Cognadev Technical Report Series

2

20th August, 2014

The Retest Reliability of the Values Orientations (VO)

Assessing the retest reliability of the VO, where the test results to be compared are two ordered-category sequences of selected and rejected orientations, consisting of up to three orientations per sequence (*Accepted Values*) and one or two orientations (*Rejected Values*). A new computational comparison analysis algorithm was constructed to work with ordered category sequences, generating a percentage match index varying between 0% (no agreement) to 100% (absolute identity).

cognadev

Paul Barrett, PhD.
Chief Research Scientist

18B Balmoral Ave
Hurlingham Sandton 2196
Johannesburg
South Africa

Tel: +27 11 884-0878
Fax: +27 11 884-0910

19 Carlton Road
Pukekohe
Auckland 2120
New Zealand

Tel: +64 9 889-2630
Mob: +64 21-415625
Fax: +64 9 280-6121
Skype: pbar088

www.cognadev.com paul@cognadev.com

Executive Summary

1 In total, 101 cases of data formed the dataset for these analyses. The majority of respondents were candidates for job roles where the VO happened to form part of an assessment strategy for an organization. 43 of these possessed values strength and separability indices.

2 For the purposes of retest analyses, the 101 sample was divided into 4 interval duration groups:

- < 6 months (n=28)
- < 1 year (n=38)
- between 1 and 2 years (n=36)
- between 2 and 4 years (n=27)

3 Exact class-category agreement of values orientation sequences is somewhat unrealistic, as transitional boundaries between values orientations are more likely to be somewhat fuzzy rather than discrete. To allow for this, a matching index was constructed which only took into account the proximity of the constituents of orientation sequences to one another, with a match recorded (or not) based upon the application of 8 conservative rules to the two-occasion results.

4 Orientation sequence matches (reliability) were:

Duration	Sample Size	Accepted Orientations	Rejected Orientations
< 6 months	28	68%	71%
< 1 year *	38	68%	63%
Between 1 and 2 years	36	64%	50%
Between 2 and 4 years	27	59%	41%

* includes the cases of < 6 months duration

5 Strength and Separability retest reliability (over 43 cases, interval duration spanning 0 – 1043 days):

	Occasion 2 Strength				Occasion 2 Separability			
	Accepted		Rejected		Accepted		Rejected	
	r	g	r	g	r	g	r	g
Occasion 1	.72	.91	.66	.88	.34	.88	.41	.89

* **r** = Pearson correlation, **g** = Gower agreement (see Appendix 1)

Both separability and strength indices are very similar to one another in magnitude across occasions, although the separability indices show decreasing monotonicity as the interval duration increases.

Overall Conclusions

- The short-term retest analysis (6 months or less) shows a reasonable consistency between orientation sequences, but as the duration increases, so do the overall number of matches decrease. Of particular relevance is the more rapid decline of rejected orientation matches as retest interval increases.

- The strength of accepted and rejected orientations is shown to be reasonably stable over at least one year's duration. Likewise the relative separability of the orientations

- It is likely that current matching rules are too conservative, as they are based solely on values-proximity relations rather than a combination of these and a theory-based semantic rule-set. This is an area for further investigation.

Contents

Tables

Figures

1. Values Orientations

1.1 Their description

The Value Orientations (VO) reveals an individual's worldviews, their assumptions about life and perceptual orientations. Value systems represent "core intelligences" and act as a decision-making framework that guides behaviour and life choices. Value systems thus provide a structure for thinking, act as organizing principles, and guide an individual's modes of adaptation to the world.

Figure 1: The Values Orientations and their description

Valuing system	Acceptance	Rejection
Purple	Values group belonging; finds safety and security in the familiar; tends to be attached to traditions/customs; us-versus-them orientation.	Questions the tendency to be too reliant on in-groups; not concerned with the preservation of traditions/customs; sceptical of the 'us-versus-them' mentality.
Red	Energetic, forceful, could be impulsive; ego-centric; wants to be recognised and respected.	Rejects a forceful, impulsive and dominant approach; does not see life as battle to secure an own share; questions self-centred behaviour.
Blue	Controlled; values order and discipline; dutiful and diligent; wants to do the 'right' thing; values and ultimate truth.	Rejects the overemphasis on conformity, order, discipline and authority; guards against absolutist and judgemental inclinations.
Orange	Achievement / performance oriented; self-reliant; values success and 'the good life'; works with perceptions; motivated by challenge and opportunity; takes calculated risks.	Rejects an overemphasis on personal achievement, status symbols, competition and material wealth; may find the quest for the 'good life' superficial; dislikes manipulation.
Green	Humanistic; energised by interpersonal relationships; sensitive; compassionate; philosophical; relativist; open-minded; idealistic.	Questions an over-emphasis on the human factor; not energised by interpersonal relationships; not motivated by charitable endeavours; not gullible or overly accepting.
Yellow	Individualistic; has an intellectual perspective; often emotionally detached; capacity to deal with unstructured situations; systems thinking; focuses on practical utility.	Not particularly learning oriented; not comfortable with disorder and unstructured situations; not particularly individualistic; not inclined to take a detached, intellectual stance.
Turquoise	Self-transcendent; reflective; holistic thinker; spiritual; guided by a higher consciousness; planetary concerns; focuses on experiencing life.	Rejects an essentially spiritual and abstract approach to life; not inclined to adopt a philosophical-existential view on reality; not interested in the meta-physical realm.

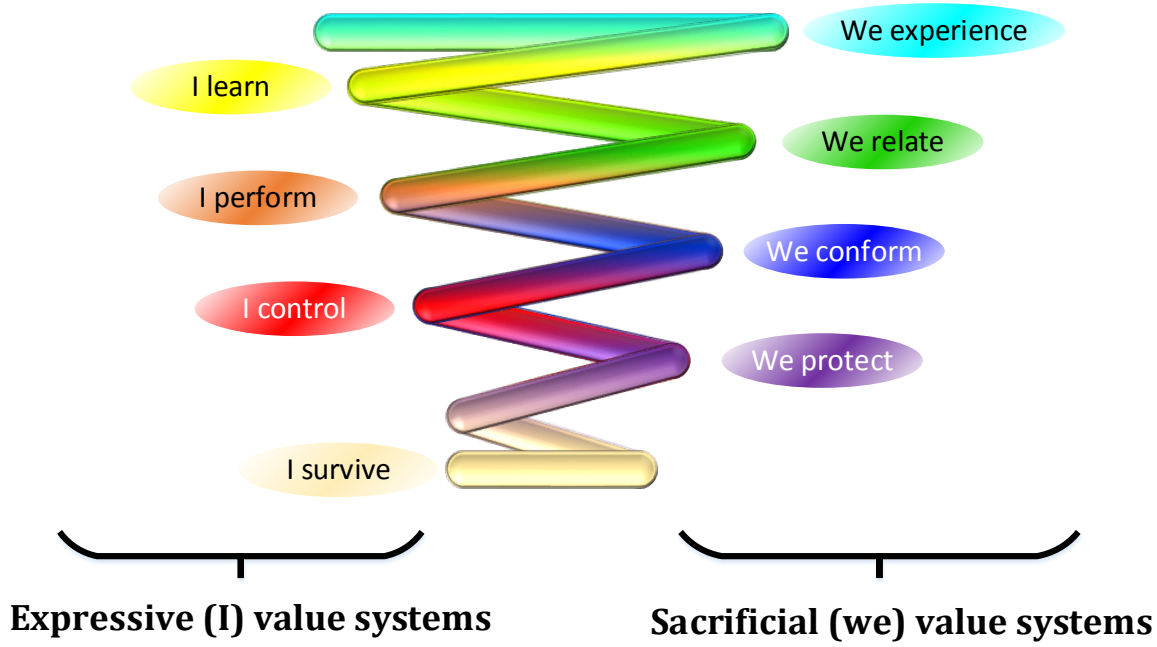
The VO draws – albeit not exclusively – from a body of knowledge (broadly referred to as "Spiral Dynamics") generated by Prof. Clare Graves[1], refined and popularised by Don Beck and Chris Cowan, and critically discussed by various theoreticians (e.g. Ken Wilber).

Spiral Dynamics (SD) focuses on the evolution or development of individuals, organisations and societies specifically in terms of value systems. According to this approach, human society has undergone a number of fundamental changes, evolving from values centred on mere survival, to, for example, value systems supporting a more holistic, integral vision of the world. In fact, a central tenet of SD is that, development-wise, a number of levels or stations can be pinpointed, representing different value systems. These levels or stations are represented in terms of a spiral. To the proponents of SD, the development of human societies can be likened to a spiralling evolution from one station to another. Although the notion of development or evolution is inherent in SD, its authors are quick to point out that one station is not necessarily "better" or "worse" than another. This prompted them to refrain from *ranking* the levels/stations.

To simply the alphabetic annotation used within Spiral Dynamics Theory, Colours were introduced to stand for the levels or 'stations'. The VO assesses seven of these as the values orientations/valuing systems for an individual, as noted in Figure 1 above.

Each of these colours has a particular credo referring to either an expressive, internally controlled "I" (the "warm" colours), or a self-sacrificing, externally anchored "we" side (the "cooler" colours). Superimposed on a spiral, a model for these eight value systems appears as follows:

Figure 2: The values spiral



Of relevance to the issue of retest reliability assessment is the theory-claim that the ranking or ordering of these value systems should not be taken too literally or seen as rigid, linear, step-by-step progressions, for the following reasons:

[1] As a theoretical approach, Spiral Dynamics was devised by Prof. Clare W. Graves during the 1950s and further elaborated on by Don Beck and Chris Cowan. The aim of this discussion is not to give a full account of Graves' entire theory, but to extract those elements relevant to the VO. Graves' theory goes well beyond the focus on the individual's value systems (as discussed here) to include the evolution of societies over time – an aspect not included here.

- adherence to any value system is dynamic and individuals can move up *or* down the spiral depending on the particular problem of existence and coping mechanisms
- more often than not an individual adheres to a complex combination of value systems. It is, therefore, overly simplistic to categorise an individual as, for example, *Blue* or *Red*
- stress or trauma could have the effect that an individual abandons a particular valuing system (e.g. *Yellow*) temporarily in order to cater for emotional needs that are better met at a "lower" level (e.g. *Red*)

1.2 How the VO assessment results are reported

1.2.1 Accepted and Rejected orientations

The test report identifies from one to three "Accepted" values orientations. The mean orientation scores are used to form the Accepted and Rejected orientation-selections via a threshold algorithm (i.e. the highest valued orientation is chosen first, then the next selected whose magnitude is within a **10%** percentage discrepancy from the highest orientation value). Using this simple threshold rule:

- a maximum of 3 values orientations are selected for the Accepted Orientations, and
- a maximum of 2 for the Rejected Orientations.

More details on the scoring are provided in Technical Manual for the VO.

1.2.2 Values strength and separability

Two new indices are computed from the mean orientation scores, **strength** and **separability**; where strength is an indication of the magnitude of the set of selected accepted or rejected values, and separability is an indication of the spread/variability of the entire sets of accepted or rejected values. Both are expressed within a 0 to 100 range, where for strength, higher values indicate more strongly expressed preferences for a values orientation. A high separability score indicates more clarity of separation between the score magnitudes of values orientations. That is, some individuals show very little variation between the magnitudes of their values orientations, whereas others show marked preferences between them. This information is helpful for the practitioner when discussing selected and rejected orientation choices with an individual. More details on the computation of these two indices is provided in Technical Manual for the VO.

---

2. The sample and assessment occasion durations

In total, 101 cases of data formed the dataset for these analyses. The majority of respondents were candidates for job roles where the VO happened to form part of an assessment strategy for an organization. Searching through the entire VO database, these 101 cases were identified and confirmed as 'the same individual'.

Table 1: Descriptive statistics of the retest sample

Variable	Descriptive Statistics (VO retest, n=101, occasion ordered - Final, 14-Aug-14.sta)								
	Valid N	Mean	Median	Minimum	Maximum	Lower Quartile	Upper Quartile	Std.Dev.	Skewness
Interval	101	504.7	478	0	1384	119	754	389.6	0.41
Age at occ1	41	36.2	34	25	58	30	43	8.4	0.92

* The median duration of the interval (in days) between the two test occasions was 478 days.
Ages at the 1st occasion testing were only available for 43 respondents.

Table 2: Gender frequencies in the retest sample

Category	Frequency table: gender (VO retest, n=101, occasion ordered - Final, 14-Aug-14.sta)			
	Count	Cumulative Count	Percent	Cumulative Percent
Male	58	58	57.43	57.43
Female	43	101	42.57	100.00
Missing	0	101	100.00	100.00

Table 3: Frequency distribution of assessment interval durations

From To	Frequency table: Interval (VO retest, n=101, occasion ordered - Final, 14-Aug-14.sta)			
	Count	Cumulative Count	Percent	Cumulative Percent
x=0	3	3	2.97	3.0
0 <x<=100	19	22	18.81	21.8
100 <x<=200	7	29	6.93	28.7
200 <x<=300	5	34	4.95	33.7
300 <x<=400	9	43	8.91	42.6
400 <x<=500	11	54	10.89	53.5
500 <x<=600	7	61	6.93	60.4
600 <x<=700	10	71	9.90	70.3
700 <x<=800	8	79	7.92	78.2
800 <x<=900	4	83	3.96	82.2
900 <x<=1000	4	87	3.96	86.1
1000 <x<=1100	6	93	5.94	92.1
1100 <x<=1200	2	95	1.98	94.1
1200 <x<=1300	3	98	2.97	97.0
1300 <x<=1400	3	101	2.97	100.0
1400 <x<=1500	0	101	0.00	100.0
Missing	0	101	0.00	100.0

* Three cases were assessed twice on the same day.

The maximum duration interval between test occasions was 1384 days (3.79 years).

For the purposes of retest analyses, the 101 sample was divided into 4 interval duration groups:

1. < 6 months (n=28)
2. < 1 year (n=38)
3. between 1 and 2 years (n=36)
4. between 2 and 4 years (n=27)

Clearly, with only 101 cases in the total sample being spread across such a wide set of occasion intervals, any sub-group analysis must be considered as indicative rather than definitive.

3. Computational Details

3.1 The reliability of ordered class sequences

Reliability here is defined by repeatability. That is, the extent to which a second assessment of values orientations and the accepted or rejected orientation sequence deviates from the first occasion results. If they are exactly the same, there is perfect reliability. In engineering terms, this direct assessment of reliability is referred to as repeatability. But, there is no known coefficient for indexing the agreement between ordered-class sequences, except that which indicates the exact agreement between orientation selections over occasions. However, as noted above in section 1.1, exact class-category agreement is somewhat unrealistic, as transitional boundaries between values orientations are more likely to be somewhat fuzzy rather than discrete. To allow for this, a conservative matching index was constructed which only took into account the proximity of the constituents of orientation sequences to one another, with a match recorded (or not) based upon the application of 8 rules to the two-occasion results. The overall match agreement was expressed as the percentage of cases assigned as 'matching' (100% would indicate all cases match across occasions, according to the matching 'rules').

3.1.1The fuzzy-match comparison algorithm

❶ Find exact sequence matches, count these as a match.

❷ Where there is only one accepted orientation for the target and comparison orientation, but they don't match exactly: If they are adjacent colours, count as match.

❸ Where two of two target accepted orientations match two out of three in a comparison sequence, then count as a match.

❹ Where only one accepted orientation is in the target sequence, but there are two in the comparison sequence, and the target matches one of the pair in the comparison sequence: If the non-matching comparison is not an adjacent colour to the second 'matched' comparison, count as mismatch.

❺ Where there are two accepted orientations in target sequence, but just one in the comparison sequence, and the comparison matches one of pair in the target sequence: If the non-matching target is not an adjacent colour to the second 'matched' target, count as mismatch.

❻ Where there are two accepted orientations in both target and comparison pairs, but only one of the pairs matches, then: if the difference between the spiral positions of the two non-matching orientations is > 1 (i.e. non-adjacent), count as mismatch.

❼ Where there are three accepted orientations in either the target and comparison pairs, but there is only one orientation match between them: count as mismatch.

❽ Where there are three accepted orientations in either the target and comparison pairs, there are two matches between two of them, then: If the non-matching comparison orientations are not an adjacent colour to each other, count as mismatch.

The criteria for mismatch were strict and did not employ semantic rules; but according to the VO theory, the "higher" level worldviews largely dominate the lower level perceptions - except in stressful situations. Very often some of the lower level worldviews emerge because of stressful situations but these are not the dominant value systems of the person. People also tend to grow to the next higher level – either from Individualistic e.g. Red to group-oriented e.g. Blue; or from Individualistic e.g. Red to Individualistic e.g. Orange. When exposed to trauma, the worldview usually temporarily drops into a more defensive and fear driven orientation.

3.1.2 The reliability of strength and separability

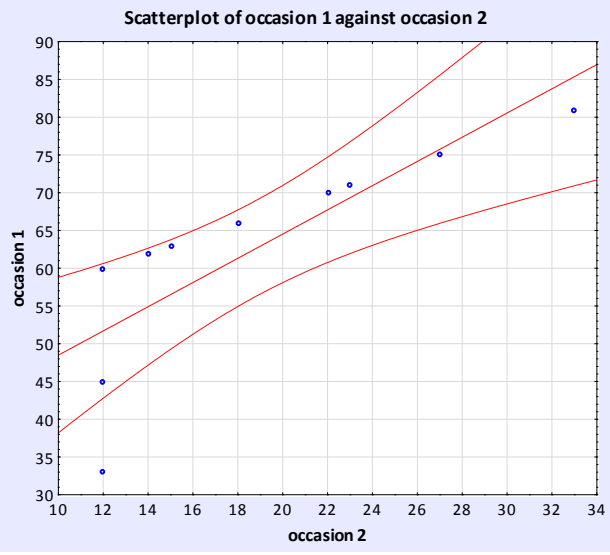
Both orientations strength and separability are expressed over a 0 to 100 real-valued number range. Although the implied precision of this metric seems impressive, it has arisen as a result of the inertial-function scoring algorithm applied to the assessment responses, and the subsequent derivations of both strength and separability indices. It is important to bear in mind the rather more diffuse nature of the psychological attributes being assessed, where the numerical operations result in convenient representations of magnitudes, but where their precision (the real-value number system) most likely exceeds what is actually measurable as 'quantities'. However, these arithmetic operations do enable objective and sense-making scoring or responses, and derived parameter estimations of indices such as strength and separability.

Therefore, given the metric of these two attributes, the monotonic relationship between the two-occasion estimates of strength and separability, respectively, was assessed using a Pearson correlation. In a very real sense, the Pearson correlation is only of marginal interest in a retest scenario, as it can only speak to the monotonic relationship of transformed observations, where observation magnitude is removed by the standardization transformation. Whereas, the critical question to be answered within the retest scenario is: "how similar are the observations (attribute scores) to one another over the two occasions?" That question requires the preservation of the observation metric, because actual observed magnitude matters now. The index of choice for answering that question is the Gower[2] coefficient. See Appendix 1. For example, consider the data in Table 4:

Table 4: Example dataset showing the contrast between monotonicity and agreement

	1 occasion 1	2 occasion 2
1	60	12
2	63	15
3	62	14
4	71	23
5	33	12
6	45	12
7	66	18
8	70	22
9	81	33
10	75	27

Figure 3: The example dataset scatterplot



The Pearson correlation for these data is: **0.82**. statistically significant at $p < 0.01$ (2-tail).

The Gower index for the same data is **0.56**; statistically non-significant $p \sim 0.92$.

The Gower coefficient is indicating that relative to the maximum possible absolute (*unsigned*) discrepancy between them, the observations agree on average to within 56% of each other's values. The bootstrap 'significance' indicates that a value of 0.56 or greater can be observed with random data at least 92% out of 10,000 same-size samples (of 10 cases who values can range between 0 and 100) of random data

The mean absolute deviation between occasion observations is 43.8; clearly there is no retest 'reliability' here. The discrepancy between occasion observations is huge.

This admittedly constrained example shows why a Pearson correlation can be misleading in retest scenarios, except as an indicator of monotonicity.

[2] Gower, J.C. (1971). A general coefficient of similarity and some of its properties. Biometrics, 27, 857-874.

4. The Results

4.1 Orientation Sequences

Table 5 reports the results from applying the matching rules presented in section 3.1.1.

Table 5: Retest match % for accepted and rejected orientation sequences, 4 durations

Duration	Sample Size	Accepted Orientations	Rejected Orientations
< 6 months	28	68%	71%
< 1 year *	38	68%	63%
Between 1 and 2 years	36	64%	50%
Between 2 and 4 years	27	59%	41%

* includes the cases of < 6 months duration

The short-term retest analysis (6 months or less) shows a reasonable consistency between orientation sequences, but as the duration increases, so do the overall number of matches decrease. Of particular relevance is the more rapid decline of Rejected Orientation matches as retest interval increase.

It is of interest to view a sample of the mismatches (for the shortest and longest durations), as some might be considered 'close' (in terms of interpretation), although not meeting the more restricted rule-set for defining a match. The current matching rules are based upon 'closeness' being defined by spatial proximity within the values spiral, rather than on semantic considerations.

Table 6: Mismatch cases, accepted orientations, < 6 months retest duration

	Mismatch case #	occ 1 accepted	occ 1_ accepted	occ 1 accepted	occ 2 accepted	occ 2 accepted	occ 2 accepted
1	1	Yellow	Blue		Red	Orange	Blue
2	3	Orange			Orange	Red	
3	4	Orange	Red		Blue	Green	
4	8	Green	Turquoise		Turquoise		
5	9	Red	Orange		Red		
6	13	Red	Blue	Orange	Orange		
7	18	Blue			Purple	Red	Blue
8	25	Orange			Orange	Green	Turquoise
9	28	Blue			Purple	Blue	

Table 7: Mismatch cases, rejected orientations, < 6 months retest duration

#	Mismatch case #	occ 1 rejected	occ 1 rejected	occ 2 rejected	occ 2 rejected
1	1	Purple		Purple	Turquoise
2	6	Turquoise		Turquoise	Purple
3	14	Green		Orange	Turquoise
4	15	Blue		Purple	
5	19	Orange		Yellow	
6	20	Green	Turquoise	Turquoise	
7	24	Purple	Red	Red	Turquoise
8	27	Red		Turquoise	

Table 8: Mismatch cases, accepted orientations, between 2 and 4 years' retest duration

	Mismatch case #	occ 1 accepted	occ 1 accepted	occ 1 accepted	occ 2 accepted	occ 2 accepted	occ 2 accepted
1	2	Orange	Purple		Blue	Turquoise	Red
2	4	Green	Blue		Green		
3	11	Blue	Green		Blue		
4	12	Red	Orange		Orange	Green	
5	13	Orange	Yellow		Orange		
6	14	Blue			Green		
7	17	Red	Yellow		Green	Yellow	
8	19	Green	Yellow		Orange		
9	21	Red	Yellow		Orange	Yellow	
10	22	Orange	Yellow		Orange		
11	23	Blue			Blue	Yellow	

Table 9: Mismatch cases, rejected orientations, between 2 and 4 years' retest duration

#	Mismatch case #	occ 1 rejected	occ 1 rejected	occ 2 rejected	occ 2 rejected
1	1	Purple		Purple	Turquoise
2	2	Yellow		Turquoise	Red
3	3	Yellow		Purple	
4	4	Yellow		Red	Purple
5	8	Purple		Turquoise	
6	10	Purple	Turquoise	Turquoise	
7	11	Red	Yellow	Green	Yellow
8	12	Purple	Turquoise	Purple	
9	14	Yellow		Red	
10	16	Purple	Turquoise	Red	
11	17	Purple	Turquoise	Red	
12	20	Purple	Yellow	Yellow	Turquoise
13	21	Purple		Purple	Turquoise
14	22	Purple		Purple	Turquoise
15	23	Purple	Turquoise	Purple	
16	26	Green		Turquoise	

4.2 Strength and Separability

Only 43 cases possessed the new derived attributes indices (as these are new features of the VO scoring system). Interval durations spanned. Figure 4 provides the histogram of interval durations within this dataset. As can be seen from this figure, the majority of cases (70%) possess interval durations exceeding 7 months.

Figure 4: Histogram of interval durations for n=43 sample possessing strength and separability indices.

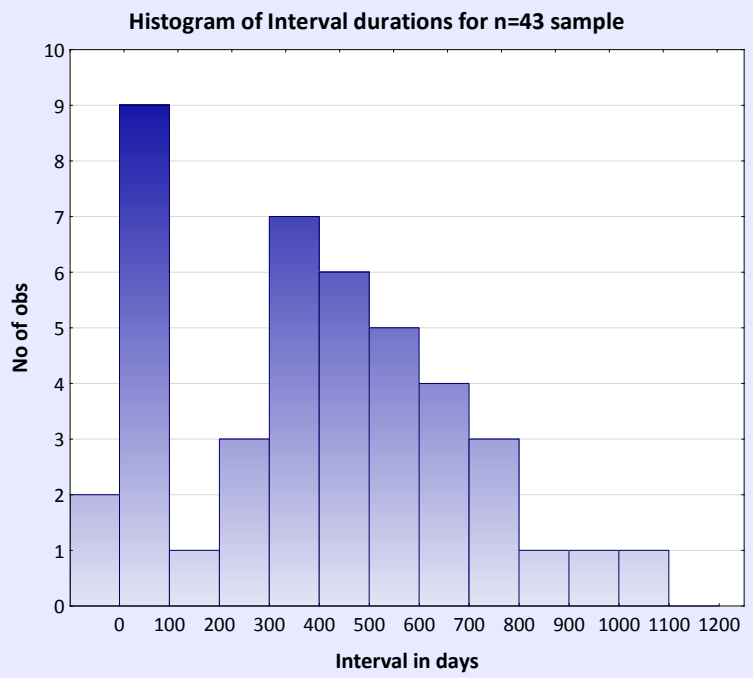
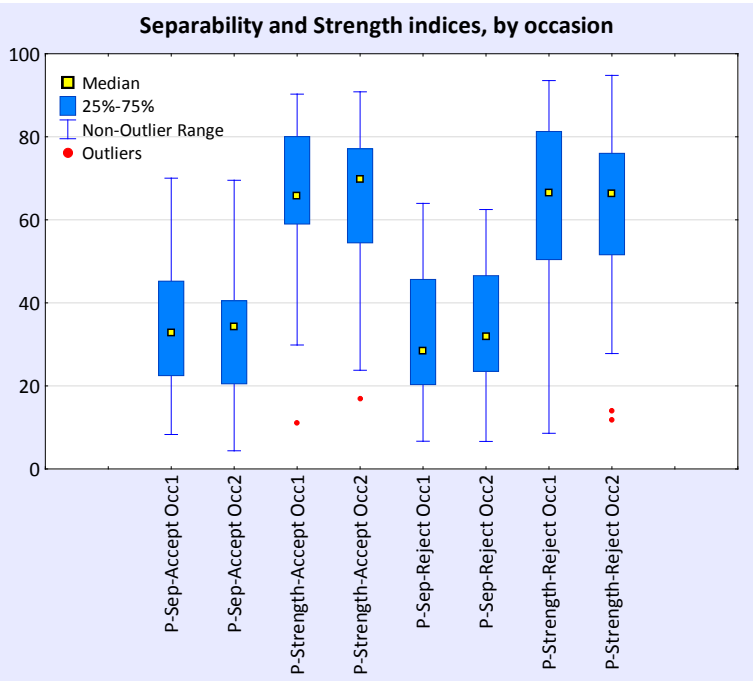


Table 10: Pearson and Gower retest indices for Strength and Separability (n=43 cases)

	Occasion 2 Strength				Occasion 2 Separability			
	Accepted		Rejected		Accepted		Rejected	
	r	g	r	g	r	g	r	g
Occasion 1	.72	.91	.66	.88	.34	.88	.41	.89

* **r** = Pearson correlation, **g** = Gower agreement

Figure 5: Box plot of separability and strength indices by occasion



Both separability and strength indices are very similar to one another in magnitude across occasions, although the separability indices show decreasing monotonicity as the interval duration increases.

Appendix 1: The Gower Agreement Coefficient

Relative to the maximum possible absolute (*unsigned*) discrepancy between the two pairs of observations, the Gower *discrepancy* coefficient indicates the % average absolute discrepancy between all pairs of observations. When expressed as a similarity coefficient (by subtracting it from 1), it indicates the % average similarity between all pairs of observations.

So, a Gower similarity coefficient of say 0.90 indicates that relative to the maximum possible absolute (*unsigned*) discrepancy between them, the observations agree on average to within 90% of each other's values.

If you change the value of that maximum possible discrepancy, then the Gower coefficient will change to reflect this, as the discrepancies between pairs of observations are divided (scaled) by that maximum possible discrepancy value. E.g. if two observations differ by 5, and the measurement range of each observation is 10, then the relative discrepancy is 0.5. However, if the measurement range for each observation was say 100, then the relative discrepancy would be just 0.1.

But that's the whole point of the Gower, it tells you how discrepant (or similar) observations are, RELATIVE to how maximally discrepant they could have been.
A 5-point difference in a 10-point maximum measurement range is substantial.
A 5-point difference between observations within a 100-point measurement range is trivial.

The equation for the Gower similarity index is:

$$Gower_{similarity} = 1 - \left[\frac{\sum_{i=1}^{n} \left(\frac{|obs_{1i} - obs_{2i}|}{range} \right)}{n} \right]$$

$n =$ the number of cases

$range =$ the maximum possible discrepancy between the two attribute/variable magnitudes (100-0)

$obs_{1i} =$ the observed value for case i of n on the first occasion

$obs_{2i} =$ the observed value for case i of n on the second occasion

A free-to-download computer program for computing the Gower, along with a free bootstrap program to compute its statistical significance *(in terms of the likelihood of observing a coefficient as large as computed by chance alone)* are available from:

http://www.pbarrett.net/Gower/Gower.html and http://www.pbarrett.net/Bootstrap/Bootstrap.html