



9. Facet Live - Reliable Results

How Reliable is a Facet Profile?

A profile like Facet is only useful if the information it gives is generally reliable. If a person completes the questionnaire a second time will they come up with the same scores (broadly). More importantly will the results be interpreted in broadly the same way and would similar conclusions be drawn?

There are many technical definitions of reliability each of which is subtly different from the others. For example we can divide the test into two halves and calculate scores for each scale based on only half the items - are they similar. This tells us whether the items are evenly spread throughout the questionnaire. This is "Split-Half" reliability. (If there is an odd number of items in each half there are clever adjustments to be made to balance them up.)

A different type of reliability is calculated by sequentially deleting each item in turn and seeing what the effect on the overall score is. This is referred to as Cronbach's α and is a favourite of psychometricians.

Finally there is the one which most people would think of which is to look at two sets of scores on two different occasions. This may appear to be the most intuitively obvious measure but is frequently sneered at by psychometricians. There are after all a number of factors which can influence things.

The most obvious and significant is the time between the first and second administration. If the time delay is very short (say a week or two) then you would expect the person to be in a similar frame of mind and not to have changed his/her core views very much. Therefore the results should be quite similar. This has therefore been called "dependability" rather than reliability. Reliability (or Stability) is generally retained for longer re-test intervals.

The original Facet research sample included a small number of people who completed the questionnaire twice with a reasonable gap of at least 3 months. The results as shown in the Facet manual were very good however the sample was very small ($n=20$). We needed to collect a larger sample and see how the results differed

Doing it again

People complete Facet twice for any number of reasons including:

- ◆ Mistake - they forgot that they had done it before. This is not uncommon with selection candidates.
- ◆ Deception - they just want to see if they can sneak a "better" set of scores. Again selection candidates are more likely to try this especially if they were not appointed the first time - they may believe that the Facet profile influenced the decision.
- ◆ Interest - they want to see if they've changed in some way since the first time
- ◆ Time - it has been a long time since they did it the first time and we wanted to work with more up-to-date information. This would normally be the case if we knew that a person had completed it more than 3 years previously.
- ◆ Mischief - they want to see if they can move the scores in a specific direction. This is often the case with observers on assessment and development programmes where Facet is being used and they have a little spare time.

In an ideal world we would like to see a sample of people who completed the questionnaire and then had nothing to do with Facet until they were asked to complete it again at least year later. This doesn't happen. Again in a perfect world we would like to have nobody who was "playing" with the data or attempting to deliberately distort it. In our sample we knew we had people from each of these groups - we were not always sure who was who.

Creating the Sample

When we amalgamated the databases which we had collected we were able to identify 107 people who had completed the questionnaire twice. Since we actually knew the circumstances of most of these we could say with confidence that the time delay was at least 3 months in all cases and as long as 3 years in some. We were also able to identify some people from the last group who had





9. Facet Live - Reliable Results

deliberately tried to distort the data (we knew because they told us).

We therefore ran the re-test reliability statistics twice. The first time was on the basis of the whole sample of 107 people. The second time was on a reduced sample of 68 with the "mischief-makers" removed. Both sets of results are shown in Table 1.

Table 1

	Will	Energy	Affection	Control	Emotion
Full sample	.70	.64	.63	.73	.61
Reduced sample	.86	.81	.84	.79	.85
Original data	.84	.85	.81	.85	.92

Is this good or bad?

Good question! Statisticians are the worst people to ask about this since they will always say the results need more data/ more analysis or more time. However we need to draw a line somewhere. In the social sciences re-test reliability figures (which are correlations and can therefore range from -1.00 to +1.00) are usually expected to be in excess of 0.6 and if they are too high then the data may be held to be suspect. Out of interest we can compare some other well known instruments and highly reputable which publish their figures. They are shown in Table 2 below.

Table 2

Source	Range	Average
Facet 5.0 (full sample)	.61 - .73	.67
Facet 5.0 (reduced sample)	.79 - .86	.83
MBTI ¹	.60 - .75	.71
MBTI ²	.64 - .78	.70
16PF ³	.28 - .63	.47
16PF ⁴	.21 - .64	.47

We were unable to get long term re-test reliability data for the OPQ although it will certainly be available.

¹ Briggs-Myers, I and McCaulley, MH. "Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator", CPP 1989, Table 10.5 Amherst Class of 1963 - 14-16 months

² Ibid Table 10.5 University of New Mexico - 9-21 months

³ Cattell, RB, Eber, HW and Tatsuoka, MM, "The Handbook for the Sixteen Personality Factor Questionnaire (16PF)", 1970, IPAT p31 Table 5.2 Males - 4 year

⁴ Ibid , P31 Table 5.2 Females - 4 years