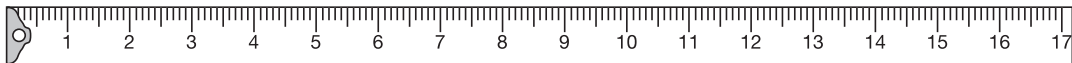


Reliability

Whenever we set out to measure something, we cannot be sure that we are 100% accurate. If we set out to measure the distance between two points, we can only be as accurate as the tape measure we use. We can attempt to use more and more accurate measuring devices, but there will still be some degree of error in our measurement, no matter how small. Any measurement must be seen as an estimate of the true distance, rather than a statement of absolute truth.



If we are making decisions on the basis of our measurements, it is important to know just how accurate our 'tape measure' is. If we know this, we can decide whether our measure is sufficiently precise and consistent for our needs.

Like any measurement device, personality questionnaires are not perfectly accurate. The reliability of a questionnaire is concerned with just how accurate, precise and free from error it is. By knowing the reliability of a personality questionnaire, we can decide whether it is accurate enough for our needs. We can then allow for any potential error when making selection or other decisions.

Reliability is concerned with the extent to which the questionnaire score is a reflection of the individual's true score. The higher the reliability of the questionnaire, the more likely it is to measure the individual's true score.

Factors affecting reliability

The reliability of a personality questionnaire is affected by a number of factors that have an impact on the way the respondent completes the questionnaire. This variation can come from the way the questionnaire has been constructed, and also the way the questionnaire is administered.

When using psychometric questionnaires, there are various sources of error that will lead to inaccuracy in questionnaire results. These can be broadly categorised into three areas:

1. Condition of the respondent

- > The respondent's mood, their level of experience and how anxious they are
- > Their understanding of the process and how the data is going to be used
- > Their frame of reference (for example, thinking of themselves with their 'work hat' or 'home hat' on)

2. Questionnaire administration and scoring

- > The degree to which instructions are clear, unambiguous and follow the standard format
- > The mood and tone set by the individual carrying out the administration of the questionnaire
- > Environment: lighting, noise and other physical conditions
- > Careful attention to scoring

3. Questionnaire content

- > There should be enough questions in each scale
- > The questions should be:
 - Easily understood
 - Unambiguous
 - Without double negatives
 - Relevant to what the questionnaire is trying to measure

Good questionnaire administration practice by the administrator (including thorough preparation of candidates, careful attention to a suitable introduction, clear questionnaire administration and accurate scoring) is therefore crucial in reducing error and producing accurate results (best practice in administration and scoring of the FIRO questionnaire are covered in Chapter 1). It is also important that the questionnaire publisher has carefully reviewed the questionnaire content and statistically measured its reliability.

Measuring reliability

There are a number of methods for calculating the reliability of a questionnaire, each looking at a different aspect of accuracy. It is important when interpreting reliability information to know which type of reliability you are looking at and what it measures. The methods generally used to estimate the reliability of a questionnaire involve the calculation of the correlation between two or more sets of scores on the same or similar questionnaires for the same group of individuals.

There are set standards for correlations in reliability studies. These reliability coefficients judge how 'good' the correlation coefficients are.

The standards are as follows:

- > 0.80 and above: excellent
- > 0.70–0.79: good

- > 0.60–0.69: fair
- > below 0.60: poor

Probably the most common kind of reliability measure is known as **test-retest reliability**. This concerns the likelihood of coming out with the same profile when the questionnaire is taken again. Another form of reliability is known as **internal consistency**. This looks at how people respond on just one occasion, and concerns the degree to which people's responses to the items relating to a single scale are consistent.

The 'popcorn machine' explanation of reliability

Test-retest

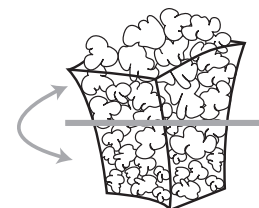
If we are testing the reliability of a popcorn machine, we can test the popcorn it produces at Time 1 and again at Time 2 to see whether the machine is consistently producing the same popcorn over a period of time.



Internal consistency

We can also test whether the popcorn in one pot is the same by:

- a) splitting the pot in half and checking that one half of the popcorn is consistent with the other half (split-half reliability).



OR

- b) by checking each individual piece of popcorn is consistent with each of the other pieces (alpha coefficient).



Some questionnaires have more than one version – two different sets of questions that measure the same attributes. These are usually known as alternate forms or parallel forms. We can establish the alternate form reliability of these by getting a group of people to complete both questionnaires and then correlating the two scores together. There is only one European English form of the FIRO-B questionnaire and one version of the FIRO Business questionnaire, so alternate form reliability is not relevant for this questionnaire.

Internal consistency

One key question when we look at the reliability of a questionnaire concerns the content of the questions in the questionnaire. Do all of these items ask questions about the same kinds of concepts, or do they look at widely different things? In other words, do all the items measure the same thing? The more disparate the item content, the less precise a measuring device the questionnaire will be, and hence the lower the internal consistency reliability.

The simplest way of calculating internal consistency reliability is to use the method known as **split-half reliability**. This calculation requires that a group of people take the questionnaire; their score on one half of the questionnaire is then calculated and correlated with their score on the other half. Again, the higher the correlation, the more reliable the questionnaire is deemed to be. One of the issues with split-half reliability is working out where to split the questionnaire. Often, this is done by correlating the score on 'odd-numbered' items with the score on 'even-numbered' items. However, the questionnaire could be split in many different ways, and each one would give a slightly different result. More sophisticated methods of calculating internal consistency, such as **Cronbach's Alpha** (also known as the alpha coefficient), effectively give the same result as would be obtained if the questionnaire was split in all possible ways and the average calculated.

Split-half reliability of the FIRO-B questionnaire

The FIRO-B instrument has been tested for internal consistency, and the split-half reliability coefficients for all scales reached fair to excellent standards of reliability. This data is presented in Table 2.1 of the *FIRO-B European Data Supplement* (2nd edition, June 2016, OPP Ltd – see www.opp.com).

Split-half or alpha coefficients are the usual way of establishing the internal consistency of a questionnaire. The FIRO-B questionnaire is constructed using Guttman scaling (see Chapter 2 for more information on the questionnaire construction). With Guttman scales it is more appropriate to use **reproducibility** to look at the internal consistency of the scales. Reproducibility is a more stringent measure of reliability as it not only looks at all items measuring the same dimension, but also looks at the order of the items.

Reproducibility of the FIRO-B questionnaire

The usual requirement for reproducibility is that 90% of all responses are predictable from knowledge of the overall scale scores. The reproducibility of all FIRO-B scales achieved the 0.90 standard or better. These are reported in Table 2.2 in the EDS. These reproducibility scores are coefficients of internal consistency for the FIRO-B instrument.

FIRO-B test-retest reliability

Test-retest reliability is a measurement of the consistency of results on the same questionnaire scale over time. It is calculated by testing a group of individuals, waiting a suitable interval, testing them again, and then correlating the results on the first occasion with the results on the second. The strength of this correlation is a measure of how consistent a particular test is over time.

Test-retest reliability of the FIRO-B questionnaire

The data found in Table 2.3 of the EDS shows that over the time interval of 2–4 weeks the data remained consistent, and when compared with the reliability coefficients it was good to excellent for each scale.

FIRO Business test-retest reliability

Table 5.1

FIRO Business Scale	Test-Retest Correlation
Expressed Involvement	.79
Wanted Involvement	.72
Expressed Influence	.72
Wanted Influence	.66
Expressed Connection	.74
Wanted Connection	.70
Mean	.72

Note: N = 241; $p < .01$. Elapsed time = 7-70 weeks.

Reference: *FIRO Business: Technical Guide*. Herk, N., Thompson, R., Morris, M., Schaubhut, N. (2009, CPP, Inc.).

The effect of different frames of reference on FIRO-B scores

As the FIRO-B questionnaire looks at behaviour, it is not surprising that research indicates that people respond differently when considering themselves at home or at work. This is known as the 'frame of reference' used when completing the questionnaire. The effect of this is consistent with the view that most people will adapt somewhat to the situation they are in.

The EDS reports two studies that specifically look at the frame of reference that respondents use when completing the FIRO-B questionnaire. These studies show that scores on Inclusion and Affection are typically higher within a home frame of reference and scores on Control are typically higher within a work frame of reference (Chapter 2 in the EDS).

The important thing to be aware of is that frames of reference can have an impact on the scores in the FIRO-B questionnaire and that the most appropriate frame of reference should be specified during administration. See Chapter 1 for an administration checklist and a sample email/letter for respondents.

Comparing the FIRO Business with the FIRO-B instrument

The results from the FIRO-B instrument are expected to differ from those from the FIRO Business instrument, because the raw scores are being compared against different norm groups. The FIRO-B instrument is compared against a UK representative sample, whilst the FIRO Business instrument is compared against the international employed sample.

Below are the correlations between the FIRO-B scale scores and the corresponding FIRO Business scale scores from research conducted using the FIRO Business International Sample.

Table 5.2

FIRO Business/FIRO-B category	Correlation
Expressed Involvement/Inclusion	0.86
Wanted Involvement/Inclusion	0.87
Expressed Influence/Control	0.91
Wanted Influence/Control	0.90
Expressed Connection/Affection	0.83
Wanted Connection/Affection	0.86

There is a high level of correlation between the two sets of results, but there will be some differences in the interpretation categories into which results fall for one versus the other instrument.

Reference: *FIRO Business: Technical Guide*. Herk, N., Thompson, R., Morris, M., Schaubhut, N. (2009, CPP, Inc.).

Reliability – summary

Test-retest reliability looks at the consistency in results of the FIRO-B instrument over time. We can ask whether people come out with the same profile scores when completing the questionnaire for a second time within a pre-determined time interval. Such studies show that the test-retest reliability of the FIRO-B instrument is typically good to excellent.

Internal consistency looks at the consistency of results on a single occasion, by correlating the results of two halves of each continuous scale (split-half) or by doing a similar but more rigorous procedure (alpha coefficient). Such studies suggest that the internal consistency of the FIRO-B instrument is fair to excellent.

It is more appropriate with the FIRO-B instrument, which is designed using Guttman scaling, to look at reproducibility. The reproducibility of all six scales in the FIRO-B questionnaire achieves or exceeds the 0.90 standard.