Likert Response Range and Correlation Attenuation

Some questions were asked of me the other day…

- What is the likely attenuation on a correlation between two items on a test, if we use a 2-choice rather than say a 3, 4, 5, up to a 9-choice Likert format?
- What happens to the item variances under these conditions?
- Would using more Likert categories automatically increase alpha reliability for a test?

To answer these, I put the issue of psychological meaning to one side (i.e. what happens to a person's judgement when it is constrained to a 2, 3, 4, 5, 6, 7, 8, or 9-category rating scale), and concentrated solely on the mathematical-statistical issue. Specifically, I was curious myself as to what happens exactly when say real-valued continuous number responses are categorised.

Two papers I have read speak directly to the issues at hand,

Alwin, D. F. (1992) Chapter 3: **Information transmission in the survey interview: Number of response categories and the reliabilty of attitude measurement**. In Marsden, P.V. (Eds.). Sociological Methodology. Basil Blackwell.

and very specifically

Bollen, K.A., & Barb, K.H. (1981) **Pearson's R and Coarsely Categorized Measures**. *American Sociological review*, 46, Apr, 232-239.

This latter paper computes the attenuation in a Pearson correlation directly as a function of the process of categorization, for several values of a "population" correlation coefficient value.

However, I thought it might be of interest for students to see how a simulation like this might be attempted – using STATISTICA – and allow the STATISTICA Basic program for doing this to be downloaded etc. by any interested reader (also attached here as Appendix 1 for anyone interested in it but who does not have access to STATISTICA 6). Also, it allowed me to look at what happens to the item variances as I increased the categories.

Method

I randomly sampled a Continuous Unit Normal Distribution, with mean 0 and SD of 1.0, 1000 times, converting the real-valued numbers into those from a distribution of mean 4 and SD of 2 (merely for visual convenience). I placed these numbers into a variable called Real_1. I did the same for a second variable, and called this Real_2.

At the moment, these variables correlate at very near 0.0 (0.0 as the population value), since both variables contain observations sampled at random from a Normal Distribution. I now transformed the Real_2 variable to correlate 0.5 with Real_1, using the transformation:

$$\text{New\_Real\_2} = r \cdot \text{Real\_1} + \sqrt{\left(1 - r^2\right)} \cdot \text{Real\_2}$$

where:
r = the desired correlation (0.5 in my case)
New_Real_2 = the transformed random number contained in Real_2

Now, because the New_Real_2 observations are increased in size relative to those in Real_1, I simply subtracted the difference in the means between New_Real_2 and Real_1 from all of New_Real_2's values, and put the final number back into the variable called Real_2 in my STATISTICA spreadsheet. This is because of how I am creating my categorised variables … see below!

A Technical Digression (only read this if you are interested in comparing the formula I'm using to one that is used by others)

I actually cannot remember where I picked up my formula from! I've been using it for probably the last 10 years or so – I have no idea who gave it to me or where I saw it!

For those of you who have seen David Howell's web-page entitled: **Generating Data with a specified Correlation** (at http://www.uvm.edu/~dhowell/StatPages/More_Stuff/CorrGen.html), you'll have seen that the formula he provides looks different from the above…

My formula

$$\text{New\_Real\_2} = r \cdot \text{Real\_1} + \sqrt{\left(1 - r^2\right)} \cdot \text{Real\_2}$$

David Howell's formula

$$\text{New\_Real\_2} = \left(\frac{r}{\sqrt{\left(1 - r^2\right)}} \right) \cdot \text{Real\_1} + \cdot \text{Real\_2}$$

These formulas are **NOT** equivalent as they stand, but if we expand the algebra of Howell's formula … we obtain…

$$\text{New\_Real\_2} = \frac{r \cdot \text{Real\_1} + \sqrt{\left(1 - r^2\right)} \cdot \text{Real\_2}}{\sqrt{\left(1 - r^2\right)}}$$

As you can see, Howell's formula is equivalent to mine except that I do not divide my result by $\sqrt{\left(1 - r^2\right)}$. Because this value acts a constant, it is actually unnecessary to divide the formula $r \cdot \text{Real\_1} + \sqrt{\left(1 - r^2\right)} \cdot \text{Real\_2}$ by this value. But, to be safe, I quickly simulated some data using both my formula and Howell's, sampling 5000 observations on two variables, and checked the resultant correlation and descriptives…

As expected, the correlations between New_Real_2 and Real_1 are exactly equivalent, whether the New variable observations are computed via mine or via Howell's formula. The Descriptives indicate that the normalised skewness and kurtosis are exactly equivalent. But, because of the division by a constant, the means, medians, and standard deviations are not.

	Descriptive Statistics (Spreadsheet1)							
Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.	Skewness	Kurtosis
Real_1	5000	4.008850	3.990465	-2.95273	12.57274	2.019212	0.007812	0.013530
Real_2	5000	3.951422	3.942113	-2.96644	11.21918	1.996133	0.037379	0.068407
New_Real_2 - Mine	5000	5.426457	5.430345	-1.58533	13.13713	1.991109	0.026094	0.030885
New_Real_2 - Howell	5000	6.265933	6.270422	-1.83058	15.16945	2.299134	0.026094	0.030885

So – either formula will do the trick …! You can also see why I am subtracting the difference in the means between New_Real_2 and Real_1 from all of New_Real_2's values, and putting the final number back into the variable called Real_2 in my STATISTICA spreadsheet for the categorisation simulation – this is to bring the values into simple alignment with Real_1's.

Back to the substantive issue …
OK – so I have my simulated data on two variables which correlates at 0.5 (or near-about given the sampling error).

I now progressively categorize the data into new variables …

	Simulation Dataset varying the number of Likert response categories per item response - Paul Barrett 22-5-2003	
	1 Variable Name	2 Long Name -Description
1	Real1	
2	Real2	
3	Two_1	Two choices - person 1
4	Two_2	Two-choices - person 2
5	Three_1	Three choices - person 1
6	Three_2	Three choices - person 2
7	Four_1	Four choices - person 1
8	Four_2	Four choices - person 2
9	Five_1	
10	Five_2	
11	Six_1	
12	Six_2	
13	Seven_1	
14	Seven_2	
15	Eight_1	
16	Eight_2	
17	Nine_1	
18	Nine_2	

I do this by a simple categorisation process … for example, to create the Two-choice variable values, assigning codes of 0 to all Real numbers less than 4, and 1 to all numbers equal to 4 and above, my program (or "recode") statements are:

If s.Cells (i,1)< 4.0 Then
 s.Cells (i,3)=**0**
Else
 s.Cells (i,3)=**1**
End If

For Three choices,
If s.Cells (i,1)< 2.67 Then s.Cells (i,5)=**0**
If (s.Cells (i,1)>= 2.67) And (s.Cells (i,1)< 5.33) Then s.Cells (i,5)=**1**
If s.Cells (i,1) >=5.33 Then s.Cells (i,5)=**2**

And so on up to 9 choices. I've not bothered optimising the code in the program, preferring instead to maximise clarity and logic of each function over execution speed. The categorisation caused me some heartache! On an even number of categories, the mean was the threshold for adjoining categories. For an odd-number of categories, the mean was centred within the middle category. All categories were of equal range, which ensured discrete normal distributions of categorised responses. Check the program to see the exact decision rules.

So, I now have my set of differentially categorised responses computed form my two Real-valued variables (Real_1 and Real_2).

I then compute the correlations between every relevant pair of variables … Real_1 and Real_2, Two_1 and Two_2, Three_1 and Three_2, up to Nine_1 and Nine_2. I also compute the standard deviation for each categorised variable. I put these parameters into a new Spreadsheet.

Then I take a new sample of 1000 observations for Real_1 and Real_2, and do the whole lot again … in all 100 times. So, my final parameter summary file contains data from 100 simulations.

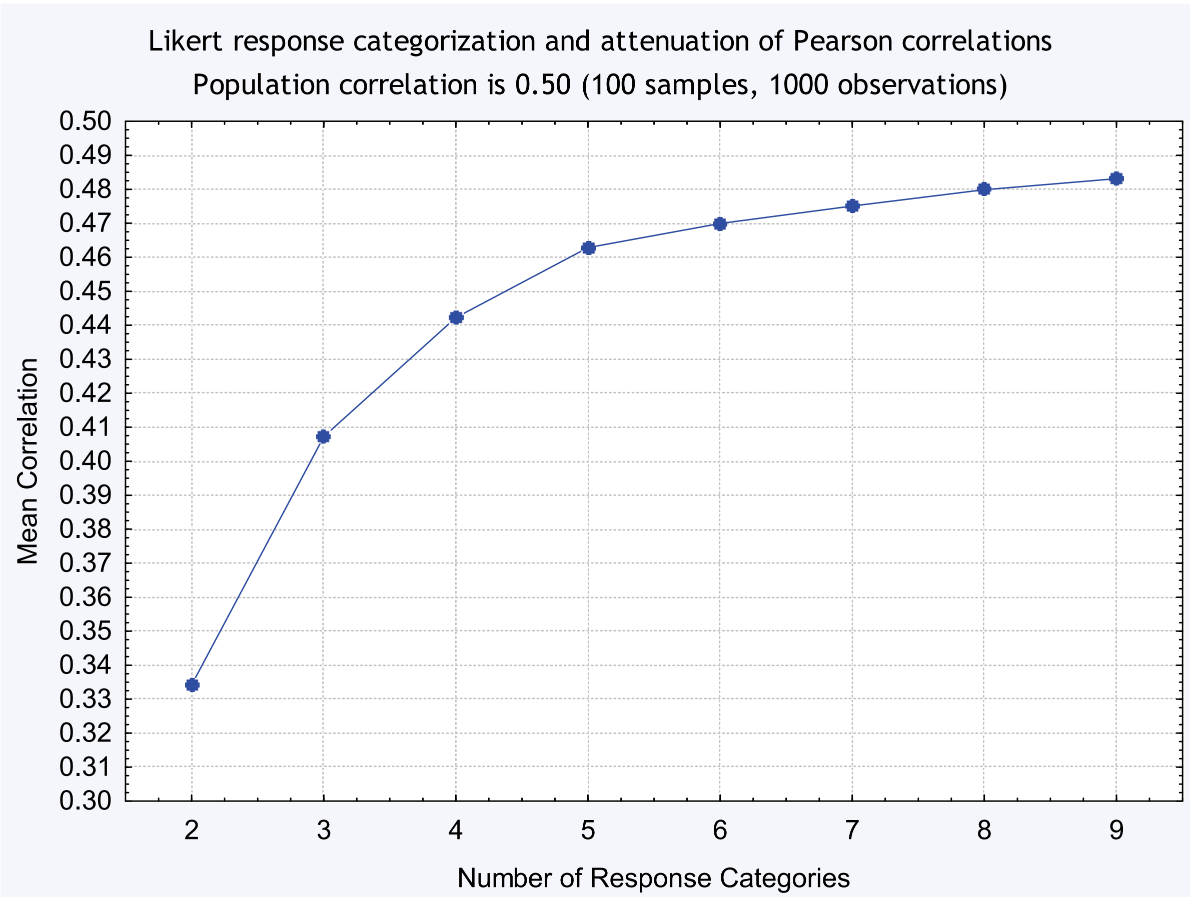
Now for the results …

Results

First, a table of the descriptives for each correlation …

Variable	Descriptive Statistics (Bootstrap Results.sta)							
	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.	Skewness	Kurtosis
Corr_Real_1_2	100	0.499678	0.497902	0.445120	0.565613	0.021814	0.468204	0.884321
Corr_Two_1_2	100	0.334196	0.335346	0.283917	0.398005	0.028359	0.198970	-0.502534
Corr_Three_1_2	100	0.407222	0.409064	0.361783	0.476770	0.025731	0.382186	0.144225
Corr_Four_1_2	100	0.442276	0.440764	0.397008	0.490262	0.023814	0.151734	-0.712558
Corr_Five_1_2	100	0.462875	0.461291	0.400194	0.549506	0.025018	0.525061	0.968331
Corr_Six_1_2	100	0.469932	0.467125	0.420095	0.530658	0.022805	0.299903	-0.041759
Corr_Seven_1_2	100	0.475053	0.474197	0.423322	0.540686	0.022821	0.545906	0.654058
Corr_Eight_1_2	100	0.479866	0.478631	0.438170	0.537547	0.022633	0.304170	-0.084256
Corr_Nine_1_2	100	0.483108	0.481277	0.436378	0.551233	0.022261	0.673341	0.725669

And the graph of correlation size against categorisation …



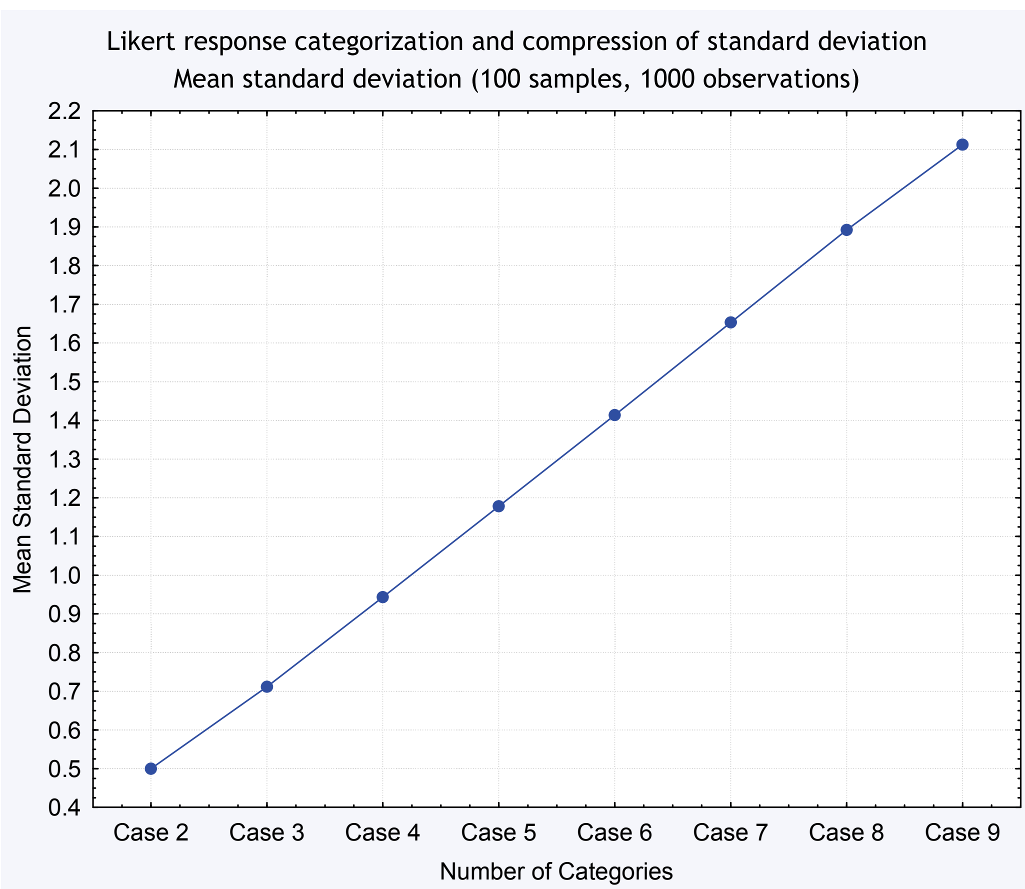
Interesting – and extremely similar to Bollen and Barb's (1992) data – as expected.

Let's look at the standard deviation for each variable as a function of categorization …

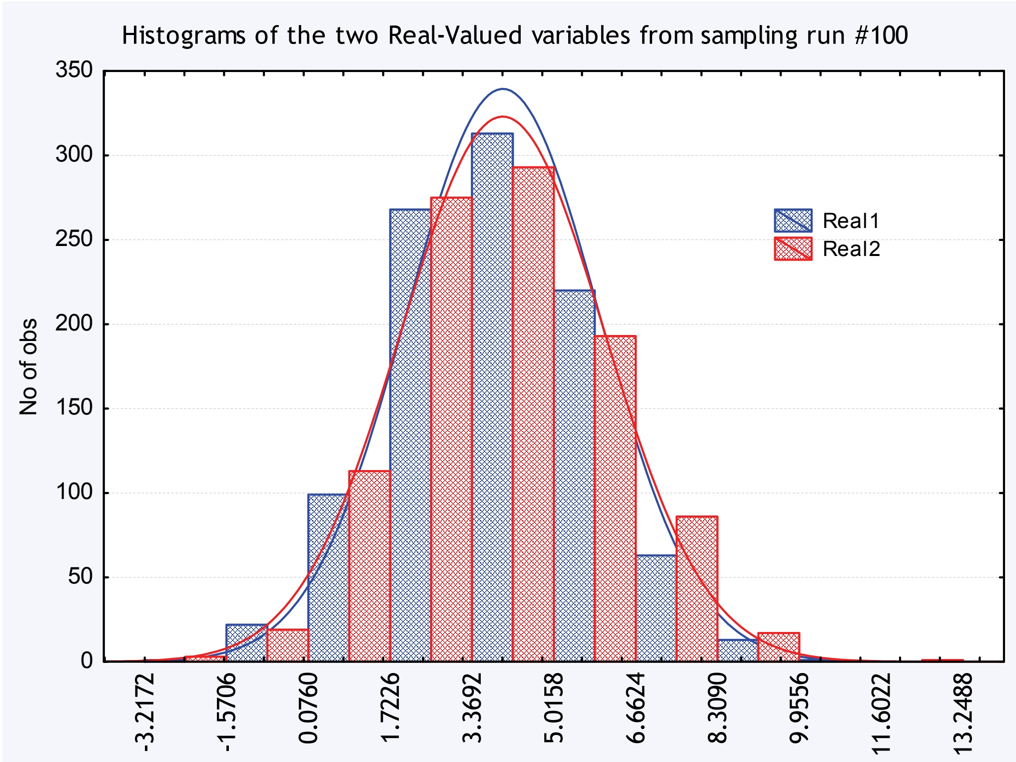
The descriptives for the mean of the standard deviations are:

Variable	Descriptive Statistics (Bootstrap Results.sta)							
	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.	Skewness	Kurtosis
SD_Real_1	100	1.995453	1.992633	1.911517	2.106360	0.040033	0.16613	-0.052396
SD_Real_2	100	2.010713	2.011169	1.914562	2.101897	0.039795	0.09568	-0.240868
SD_Two_1	100	0.499991	0.500150	0.498220	0.500249	0.000353	-2.24379	6.500907
SD_Two_2	100	0.499925	0.500081	0.498129	0.500250	0.000401	-1.74513	3.466620
SD_Three_1	100	0.710405	0.710383	0.681513	0.742611	0.011838	-0.06581	0.094118
SD_Three_2	100	0.711729	0.712868	0.688093	0.737253	0.009948	-0.03171	-0.101889
SD_Four_1	100	0.938106	0.938840	0.904851	0.969525	0.014608	0.07101	-0.618888
SD_Four_2	100	0.943220	0.944162	0.914630	0.981176	0.014087	0.27854	-0.155126
SD_Five_1	100	1.172396	1.170820	1.127500	1.216567	0.019559	0.11785	-0.311619
SD_Five_2	100	1.177850	1.177009	1.137097	1.232281	0.020093	0.38918	-0.045650
SD_Six_1	100	1.407815	1.407067	1.360810	1.469377	0.023145	0.11174	-0.260243
SD_Six_2	100	1.413954	1.413380	1.361491	1.469736	0.024380	0.24917	-0.563783
SD_Seven_1	100	1.644896	1.647397	1.583503	1.723720	0.028006	0.08990	-0.121626
SD_Seven_2	100	1.652960	1.650514	1.593590	1.712743	0.026623	0.24807	-0.355670
SD_Eight_1	100	1.880302	1.878755	1.810170	1.965947	0.032093	0.08033	-0.153302
SD_Eight_2	100	1.891984	1.891149	1.837676	1.973676	0.031400	0.36666	-0.290268
SD_Nine_1	100	2.099948	2.096137	2.028051	2.193657	0.036277	0.26991	-0.270533
SD_Nine_2	100	2.112696	2.114160	2.031679	2.195691	0.034327	0.23142	-0.123897

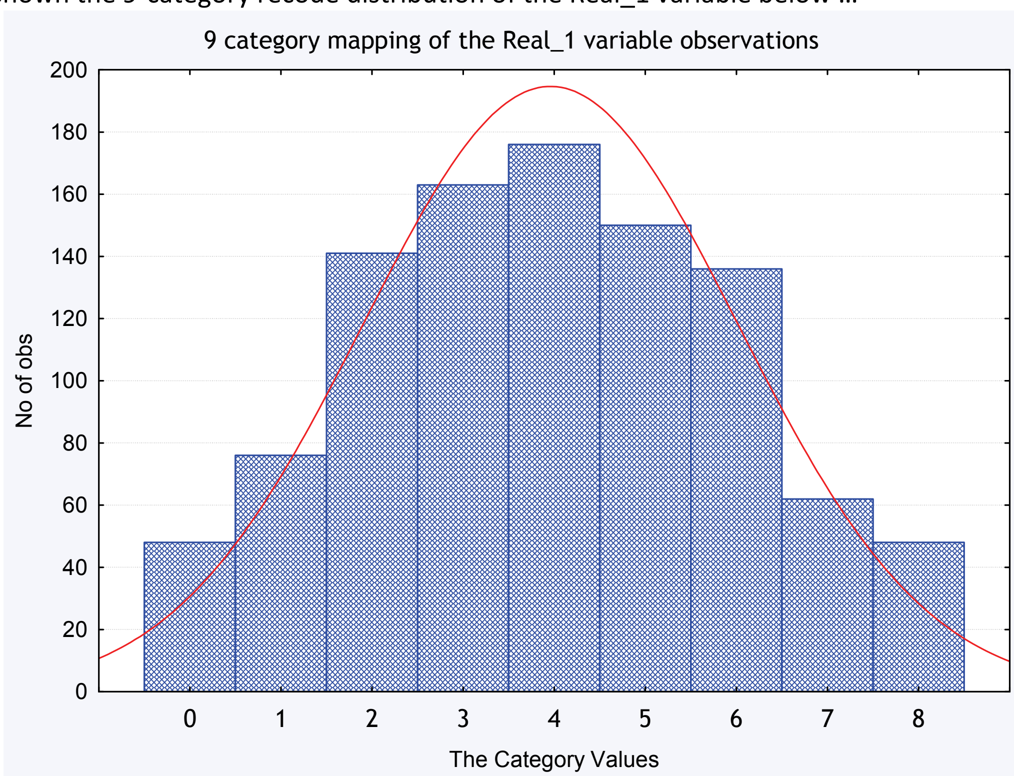
The standard deviations are being compressed as a function of categorization – the graph below shows this (I've just used the values for each _2 variable, as the _1 corresponding values are so similar) – and also shows that the compression is clearly a linear function of the number of categories.



To give you a flavour of the shape of the distributions for the real-valued observations, I've shown a dual-plot histogram below …



I've also shown the 9-category recode distribution of the Real_1 variable below …



Would using more Likert categories automatically increase alpha reliability for a test?

The upshot of all this work is that, purely from a mathematical perspective, as the number of categories of a response decrease from real-valued to dichotomous responses, so do correlations between these coded-response observations decrease. Likewise the variance for each set of coded observations.

Given alpha can be expressed ultimately as the average mean inter-item correlation between all possible items in the "item-universe" for a test, then, from this mathematical perspective, the alpha for dichotomised items will **ALWAYS** be less than that for an increasing number of coded item responses, GIVEN that the responses can be conceived of as actually being mapped from a continuous, real-valued, normally distributed, and with a perfectly linear relationship at all levels of response on each item!

Remember though, this work only speaks to the mathematical costs of constraining response codes. There is another aspect to increasing Likert response codes which is all about the psychology of the response - i.e. can people use scales that are much longer than say 5 choices etc... - and even if they do, are any differences in usage trivial for all practical purposes? I suspect so - and using some stats for "difference analysis" (e.g. statistical significance for difference between alphas etc.), it is possible to more systematically look for statistically significant results.

But, let’s say we observe an increase in alpha from 0.8 to 0.9. What really is the practical difference between these alphas? Well, the best way of viewing this is to look at the most critical use of alpha in practice - for the standard error of measurement of any test score.

$$SEM = s_t \cdot \sqrt{(1 - r_{tt})}$$

where $s_t =$ the standard deviation of test scores

$r_{tt} =$ the alpha reliability for the test scale

So, given the standard deviation (SD) of a test score will increase with increasing response range, and alpha will most likely increase correspondingly (via the increase in response range), the issue is that the SEM for a test constructed from 2-choice or 9-choice response scales may be effectively equivalent in meaning - given the difference in range of possible test scores.

For example, let's take the EPQR short form for E. We'll assume an SD of 3, and alpha of 0.8, with a maximum score of 12. The SEM for such data would be ... 1.34 - this is the standard error of the observed test score. The ratio of SEM to maximum score range is **0.11** (1.34/12)

Now, we use a 9-point scale - ranging from 0 to 8 in response categories. Total test score range thus goes from 0 to 96! Let's assume the SD is now 30 for this scale, and alpha = 0.90. The SEM is ... 9.5. The ratio of SEM to total score range is **0.10**. If however the SD was 24, with alpha of 0.9, then the SEM would be 7.6. So - the change in SD of the scale relative to the change in alpha is critical!

Anyway, you can see the point I'm making - you need both alpha and SD to really gauge whether increasing the response range really has a substantive, meaningful effect on the psychometric properties of a test score, or whether the increases in isolated parameters like alpha or variance are indeed merely "artifactual" and of no real practical value. The trick is to gain an increase in alpha over and above any relative increase in test score SD!

From another perspective though, a high alpha will produce less measurement error when correlating two variables together, say EPQR-E and number of parties attended in a year. As EPQR-E increases in reliability, so we might expect the observed correlation between it and any other variable to be less attenuated by measurement error. Let's say we observe a correlation between EPQR-E and some other variable of 0.6 (for convenience, we'll assume this other variable has perfect reliability). If the alpha of E is 0.8, we can correct this correlation for the unreliability of E, which yields 0.67. If the alpha was 0.9, with the same observed correlation of 0.6, we would have a corrected value of 0.63. Hardly ground-shaking stuff!

Finally, a recent message on SEMNET from Nic Allum listed some useful references on the topic – these are not comprehensive as my two "old faithfuls" are not even listed! But – this is a very useful list of references all the same!

[**2010**]: I've added two new references here which also review/address the problem – and provide a handy correction formula (the first paper).

Aguinis, H., Pierce, C.A., Culpepper, S.A. (2009) Scale coarseness as a methodological artifact correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods*, 12, 4, 623-652.

Wu, C-H. (2007) An empirical study on the transformation of Likert-scale data to numerical scores. *Applied Mathematical Sciences*, 1, 58, 2851-2862.

Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly*, 48, 409-442.

Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology*, 59, 313-320.

Bendig, A. W. (1953). The reliability of self-ratings as a function of amount of verbal anchoring and of the number of categories on the scale. *Journal of Applied Psychology*, 37, 38-41.

Bendig, A. W. (1954). Transmitted information and the length of rating scales. *Journal of Experimental Psychology*, 47, 303-308.

Bendig, A. W., & Hughes, J. B. (1953). Effect of amount of verbal anchoring and number of rating-scale categories upon transmitted information. *Journal of Experimental Psychology*, 46, 87-90.

Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51, 220-232.

Birkett, N. J. (1986). Selecting the number of response categories for a Likert-type scale. *Proceedings of the American Statistical Association 1987 Annual Meetings, Section on Survey Research Methods*.

Champney, H., & Marshall, H. (1939). Optimal refinement of the rating scale. *Journal of Applied Psychology,* 23, 323-331.

Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, 17, 407-422.

Eagly, A. H., & Steffen, V. J. (1988). A note on assessing stereotypes. *Personality and Social Psychology Bulletin*, 14, 676-680.

Finn, R. H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 32, 255-265.

Garner, W. R. (1960). Rating scales, discriminability, and information transmission. *Psychological Review*, 67, 343-352.

Ghiselli, E. E. (1939). All or none versus graded response questionnaires. *Journal of Applied Psychology*, 23, 405-413.

Green, P. E., & Rao, V. R. (1970). Rating scales and information recovery - how many scales and response categories to use? *Journal of Marketing*, 34, 33-39.

Kalton, G., Roberts, J., & Holt, D. (1980). The effects of offering a middle response option with opinion questions. *Statistician*, 29, 65-78.

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of scales. Educational and Psychological Measurement, 25, 897-995.

Lehmann, D. R., & Hulbert, J. (1972). Are three-point scales always good enough? *Journal of Marketing Research*, 9, 444-446.

Lipset, S. M., & Schneider, W. (1983). *The confidence gap*. New York: The Free Press, p. 89-96.

Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A monte carlo approach. *Journal of Applied Psychology*, 60, 10-13

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, 31, 657-674.

Matell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, 56, 506-509.

McGuire, W. J. (1981). *The probabilogical model of cognitive structure and attitude change*. In R. E. Petty, T. M. Ostrom, and T. C. Brock (Eds.), Cognitive responses in persuasion. Hillsdale, NJ: Erlbaum.

Peterson, B. L. (1985). Confidence: Categories and confusion. *GSS Technical Report No. 50, National Opinion Research Center, University of Chicago*.

Ramsay, J. O. (1973). The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika*, 38, 513-532.

Rosenstone, S. J., Hansen, J. M., & Kinder, D. R. (1986). Measuring change in personal economic well-being. *Public Opinion Quarterly*, 50, 176-192.

*Schuman, H., & Presser, S. (1981). Questions and answers in attitude surveys. New York: Academic Press. Chapters 6, "*Measuring a middle position*."

Smith, T. W., & Peterson, B. L. (1985). The impact of number of response categories on inter-item associations: Experimental and simulated results. *Paper presented at the American Sociological Association annual meetings, Washington, D.C*.

Stember, H., & Hyman, H. (1949-1950). How interviewer effects operate through question form. *International Journal of Opinion and Attitude Research,* 3, 493-512.

Symonds, P. M. (1924). On the loss of reliability in rating due to coarseness of the scale. *Journal of Experimental Psychology*, 456-461.

Warr, P., Barter, J., & Brownridge, G. (1983). On the independence of positive and negative affect. *Journal of Personality and Social Psychology*, 44, 644-651.

Watson, D. (1988). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. *Journal of Personality and Social Psychology*, 55, 128-141.

Wedell, D. H., & Parducci, A. (1988). The category effect in social judgment: Experimental ratings of happiness. *Journal of personality and Social Psychology*, 55, 341-356.

Appendix 1: The STATISTICA Basic program that implemented the monte-carlo examination

{comments are in blue italics}

```
Sub Main
Dim s As Spreadsheet
Set s = ActiveSpreadsheet

boot=100 'the number of simulations to be implemented

'Setup the new Results Spreadsheet - with 27 variables to hold the correlations and standard deviations
Dim rs As New Spreadsheet
rs.SetSize(boot,27)

'This is the header title for my new spreadsheet
rs.Header = "Bootstrap Results Analysis - Correlation and Variance magnitudes as a function of Likert
Coding"

'Here I set the variable widths
For i = 1 To 27
rs.VariableWidth(i)=1.1
Next

'Here are my new variable titles for this spreadsheet
rs.VariableName(1)="Corr_Real_1_2"
rs.VariableName(2)="SD_Real_1"
rs.VariableName(3)="SD_Real_2"

rs.VariableName(4)="Corr_Two_1_2"
rs.VariableName(5)="SD_Two_1"
rs.VariableName(6)="SD_Two_2"

rs.VariableName(7)="Corr_Three_1_2"
rs.VariableName(8)="SD_Three_1"
rs.VariableName(9)="SD_Three_2"

rs.VariableName(10)="Corr_Four_1_2"
rs.VariableName(11)="SD_Four_1"
rs.VariableName(12)="SD_Four_2"

rs.VariableName(13)="Corr_Five_1_2"
rs.VariableName(14)="SD_Five_1"
rs.VariableName(15)="SD_Five_2"

rs.VariableName(16)="Corr_Six_1_2"
rs.VariableName(17)="SD_Six_1"
rs.VariableName(18)="SD_Six_2"
```

```
rs.VariableName(19)="Corr_Seven_1_2"
rs.VariableName(20)="SD_Seven_1"
rs.VariableName(21)="SD_Seven_2"

rs.VariableName(22)="Corr_Eight_1_2"
rs.VariableName(23)="SD_Eight_1"
rs.VariableName(24)="SD_Eight_2"

rs.VariableName(25)="Corr_Nine_1_2"
rs.VariableName(26)="SD_Nine_1"
rs.VariableName(27)="SD_Nine_2"



'Here I set the format to be a real number with two displayed decimal digit precision
For i = 1 To 27
rs.Variable(i).NumberFormat = "#.00"
Next


'Here I set the displayed width of my new spreadsheet
'and then I display it!
rs.Window.Width=700
rs.Visible=True

'This is the control statement for the number of samples to construct and use to acquire parameter
estimates
For mast = 1 To boot

s.Redraw = False

N = s.Cases.Count
sumx=0
sumy=0

'For the first real-valued random variable
For i = 1  To N
s.Cells(i,1)=(RndNormal(2)+4) 'this commands generates random normal variates from a
                                 distribution with mean 4 and SD of 2
sumx=sumx+s.Cells(i,1)
Next
mn=sumx/N

'For the second real-valued random variable
For i = 1  To N
s.Cells(i,2)=(RndNormal(2)+4) 'this commands generates random normal variates from a
                                 distribution with mean 4 and SD of 2
Next
```

```
'Now create correlated variable  {Real_2} with a 0.5 correlation with the first {Real_1}
For i=1 To N
s.Cells(i,2)= 0.5 * s.Cells(i,1) + (Sqr(1-0.5*0.5))*s.Cells(i,2)
sumy=sumy+s.Cells(i,2)
Next
mn1=sumy/N

diff=mn1-mn

'this is where I subtract the difference between means (Real_1-Real_2) from Real_2 values
For i=1 To N
s.Cells(i,2)=s.Cells(i,2)-diff
Next


'This is where I create my categorised variables
'===============================================

'Now produce Likert Integers - two choice - split around mean
For i = 1 To N
If s.Cells (i,1)< 4.0 Then
 s.Cells (i,3)=0
Else
 s.Cells (i,3)=1
End If

If s.Cells (i,2)< 4.0 Then
 s.Cells (i,4)=0
Else
 s.Cells (i,4)=1
End If
Next

'Now produce Likert Integers - Three choice
For i = 1 To N
If s.Cells (i,1)< 2.67 Then s.Cells (i,5)=0
If (s.Cells (i,1)>= 2.67) And (s.Cells (i,1)< 5.33) Then s.Cells (i,5)=1
If s.Cells (i,1) >=5.33 Then s.Cells (i,5)=2

If s.Cells (i,2)< 2.67 Then s.Cells (i,6)=0
If (s.Cells (i,2)>= 2.67) And (s.Cells (i,2)< 5.33) Then s.Cells (i,6)=1
If s.Cells (i,2) >=5.33 Then s.Cells (i,6)=2
Next

'Now produce Likert Integers - Four choice
For i = 1 To N
If s.Cells (i,1)< 2 Then s.Cells (i,7)=0
```

```
If (s.Cells (i,1)>= 2) And (s.Cells (i,1)< 4) Then s.Cells (i,7)=1
If (s.Cells (i,1)>= 4) And (s.Cells (i,1)< 6) Then s.Cells (i,7)=2
If s.Cells (i,1) >=6 Then s.Cells (i,7)=3

If s.Cells (i,2)< 2 Then s.Cells (i,8)=0
If (s.Cells (i,2)>= 2) And (s.Cells (i,2)< 4) Then s.Cells (i,8)=1
If (s.Cells (i,2)>= 4) And (s.Cells (i,2)< 6) Then s.Cells (i,8)=2
If s.Cells (i,2) >=6 Then s.Cells (i,8)=3
Next

'Now produce Likert Integers - Five choice
For i = 1 To N
If s.Cells (i,1)< 1.6 Then s.Cells (i,9)=0
If (s.Cells (i,1)>= 1.6) And (s.Cells (i,1)< 3.2) Then s.Cells (i,9)=1
If (s.Cells (i,1)>= 3.2) And (s.Cells (i,1)< 4.8) Then s.Cells (i,9)=2
If (s.Cells (i,1)>= 4.8) And (s.Cells (i,1)< 6.4) Then s.Cells (i,9)=3
If s.Cells (i,1) >= 6.4 Then s.Cells (i,9)=4

If s.Cells (i,2)< 1.6 Then s.Cells (i,10)=0
If (s.Cells (i,2)>= 1.6) And (s.Cells (i,2)< 3.2) Then s.Cells (i,10)=1
If (s.Cells (i,2)>= 3.2) And (s.Cells (i,2)< 4.8) Then s.Cells (i,10)=2
If (s.Cells (i,2)>= 4.8) And (s.Cells (i,2)< 6.4) Then s.Cells (i,10)=3
If s.Cells (i,2) >= 6.4 Then s.Cells (i,10)=4
Next

'Now produce Likert Integers - Six Choice
For i = 1 To N
If s.Cells (i,1)< 1.33 Then s.Cells (i,11)=0
If (s.Cells (i,1)>= 1.33) And (s.Cells (i,1)< 2.66) Then s.Cells (i,11)=1
If (s.Cells (i,1)>= 2.66) And (s.Cells (i,1)< 4.0) Then s.Cells (i,11)=2
If (s.Cells (i,1)>= 4.0) And (s.Cells (i,1)< 5.33) Then s.Cells (i,11)=3
If (s.Cells (i,1)>= 5.33) And (s.Cells (i,1)< 6.66) Then s.Cells (i,11)=4
If s.Cells (i,1) >= 6.66 Then s.Cells (i,11)=5

If s.Cells (i,2)< 1.33 Then s.Cells (i,12)=0
If (s.Cells (i,2)>= 1.33) And (s.Cells (i,2)< 2.66) Then s.Cells (i,12)=1
If (s.Cells (i,2)>= 2.66) And (s.Cells (i,2)< 4.0) Then s.Cells (i,12)=2
If (s.Cells (i,2)>= 4.0) And (s.Cells (i,2)< 5.33) Then s.Cells (i,12)=3
If (s.Cells (i,2)>= 5.33) And (s.Cells (i,2)< 6.66) Then s.Cells (i,12)=4
If s.Cells (i,2) >= 6.66 Then s.Cells (i,12)=5
Next

'Now produce Likert Integers - Seven Choice
For i = 1 To N
If s.Cells (i,1)< 1.1429 Then s.Cells (i,13)=0
If (s.Cells (i,1)>= 1.1429) And (s.Cells (i,1)< 2.2857) Then s.Cells (i,13)=1
If (s.Cells (i,1)>= 2.2857) And (s.Cells (i,1)< 3.4286) Then s.Cells (i,13)=2
If (s.Cells (i,1)>= 3.4286) And (s.Cells (i,1)< 4.5714) Then s.Cells (i,13)=3
```

```
If (s.Cells (i,1)>= 4.5714) And (s.Cells (i,1)< 5.7143) Then s.Cells (i,13)=4
If (s.Cells (i,1)>= 5.7143) And (s.Cells (i,1)< 6.8571) Then s.Cells (i,13)=5
If s.Cells (i,1) >= 6.8571 Then s.Cells (i,13)=6

If s.Cells (i,2)< 1.1429 Then s.Cells (i,14)=0
If (s.Cells (i,2)>= 1.1429) And (s.Cells (i,2)< 2.2857) Then s.Cells (i,14)=1
If (s.Cells (i,2)>= 2.2857) And (s.Cells (i,2)< 3.4286) Then s.Cells (i,14)=2
If (s.Cells (i,2)>= 3.4286) And (s.Cells (i,2)< 4.5714) Then s.Cells (i,14)=3
If (s.Cells (i,2)>= 4.5714) And (s.Cells (i,2)< 5.7143) Then s.Cells (i,14)=4
If (s.Cells (i,2)>= 5.7143) And (s.Cells (i,2)< 6.8571) Then s.Cells (i,14)=5
If s.Cells (i,2) >= 6.8571 Then s.Cells (i,14)=6
Next


'Now produce Likert Integers - Eight Choice
For i = 1 To N
If s.Cells (i,1)< 1 Then s.Cells (i,15)=0
If (s.Cells (i,1)>= 1) And (s.Cells (i,1)< 2) Then s.Cells (i,15)=1
If (s.Cells (i,1)>= 2) And (s.Cells (i,1)< 3) Then s.Cells (i,15)=2
If (s.Cells (i,1)>= 3) And (s.Cells (i,1)< 4) Then s.Cells (i,15)=3
If (s.Cells (i,1)>= 4) And (s.Cells (i,1)< 5) Then s.Cells (i,15)=4
If (s.Cells (i,1)>= 5) And (s.Cells (i,1)< 6) Then s.Cells (i,15)=5
If (s.Cells (i,1)>= 6) And (s.Cells (i,1)< 7) Then s.Cells (i,15)=6
If s.Cells (i,1) >= 7 Then s.Cells (i,15)=7

If s.Cells (i,2)< 1 Then s.Cells (i,16)=0
If (s.Cells (i,2)>= 1) And (s.Cells (i,2)< 2) Then s.Cells (i,16)=1
If (s.Cells (i,2)>= 2) And (s.Cells (i,2)< 3) Then s.Cells (i,16)=2
If (s.Cells (i,2)>= 3) And (s.Cells (i,2)< 4) Then s.Cells (i,16)=3
If (s.Cells (i,2)>= 4) And (s.Cells (i,2)< 5) Then s.Cells (i,16)=4
If (s.Cells (i,2)>= 5) And (s.Cells (i,2)< 6) Then s.Cells (i,16)=5
If (s.Cells (i,2)>= 6) And (s.Cells (i,2)< 7) Then s.Cells (i,16)=6
If s.Cells (i,2) >= 7 Then s.Cells (i,16)=7

Next

'Now produce Likert Integers - Nine Choice {0.9}
For i = 1 To N
If s.Cells (i,1)< 0.85 Then s.Cells (i,17)=0
If (s.Cells (i,1)>= 0.85) And (s.Cells (i,1)< 1.75) Then s.Cells (i,17)=1
If (s.Cells (i,1)>= 1.75) And (s.Cells (i,1)< 2.65) Then s.Cells (i,17)=2
If (s.Cells (i,1)>= 2.65) And (s.Cells (i,1)< 3.55) Then s.Cells (i,17)=3
If (s.Cells (i,1)>= 3.55) And (s.Cells (i,1)< 4.45) Then s.Cells (i,17)=4
If (s.Cells (i,1)>= 4.45) And (s.Cells (i,1)< 5.35) Then s.Cells (i,17)=5
If (s.Cells (i,1)>= 5.35) And (s.Cells (i,1)< 6.25) Then s.Cells (i,17)=6
If (s.Cells (i,1)>= 6.25) And (s.Cells (i,1)< 7.15) Then s.Cells (i,17)=7
If s.Cells (i,1) >= 7.15 Then s.Cells (i,17)=8
```

```
If s.Cells (i,2)< 0.85 Then s.Cells (i,18)=0
If (s.Cells (i,2)>= 0.85) And (s.Cells (i,2)< 1.75) Then s.Cells (i,18)=1
If (s.Cells (i,2)>= 1.75) And (s.Cells (i,2)< 2.65) Then s.Cells (i,18)=2
If (s.Cells (i,2)>= 2.65) And (s.Cells (i,2)< 3.55) Then s.Cells (i,18)=3
If (s.Cells (i,2)>= 3.55) And (s.Cells (i,2)< 4.45) Then s.Cells (i,18)=4
If (s.Cells (i,2)>= 4.45) And (s.Cells (i,2)< 5.35) Then s.Cells (i,18)=5
If (s.Cells (i,2)>= 5.35) And (s.Cells (i,2)< 6.25) Then s.Cells (i,18)=6
If (s.Cells (i,2)>= 6.25) And (s.Cells (i,2)< 7.15) Then s.Cells (i,18)=7
If s.Cells (i,2) >= 7.15 Then s.Cells (i,18)=8
Next

'Compute sums of squares etc. for Sd and Pearson r
kon=0
kon1=1
For j = 1 To 18 Step 2
sumx=0
sumy=0
sumxy=0
sumx2=0
sumy2=0

col1=j
col2=j+1
kon=kon+1

For k = 1 To N
sumx=sumx+s.Cells(k,col1)
sumx2=sumx2+s.Cells(k,col1)*s.Cells(k,col1)
sumy=sumy+s.Cells(k,col2)
sumy2=sumy2+s.Cells(k,col2)*s.Cells(k,col2)
sumxy=sumxy+s.Cells(k,col1)*s.Cells(k,col2)
Next

'put the parameter values in the appropriate cells of my "results" spreadsheet
pr=(N*sumxy-sumx*sumy)/Sqr((N*sumx2-sumx^2)*(N*sumy2-sumy^2))
rs.Cells(mast,kon1)=pr
mn=sumx/N
rs.Cells(mast,kon1+1)=Sqr((sumx2-N*mn*mn)/(N-1))
mn=sumy/N
rs.Cells(mast,kon1+2)=Sqr((sumy2-N*mn*mn)/(N-1))
kon1=kon1+3
Next 'j

Next 'mast

s.Redraw = True

End Sub
```