BlogPost #22: CRONBACH'S ALPHA

Applying a test for internal reliability to my e-Questionnaire

Everyone uses Cronbach's Alpha (α) to establish the supposed internal reliability of their data collection scales. It is important to take into account, however, that the coefficient is a measure for determining the extent to which scale items reflect the consistency of scores obtained in specific samples and isn't assessing the reliability of the scale per se (Boyle et al, 2015) because it is reporting a feature or property of the individuals' responses who have actually taken part in the questionnaire process. This means that although the alpha value provides some indication of internal consistency it isn't necessarily evaluating the HOMOGENEITY, that is, the UNIDIMENSIONALITY of a set of items that constitute a scale.

Nevertheless and with this caveat in mind, the Cronbach's Alpha process has been applied to the scales in my datasets using the 'Scale-Analyse' feature in SPSS. This feature of the application not only provides the alpha coefficient, but because it uses a measure of correlation between scale items which are then also listed in the SPSS output, item redundancy is also highlighted – that is, the impact on the alpha coefficient that would be observed if specific scale items are omitted from the overall alpha coefficient calculation.

The first table below summarizes the outputs that were generated for all 8 scales contained in my eQNR:

SCALE	CRONBACH'S ALPHA VALUE (A) / INTER-CORRELATION MATRIX
Learning related emotions (LRE)	0.547
Anxiety regulation and motivation (ARM)	0.602
Academic self-efficacy (ASE)	0.579
Self esteem (SE)	0.644
Learned helplessness (LH)	0.753
Academic procrastination (AP)	0.748

Dyslexia Index (Dx)	0.842	
Academic Behavioural Confidence (ABC)	0.914	

What do these outputs mean?

According to Kline (1986) an alpha value within the range 0.3 < α < 0.7 is pretty good. Kline proposed that a value of α < 0.3 is indicating that the internal consistency of the scale is fairly poor whilst a value of α > 0.7 may be indicating that the scale contains redundant items whose values aren't providing much new information. However an interesting paper by Schmitt (1996) highlights research weaknesses that are exposed by relying on Cronbach's Alpha alone to inform the reliability of questionnaires' scales, and proposes that additional evaluators about the inter-relatedness of scale items should also be reported, particularly, inter-correlations. SPSS has been used to generate the α values above and the extensive output window that accompanies the root value also presents a complete matrix of inter-correlations. These can be viewed for each of the α values above and make interesting viewing about which more will be written below in due course.

So on face value at least, the alpha coefficients shown above appear to be indicating a reasonably good level of internal consistency in the scales I've developed and which my eQNR is trying to measure. The value of α for Dyslexia Index (Dx) is showing a particularly high level of internal consistency with a value of α = 0.842 and minded my comments above, the output from the SPSS analysis was more keenly scrutinized: it is helpful that one of the summary outputs lists the fresh values of α if any particular scale item is omitted from the analysis. So the second table (below) presents the range of values for α if ONE particular item is deleted from the analysis and also shows which item this is in order to generate the highest value for α.

It is interesting to note that the majority of the scales have quite tight α ranges for a single-item deletion which suggests that omitting no single scale item upsets the internal consistency of the scale by any significant margin. Indeed, the α range for the Dyslexia Index is particularly small which I am taking to mean that all of the items properly contribute to the overall Dx metric.

SCALE	A RANGE	SCALE ITEM DELETED TO ACHIEVE ALPHA-MAX
Learning related emotions (LRE)	$0.387 < \alpha < 0.630$	'I am able to settle down to my work anytime, anyplace'
Anxiety regulation and motivation (ARM)	$0.435 < \alpha < 0.640$	'I enjoy my studies even more when the work becomes difficult'
Academic self-efficacy (ASE)	$0.509 < \alpha < 0.557$	NO SINGLE-ITEM DELETION PRODUCED A HIGHER A VALUE THAN FOR THE COMPLETE SCALE
Self esteem (SE)	$0.529 < \alpha < 0.657$	'IF I TRY HARD, I CAN ACHIEVE JUST AS MUCH AS ANYONE ELSE'
Learned helplessness (LH)	$0.691 < \alpha < 0.759$	'WHEN I START A NEW COURSE OR TOPIC I USUALLY THINK IT WILL BE TOO DIFFICULT FOR ME'
Academic procrastination (AP)	$0.651 < \alpha < 0.762$	'FOR ONE REASON OR ANOTHER, I OFTEN HAVE TO REQUEST EXTRA TIME TO COMPLETE MY WORK'
Dyslexia Index (Dx)	$0.822 < \alpha < 0.858$	'I FIND FOLLOWING DIRECTIONS TO GET TO PLACES QUITE STRAIGHTFORWARD'

However, the matrix of inter-correlations for the metric Dx presents a wide range of correlation coefficients. These range from **R** = -0.446, between scale item statements: 'I THINK I'M A HIGHLY ORGANIZED LEARNER' and 'I FIND IT VERY CHALLENGING TO MANAGE MY TIME EFFICIENTLY' – which might be expected, to **R** = 0.635, between scale item statements: 'I GET REALLY ANXIOUS IF I'M ASKED TO READ 'OUT LOUD' ' and 'WHEN I'M READING, I SOMETIMES READ THE SAME LINE AGAIN OR MISS OUT A LINE ALTOGETHER' – which we also might expect. So with the Cronbach alpha value of **α =** 0.842, indicating perhaps a suspiciously high level of internal reliability consistency, according to Kline's proposal at least, this is causing me to re-consider which scale items I should reverse-code prior to their contribution to my overall Dyslexia Index value. At present, the only scale item statement that this process is applied to is: 'MY SPELLING IS GENERALLY GOOD'. Clearly adjusting scores in this way will not affect the correlation coefficient (aside from reverse its +/- sign) nor the value of α as this is essentially based on correlation coefficients.

Reporting more than Cronbach's α

Further reading about internal consistency reliability coefficients has led me to some interesting papers by Henson (2001) and Onwuegbuzie et al (2002). Both researchers firstly identify persistent weaknesses in the reporting of data reliability in research, particularly in their fields of, broadly speaking, social sciences research. Secondly, useful frameworks are provided for reporting and interpreting internal consistency reliability estimates which, it is argued, then present a more comprehensive picture of the reliability of data collection procedures, particularly data elicited through self-report questionnaires. Henson (op cit) strongly emphasizes the point that 'INTERNAL CONSISTENCY COEFFICIENTS ARE NOT DIRECT MEASURES OF RELIABILITY, BUT RATHER ARE THEORETICAL ESTIMATES DERIVED FROM CLASSICAL TEST THEORY' (2001, p177), which connects with Boyle's (2015, above) interpretation about the sense of this measure. However Boyle's view relating to the scale item homogeneity appears to be different from Henson's who DOES state that internal consistency measures do indeed offer an insight into whether or not scale items are combining to measure the same construct. This difference of view isn't helpful. Henson strongly advocates that when (scale) item relationship correlations are of a high order, this indicates that the scale as a whole is gauging the construct of interest with some degree of consistency – that is, that the scores obtained from this sample at least, are reliable (Henson, 2001, p180).

Onwuegbuzie and Daniel (2002) base their paper on much of Henson's work but go further by presenting recommendations to researchers which proposes that they/we should always estimate and report:

- o internal consistency reliability coefficients for the current sample;
- o confidence intervals around internal consistency reliability coefficients – but specifically upper tail limit values;
- o internal consistency reliability coefficients and the upper tail confidence value for each sample subgroup (ibid, p92).

I like the idea of providing a confidence interval for Cronbach's α since, as being discussed here, we now know that the value of the coefficient is relating information about the internal consistency of scores for items making up a scale that pertains to that particular sample. Hence it then represents a POINT estimate of the likely internal consistency reliability of the scale, and hence the construct of interest, for all samples taken from the background population. But interval estimates are better, especially as the point estimate value, α, is claimed by Cronbach himself in his original paper (1951) to be most likely a lower-bound estimate of score consistency. So Onwuegbuzie and Daniel's suggestion that one-sided confidence intervals (the

upper bound) are reported in addition to the value of Cronbach's α is a good guide for more comprehensively reporting the internal consistency reliability of data.

Calculating the upper-limit confidence value for Cronbach's α

Confidence intervals are most usually specified to provide an interval estimate for the population mean using sample data to do this by using a sample mean – which is a POINT ESTIMATE for the population mean – and building the confidence interval estimate based on the assumption that the background population follows the normal distribution. We won't go into a further discussion about this assumption and the Central Limit Theorem, but I know about it.

So it follows that any POINT ESTIMATE of a population parameter might also have a confidence INTERVAL ESTIMATE constructed around it provided the most underlying assumption that the distribution of the parameter is normal. For a correlation coefficient between two variables in a sample, this is a point estimate of the correlation coefficient between the two variables in the background population and if we took a separate sample from the population we might expect a different correlation coefficient to be produced. Hence a distribution of correlation coefficients would emerge in much akin to the distribution of sample means that constitutes the fundamental tenet of the Central Limit Theorem and which permits us to generate confidence intervals for a background population mean based on sample data. Fisher (1915) explored this idea to arrive at a transformation that mapped the Pearson Product-Moment Correlation Coefficient, **R** , onto a value, **Z'**, which is approximately normally distributed and hence, confidence interval estimates could be constructed. Given that Cronbach's **α** is essentially based on values of **r**, we can use Fisher's **Z'** to transform Cronbach's **α** and subsequently apply the standard processes for creating our confidence interval estimates for the range of values of α we might expect in the background population. Fisher showed that the standard error of **Z'**, which is obviously required in the construction of confidence intervals, to be solely related to the sample size: $SE = 1/\sqrt{(n-3)}$, with the transformation process for generating **Z'** shown (below).
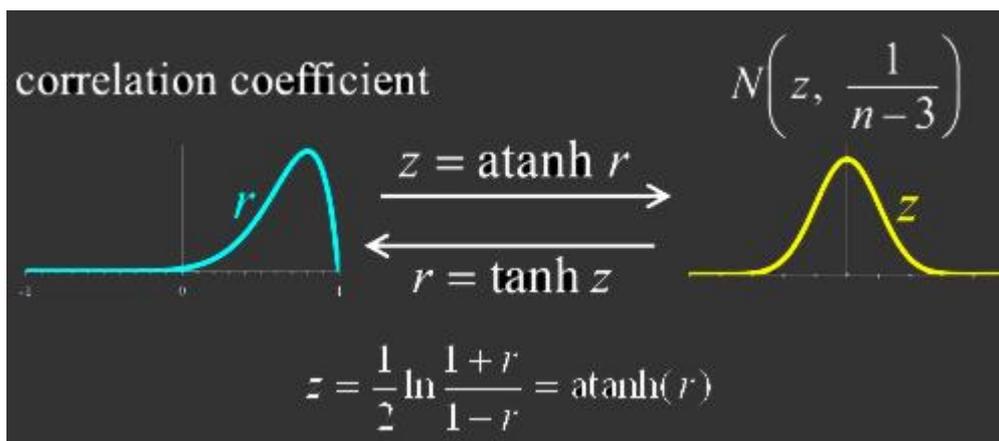
So now I can generate the upper-tail 95% confidence interval limit for my Cronbach alpha values and to do this, I followed the step-by-step process described by Onwuegbuzie and Daniel (op cit) and worked through in a useful example by Lane (2013):

- Transform the value for Cronbach's α to Fisher's **Z'**
- Calculate the Standard Error (SE) for **Z'**
- Calculate the upper 95% confidence limit for **Z'** + (SE)**Z** [for the upper tail of 95% two-tail confidence interval, **Z** = 1.96]
- Transform back the upper confidence limit for **Z'** back to a Cronbach's α internal consistency reliability coefficient.

There are a number of online tools for transforming to Fisher's **Z'** but I preferred to set this up in Excel using the formula in the graphic above. The table below shows the set of cell calculations from my Excel spreadsheet and particularly, the upper 95% confidence limit for **α** for each of the seven scales that I developed for my eQNR:

Cronbach's Alpha

Scale:	LRE	ARM	ASE	SE	LH	AP	Dx	DxR	ABC24
Cronbach's Alpha	0.547	0.602	0.579	0.644	0.753	0.748	0.852	0.886	0.914
corresponding Fisher's Z'	0.614	0.696	0.661	0.765	0.980	0.968	1.263	1.403	1.551
standard error for Z', 166	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078
95% upper tail boundary Z =	1.960	1.960	1.960	1.960	1.960	1.960	1.960	1.960	1.960
upper tail 95% CI for Z'	0.768	0.850	0.814	0.918	1.133	1.122	1.417	1.557	1.705
upper tail for Cronbach's Alpha	0.646	0.691	0.672	0.725	0.812	0.808	0.889	0.915	0.936

So I have completed the first part of Onwuegbuzie & Daniel's additional recommendation by reporting not only the internal consistency reliability coefficient for each of my scales but also included the upper tail 95% confidence interval value. All that remains is to first use SPSS to generate the values for Cronbach's α for my sample's subgroups, that is, research groups DI, ND and DNI, repeat the calculations to derive the upper CI value then reflect on what this is all telling me. The table below shows the first part of this analysis, presenting a summary of the results for the two primary research subgroups: ND and DI.

Cronbach's α for Dyslexia Index (Dx - ND)

Scale:	Dx
Cronbach's α	0.842
corresponding Fisher's Z'	1.228
standard error for Z', n = 98	0.103
95% upper tail boundary Z =	1.960
upper tail, 95% CI for Z'	1.429
upper tail, 95% CI for Cronbach's α	0.891

Cronbach's α for Dyslexia Index (Dx - DI)

Scale:	Dx
Cronbach's α	0.689
corresponding Fisher's Z'	0.846
standard error for Z', n = 66	0.126
95% upper tail boundary Z =	1.960
upper tail, 95% CI for Z'	1.093
upper tail, 95% CI for Cronbach's α	0.798

These tables show root values of α to be 0.842 and 0.689 respectively which are both 'respectable' values for the internal consistency reliability of my Dyslexia Index Scale although at the moment I can't explain why the value of α = 0.852 for the COMPLETE research datapool is higher than either of these values.

However it clear to see that, assuming a satisfactory explanation can be found to explain minor discrepancies (most likely calculating errors), the upper tail confidence interval boundaries for not only the complete research datapool but for both subgroups all present an α value that is indicating a strong degree of internal consistency reliability for the Dyslexia Index Scale, notwithstanding Kline's earlier caveats mentioned above.

References

Boyle, G.J., Saklofske, D.H., Matthews, G., 2015, MEASURES OF PERSONALITY AND SOCIAL PSYCHOLOGICAL CONSTRUCTS, London, Academic Press.

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

Fisher, R.A., 1915, Frequency distribution of the balues of the correlation coefficient in samples from an infinitely large population, BIOMETRIKA, 10(4), 507-521.

Henson, R.K., 2001, Understanding internal reliability estimates: a conceptual primer on coefficient alpha, MEASUREMENT AND EVALUATION IN COUNSELLING AND DEVELOPMENT, 34, 177-189.

Lane, D.M., 2013, Hyperstat Online Statistics Textbook, available at: http://davidmlane.com/hyperstat/confidence_intervals.html, accessed on: 29th August 2016.

Kline, P., 1986, A HANDBOOK OF TEST CONSTRUCTION: INTRODUCTION TO PSYCHOMETRIC DESIGN. London, Methuen.

Onwuegbuzie, A.J., Daniel, L.G., 2002, A framework for reporting and interpreting internal consistency reliability estimates, MEASUREMENT AND EVALUATION IN COUNSELLING AND DEVELOPMENT, 35, 89-103.

Schmitt, N., 1996, Uses and abuses of Coefficient Alpha, PSYCHOLOGICAL ASSESSMENT, 8(4), 350-353.