

02b: Internal Consistency and Item Analysis

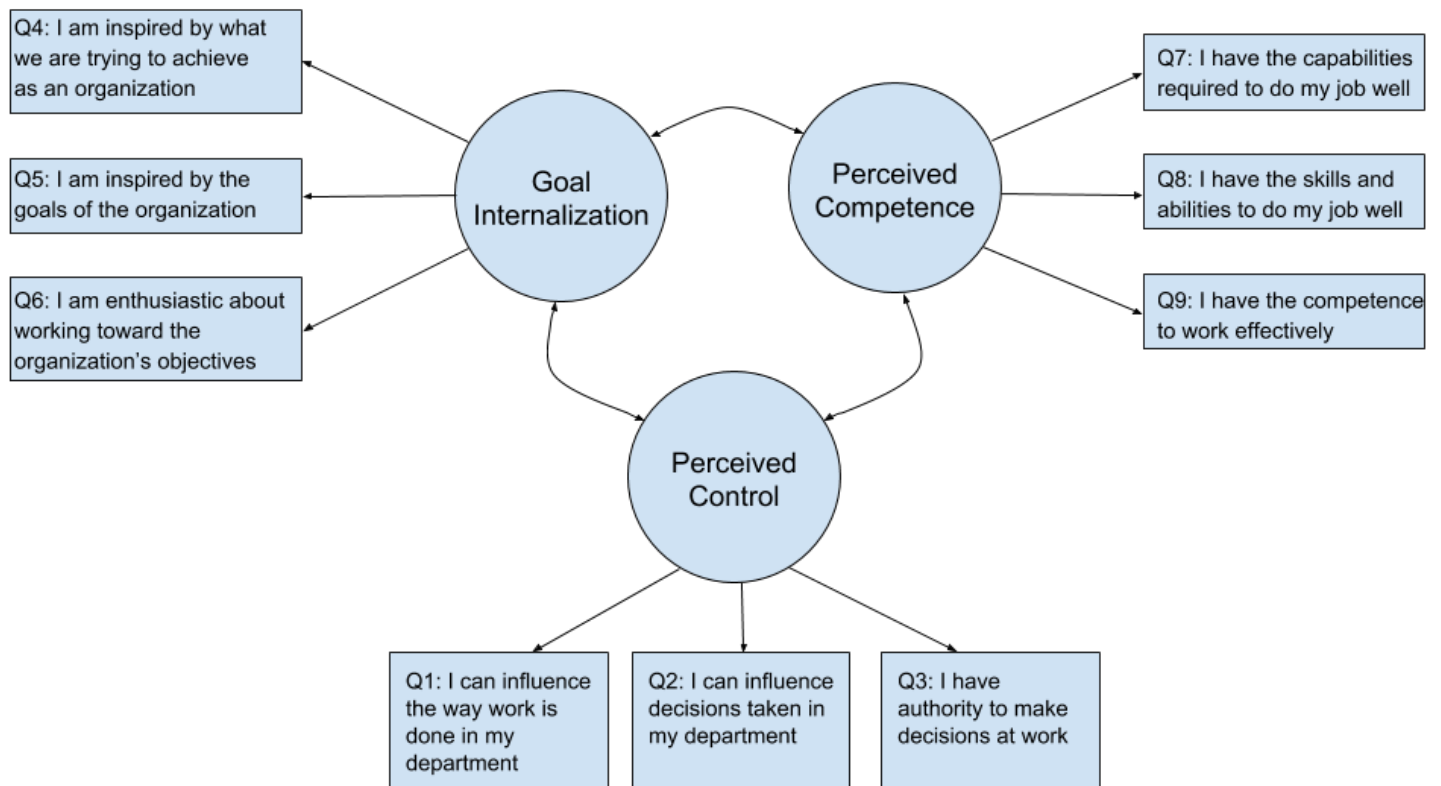
Topics

1. Internal Consistency Logic
2. Employment Thoughts Data
3. Cronbach's Alpha and Item Analysis
4. SPSS: Alpha and Item Analysis Statistics for Perceived Control
5. Comparison of Results with Menon
6. Dimensionality
7. Reverse Scoring and Composite Scores
8. Published Examples

1. Internal Consistency Logic

Indicators of latent variables, or constructs, should demonstrate internally consistent responses. This means responses on one indicator should correlate positively, and moderately or strongly, with responses to another indicator unless one indicator is reverse coded, then the correlation should be negative and of moderate or strong strength. The logic here suggests that a respondent's answer to one item should predict how that respondent is likely to answer similar items designed to measure the same construct.

Figure 1: Menon's Measurement Model for Employee Empowerment



Above, in Figure 1, is a model employee empowerment proposed by Menon (2001) representing three latent variables (or constructs): perceived control, goal internalization, and perceived competency. Each latent variable is measure by three indicators. If the three indicators are internally consistent, then how one responds to Q1 (I can influence the way work is done in my department) should be similar to how one responds to Q2 (I can influence decisions taken in my department) and to Q3 (I have authority to make decisions at work).

When indicators, or items, are designed to measure the same construct, it is possible to logically identify items that are inconsistent by

- assuming an extreme position (e.g., strongly for or against; extremely happy or sad; do something very frequently or infrequently, etc.) on the measured construct;
- pretend you are a respondent and respond from an extreme position to each item measuring that construct; then
- examine whether responses are consistent across all indicators, i.e., do they measure same thing, same dimension or domain? Keep in mind that reversed coded items should have a negative correlation, so responses should be reversed to other items.

Reversed Coded Items Explained

Reversed items have been mentioned several times in these notes and in others. A reversed item is illustrated below with three items designed to measure life satisfaction. Note that item 3 is reversed.

Assume that you are very satisfied with your life. How would you respond to items 1, 2, and 3? Likely for items 1 and 2 your response would be 4, Strongly Agree. However, for item 3, your response would likely be 1, Strongly Disagree. Since item 3 is reversed coded, it should elicit responses that are placed on the opposite end of the response scale from other items.

| Item | Strongly Disagree | Mildly Disagree | Mildly Agree | Strongly Agree |
|--|-------------------|-----------------|--------------|----------------|
| 1. In most ways my life is close to idea. | 1 | 2 | 3 | 4 |
| 2. I am satisfied with my life. | 1 | 2 | 3 | 4 |
| 3. The conditions of my life are very bad. | 1 | 2 | 3 | 4 |

Reverse coded items (or negative valence items) are used to break what is known as response set (mindlessly answering items or repeating answers without serious consideration to what the item queries), acquiescence, and boredom. It is unclear whether reversed coded items perform this task well, or whether reversed coded items enhance or reduce reliability and validity – the research is mixed, but there are those who favor and oppose use of reverse scored items. In general, most seem to recommend against the practice of using reverse coded items.

Example: Consider Menon's measure of perceived control. Assume that you have no power to control your work environment or influence others. For this situation, how might you respond to each of these three items?

| Item | Strongly Disagree | Moderately Disagree | Mildly Disagree | Mildly Agree | Moderately Agree | Strongly Agree |
|--|-------------------|---------------------|-----------------|--------------|------------------|----------------|
| Q1. I can influence the way work is done in my department. | 1 | 2 | 3 | 4 | 5 | 6 |
| Q2. I can influence decisions taken in my department. | 1 | 2 | 3 | 4 | 5 | 6 |
| Q3. I have authority to make decisions at work. | 1 | 2 | 3 | 4 | 5 | 6 |

Someone with little power would likely select a score of 1 for each of the three items. This demonstrates logical consistency among the three items measuring perceived control.

As a second example, consider the scale below designed to assess Course Satisfaction. Assume you are answering this scale for EDUR 8331. Are there any items that appear to be internally inconsistent with a measure of one's Course Satisfaction with EDUR 8331?

| Item | Not at all | To a small degree | Moderately | Mostly | Very Much |
|--|------------|-------------------|------------|--------|-----------|
| Q1. Do you like this course? | 1 | 2 | 3 | 4 | 5 |
| Q2. Do you like the way this course is taught? | 1 | 2 | 3 | 4 | 5 |
| Q3. Do you like the way your other courses are taught? | 1 | 2 | 3 | 4 | 5 |

Likely Q3 will not be internally consistency with Q1 and Q2 because Q1 and Q2 focus on EDUR 8331 while Q3 focuses upon other courses. You may strongly dislike EDUR 8331, but strongly favor your other courses, or vice versa. This means Q3 may produce random scores relative to items Q1 an Q2, hence be inconsistent with responses to Q1 and Q2.

2. Employment Thoughts Data

Recall the employment data in the test-retest notes with construct indicators selected from Menon (2001) and displayed above in Figure 1. These 9 items were designed to measure three employment related constructs. Responses to each item is scaled from Strongly Disagree (1) to Strongly Agree (6). The nine items are repeated below.

Perceived Control

Q1: I can influence the way work is done in my department

Q2: I can influence decisions taken in my department

Q3: I have the authority to make decisions at work

Goal Internalization

Q4: I am inspired by what we are trying to achieve as an organization

Q5: I am inspired by the goals of the organization

Q6: I am enthusiastic about working toward the organization's objectives

Perceived Competence

Q7: I have the capabilities required to do my job well

Q8: I have the skills and abilities to do my job well

Q9: I have the competence to work effectively

None of these are reverse coded, so composite variables can be computed directly by taking the mean across the three indicators for each construct.

SPSS data file link is provided below.

<http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-content/06-reliability/06-EDUR9131-EmploymentThoughts-Merged.sav>

Recall from the test-retest notes that items with `_1` are from the first administration, and those with `_2` are from the second administration.

3. Cronbach's Alpha and Item Analysis

Cronbach's alpha is one measure of internal consistency and perhaps the most used and reported. As explained in the readings, Cronbach's alpha is a general form of internal consistency and other formulas (e.g., KR 20, split-half) are either subsumed under alpha or produce similar results to alpha. For example, the KR-20 formula holds only for dichotomous items (scored 0, 1) while alpha works for 0,1 coding or ordinal, interval, or ratio data; similarly, alpha is the average of all possible split half estimates (Cortina, 1993). Formulas for calculating alpha won't be reported here; instead, we will focus on using software (i.e., SPSS) to obtain Cronbach's alpha.

Cronbach's alpha interpretation:

"Cronbach's alpha can be interpreted as the percent of variance the observed scale would explain in the hypothetical true scale composed of all possible items in the universe. Alternatively, it can be interpreted as the correlation of the observed scale with all possible other scales measuring the same thing and using the same number of items." Garson (2016) p.43

Alpha also represents the proportion of shared variance among items. The higher alpha, the more variance shared among items. If items tend to measure the same construct, they share common variance in questionnaire responses about that construct; if they measure different things, then they share less common variance and alpha will be lower.

Cronbach's alpha ranges from 0.00 to 1.00 (although it is possible to obtain negative values for alpha). Although some argue that guidelines for interpreting alpha can be misleading and harmful to research practice, many follow the accepted standard that alpha value of .70 and above for research, and .90 and above for professionally developed scales and tests (e.g., standardize tests used in schools).

Item analysis refers to the assessment of the contribution of each item to the measurement of the latent variable. This includes examining the logical or theoretical fit of the item, and the item's contribution to both reliability and validity. This analysis applies to both tests and scales, however, later we will cover item analysis from a testing perspective. While the process is the same, there are differences that make item analysis more practical for those in education settings.

Below is outlined steps one takes to assess internal consistency for a latent variable and, also, item analysis of reliability contribution for each item (i.e. examination of how each item contributes to internal consistency, i.e., do scores from the item examined correlate as expected with other items designed to measure the same latent variable).

Steps in assessment of internal consistency:

- Check for outliers and missing data (i.e., scatterplots, frequency displays, etc.)
- Examine correlations among indicators
 - correlations should be moderate or strong
 - negative correlations signal scores may need to be reverse scored
 - weak correlations may indicate items that do not function well with other indicators of the construct, and could be due to sample data or theoretical issues with wording of items
- Calculate Cronbach's alpha
- SPSS reports two Cronbach's alpha values if correlations are requested

- Cronbach's alpha = calculated on raw data items that usually have different variances, unequal variances, and alpha is based upon covariances, not correlations
- Cronbach's alpha based on standardized items = estimate of reliability if all items have equal variances; called Spearman-Brown stepped-up reliability coefficient; value of alpha obtained if all variables are standardized to have equal variances
- Since equal variances among items rarely reflects real-world situations, I recommend using the unstandardized alpha
- Item Analysis Information
 - Logical and theoretical assessment: How well does this item contribute to measurement of the construct/dimension relative to other items?
 - Examine correlations among items: correlations should be moderate to strong; an example with Perceived Control items is discussed below
 - Corrected item-total correlation: correlation of scores from one item with total score obtained from all other items summed, this correlation should be moderate to strong; an example with Perceived Control items is discussed below
 - Alpha if item deleted: this shows how Cronbach's alpha would change if a given items were removed from the scale and alpha recalculated for the remaining items, if alpha is lowered that is an indication that the removed item contributes to internal consistency and should not be removed if possible; an example with Perceived Control items is discussed below
 - Important Considerations
 - Avoid making mechanical decisions about retaining or removing items (e.g., don't remove items only because they reduce Cronbach's alpha or have low correlations with other items; think critically about the total contribution of each item – theoretic important, reliability contribution, and validity contribution)
 - Consider carefully theoretical contribution of item to measurement of targeted construct
 - Goal is not just to maximize internal consistency (e.g. alpha) because this leads to narrow content items which may limit validity of construct measured
 - Goal is to create scale that maximizes validity with acceptable, good, or excellent reliability – better to have items that offer wider assessment of construct than narrow items that omit important components of a construct
- Repeat above with items for
 - Goal Internalization
 - Perceived Competence

4. SPSS: Alpha and Item Analysis Statistics for Perceived Control

Using the Menon data linked above, calculate internal consistency with Cronbach's alpha for each of the three constructs and compare the values obtained against those reported by Menon on page 171 of his article. I report below the Cronbach's alphas Menon provided.

| | | Menon Alpha | Course Data Alpha |
|----------------------|---|-------------|-------------------|
| P. Control | = | .83 | |
| Goal Internalization | = | .88 | |
| P. Competence | = | .80 | |

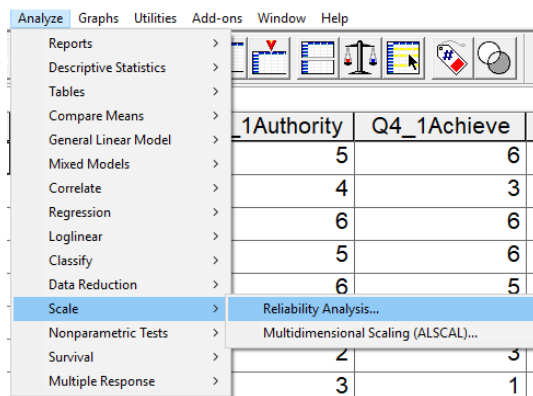
Menon's (2001) article is linked below for those interested.

<http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-assignments/02-Menon-ST-2001.pdf>

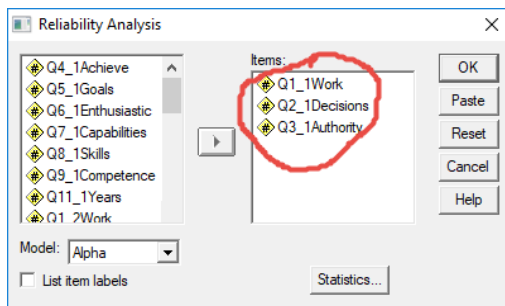
SPSS Data Entry (when data file is downloaded and opened, data should look as shown below)

| | Q1_1Work | Q2_1Decisions | Q3_1Authority |
|----|----------|---------------|---------------|
| 1 | 5 | 5 | 5 |
| 2 | 4 | 4 | 4 |
| 3 | 6 | 6 | 6 |
| 4 | 5 | 5 | 5 |
| 5 | 6 | 6 | 6 |
| 6 | 3 | 1 | 2 |
| 7 | 5 | 4 | 2 |
| 8 | 4 | 4 | 3 |
| 9 | 5 | 5 | 5 |
| 10 | 5 | 6 | 5 |
| 11 | 6 | 6 | 6 |
| 12 | 6 | 6 | 6 |

SPSS Commands to obtain Cronbach's alpha: Analyze → Scale → Reliability Analysis



Find Cronbach's alpha for Perceived Control, measured by items Q1, Q2, and Q3. We use the first administration data, thus Q1_1, Q2_1, and Q3_1 as indicated in the screenshot below. Move these items from the variable box to the Items box.



Once the variables, or items, are identified in the Items box, next select which statistics to calculate. Note the options shown below: descriptive for item, scale, and scale if item deleted; and inter-item correlations.

SPSS Results

Overall alpha – use Cronbach’s alpha; recall that alpha based on standardized items is usually unrealistic in real-world practice since it assumes all items have the same variance and SD. For Perceived Control, $\alpha = .923$.

Reliability Statistics

| | Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|--|------------------|--|------------|
| | .923 | .942 | 3 |

Item statistics and correlations are shown below. For each item the M and SD are reported.

Item Statistics

| | Mean | Std. Deviation | N |
|---------------|------|----------------|----|
| Q1_1Work | 5.00 | .970 | 18 |
| Q2_1Decisions | 4.94 | 1.305 | 18 |
| Q3_1Authority | 4.61 | 1.501 | 18 |

Inter-Item Correlation Matrix

| | Q1_1Work | Q2_1Decisions | Q3_1Authority |
|---------------|----------|---------------|---------------|
| Q1_1Work | 1.000 | .883 | .849 |
| Q2_1Decisions | .883 | 1.000 | .799 |
| Q3_1Authority | .849 | .799 | 1.000 |

The covariance matrix is calculated and used in the analysis.

The above correlation matrix shows that all inter-item correlations are strong, with the weakest being above .79. This suggests all items correlate very well together, and good condition for strong internal consistency.

Below is the Item-total Statistics table.

Item-Total Statistics

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Squared Multiple Correlation | Cronbach's Alpha if Item Deleted |
|---------------|----------------------------|--------------------------------|----------------------------------|------------------------------|----------------------------------|
| Q1_1Work | 9.56 | 7.085 | .911 | .836 | .884 |
| Q2_1Decisions | 9.61 | 5.663 | .864 | .789 | .872 |
| Q3_1Authority | 9.94 | 4.879 | .845 | .732 | .916 |

The table above contains five columns of statistics. Each are explained below.

1. Scale Mean if Item Deleted: One way to construct the Perceived Control variable for each respondent is to take the scale total across the three items. For example, if a respondent answered with a 6 (Strongly Agree) for all three items, the total score for that respondent would be $6 + 6 + 6 = 18$. The Scale Mean if Item Deleted shows the sum of the remaining two items. For example, if item Q1 were deleted, then the total score for they example respondent would then be $6 + 6 = 12$. This column helps show whether the overall mean across respondents remains about the same if each item is deleted one at a time. If the means are similar, this suggests the items are all contributing about the same to the total score.

2. Scale Variance if Item Deleted: Same as above, but with variance rather than total score calculated. This column helps to show which items contribute to variability in the total score. In general, more variance means more discrimination by items, which is good since it helps measure the construct with a range of scores rather than all scores appearing to be similar and failing to show differences among respondents. With the Perceived Competence we see variance is greatest if Q1 is deleted and least when Q3 is deleted.

3. Corrected Item-total Correlation: This column presents the correlation between each item and the total score for the remaining item. For example, for Q1 the corrected item-total correlation is .911; this means Q1 correlates .911 with the total score of items Q2 and Q3 combined.

Below is illustrated the corrected item-total correlation for Q1.

a. First, sum the items that will be totaled, Q2 and Q3. The SPSS screenshot below shows descriptive statistics for the sum of Q2 and Q3. Note that the statistics match those reported by SPSS above in the Item-Total Statistics table when Q1 is deleted (i.e., $M = 9.56$ or 9.555 and $Variance = 7.085$).

Questions 2 and 3 added (summed)

| Q2_and_3_summed | | |
|-----------------|---------|---------|
| N | Valid | 18 |
| | Missing | 0 |
| Mean | | 9.5556 |
| Median | | 10.0000 |
| Std. Deviation | | 2.66176 |
| Variance | | 7.085 |
| Minimum | | 3.00 |
| Maximum | | 12.00 |

b. Next, correlate Q1 with the sum of Q2 and Q3. This correlation is shown below. Note $r = .911$, the same value reported by SPSS in the Item-Total Statistics table.

Correlation between Question 1 and sum of Questions 2 and 3
Correlations

| | | Q1_1Work | Q2_and_3_summed |
|-----------------|---------------------|----------|-----------------|
| Q1_1Work | Pearson Correlation | 1 | .911** |
| | Sig. (2-tailed) | | .000 |
| | N | 18 | 18 |
| Q2_and_3_summed | Pearson Correlation | .911** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 18 | 18 |

** . Correlation is significant at the 0.01 level (2-tailed).

4. Squared Multiple Correlation: How well the other items predict scores on the targeted item. For Q1 the squared multiple correlation is $R^2 = .836$. This means items Q2 and Q3 can predict about 83.6% of the variance in Q1. The squared multiple correlation ranges from 0.00 to 1.00. The closer to 1.00, the more variance predicted. This is good because it tells us the items are highly correlated. A low value indicates low relation and therefore potential problems for that item because it does not appear to be internally consistent with the other items. The squared multiple correlation, R^2 , is obtained using the regression equation $Q1 = b_0 + b_1Q2 + b_2Q3 + e$.

5. Cronbach's Alpha if Item Deleted: This last column reports the recalculated Cronbach's alpha if the given item is deleted from the scale. For Q1 alpha would drop from the total scale alpha of .923 to .884 if Q1 were not included in the scale. This indicates that inclusion of Q1 contributes to increase internal consistency and an enhanced alpha value. If removing an item increases alpha, then that item detracts from internal consistency for the overall scale. If the item is poorly worded or fits poorly with the latent variable for which it was designed to measure, it should be considered for removal from the scale.

Below is an SPSS screenshot showing the new Cronbach alpha estimate for Q2 and Q3 only. Note that it provides the same alpha as reported in the Cronbach's Alpha if Item Deleted column.

Reliability Statistics

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|------------------|--|------------|
| .884 | .889 | 2 |

5. Comparison of Results with Menon

Note that estimates of reliability are sample specific so one should always check reliability for each sample. It can be useful to know how well an instrument behaves across samples. How do the results for the data linked above compare with results reported by Menon (2001)? Use results from the 1st administration of the employment thoughts questionnaire and find Cronbach's alpha for Goal Internalization and Perceived Competence. Answers are provided below; run the analyses and compare your answers to those posted below.

EDUR Course Data Alpha Estimates

| | | Menon Alpha | EDUR 9131 Alpha |
|----------------------|---|----------------|--------------------|
| P. Control | = | .83 | .923 |
| Goal Internalization | = | .88 | .986 |
| P. Competence | = | .80 | .947 |

Results show that for the EDUR students sampled the alpha estimates are all consistently higher than those reported by Menon. One possible explanation is that EDUR students who completed this questionnaire, primarily doctoral students in Ed. Leadership, are a more homogeneous group of participants than those sampled by Menon, so they may tend to respond similarly to each of the employment scale items.

6. Dimensionality

Dimensionality refers to the number of dimensions measured by a scale. For example, the employee empowerment scale developed by Menon contains three dimensions: perceived control, goal internalization, and perceived competence.

Sometimes scales with multiple dimensions are developed to be combined. With test anxiety, for example, there are at least two dimensions, physiological and emotionality, but scales designed to measure test anxiety combine these to form an overall measure, or scale, of test anxiety. For this type of scale Cronbach's alpha would be suitable for all test anxiety items, that is, it would be suitable to assess internal consistency for both physiological and emotionality dimensions.

With other scales dimensions are not designed to be combined; each dimension should be measured and analyzed separately. Menon's employee empowerment is such a scale. Each of the three dimensions should be calculated and analyzed independently. With this type of scale Cronbach's alpha should be calculated separately for each dimension. It would be a mistake to combine these three with one measure of alpha because the three scales measure independent, but related, constructs, so one would not expect all items to correlation strongly or be internally consistent.

With dimensionality in mind, below are some points to consider with Cronbach's alpha.

- Cronbach's alpha is not a measure of dimensionality; it does not assess the internal structure of a scale. It cannot be used to learn whether a scale contains one or more dimensions. For internal structure, use factor analysis instead.
- Cronbach's alpha should not be used as an overall internal consistency measure on instruments with diverse constructs:
 - calculate alpha only for uni-dimensional constructs (constructs with one dimension)
 - if more than one construct present, calculate alpha separately for each
 - alpha is not appropriate for items not designed to form a single construct, e.g., inappropriate to calculate alpha for sex, race, and age combined.
- What happens if we calculate overall alpha on the nine items of Employment Thoughts data?
 - Might be okay if those three dimensions are part of an overall measure of employment empowerment otherwise overall alpha should be used for each construct separately
 - Likely lower alpha despite alpha becoming larger with more items due to different nature of constructs and constructs may not be strongly related

Below is part of a post I made to a discussion group explaining why Cronbach's alpha cannot be used as an assessment for dimensionality, or, more specifically, uni-dimensionality.

Cronbach's alpha is not designed to measure internal structure (think in terms of factor analysis here), but can provide a measure of internal consistency (think in terms of mean inter-item correlations here), although that appears to be questionable too.

As noted by Bruce above alpha is a function of covariances (and correlations), and it is also a function of number of items. Here is a formula for Cronbach's alpha in terms of mean inter-item correlations ($m[r]$) and the number of items (k):

$$\alpha = (k * m[r]) / (1 + (k - 1) * m[r])$$

where

k = number of items on instrument used to calculate alpha,

$m[r]$ = mean correlation among the k items.

Given this formula, the following two scenarios are possible:

1. Researcher has instrument with 4 items designed to measure the same construct, so there should be one factor here. The mean correlation among items is $m[r] = .5862$. Using the formula above:

$$\alpha = (k * m[r]) / (1 + (k - 1) * m[r])$$

$$\alpha = (4 * .5862) / (1 + (4 - 1) * .5862) \approx .85$$

2. Researcher has an instrument designed to measure 4 unrelated or weakly related constructs. Factor analysis reveals that the internal structure to contain four distinct factors. There is a total of 50 items on this instrument (10 items for factor A, 8 items for factor B, 17 items for factor C, and 15 items for factor D). The mean correlation among all 50 items is $m[r] = .1018$. If one erroneously applies the alpha reliability formula to these 50 items, the result would be:

$$\alpha = (k * m[r]) / (1 + (k - 1) * m[r])$$

$$\alpha = (50 * .1018) / (1 + (50 - 1) * .1018) \approx .85$$

Note that Cronbach's alpha is the same, within rounding error, in both situations, yet the internal structure is very different in both cases. These two examples demonstrate that Cronbach's alpha is not designed to reveal internal structure of items. Better to use EFA or CFA to assess structure.

These examples also illustrate that Cronbach's alpha also does not reveal much about the mean correlations among items because alpha is so influenced by the number of items.

7. Reverse Scoring and Composite Scores

Above Reverse Coded items were discussed. When a scale contains reversed items, one must first reverse score those items before analyses can occur otherwise misleading results will be obtained. Reverse scoring will be discussed in more detail later, but below are a few summary points and a data example with reversed items.

Brief Summary

- Formula:
 - Reversed Score = (minimum score) + (maximum score) – actual score
- Calculation Check: Correlate original and reversed item, $r = -1.00$

- Check frequency display of original and reversed version to ensure scores were flipped or reversed.
- Composite: Sum vs. Mean (I recommend using mean rather than summed scores; explained in more detail later)
- How affect reliability?
 - Weakens alpha if items not reversed scored
 - Composite score reliability and validity lowered
 - Composite score not interpretable, so statistics with composite scores are not interpretable (e.g., correlations, t-tests, M, SD, etc.)

How to determine which items are reversed

- Logical check reversed items – assume extreme position and answer each item
- Mechanical check for reversed items – examine correlation matrix; negative correlations suggest reversed items present

Example: Leisure Attitude

Below are four items that form a scale that measure the cognitive component of Leisure Attitudes. Scale responses to all items range from low of 1 = Strongly Disagree to 6 = Strongly Agree. Which item is reverse coded?

Item 1. Leisure activities can be a means for self-improvement.

Item 2. Leisure activities can increase one's happiness.

Item 3. People rarely need leisure activities.

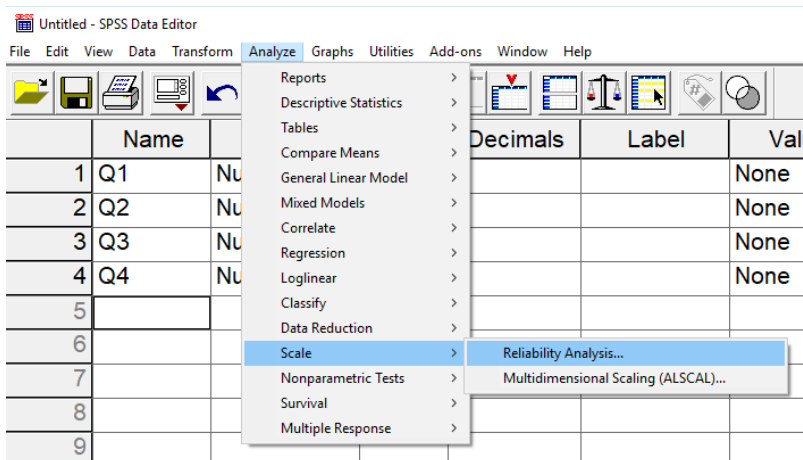
Item 4. Leisure activities can contribute to one's health.

We will use the following Leisure Attitude data to illustrate internal consistency with reversed coded items.

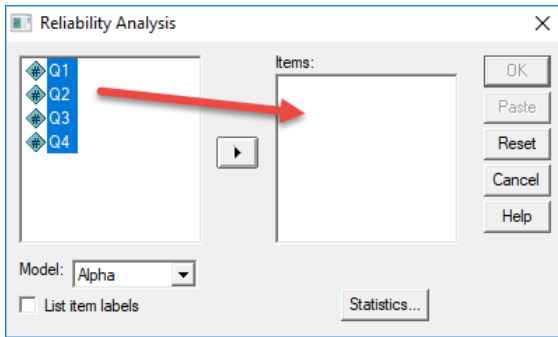
- Complete Leisure Activities questionnaire: <http://goo.gl/forms/JrMtZmVHF5>
- Copy and paste the Leisure Activities data into SPSS: <https://tinyurl.com/yb7ted67>
- Assess internal consistency reliability, if needed, reverse score the item and re-assess internal consistency reliability

SPSS Commands for Cronbach's Alpha

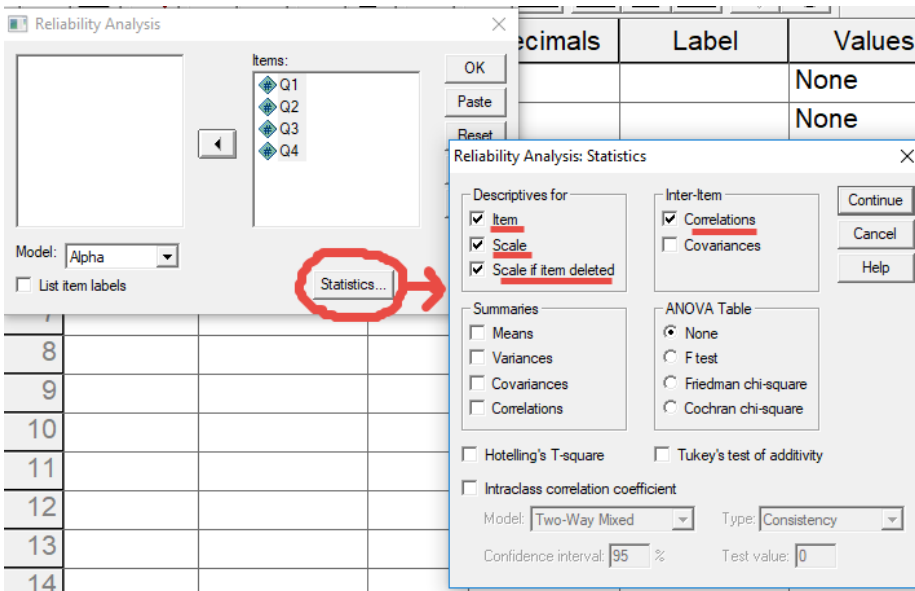
SPSS command to obtain Cronbach's alpha shown below.



For the Leisure Attitude scale, the four items should be moved from the variable box on the left to the Items box on the right – see below.



Next, select Statistics then select Descriptives for Item, Scale, and Scale if item deleted; also obtain Inter-item correlations for reference if needed. These should be the same correlations obtained by the Pearson command in SPSS.



SPSS Results

The results shown below were obtained from the data when the sample size was 33 respondents; it is possible the data you use, since it may contain a larger sample, will produce different correlations and alpha values than those reported below.

The overall Cronbach's alpha for these four items is .270 (see tables below), which is unacceptably low.

Reliability Statistics

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|------------------|--|------------|
| .270 | .420 | 4 |

Item Statistics

| | Mean | Std. Deviation | N |
|----|--------|----------------|----|
| Q1 | 5.3939 | .65857 | 33 |
| Q2 | 5.8182 | .46466 | 33 |
| Q3 | 1.4545 | .86930 | 33 |
| Q4 | 5.6970 | .52944 | 33 |

Inter-Item Correlation Matrix

| | Q1 | Q2 | Q3 | Q4 |
|----|-------|-------|-------|-------|
| Q1 | 1.000 | .446 | -.159 | .622 |
| Q2 | .446 | 1.000 | -.098 | .277 |
| Q3 | -.159 | -.098 | 1.000 | -.167 |
| Q4 | .622 | .277 | -.167 | 1.000 |

The covariance matrix is calculated and used in the analysis.

As the tables above show, item Q3 does not behave in a manner that is consistent with the other items, i.e., it is not internally consistent with the other items. Note that the mean for Q3 is much lower, and that it correlates negatively with the other items. This is a clear indication that (a) either this item does not belong with this group of items, or (b) it is reverse coded and that needs to be considered when calculating alpha. Because of these negative correlations, alpha will be suppressed.

See the last column in the table below (Cronbach's Alpha if Item Deleted). If Q3 is removed, Cronbach's alpha will increase from .27 to .713. This further suggests something is odd about Q3.

Item-Total Statistics

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Squared Multiple Correlation | Cronbach's Alpha if Item Deleted |
|----|----------------------------|--------------------------------|----------------------------------|------------------------------|----------------------------------|
| Q1 | 12.9697 | 1.155 | .371 | .469 | -.125 ^a |
| Q2 | 12.5455 | 1.568 | .283 | .199 | .094 |
| Q3 | 16.9091 | 1.773 | -.179 | .034 | .713 |
| Q4 | 12.6667 | 1.417 | .331 | .392 | .012 |

a. The value is negative due to a negative average covariance among items. This violates reliability model assumptions. You may want to check item codings.

Reverse Scoring in SPSS

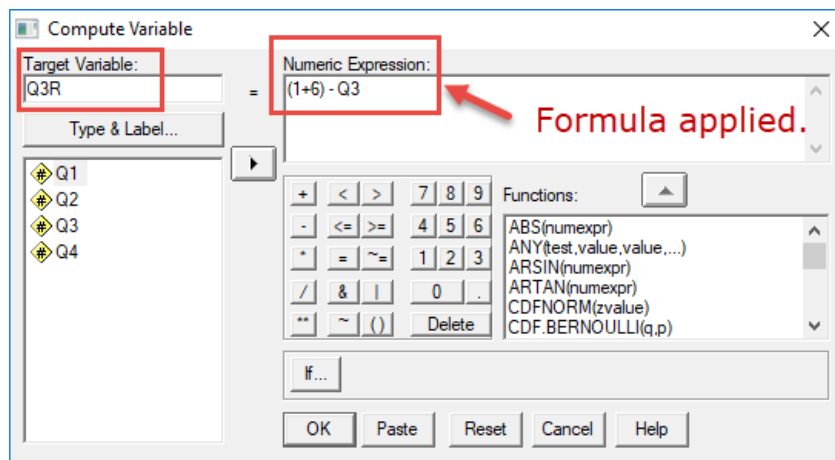
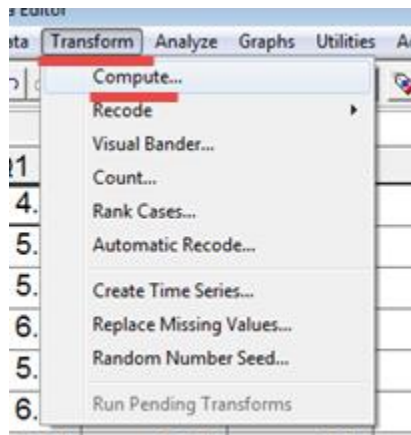
We know that Q3 requires reverse scoring, and that is why it is not internally consistent with the other items as documented above. In SPSS use the "Transform → Compute" command then enter the reverse scoring equation listed above and add a name for the new variable to be created; normally I use the original name plus "R" to signify reverse scoring, e.g., Q1 becomes Q1R.

The equation for reverse scoring is given below. There are three parts: (a) minimum score, (b) maximum score, and (c) actual score.

The scale for these items range from 1 = “Strongly Disagree” to 6 “Strongly Agree,” so the minimum score possible on the scale is 1.00, and the maximum score possible is 6. The actual score refers to the score obtained on the item to be reversed, which for this example is Q3.

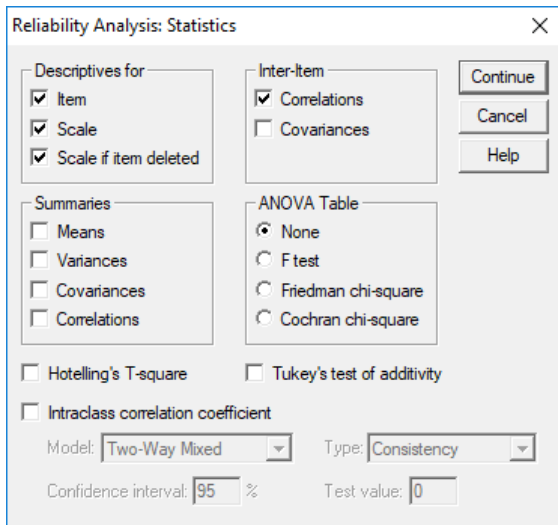
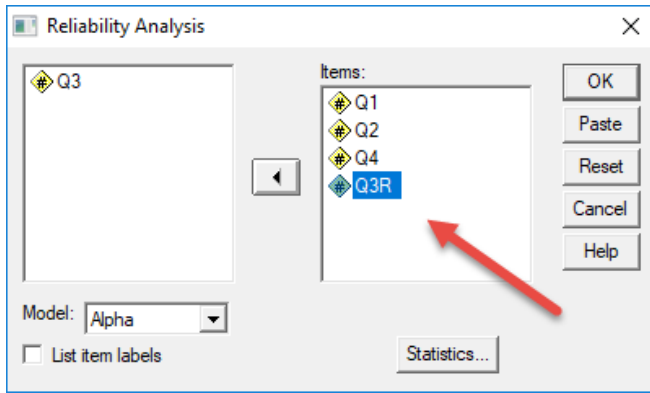
$$\text{Reversed Score} = (\text{minimum score}) + (\text{maximum score}) - \text{actual score}$$

You can see the formula applied in the second screenshot below.



Recalculating Cronbach's Alpha with Reversed Scores

Now that Q3 is reverse scored, use it to re-estimate Cronbach's alpha. Originally, we used Q1, Q2, Q3, and Q4; now we should use Q1, Q2, Q3R, and Q4 – note the Q3R, not Q3. We incorporate the reversed Q3R instead of the original Q3. SPSS results are provided below.



Reliability Statistics

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|------------------|--|------------|
| .570 | .626 | 4 |

Item Statistics

| | Mean | Std. Deviation | N |
|-----|--------|----------------|----|
| Q1 | 5.3939 | .65857 | 33 |
| Q2 | 5.8182 | .46466 | 33 |
| Q4 | 5.6970 | .52944 | 33 |
| Q3R | 5.5455 | .86930 | 33 |

Inter-Item Correlation Matrix

| | Q1 | Q2 | Q4 | Q3R |
|-----|-------|-------|-------|-------|
| Q1 | 1.000 | .446 | .622 | .159 |
| Q2 | .446 | 1.000 | .277 | .098 |
| Q4 | .622 | .277 | 1.000 | .167 |
| Q3R | .159 | .098 | .167 | 1.000 |

The covariance matrix is calculated and used in the analysis.

Note that Q3R now correlates positively with the other items and has a mean score more consistent with the other items too. Unfortunately, Cronbach's alpha at .57 is still too low for solid research use.

Item-Total Statistics

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Squared Multiple Correlation | Cronbach's Alpha if Item Deleted |
|-----|----------------------------|--------------------------------|----------------------------------|------------------------------|----------------------------------|
| Q1 | 17.0606 | 1.621 | .530 | .469 | .342 |
| Q2 | 16.6364 | 2.239 | .351 | .199 | .515 |
| Q4 | 16.7576 | 1.939 | .491 | .392 | .413 |
| Q3R | 16.9091 | 1.773 | .179 | .034 | .713 |

The Item-Total Statistics table shows Q3R does not seem to work well with Q1, Q2, and Q3 despite reversing the score. Dropping Q3R would improve alpha and may also improve the scale. Likely the problem with wording of Q3 or the reversed nature of Q3.

8. Published Examples

Below are linked several example publications that demonstrate how internal consistency reliability is typically presented in research.

- [Menon](#) 2001: Reported in text format; p. 164 Cronbach's alpha and test-retest reliabilities reported on 15 item scale; p. 171 Cronbach's alpha and test-retest reliabilities reported on 9 item scale
- [Kanning, Böttcher & Herrmann](#) 2012: Reported in table format; see Table 2 p. 145 (alpha and test-retest)
- [Frey & Bos](#) 2012: Reported in table format; see Table 4 p. 34 (alpha, item-total correlations minimum and maximum)
- [Fassinger](#) 1994: Development and Testing of the Attitudes Toward Feminism and The Women's Movement (FWM) Scale. Item-total correlation, Table 1 p. 395

References

- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
- Garson, G. D. (2016). *Validity & Reliability*. Statistical Associates Publishing. Asheboro, NC.
- Menon, S.T. (2001). Employee empowerment: An integrative psychological approach. *Applied Psychology: An International Review*, 50, 153-180.
- Ragheb, M. G., & Beard, J. G. (1982). Measuring leisure attitude. *Journal of leisure Research*, 14(2), 155-167.