Schmidt, 1996). Nevertheless, the macro can be used to shorten multidimensional scales by focusing each run of the macro on items that are known to load on a common factor. Once the number of items for each factor has been reduced, a factor analysis can be used to verify that the shortened multidimensional scale retains its factor structure.

Practical Limitations on the Number of Items

In the example presented here, ALPHAMAX was been applied to a multi-item measurement scale with only 8 items. In principle, there is no limit to the number of items a scale submitted to the macro can handle. But in practice, time and computer memory both impose limits. As mentioned earlier, there are $2^k - k - 2$ possible subscales

of length no less than 2 that can be constructed from a scale with k items. As can be seen most clearly in Table 2, the number of possible subscales explodes as k increases above 20 or so. Some computers may have problems storing the data file that results (which will contain $[2^k - k - 2][k + 3]$ cells), and depending on the speed of the computer, enumerating all the possible subscales may simply take too long for the user to wait for the result.

Conclusion

Because of the constraints that public opinion researchers face, it is difficult for public opinion researchers to capitalize on the work of personality and social psychologists who have developed reliable, valid measures of attitudes and individual differences. Most of the measurement instruments that have been developed are too long to be implemented successfully in telephone, face-to-face, and mail contexts where time and administrative costs loom much larger than they do when collecting data using conveniently available, captive participants such as college students. I argued that it is worthwhile to attempt to ride the coattails of those before us by taking existing measures of personality, attitudes, and opinions and developing ultra-shortened versions that can be included in public opinion polls. In this manuscript I introduced and described the use of a computational tool in the form of a macro for SPSS and SAS to aid in the construction of short forms of composite attitude, personality and opinion scales. This macro can also be used to facilitate the construction of new measurement instruments that are as short as need be to meet minimum psychometrics standards for inclusion in public opinion research.

A final caveat must be acknowledged. It takes time and substantial research before a measurement procedure is established and accepted as reliable and valid. We

cannot just willy-nilly modify instruments that may be based on decades of research and assume that a short form is also reliable and valid. Although one can logically argue that a short form that correlates strongly with the full form is probably equally “good” in a psychometric sense, this is simply an argument and an assumption being made.

Whenever possible, short forms should go through a rigorous process of validation to establish that they, too, are reliable and valid measures of the desired construct. At a minimum, a researcher using a short form developed by the procedure described here should attempt to make the argument, preferably empirically, that the measurement instrument used is yielding high quality measurement of the intended construct.

References

- Berensky, A. J. (2004). Can we talk? Self-presentation and the survey response. *Political Psychology, 25*, 643-659.
- Bizer, G. Y., Krosnick, J. A., Petty, R. E., Rucker, D. D., & Wheeler, S. C. (2000). *Need for cognition and need to evaluate in the 1998 National Election Survey pilot study*. NES Pilot Study Report, No. nes008997, downloaded November 7, 2005 from <http://www.umich.edu/~nes/resources/papers/pilotrpt.htm>
- Bizer, G. Y., Krosnick, J. A., Holbrook, A. L., Wheeler, S. C., Rucker, D. D., & Petty, R. E. (2004). The impact of personality on cognitive, behavioral, and affective political processes: The effects of need to evaluate. *Journal of Personality, 72*, 995-1027.
- Brians, C. L., & Wattenberg, M. P. (1996). Campaign issue knowledge and salience: Comparing reception from TV commercials, TV news, and newspapers. *American Journal of Political Science, 40*, 172-193.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.
- Clancy, K. J., & Wachsler, R. A. (1971). Positional effects in shared-costs surveys. *Public Opinion Quarterly, 35*, 258-265.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

- Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.
- Delli Carpini, M. X., & Keeter, S. (1993). Measuring political knowledge: Putting first things first. *American Journal of Political Science*, *37*, 1179-1206.
- Glynn, C. J., Hayes, A. F., & Shanahan, J. (1997). Perceived support for one's opinions and willingness to speak out: A meta-analysis of survey studies on the "spiral of silence." *Public Opinion Quarterly*, *61*, 452-463.
- Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hayes, A. F., Scheufele, D. A., & Huge, M. (in review). Nonparticipation as self-censorship: Publicly-observable political activity in a polarized opinion climate. *Manuscript submitted for publication*.
- Hayes, A. F., Glynn, C. J., & Shanahan, J. (2005). Willingness to self-censor: A construct and measurement tool for public opinion research. *International Journal of Public Opinion Research*, *17*, 298-323.
- Hayes, A. F., Glynn, C. J., & Shanahan, J. (in press). Validating the willingness to self-censor scale: Individual differences in the effect of the climate of opinion on opinion expression. *International Journal of Public Opinion Research*.
- Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, *43*, 447-462.
- Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*, 549-559.

- Huge, M. (2004). *Measurable hesitation as a precursor to self-censorship: Replication and extension of the minority slowness effect*. Paper presented at the annual meeting of the Midwestern Association for Public Opinion Research, Chicago, IL.
- Jarvis, W. G., & Petty, R. E. (1996). The need to evaluate. *Journal of Personality and Social Psychology, 70*, 172-194.
- Johnson, W. R., Sieveking, N. A., & Clanton, E. S. (1974). Effects of alternating positioning of open-ended questions in multiple-choice questionnaires. *Journal of Applied Psychology, 59*, 776-778.
- Kanuk, L., & Berenson, C. (1975). Mail surveys and response rates: A literature review. *Journal of Marketing Research, 12*, 440-453.
- Kopalle, P. K., & Lehman, D. R. (1997). Alpha inflation?: The impact of eliminating scale items on Cronbach's alpha. *Organizational Behavior and Human Decision Processes, 70*, 189-197.
- Kraut, A. I., Wolfson, A. D., & Rothenberg, A. (1975). Some effects of position on opinion survey items. *Journal of Applied Psychology, 60*, 774-776.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567.
- Marcus, G. E., Sullivan, J. L., Theiss-Morse, E., & Wood, S. L. (1995). *With malice toward some: How people make civil liberties judgments*. Cambridge, England: Cambridge University Press.
- Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlations. *Psychological Bulletin, 111*, 172-175.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1999). *Measures of political attitudes*. San Diego, CA: Academic Press.
- Rubin, R. B., Palmgreen, P., & Sypher, H. E. (2004). *Communication research measures: A sourcebook*. Mahwah, NJ: Erlbaum.
- Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, *94*, 18-38.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350-353.
- Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior. A meta-analysis. *Public Opinion Quarterly*, *55*, 613-639.

Footnote

¹ Using the data described in the original paper introducing the Willingness to Self-Censor scale (Hayes et al., 2005), the correlation between Willingness to Self-Censor short form scores and the same individual differences used to assess the convergent validity of the full scale were calculated and compared with the same correlations based on the full scale. Those correlations were (with full-scale correlations reported second): fear of negative evaluation, $r = .491$ vs $.542$, $p = .033$, $n = 270$; public self-consciousness, $r = .239$ vs $.244$, $p = .821$, $n = 496$; shyness, $r = .544$ vs $.631$, $p < .0001$, $n = 1622$; social anxiety, $r = .487$ vs $r = .523$, $p = .137$, $n = 232$; communication apprehension, $r = -.486$ vs $-.556$, $p = .074$, $n = 85$; self-esteem, $r = -.455$ vs $-.472$, $p = .498$, $n = 231$; argumentativeness, $r = -.511$ vs $-.512$, $p = .968$, $n = 84$. The p -values for the difference were derived using the Meng, Rosenthal, & Rubin (1992) test for comparing statistically dependent correlations.

Table 1.

The Full Willingness to Self-Censor Scale

Item	Statement
1	It is difficult for me to express my opinion if I think others won't agree with what I say. *
2	There have been many times when I thought others around me were wrong but I didn't let them know. *
3	When I disagree with others, I'd rather go along with them than argue about it. *
4	It is easy for me to express my opinion around others who I think will disagree with what I say. (R)
5	I'd feel uncomfortable if someone asked my opinion and I knew that he or she wouldn't agree with me.
6	I tend to speak my opinion only around friends or other people I trust.
7	It is safer to keep quiet than publicly speak an opinion that you know most others don't share. *
8	If I disagree with others, I have no problem letting them know it. (R)

(R) = reverse scored.

* Item in the final Willingness to Self-Censor scale Short Form

Responses are made on a 1 (strongly disagree) to 5 (strongly agree) scale

Table 2.

*Number of Possible Subscales with at Least Two**Items that can be Constructed from a Set of k Items*

k	Possible Subscales
3	3
4	10
5	25
6	56
7	119
8	246
9	501
10	1,012
11	2,035
12	4,082
13	8,177
14	16,368
15	32,571
16	65,518
17	131,053
18	262,124
19	524,267
20	1,048,554
21	2,097,129
22	4,194,280
23	8,388,583
24	16,777,190
25	33,554,405
26	67,108,836
27	134,217,699
28	268,435,426
29	536,870,881
30	1,073,741,792
31	2,147,483,615
32	4,294,967,262
33	8,589,934,557
34	17,179,869,148
35	34,359,738,331
40	1,099,511,627,734

Note: Possible subscales = $2^k - k - 2$

Figure 1.

Subscale Spreadsheet from the SPSS ALPHAMAX Macro, Development Sample

	wpsc1	wpsc2	wpsc3	wpsc4r	wpsc5	wpsc6	wpsc7	wpsc8r	a_sub	r_sub	items
1	1	1	1	1	1	1	0	1	.8126	.9736	7
2	1	1	1	1	1	1	1	1	.9123	1.0000	8
3	1	1	1	1	1	0	0	1	.8058	.9623	6
...
10	1	0	1	1	1	0	0	1	.7900	.9424	5
11	1	1	1	1	1	0	1	0	.7396	.9635	6
...
30	1	0	1	1	0	1	1	1	.7657	.9731	6
31	1	0	0	1	1	0	0	1	.7655	.9114	4
...
60	1	1	0	1	0	1	0	1	.7427	.9558	5
61	1	0	0	1	1	0	0	0	.7424	.8770	3
...
81	1	1	1	0	1	0	0	0	.7273	.9158	4
82	0	1	1	0	1	1	1	1	.7249	.9715	6
...
92	0	1	1	1	1	0	0	0	.7106	.9207	4
93	1	0	1	1	0	0	0	0	.7178	.9019	3
94	1	1	0	1	0	1	1	0	.7172	.9590	5
...
117	1	1	1	0	0	0	1	0	.6990	.9110	4
118	1	1	1	0	0	1	1	0	.6988	.9557	5
...
129
130	1	1	0	0	0	1	1	1	.6077	.9600	5
...
151	0	0	1	1	1	1	0	0	.6718	.9292	4
152	1	1	1	0	0	1	0	0	.6688	.9337	4
...
247	0	0	0	0	0	1	1	0	.3519	.7543	2

Cronbach's alpha
Correlation with full scale

- ← Largest alpha, 5 indicators
- ← Largest alpha, 4 indicators
- ← Largest alpha, 3 indicators
- ← A potential 4 item short-form with only 1 indicator from (1, 4, and 8)
- ← Largest correlation with full scale, 3 indicators
- ← A potential 4 item short-form with only 1 indicator from (1, 4, and 8)
- ← Largest correlation with full scale, 5 indicators
- ← Largest correlation with full scale, 4 indicators

Figure 2.

Summary Output from the SPSS ALPHAMAX Macro Applied to the Development
Sample for Constructing the Willingness to Self-Censor Scale Short Form.

alphamax wtsc1 wtsc2 wtsc3 wtsc4r wtsc5 wtsc6 wtsc7 wtsc8r.

Run MATRIX procedure:

*** Subtest Statistics are in Data Window ***

Mean Scale Alpha for All Possible Scales With and Without Item

	Without	With	Gain
wtsc1	.6283	.7168	.0885
wtsc2	.6546	.6920	.0375
wtsc3	.6421	.7038	.0618
wtsc4r	.6308	.7144	.0836
wtsc5	.6486	.6976	.0490
wtsc6	.6773	.6705	-.0068
wtsc7	.6666	.6806	.0140
wtsc8r	.6486	.6977	.0491

Maximum alpha (a_sub) as a Function of Number of Items

Items	a_sub	r_sub
2.0000	.7012	.8438
3.0000	.7424	.8770
4.0000	.7655	.9114
5.0000	.7900	.9424
6.0000	.8058	.9623
7.0000	.8126	.9766
8.0000	.8123	1.0000

* r_sub is correlation between scale with maximum alpha and full scale total.

Maximum Subscale-Full Scale Correlation as a Function of Number of Items

Items	r_sub	a_sub
2.0000	.8472	.5905
3.0000	.9019	.7178
4.0000	.9337	.6688
5.0000	.9608	.6877
6.0000	.9777	.7556
7.0000	.9899	.7858
8.0000	1.0000	.8123

Number of cases:

359

----- END MATRIX -----

Appendix

The ALPHAMAX Macro

SPSS and SAS have matrix languages (MATRIX in SPSS and PROC IML in SAS) that, when combined with the macro facility (DEFINE..!END DEFINE in SPSS and %MACRO...%MEND in SAS), allow the user to create new commands to conduct analyses the programs otherwise cannot accomplish. This appendix describes how to activate the ALPHAMAX macro and the syntax structure of the ALPHAMAX command.

SPSS

The command set below creates the ALPHAMAX macro. This command set must be run in a SPSS syntax window. The macro can typed into a syntax window directly or downloaded from <http://www.comm.ohio-state.edu/ahayes/macros.htm>. Once the command set has executed, the user will have a new command called ALPHAMAX. Macros are temporary commands and deleted when the user quits SPSS.

The syntax is

```
ALPHAMAX varlist.
```

where *varlist* is a list of k variables in the active SPSS data file containing the responses to the k items in the full scale. The macro assumes that missing data are represented as “system missing” with an empty cell (“.”). Listwise deletion is used to purge the data of respondents (rows in the data matrix) who are missing on any of the variables in *varlist*. The macro does not recognize user defined missing value codes and it will treat user defined missing values as valid data. Thus, the user should change user missing to system missing before running the ALPHAMAX command.

The macro will replace the data file containing the responses to the k items in the full scale with a spreadsheet containing the possible subscales, as described in the

manuscript. To avoid data loss, the data file should be saved before running the macro.

The macro has no error-checking routines. Various problems with the data can produce a lengthy string of errors in the output window, including a subscale or item with no variance in responses.

The time it takes the macro to execute increases substantially as k increases. SPSS will display “Running matrix...” when the macro is executing. With large k , it may be necessary to increase the size of the workspace and data matrix with the SET WORKSPACE and SET MXCELLS commands in order for the macro to execute. See the SPSS syntax reference manual for guidance.

```

DEFINE alphamax (!POSITIONAL !CHAREND ('/')).
PRESERVE.
SET MXLOOPS = 100000000.
SET LENGTH = NONE.
matrix.
get dt/variables = !1/names = nm/missing = omit.
compute total = rsum(dt).
compute it = ncol(dt).
compute n = nrow(dt).
compute dt = {dt,total}.
compute ones = make(n,1,1).
compute dtc = dt-(ones*(csum(dt)&/n)).
compute vcov = (1/(n-1))*(t(dtc)*dtc).
compute ivar = diag(vcov).
compute dt = dt(:,1:it).
compute x = make(1,it,0).
compute k = 0.
compute y = make(((2&**it)-1)-it,it+3,0).
loop j = 1 to ((2&**it)-1).
  compute x(1,1)=x(1,1)+1.
  loop i = 1 to it.
    do if x(1,i) = 2.
      compute x(1,i)=0.
      compute x(1,i+1)=x(1,i+1)+1.
    end if.
  end loop.
do if rsum(x) > 1.
  compute ttl = dt*t(x).
  compute tmp = {ttl,total}.
  compute tmpc = tmp-(ones*(csum(tmp)&/n)).
  compute vcov = (1/(n-1))*(t(tmpc)*tmpc).
  compute var = diag(vcov).
  compute r = mdiag(1/sqrt(var))*vcov*mdiag(1/sqrt(var)).
  compute r = r(2,1).
  compute alpha=(rsum(x)/(rsum(x)-1))*(1-(x*ivar(1:it,1))/var(1,1))).
  compute k = k + 1.
  compute y(k,:)= {x, alpha, r, rsum(x)}.
end if.
end loop.
compute res = make(it,3,0).
loop i = 1 to it.

```

```

compute pre = {make(nrow(y),1,1), y(:,i)}.
compute b = inv(t(pre)*pre)*t(pre)*y(:,(it+1)).
compute res(i,:)={b(1,1), (b(1,1)+b(2,1)),b(2,1)}.
end loop.
compute tp = make(it,1,-999).
compute tp2 = tp.
compute tp3 = tp.
compute tp4 = tp.
loop i = 1 to (nrow(y)).
  do if (tp(y(i,it+3),1) < y(i,it+1)).
    compute tp(y(i,it+3),1) = y(i,it+1).
    compute tp2(y(i,it+3),1) = y(i,it+2).
  end if.
  do if (tp3(y(i,it+3),1) < y(i,it+2)).
    compute tp3(y(i,it+3),1) = y(i,it+2).
    compute tp4(y(i,it+3),1) = y(i,it+1).
  end if.
end loop.
compute tmp = {t({2:it}), tp(2:it,1), tp2(2:it,1)}.
print/title = "*** Subtest Statistics are in Data Window ***".
print res/title = " Mean Scale Alpha for All Possible Scales With and
  Without Item "/clabels = "Without" "With" "Gain"/rnames = nm/format = F8.4.
print tmp/title = "Maximum alpha (a_sub) as a Function of Number of
  Items"/clabels = "Items" "a_sub" "r_sub"/format = F8.4.
print/title = "* r_sub is correlation between scale with maximum alpha and full
  scale total.".
compute tmp = {t({2:it}), tp3(2:it,1), tp4(2:it,1)}.
print tmp/title = "Maximum Subscale-Full Scale Correlation as a Function of
  Number of Items"/clabels = "Items" "r_sub" "a_sub"/format = F8.4.
compute nms = {nm, "a_sub", "r_sub", "items"}.
print n/title = "Number of cases:".
save y/outfile = */names = nms.
END MATRIX.
format !1 items (F2.0).
format a_sub r_sub (F6.4).
variable labels a_sub "Cronbach's alpha".
variable labels r_sub 'Correlation with total from full scale'.
variable labels items 'Number of Items'.
sort cases by a_sub (d).
RESTORE.
!END DEFINE.

```

SAS

With a few exceptions, the SAS macro works just like the SPSS version. First, the command set below must be executed in SAS by typing it in or downloading it from <http://www.comm.ohio-state.edu/ahayes/macros.htm>. This will activate the ALPHAMAX command. The syntax for the ALPHAMAX command is

```
%alphamax (data = filename, vars = varlist);
```

where *filename* is the name of the SAS data file containing the responses to the *k* items in the full scale and *varlist* is a list of *k* variable names holding the responses to the items.

Any errors that result will appear in the SAS log. The data file containing the subscale spreadsheet will be saved as a SAS datafile name “subscales.” The macro automatically loads the subscale file into memory. To see the spreadsheet, the user must print it with PROC PRINT or, in Windows versions of SAS, by opening up the data editor.

The macro uses listwise deletion of cases missing any response on the *k* items, and it assumes missing data are represented as “.” in the data file.

```
%macro alphamax(data=,vars=);
proc iml;
use &data;
read all var{&vars} into dt;
nm={&vars};
xx=(dt = .);xx=xx[,+];
j=1;do i = 1 to nrow(dt);if xx[i,1]=0 then;do;dt[j,]=dt[i,];j=j+1;end;end;
dt=dt[1:j-1,];
total = dt[,+];
it = ncol(dt);
n = nrow(dt);
dt = dt||total;
ones = j(n,1,1);
dtc=dt-(ones*(dt[,+]/n));
vcov = (1/(n-1))*dtc`*dtc;
ivar=vecdiag(vcov);
dt=dt[,1:it];
x=j(1,it,0);
k =0;
y=j(((2##it)-1)-it,it+3,0);
do j = 1 to (2##it)-1;
  x[1,1]=x[1,1]+1;
  do i = 1 to it;
    if x[1,i]=2 then;
      do;
        x[1,i]=0;
        x[1,i+1]=x[1,i+1]+1;
      end;
    end;
  end;
if x[,+] > 1 then;
do;
  ttl=dt*x`;
  tmp=ttl||total;
  tmpc=tmp-(ones*(tmp[,+]/n));
  vcov=(1/(n-1))*(tmpc`*tmpc);
  var=vecdiag(vcov);
  r=diag(1/sqrt(var))*vcov*diag(1/sqrt(var));
  r=r[2,1];
  alpha=(x[,+]/(x[,+]-1))*(1-(x*ivar[1:it,1])/var[1,1]);
  k=k+1;
  y[k,]=x||alpha||r||x[,+];
end;
end;
```

```

end;
res=j(it,3,0);
do i = 1 to it;
  pre=j(nrow(y),1,1)||y[,i];
  b=inv(pre`*pre)*pre`*y[, (it+1)];
  res[i,]=b[1,1]||(b[1,1]+b[2,1])||b[2,1];
end;
tp=j(it,3,0);
tp2=tp;
tp3=tp;
tp4=tp;
do i = 1 to nrow(y);
  if (tp[y[i,it+3],1] < y[i,it+1]) then;
    do;
      tp[y[i,it+3],1] = y[i,it+1];
      tp2[y[i,it+3],1] = y[i,it+2];
    end;
  if (tp3[y[i,it+3],1] < y[i,it+2]) then;
    do;
      tp3[y[i,it+3],1]=y[i,it+2];
      tp4[y[i,it+3],1]=y[i,it+1];
    end;
  end;
end;
tmp=(2:it)`||tp[2:it,1]||tp2[2:it,1];
print "*** Subtest Statistics are in Data File ***";
print "Mean Scale Alpha for All Possible Scales With and Without Item";
cn = {"Without" "With" "Gain"};
print res [rowname = nm colname = cn format = 8.4];
print "Maximum alpha (a_sub) as a Function of Number of Items";
cn = {"Items" "a_sub" "r_sub"};
print tmp [colname = cn format = 8.4];
print "Maximum Subscale-Full Scale Correlation as a Function of Number of
Items";
tmp = (2:it)`||tp3[2:it,1]||tp4[2:it,1];
print tmp [colname = cn format = 8.4];
print "Number of Cases";
print n;
cn = nm||"a_sub"||"r_sub"||"Items";
create subscale from y [colname = cn];
append from y;
quit;
proc sort data=subscale out=subscale ;
  BY descending a_sub ;
data;
  set subscale;
run;
%mend;

```