An introduction to computing and interpreting Cronbach Coefficient Alpha in SAS
Chong Ho Yu, Ph.D., Arizona State University, Tempe, AZ

ABSTRACT

In spite of the ease of computation of Cronbach Coefficient Alpha, its misconceptions and mis-applications are still widespread, such as the confusion of consistency and dimensionality, as well as the confusion of raw Alpha and standardized Alpha. To clarify these misconceptions, this paper will illustrate the computation of Cronbach Coefficient Alpha in a step-by-step manner, and also explain the meaning of each component of the SAS output.

INTRODUCTION

Reliability can be expressed in terms of stability, equivalence, and consistency. Consistency check, which is commonly expressed in the form of Cronbach Coefficient Alpha (Cronbach, 1951), is a popular method. Unlike test-retest for stability and alternate form for equivalence, only a single test is needed for estimating internal consistency. In spite of its ease of computation, misconceptions and mis-applications of Cronbach Coefficient Alpha are widespread. The following problems are frequently observed:

1. Assumptions of Cronbach Alpha are neglected by researchers and as a result over-estimation and under-estimation of reliability are not taken into consideration.

2. Some researchers believe that the standardized Alpha is superior to the raw Alpha because they believe standardization can normalize skewed data. This problem also reflects the confusion of covariance matrix with correlation matrix.

3. Additionally, some people throw out difficult or easy items based on the simple statistics of each item without taking the entire test into account.

4. Further, when a survey or test contains different latent dimensions, some researchers compute the overall Alpha only and jump to the wrong conclusion that the entire test or survey is poorly written.

5. On the other hand, when a high overall Alpha is obtained, many researchers assume a single dimension and do not further investigate whether the test carries subscales.

6. Several researchers use a pretest as the baseline or as a covariate. However, a low Alpha in the pretest may result from random guessing when the subjects have not been exposed to the treatment (e.g. training of the test content). Judging the reliability of the instrument based on the pretest scores is premature.

7. Last but not least, quite a few researchers adopt a validated instrument but skip computing Cronbach Coefficient Alpha with their sample. This practice makes subsequent meta-analysis of mean difference and Alpha impossible.

To clarify these misconceptions, this paper will illustrate the computation of Cronbach Coefficient Alpha in a step-by-step manner, and also explain the meaning of each component of the SAS output.

WHICH RELIABILITY INFORMATION SHOULD I USE?

One could compute Cronbach Coefficient Alpha, Kuder Richardson (KR) Formula, or Spilt-half Reliability Coefficient to examine internal consistency within a single test. Cronbach Alpha is recommended over the other two for the following reasons:

1. Cronbach Alpha can be used for both binary-type and large-scale data. On the other hand, KR can be applied to dichotomously scored data only.

2. Spilt-half can be viewed as a one-test equivalent to alternate form and test-retest, which use two tests. In spilt-half, you treat one single test as two tests by dividing the items into two subsets. Reliability is estimated by computing the correlation between the two subsets. The drawback is that the outcome is affected by how you group the items. Therefore, the reliability coefficient may vary from group to group. On the other hand, Cronbach Alpha is the mean of all possible spilt-half coefficients that are computed by the Rulon method (Crocker & Algina, 1986).

WHAT IS CRONBACH ALPHA?

Cronbach coefficient Alpha is a measure of squared correlation between observed scores and true scores. Put another way, reliability is measured in terms of the ratio of true score variance to observed score variance. The theory behind it is that the observed score is equal to the true score plus the measurement error ($Y = T + E$). For example, I know 80% of the material but my score is 85% because of guessing. In this case, my observed score is 85 while my true score is 80. The additional five points are due to the measurement error. It is assumed that a reliable test should minimize the measurement error so that the error is not highly correlated with the true score. On the other hand, the relationship between true score and observed score should be strong.

In addition, it is assumed that the mean of the measurement error should be zero. In other words, the error scores should be random and uncorrelated with each other. Failure of meeting this assumption may lead to an over-estimation of Cronbach Alpha though in practice this assumption cannot be fully met.

It is also assumed that items must be essentially tau equivalent, in which the true scores for any two items must be within a constant of each other for an examine. If this assumption for Cronbach Alpha is violated, Alpha may underestimate reliability. For this reason, it is generally agreed that Cronbach Alpha is a lower bound estimate of reliability because perfect essentially tau-equivalence is seldom achieved (Cortina, 1993). Using simulations, Zimmerman, and Zumbo (1993) found that the violations of these assumptions lead to substantive over-estimation and under-estimation of Cronbach Alpha. Researchers should be aware of these potential problems while applying Cronbach Alpha.

HOW TO COMPUTE CRONBACH ALPHA

The following is an example of SAS code to run Cronbach Alpha:

```
Data one;
     input post_em1-post_em5;
 cards;
 1 1 1 0 0 1
 1 0 1 1 1 0
 1 1 1 1 1 1
 0 0 0 1 1 1
 0 1 0 1 0 1
proc corr alpha nocorr nomiss;
var post_em1-post_em5; run;
```

In this example, the "nocorr" option suppresses the item correlation information. Although the correlation matrix can be used to examine whether particular items are negatively correlated with others, a more efficient way is to check the table entitled "if items are deleted…" This table tells you whether particular items are negatively correlated with the total and thus it is recommended to suppress the correlation matrix from the output. "If items are deleted…" will be explained in a later section.

It is important to include the "nomiss" option in the procedure statement. If the tester did not answer several questions, Cronbach Alpha will not be computed. In surveys, it is not unusual for respondents to skip questions that they don't want to answer. Also, if you use a scanning device to record responses, slight pencil marks may not be detected by the scanner. In both cases, you will have "holes" in your data set and Cronbach Alpha procedure will be halted. To prevent this problem from happening, the "nomiss" option tells SAS to ignore cases that have missing values.

However, in the preceding approach, even if the tester skips one question, the entire test record will be ignored by SAS. In a speeded test where testers may not be able to complete all items, the use of "nomiss" will lead to some loss of information. One way to overcome this problem is to set a criterion for a valid test response. Assume that 80 percent of test items must be answered in order to be included into the analysis. The following SAS code should be implemented:

```
Data one; infile "c:\data";
  input x1-x10;
  if nmiss(of x1-x10) > 2 then
delete;
  array x{I} x1-x10;
   do I=1 to 10;
   if X(I) =. then X(I) = 0;
proc corr alpha nocorr nomiss;
   var x1-x10; run;
```

In the preceding SAS code, if a record has more than two unanswered questions (80%), the record will be deleted. In the remaining records, the missing values will be replaced by a zero, and thus these records will be counted into the analysis.

It is acceptable to count missing responses of a test as wrong answers and assign a value of "zero" to them. But it is not appropriate to do so if the instrument is a survey such as an attitude scale. One of the popular approaches for dealing with missing data in surveys is the mean replacement method (Afifi & Elashoff, 1966), in which means are used to replace missing data. The SAS source code for the replacement is the same as the preceding one except the following line:

```
if X(I) = .
     then X(I) = mean(of x1-x10);
```

HOW TO INTERPRET THE SAS OUTPUT

Descriptive statistics

The mean output as shown in Figure 1 tells you how difficult the items are. Because in this case the answer is either right (1) or wrong (0), the mean is ranging from 0 to 1. 0.9 indicates that the question is fairly easy and thus 90% of the testers scored it. It is a common mistake that people look at each item individually and throw out the item that appears to be too difficult or too easy. Indeed, you should take the entire test into consideration. This point will be discussed later.

<u>Figure 1.</u> Simple statistics of Cronbach Coefficient Alpha's output

```
                   Simple Statistics

Variable      N       Mean    Std Dev       Sum    Minimum   Maximum

POST_EM1     93     0.9677    0.1776    90.0000        0     1.0000
POST_EM2     93     0.9677    0.1776    90.0000        0     1.0000
POST_EM3     93     0.9677    0.1776    90.0000        0     1.0000
POST_EM4     93     0.9140    0.2819    85.0000        0     1.0000
POST_EM5     93     0.9570    0.2040    89.0000        0     1.0000


                Correlation Analysis

            Cronbach Coefficient Alpha

       for RAW variables        : 0.305078
       for STANDARDIZED variables: 0.292091
```

Raw and standardized Alphas

As shown in Figure 1, Cronbach Alpha procedure returns two coefficients:

1. Raw: It is based upon item correlation. The stronger the items are inter-related, the more likely the test is consistent.

2. Standardized: It is based upon item covariance. Covariance is not a difficult concept. Variance is a measure of how a distribution of a single variable (item) spreads out. Covariance is simply a measure of the distributions of two variables. The higher the correlation coefficient is, the higher the covariance is.
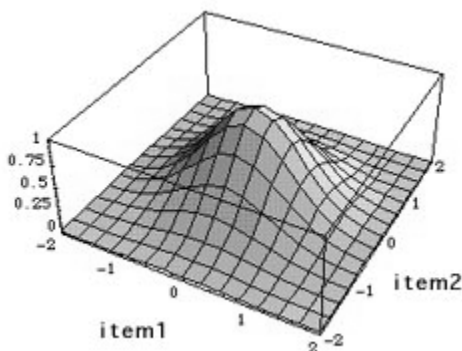
Some researchers mistakenly believe that the standardized Alpha is superior to the raw Alpha because

they thought that standardization normalizes skewed data. Actually standardization is a linear transformation, and thus it never normalizes data. Standardized Alpha is not superior to its raw counterpart. It is used when scales are comparable, because as mentioned before, variance and covariance are taken into account for computation.

Variance and covariance

The concepts of variance and covariance are better illustrated graphically. In one variable, the distribution is a bell-curve if it is normal. In a two-variable case, the normal distribution appears to be a mountain as shown in Figure 2. In this example, both item1 and item2 has a mean of zero because the computation of covariance uses standardized scores (z-score). From the shape of the "mountain," we can tell whether the response patterns of testers to item1 and item 2 are consistent. If the mountain peak is at or near zero and the slopes of all directions spread out evenly, we can conclude that the response pattern of these items is consistent.

Figure 2. Covariance

However, in order to determine whether the response pattern to the **entire** test is consistent, we must go beyond just viewing one pair. Cronbach Alpha computation examines the covariance matrix (all possible pairs) to draw a conclusion. It is noteworthy that not all the information on the matrix is usable. For example, the pairs of the item itself such as (item1, item1) can be omitted. Also, the order of the pair doesn't matter i.e. the covariance of pair (item1, item2) is the same as that of (item2, item1) (see Table 1).

Table 1. Covariance matrix table

	Item1	Item2	Item3	Item4	Item5
item1		Covariance	Covariance	Covariance	Covariance
item2			Covariance	Covariance	Covariance
item3				Covariance	Covariance
item4					Covariance
item5					

Consistency and dimensionality

Generally speaking, the higher the Alpha is, the more reliable the test is. There isn't a commonly agreed cut-off. Usually 0.7 and above is acceptable (Nunnally, 1978). It is a common misconception that if the Alpha is low, it must be a bad test. Actually your test may measure several latent attributes/dimensions rather than one and thus the Cronbach Alpha is deflated. For example, it is expected that the scores of GRE-Verbal, GRE-Quantitative, and GRE-Analytical may not be highly correlated because they evaluate different types of knowledge.
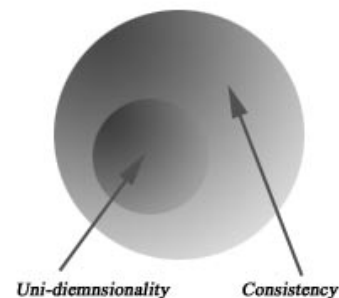
If your test is not internally consistent, you may want to perform factor analysis or principal component analysis to combine items into a few factors/components. You may also drop the items that affect the overall consistency, which will be discussed in a later section. If you know what the subscales are, you should compute the Cronbach Alpha for each subscale.

On the other hand, when the Cronbach Alpha is larger than .70, researchers may go to another extreme:. Cortina (1993) observed that many people accept a high Alpha as adequate and thus seldom make further scale modifications. Cortina explicitly criticized that this is an improper usage of statistics.

It is important to note that a low overall Alpha may indicate the existence of latent constructs, but a high overall Alpha does not necessarily imply the absence of multiple latent dimensions. One may argue that when a high Cronbach Alpha indicates a high degree of internal consistency, the test or the survey must be uni-dimensional. Thus, there is no need to further investigate its subscales.

This is a common misconception. Actually **consistency** and **dimensionality** must be assessed separately. The relationship between consistency and uni-dimensionality is illustrated in Figure 3. Uni-dimensionality is a subset of consistency. If a test is uni-dimensional, then it will show internal consistency. But if a test is internally consistent, it does not necessarily entail one construct (Gardner, 1995; 1996). This logic works like this: If I am a man, I must be a human. But if I am a human, I may not be a man (could be a woman). The logical fallacy that "if A then B; if B then A" is termed as "affirming the consequent" (Kelly, 1998). This fallacy often happens in the mis-interpretation of Cronbach Alpha.
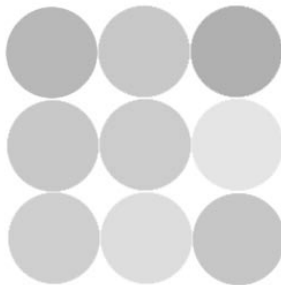
Figure 3. Uni-dimensionality and consistency

Uni-diemnsionality   Consistency

Gardner (1995) used a nine-item scale as an example to explain why a high Alpha does not necessarily indicate one dimension: Cronbach Alpha is a measure of common variance shared by test items. The Cronbach Alpha could be high when each test item shares variance with at least some other items; it does not have to share variance with all other items.

Different possible scenarios are illustrated in Figure 4a-c. As mentioned before, Cronbach Alpha can be calculated based upon item correlation. When the correlation coefficient is squared, it becomes the strength of determination, which indicates variance explained. Variance explained is often visualized by sets. When two sets are intersected, the overlapped portion denotes common variance. The non-overlapped portion indicates independent information.

Figure 4a. Inconsistent and no uni-dimension

In Figure 4a, all nine sets have no overlapped area, and thus all nine items share no common variance. They are neither internally consistent nor uni-dimensional. In this situation, interpreting a low Alpha as the absence of uni-dimensionality is correct.

Figure 4b. Consistent and uni-dimensional

The scenario presented in Figure 4b is exactly opposite to that in Figure 4c. It shows the presence of a high degree of internal consistency and uni-dimensionality because all items share common variance with each other. Interpreting a high alpha as an indication of the presence of one single construct could be accepted.

Figure 4c. Consistent but not uni-dimensional.

Unlike the above two situations, the last scenario is misleading. In Figure 4c, several items share variance with some others. In other words, subscales exist inside the instrument even though the Alpha is high and the instrument is internally consistent. Interpreting a high Alpha as a sign of uni-dimensionality is problematic.

Since consistency and dimensionality should be examined by different procedures, it is recommended that PROC FACTOR is used in addition to PROC CORR ALPHA.

More on variance and dispersion

Since Cronbach Alpha takes variance into account, it is important to interpret the data in the context of dispersion. For example, when you compare the mean scores in the following two tables, you can find that both pre-test and post-test responses are consistent, respectively. However, the Alpha of post-test is only .30 (raw) and .29 (standardized) while the Alpha of pre-test is as high as .60 (raw and standardized). It is because the standard deviation (SD) of the post-test ranges from .17 to .28, but the SD of the pre-test is more consistent (.42-.48) (see Figure 5).

Figure 5. Simple statistics

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
POST_EM1	93	0.9677	0.1776	90.0000	0	1.0000
POST_EM2	93	0.9677	0.1776	90.0000	0	1.0000
POST_EM3	93	0.9677	0.1776	90.0000	0	1.0000
POST_EM4	93	0.9140	0.2819	85.0000	0	1.0000
POST_EM5	93	0.9570	0.2040	89.0000	0	1.0000

Correlation Analysis

Cronbach Coefficient Alpha

for RAW variables   : 0.305078
for STANDARDIZED variables: 0.292081

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
PRE_EM1	152	0.7171	0.4519	109.0000	0	1.0000
PRE_EM2	152	0.6118	0.4889	93.0000	0	1.0000
PRE_EM3	152	0.6382	0.4821	97.0000	0	1.0000
PRE_EM4	152	0.7632	0.4266	116.0000	0	1.0000
PRE_EM5	152	0.7105	0.4550	108.0000	0	1.0000

Correlation Analysis

Cronbach Coefficient Alpha

for RAW variables   : 0.608905
for STANDARDIZED variables: 0.604723

If the item is deleted...

As mentioned before, a good analysis of test items should take the whole test into consideration. The following table tells you how each item is correlated with the entire test and what the Alpha would be if that variable were deleted. For example, the first line shows you the correlation coefficient between post-test item 1 and the composite score of post-test item1-item5. The first item is negatively correlated with the total score. If it is deleted, the Alpha will be improved to .41 (raw) or .42 (standardized). Question 5 has the strongest relationship with the entire test. If this item is removed, the Alpha will be dropped to -.01 (raw) or .04 (standardized). This approach helps you to spot the bad apple and retain the good one (see Figure 6).

Figure 6. If the item is deleted…

| Deleted Variable | Raw Variables | | Std. Variables | |
	Correlation with Total	Alpha	Correlation with Total	Alpha
POST_EM1	-0.066825	0.411696	-0.066739	0.421235
POST_EM2	0.163909	0.232006	0.226237	0.163357
POST_EM3	0.051667	0.329931	0.090600	0.299428
POST_EM4	0.176504	0.232205	0.151551	0.235170
POST_EM5	0.419094	-0.010825	0.336565	0.049066

Once again, variance plays a vital role in Cronbach Alpha calculation. Without variance there will be no sensible result. The following questions are from another post-test. Every body scored Question 3 and 4 (1.00) but missed Question 4 (0.00). Because there is no variance, standardized Cronbach Alpha, which is based on covariance matrix, cannot be computed at all (see Figure 7).

Figure 7. Simple statistics when there is no variance

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
POST_LS1	63	0.7778	0.4191	49.0000	0	1.0000
POST_LS2	63	0.0476	0.2147	3.0000	0	1.0000
POST_LS3	63	1.0000	0	63.0000	1.0000	1.0000
POST_LS4	63	0	0	0	0	0
POST_LS5	63	1.0000	0	63.0000	1.0000	1.0000

Correlation Analysis

Cronbach Coefficient Alpha

for RAW variables   : -.300860
for STANDARDIZED variables: .

| Deleted Variable | Raw Variables | | Std. Variables | |
	Correlation with Total	Alpha	Correlation with Total	Alpha
POST_LS1	-0.239046	0.000000	.	.
POST_LS2	-0.239046	0.000000	.	.
POST_LS3	.	-0.320917	.	.
POST_LS4	.	-0.320917	.	.
POST_LS5	.	-0.320917	.	.

"I DON'T KNOW" OPTION IN THE PRETEST

In the pretest where subjects are not exposed to the treatment and thus are unfamiliar with the subject matter, a low reliability caused by random guessing is expected. One easy way to overcome this problem is to include "I don't know" as an option in multiple choices. In experimental settings where students' responses would not affect their final grades, the experimenter should explicitly instruct students to choose "I don't know" instead of making a guess if they really don't know the answer. Low reliability is a signal of high measurement error, which reflects a gap between what students actually know and what scores they receive. The choice "I don't know" can help in closing this gap.

MORE THAN ONE SHOT

Another common misconception of Cronbach Alpha is that if someone adopts a validated instrument, he/she does not need to check the reliability and validity with his/her own data. Imagine this: When I buy a drug that has been approved by FDA and my friend asks me whether it heals me, I tell him, "I am taking a drug approved by FDA and therefore I don't need to know whether it works for me or not!" A responsible evaluator should still check the instrument's reliability and validity with his/her own subjects and make any modifications if necessary.

Further, when the researchers report the reliability information of their own data, it helps other subsequent researchers to conduct meta-analyses. Meta-analysis is a research tool for synthesizing results obtained from previous research. Effect size, which is expressed in terms of standardized mean difference across groups, is used in meta-analysis. However, the effect size may be affected by the measurement error. To counteract this problem, Hunter and Schmidt (1990) adjust the effect size for measurement error by dividing the effect size by the square root of the reliability coefficient (r) of the dependent variable. The formula is shown in the following:

$$\text{Measurement error correction} = \frac{\text{Effect size}}{\text{Square root of } r}$$

Unfortunately, the absence of reliability information in many studies makes this type of meta-analysis impossible. Further, different studies yield not only different effect sizes, but also different Cronbach coefficient Alphas. Besides comparing mean differences, meta-analysis could also be employed to examine whether a particular instrument is consistently consistent. There are several approaches to accomplish this goal. One way is to transform Cronbach Alphas of the same instrument reported in past research via Fisher's Z transformation, and then to compute the Q statistics (Behrens, 1997). Another way is UX test developed by Feldt, Woodruff, and Salih (1987). UX test can be used for comparing Alphas obtained from different independent samples, as well as from the same sample. Discussion of Fisher's Z tranformation, Q statistics, and UX test is out of the scope of this paper. The important point here is that researchers should consider going beyond the Cronbach Alpha reported in one particular study and look for a farther inference.

CONCLUSION

Although Cronbach Alpha is fairly easy to compute, its application requires conceptual understanding such as true score, observed score, measurement error, variance, covariance matrix, consistency, and dimensionality. It is

hoped that this paper could clarify common misconceptions of Cronbach Alpha and improve the effective use of SAS procedures.

REFERENCES

Afifi, A. A., & Elashoff, R. M. (1966). Missing observations in multivariate statistics. Part I. Review of the literature. Journal of the American Statistical Association, 61, 595-604.

Behrens, J. (1997). Does the White Racial Identity Attitude Scale measure racial identity. Journal of Counseling Psychology, 44, 3-12.

Cortina, J. M. (1993). What is Coefficient Alpha? An examination of theory and applications. Journal of Applied psychology, 78, 98-104.

Crocker, L. M., & Algina, J. (1986). Introduction to classical and modern test theory. New York : Holt, Rinehart, and Winston.

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

Feldt, L.; Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for Coefficient Alpha. Applied Psychological Measurement, 11, 93-103.

Gardner, P. L. (1995). Measuring attitudes to science: Unidimensionality and internal consistency revisited. Research in Science Education, 25, 283-9.

Gardner, P. L. (1996). The dimensionality of attitude scales: A widely misunderstood idea. International Journal of Science Education, 18, 913-9.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage.

Kelley, D. (1998). The art of reasoning (3rd ed.). New York: W. W. Norton & Company.

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

Zimmerman, D. W., & Zumbo, B. D. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions Educational & Psychological Measurement, 53, 33-50.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:
Chong Ho Yu, Ph.D.
Arizona State University
Psychology in Education
302 Payne Hall
Tempe AZ 85287-0611
(480)965-3475
Email: alex@asu.edu
Web: http://seamonkey.ed.asu.edu/~alex/