Cognadev Technical Report Series

t

Revised
8th Aug, 2017

Retest Reliability of the Cognitive Process Profile (CPP)

Study 1: using a sample of 87 students undertaking an Accounting degree course at a South African university

Study 2: using a sample of 2,724 respondents comprised primarily of job applicants who had completed the CPP on two separate occasions, but also included some students, and attendees at CPP training courses.

Cognadev UK Ltd., 2 Jardine House, Harrovian Business Village, Harrow, Middlesex, HA1 3EX. Registered UK # 05562662

Contents

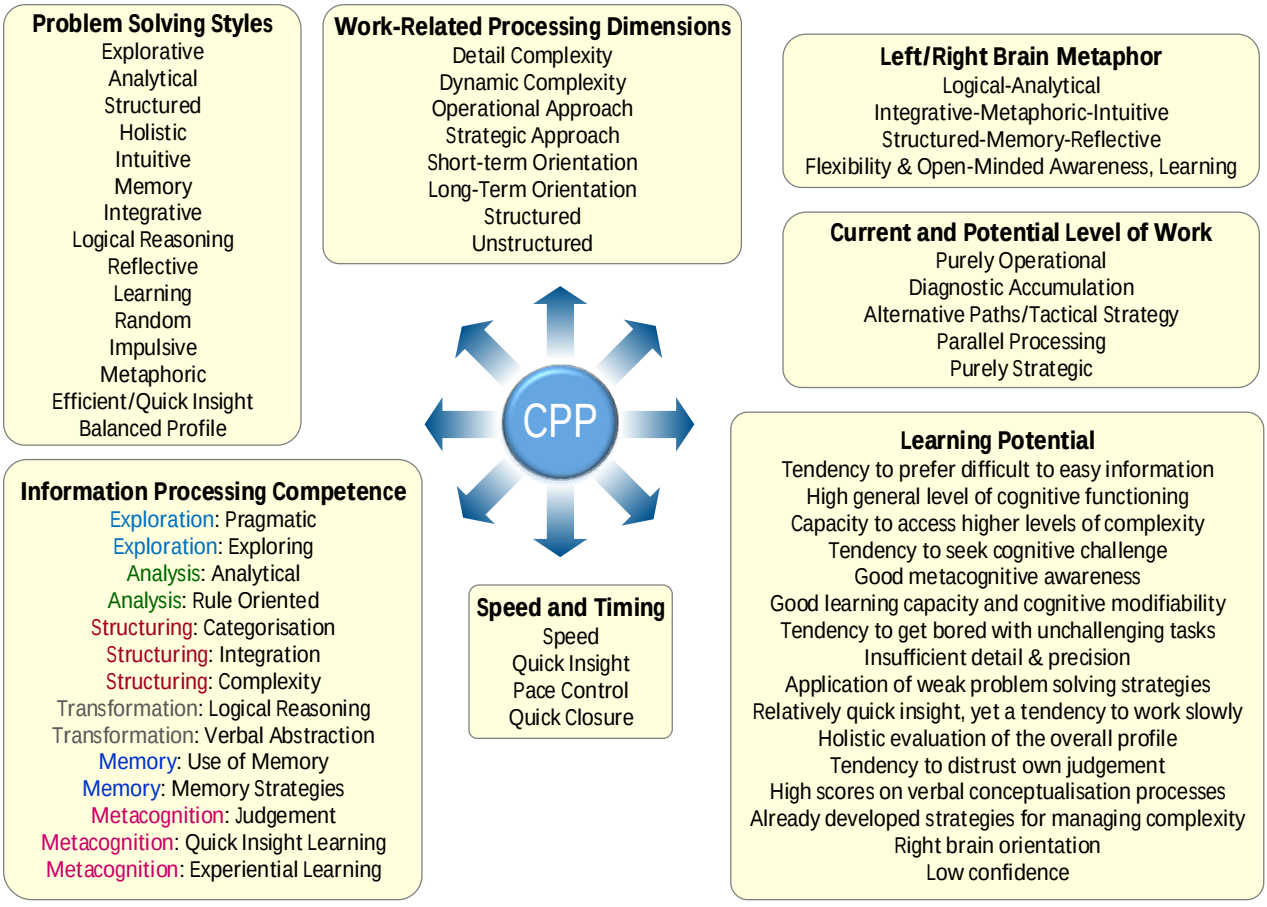
Tables

Figures

1. The Cognitive Process Profile (CPP)

The CCP is the practical application designed to assess the specific processes and concepts shown in Figure 1. It measures intellectual functioning in terms of constructs such as judgement and decision making, strategizing, generalist versus specialist orientation, creativity, complexity preferences and other thinking and problem solving factors related to professional, managerial and executive functioning. It is an advanced computerised assessment technique; using simulation exercises. Subjects are monitored in terms of their preferences and capabilities in exploring, analysing, structuring, transforming, remembering, and learning information and making decisions, or exercising their judgement. The results can serve both as a source of personal understanding and development, as well as being linked to job-related performance. Figure 1 summarises the key processing components and styles.

Figure 1: The constructs reported on by the CPP



The CPP assessment consists of a task requiring the deciphering of hieroglyphic messages. It was designed to externalise and track each of the thinking processes specified in Figure 1, and their many subcomponents. While completing the test, a person explores, links, structures, transforms, remembers, clarifies and monitors his/her actions on the computer screen using a computer mouse. All the "movements" made on the computer screen are saved as the person traverses the test. At the

end of each task, the person provides his/her interpretation of the symbolic message (normally a one-line statement) by keying it into the computer. A "scoring and statement parsing system" subsequently integrates all these movements and story interpretations, which are subsequently analysed using more algorithms to produce the CPP report. For the CPP to measure the various concepts detailed in the different sections of the report, the cognitive processes are grouped and analysed in many different ways. These intricate groupings often overlap, and the analysis performed by the software is highly complex.

2. Study 1

2.1 Sample details

87 students undertaking an Accounting degree course at a South African university comprised the sample who completed the CPP twice, along with the Learning Orientation Index (LOI) assessment. Specific gender information was not available at the time of analysis, although judging by the forenames, the majority were male students. Their ages at the first CPP assessment are provided in Table 1.

Table 1: Study 1: Sample participant ages at the first assessment

	Descriptive Statistics						
Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.	Skewness
Age at Assessment	87	24.74	24	21	32	2.461	1.141

The descriptive statistics for the retest durations (in days) between the 1st and 2nd occasion CPP assessments are provided in Table 2. The median duration is **almost 2 months** (7.9 weeks), with the range between 7 and 9.3 weeks.

Table 2: Study 1: Summary statistics for retest durations (in days)

	Descriptive Statistics						
Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.	Skewness
Retest Duration	87	54.89	55	49	65	2.678	0.832

2.2 Estimating retest reliability

Gower[1] agreement indices (see Appendix 1) were used throughout to express magnitude agreement, as what matters here is answering the simple question "*how closely do the retest occasion observations agree with one another?*", and not "*do observations on the first occasion possess a monotonic relationship with those on the second occasions?*"

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

Indexing monotonic relationship (Pearson correlation, gamma, some ICCs, Cronbach alpha) is not relevant to the assessment of retest 'reliability'; what we want to see is how closely our scores on one occasion are *reproduced* on the second. For that we need to preserve the actual magnitudes of our observations and not remove that information via a standardization transformation.

A presentation on these issues, showing empirically the inaccuracy of ICCs and Pearson coefficients (and r_{wg}^2) for assessing retest reliability, entitled: *Interrater Reliability: measuring agreement and nothing else* can be downloaded from: http://www.pbarrett.net/issid/issid2009.html.

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 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.

---

[2] James, L.R., Demaree, R.G., & Wolf, G. (1984) Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 1, 85-98.

3. Results – Study 1 (~2 months duration)

3.1 CPP Ranked Styles

CPP cognitive styles refer to broad response tendencies or patterns in thinking and problem-solving behaviour. These are measured by tracking a person's responses to unfamiliar information. A person's stylistic preferences when dealing with unfamiliar information, however, also tend to be used when working with familiar information. Some personality factors are indicated here, as these are sometime evident in the way a person thinks.

A person may develop specific stylistic preferences due to personality and emotional factors, cultural values, educational exposure, learning opportunities, work experience and fields of interest. In interpreting this report, the specific combination of preferred styles provides a useful indication of certain factors in the person's developmental history.

Various descriptive categories are reported on as indications of stylistic preference, namely: Explorative, Analytical, Logical, Structured, Reflective, Reactive, Trial-and-error, Integrative, Holistic, Intuitive, Quick Insight, Learning, Metaphoric and Memory approaches. A Trial-and-error or Quick Closure style may be an indication of performance anxiety, emotional or developmental factors. It may also be a valid reflection of the person's approach to unfamiliar problem-solving. Insight can be gained from interpreting the person's particular combination of stylistic preferences. The construct of "Style" also informs the identification of a suitable work environment.

Table 3: Study 1: Retest Reliability CPP Ranked Styles {most-preferred = 1, least preferred = 14}

Ranked Style	Gower	MAD
CPP - Explorative	.81	2
CPP-Analytical	.79	3
CPP Structured	.85	2
CPP Holistic	.85	2
CPP Intuitive	.88	2
CPP Memory	.76	3
CPP Logical	.79	3
CPP Impulsive	.81	2
CPP Random	.80	3
CPP Integrative	.87	2
CPP Learning	.79	3
CPP Quick insight	.78	3
CPP Reflective	.77	3
CPP Metaphoric	.82	2

Notes: MAD = Mean Absolute deviation between the paired observations.
The MAD values are rounded integers as the possible 'score' ranges are integer ranks

For every individual, the scores for each cognitive style are rank-ordered in their report
(1 = most preferred,
14 = least-preferred).

The median agreement index across all 14 styles is 81%, indicating that relative to the maximum possible absolute (*unsigned*) discrepancy between them, the ranks assigned to styles across occasions agree *on average* to within 81% of each other's values.

3.2 CPP Levels of Work

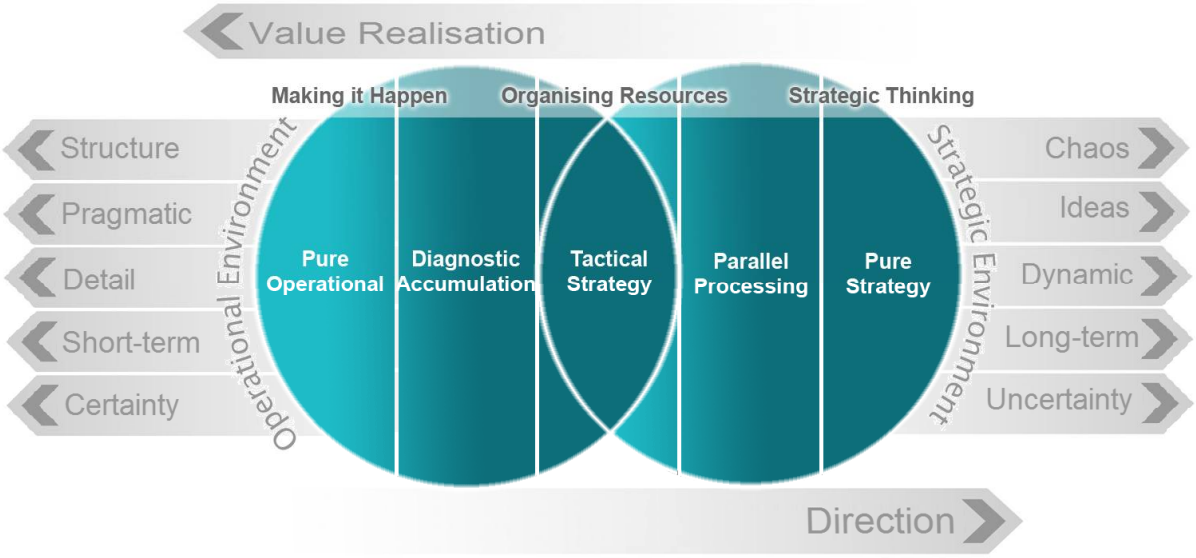
The CPP links a person's cognitive profile to the cognitive requirements of specific operational and strategic work environments. Algorithms are used to compare the qualitative and quantitative characteristics of a person's profile to the requirements of five work environments. The profile qualities considered include a person's:
(a) stylistic preferences,
(b) the units of information used in processing,
(c) judgement and decision making tendencies, as well as
(d) eight job-related processing dimensions.
The work environments specified reflect the Stratified Systems Model (SST) of Jaques, the Viable Systems Model (VSM) of Beer and Prinsloo's work on cognitive complexity.

Both 'current' and 'potential' work environments are indicated but no time frames are given to predict the person's readiness to progress from the current to the potential level as this depends on many different factors including opportunity and motivation. The CPP assigns an ordered-class 'score' to a respondent, ranging within five 'Levels of Work', for both current and potential Level of Work designations:

Table 4: The CPP Levels of Work

1	**Pure Operational**: individuals who show less interest in intellectual complexity, vagueness and cognitive challenge.
2	**Diagnostic**: can be quite analytical, but still show a need for structure in the form of technical guidelines and/or previous experience.
3	**Tactical Strategy**: no longer rely on linear processing, but prefer viewing issues in terms of tangible systems and the interaction between observable system elements.
4	**Parallel Processing**: those with the capacity to accommodate novelty, vagueness, dissonance and fragmentation, all of which require the cognitive skills of integration and innovation.
5	**Pure Strategic**: functioning is characterised by a strong Intuitive and Holistic "big picture" inclination.

Figure 2: The CPP Levels of Work within the Operational and Strategic Dimensions



In the World of Work there are two domains, the **Operational** and **Strategic**. These domains form a dimension that is divided into five overall **Work Environments**. The environments change from left to right as there is increasing **chaos**, **uncertainty** and **complexity**.

Table 5 reports the agreement indices for the current (cLOW) and potential (pLOW) Levels of Work attributes.

Table 5: Study 1: Retest Reliability CPP Levels of Work {cLOW range 1-4, pLOW range 1-5}

Level of Work (LOW)	Gower	MAD
CPP Current LOW	.87	.40
CPP Potential LOW	.87	.53

Note: MAD = Mean Absolute deviation between the paired class-categories

Indicating that relative to the maximum possible absolute (*unsigned*) discrepancy between them, the ordered class-categories agree *on average* to within 87% of each other's values.

It is important to view these data in more detail, as the summary agreement index doesn't really convey where discrepancies are occurring between occasions. Tables 6 and 7 report the occasion-1 vs occasion-2 cross-tab frequencies for Current and Potential Levels of Work attributes.

Table 6: Study 1: CPP Current Level of Work cross-tab frequencies

Summary Frequency Table : CPP Current Level of Work, 2-occasions Marked cells have counts > 10 (Marginal summaries are not marked)					
CPP-cLOW-1	CPP-cLOW-2 1	CPP-cLOW-2 2	CPP-cLOW-2 3	CPP-cLOW-2 4	Row Totals
1	17	12	0	0	29
2	2	32	8	3	45
3	0	3	4	3	10
4	0	0	1	2	3
All Grps	19	47	13	8	87

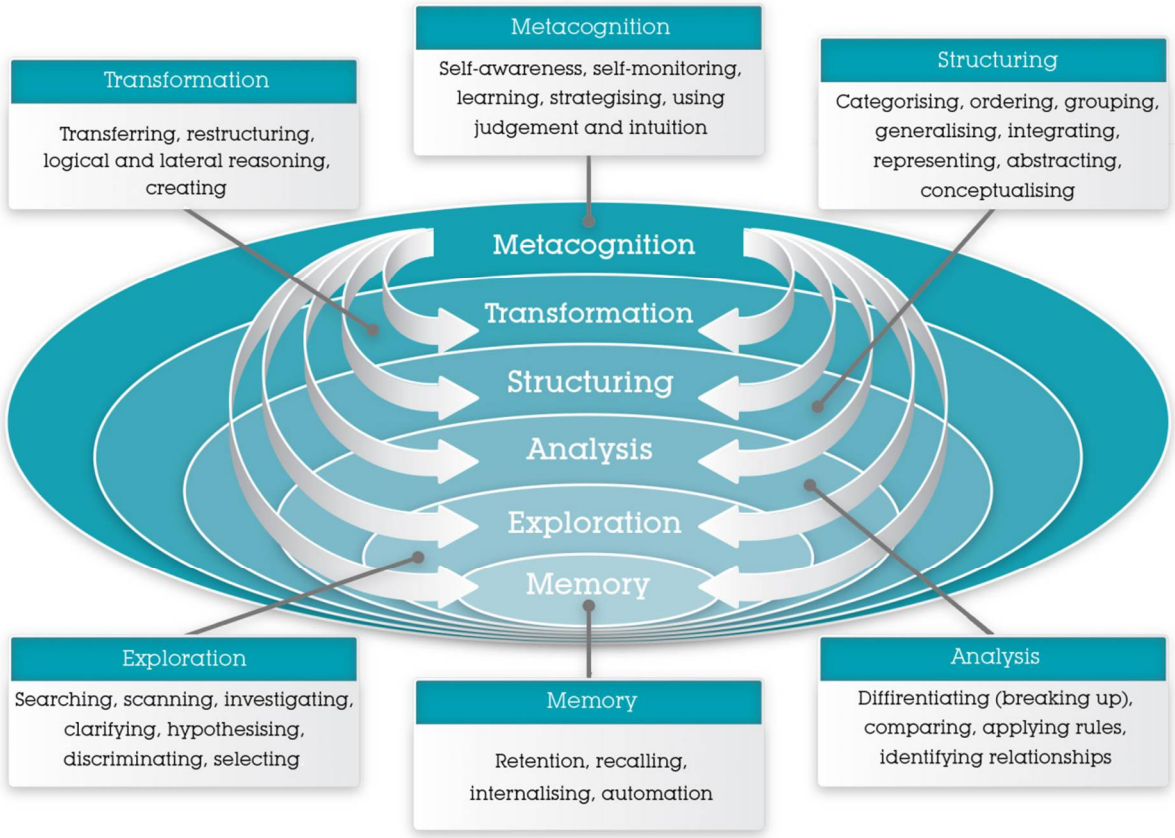
Table 7: Study 1: CPP Potential Level of Work cross-tab frequencies

Summary Frequency Table : CPP Potential Level of Work, 2-occasions Marked cells have counts > 10 (Marginal summaries are not marked)						
CPP-pLOW-1	CPP-pLOW-2 1	CPP-pLOW-2 2	CPP-pLOW-2 3	CPP-pLOW-2 4	CPP-pLOW-2 5	Row Totals
1	1	4	1	0	0	6
2	1	26	15	3	0	45
3	0	7	9	5	1	22
4	0	0	3	10	1	14
All Grps	2	37	28	18	2	87

3.3 CPP Information-Processing Competencies

These index cognitive functioning as a classification into six major processing categories. These functional processing categories can be represented as a *holon* where each successive process includes and transcends the previous one(s). Figure 3 summarises the processing categories and their subdivisions.

Figure 3: The cognitive processes assessed by the CPP



The dynamic functioning of the processes are explained by the theoretical processing model on which the CPP assessment is based. Table 8 provides the brief descriptions of the processing categories and their subdivisions.

Table 8: The 14 CPP processing competencies

Processing Competency		Description
Memory	**Use of Memory**	A tendency to rely on memory and to concentrate on the task
	Memory Strategies	Effectiveness of memory strategies
Exploration	**Pragmatic**	Practical orientation (asking whether things will work in practice). Determining relevance in structured contexts
	Exploration	The effectiveness, depth and width of exploration
Analysis	**Analysis**	Working systematically. Detailed and precise in differentiating between, and linking, elements
	Rules	A focus on rules
Structuring / Integration	**Categorisation**	Creating external order, categories and reminders. Structuring tangibles
	Integration	Synthesis of ambiguous / discrepant / fragmented information
	Complexity	The preferred level of complexity and the unit of information used
Transformation	**Logical Reasoning**	The disciplined, logical following through of reasoning processes
	Verbal Conceptualisation	Unusual / flowery / creative and/or abstract verbalisation and conceptualisation
Metacognition	**Judgment**	Capitalising on intuitive insights to clarify unstructured and vague information
	Quick Insight Learning	A tendency to grasp new concepts and acquire knowledge and understanding relatively quickly
	Gradual Improvement Learning	A preference for practical or experiential learning

Within this retest dataset, all 14 process scores were expressed on a 1 to 7 scale (low to high). Table 9 provides the Gower agreement indices between the two-occasion CPP assessments.

Table 9: Study 1: Retest Reliability CPP Process Scores {score range 1-7}

Process Score	Gower	MAD
CPP-Pragmatic	.83	1.0
CPP-Exploration	.88	0.7
CPP-Analytical	.79	1.3
CPP-Rule-Oriented	.86	0.9
CPP-Categorisation	.84	1.0
CPP-Integration	.88	0.7
CPP-Complexity	.85	0.9
CPP-Logical Reasoning	.79	1.3
CPP-Verbal Conceptualisation	.77	1.4
CPP-Use of Memory	.84	1.0
CPP-Memory Strategies	.86	0.9
CPP-Judgement	.85	0.9
CPP-Quick Insight Learning	.87	0.8
CPP-Gradual Improvement Learning	.79	1.3

Note: MAD = Mean Absolute deviation between the paired class-categories

The median agreement index across all 14 processes is 85%, indicating that relative to the maximum possible absolute (*unsigned*) discrepancy between them, the ranks assigned to processes across occasions agree *on average* to within 85% of each other's values.

The mean absolute deviation is 1.0 (rounded from 0.95). That is, each process score lies with ±1 of each other's values – bearing in mind the range of the scores is between 1 and 7.

Overall, the CPP ranked styles, Levels of Work class-categories, and processing scores showed good retest reliability over a 2-month duration; compatible with and in most cases exceeding such reliabilities found using conventional psychometric tests over this duration.

4. Study 2

4.1 Sample details

2,724 respondents completed the CPP on at least two occasions. For the purposes of this analysis, only the 1st and 2nd occasion data were utilised. The respondent sample was comprised primarily of job applicants who had completed the CPP on two separate occasions, but also included some students, and attendees at CPP training courses.

Table 10: Study 2: Sample participant ages at the first assessment

	Descriptive Statistics (Study 2, CPP Retest dataset.sta)					
Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
Age at Assessment-1	2638	34.17	34	16	60	7.932

Table 11: Study 2: Sample participant gender mix

	Frequency table: gender (Study 2, CPP Retest dataset.sta)			
Category	Count	Cumulative Count	Percent	Cumulative Percent
F	1061	1061	38.95	38.95
M	1663	2724	61.05	100.00
Missing	0	2724	0.00	100.00

Table 12: Study 2: Sample participant ethnicity mix

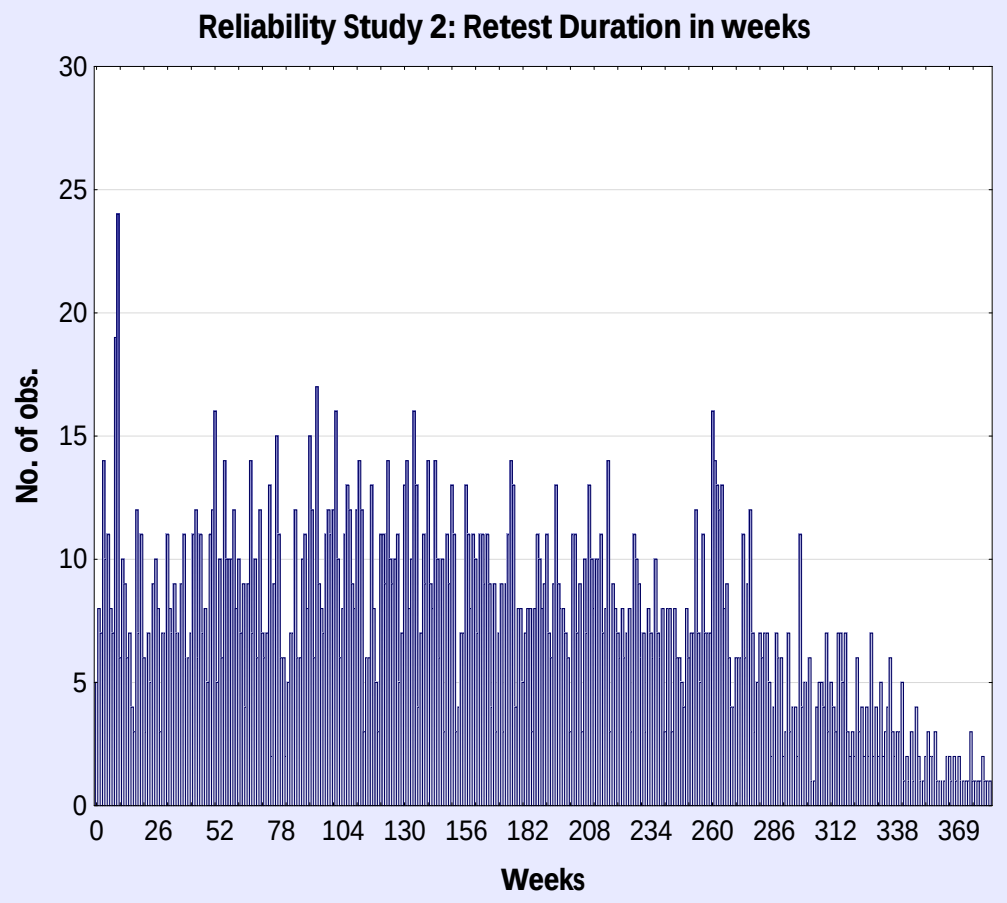
	Frequency table: ethnicity (Study 2, CPP Retest dataset.sta)			
Category	Count	Cumulative Count	Percent	Cumulative Percent
Indian	418	418	18.50	18.496
White European	636	1054	28.14	46.637
Black African	1206	2260	53.36	100.000

Table 13: Study 2: Summary statistics for retest durations (in weeks)

	Descriptive Statistics (Study 2, CPP Retest dataset.sta)					
Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
Duration	2720	157.10	149.5	0	613	95.69

Note: Four cases had invalid or missing assessment dates.

Figure 4: Study 2: Histogram of retest durations in weeks



Given the spread of durations, three relevant retest-duration groups were constructed from the weekly data, corresponding to reasonably short, medium, and long-term retest durations.

Table 14: Study 2: Retest-duration groups

	Frequency table: Duration Group (Study 2, CPP Retest dataset.sta)			
Category	Count	Cumulative Count	Percent	Cumulative Percent
0 <= 6 months	236	236	8.66	8.66
> 6 months <= 1 year	219	455	8.04	16.70
> 1 year	2265	2720	83.15	99.85
Missing	4	2724	0.15	100.00

Gower (1971) agreement indices (see Appendix 1) were used throughout to express magnitude agreement, as what matters here is answering the simple question "*how closely do the retest occasion observations agree with one another?*", and not "*do observations on the first occasion possess a monotonic relationship with those on the second occasions?*"

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 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.

Bootstrapped sampling distributions of Gower indices provided checks on the 'significance' of the observed Gower indices, using uniform random data with the same measurement range and sample size as the actual data.

Three CPP attribute sets were investigated: Ranked cognitive styles, information processing competencies, and levels of work.

5. Results – Study 2, short-medium durations

5.1 CPP Ranked Styles

For every individual, the scores for each cognitive style are rank-ordered in their report; in this dataset *(unlike in Study 1)* the rank orders range from 14 = most preferred, 1 = least-preferred. Table 15 provides the Gower agreement indices (reliabilities) for the three retest-duration groups.

Table 15: Study 2: Retest Reliability CPP Ranked Styles; three retest-duration groups

Duration	<= 6 months, n=236		> 6 months but <= 1 year, n=219		> 1 year n=2,265	
Ranked Style	**Gower**	**MAD**	**Gower**	**MAD**	**Gower**	**MAD**
CPP - Explorative	.77	3	.77	3	.77	3
CPP-Analytical	.76	3	.76	3	.75	3
CPP Structured	.82	2	.82	2	.82	2
CPP Holistic	.83	2	.82	2	.82	2
CPP Intuitive	.86	2	.87	2	.87	2
CPP Memory	.77	3	.75	3	.75	3
CPP Logical	.79	3	.75	3	.75	3
CPP Impulsive	.80	3	.80	3	.82	2
CPP Random	.78	3	.79	3	.80	3
CPP Integrative	.85	2	.86	2	.85	2
CPP Learning	.81	2	.80	3	.79	3
CPP Quick insight	.82	2	.84	2	.83	2
CPP Reflective	.82	2	.82	2	.81	2
CPP Metaphoric	.77	3	.74	3	.76	3

Notes: **MAD** = Mean Absolute deviation between the paired observations.
The MAD values are rounded integers as the possible 'score' ranges are integer ranks {1...14}

Bootstrap results for Duration 1 sample (<= 6 months)

For these data with an integer measurement range between 1 and 14, a bootstrap analysis was undertaken, generating 20,000 random samples of **n=236** cases forming an empirical sampling distribution of possible Gower indices.

- The median expected random-data Gower index was **.67**.
- The lowest observed Gower index in Table 15 is **.76**.
- None of the 20,000 random samples produced a Gower as high as .76. The highest random-data value observed was **.72**.

Bootstrap results for Duration 2 sample (> 6 months but <= 1 year)

For these data with an integer measurement range between 1 and 14, a bootstrap analysis was undertaken, generating 20,000 random samples of **n=219** cases forming an empirical sampling distribution of possible Gower indices.

- The median expected random-data Gower index was **.67**.
- The lowest observed Gower index in Table 15 is **.74**.
- None of the 20,000 random samples produced a Gower as high as .74. The highest random-data value observed was **.73**.

Bootstrap results for Duration 3 sample (> 1 year)

For these data with an integer measurement range between 1 and 14, a bootstrap analysis was undertaken, generating 20,000 random samples of **n=2,265** cases forming an empirical sampling distribution of possible Gower indices.

- The median expected random-data Gower index was **.67**.
- The lowest observed Gower index in Table 15 is **.75**.
- None of the 20,000 random samples produced a Gower as high as .75. The highest random-data value observed was **.68**.

Table 15 shows that the cognitive style rankings remain similar to one another across assessment occasions, whether over a short or long-term. The ranking position of each style varies between 2 and 3 ranking places on average.

So, for example, the rank-values of Quick Insight for duration-group 1 respondents (< = 6 months) agree on average to within 82% of each occasions' values. A rank assigned on the 1st occasion will, on average, possess a rank on occasion 2 within ±2 of the 1st occasion rank.

5.2 CPP Information-Processing Competencies

Within this retest dataset, all 14 process scores were expressed on a 0 to 100 integer scale. Table 16 provides the Gower agreement indices between the two-occasion CPP assessments for the three duration groups.

Table 16: Study 2: CPP Information Processing Competencies; three retest-duration groups

Duration	<= 6 months, n=236		> 6 months but <= 1 year, n=219		> 1 year n=2,265	
Processing Scores	**Gower**	**MAD**	**Gower**	**MAD**	**Gower**	**MAD**
CPP-Pragmatic	.89	11	.89	11	.89	11
CPP-Exploration	.94	6	.93	7	.94	6
CPP-Analytical	.84	16	.83	17	.84	16
CPP-Rule-Oriented	.90	10	.90	10	.90	10
CPP-Categorisation	.93	7	.92	8	.93	7
CPP-Integration	.92	8	.93	7	.93	7
CPP-Complexity	.90	10	.90	10	.91	9
CPP-Logical Reasoning	.86	14	.86	14	.86	14
CPP-Verbal Conceptualisation	.86	14	.88	12	.87	13
CPP-Use of Memory	.92	8	.91	9	.91	9
CPP-Memory Strategies	.92	8	.91	9	.92	8
CPP-Judgement	.90	10	.90	10	.90	10
CPP-Quick Insight Learning	.92	8	.92	8	.92	8
CPP-Gradual Improvement Learning	.89	11	.90	10	.89	11

Notes: **MAD** = Mean Absolute deviation between the paired observations.
The MAD values are rounded integers as the possible 'score' ranges are integers {0 ...100}

Bootstrap results for Duration 1 sample (<= 6 months)

For these data with an integer measurement range between 0 and 100, a bootstrap analysis was undertaken, generating 20,000 random samples of **n=236** cases forming an empirical sampling distribution of possible Gower indices.

- The median expected random-data Gower index was **.67**.
- The lowest observed Gower index in Table 16 is **.84**.
- None of the 20,000 random samples produced a Gower as high as .84. The highest random-data value observed was **.72**.

Bootstrap results for Duration 2 sample (> 6 months but <= 1 year)

For these data with an integer measurement range between 0 and 100, a bootstrap analysis was undertaken, generating 20,000 random samples of **n=219** cases forming an empirical sampling distribution of possible Gower indices.

- The median expected random-data Gower index was **.67**.
- The lowest observed Gower index in Table 16 is **.83**.
- None of the 20,000 random samples produced a Gower as high as .83. The highest random-data value observed was **.73**.

Bootstrap results for Duration 3 sample (> 1 year)

For these data with an integer measurement range between 0 and 100, a bootstrap analysis was undertaken, generating 20,000 random samples of **n=2,265** cases forming an empirical sampling distribution of possible Gower indices.

- The median expected random-data Gower index was **.67**.
- The lowest observed Gower index in Table 16 is **.84**.
- None of the 20,000 random samples produced a Gower as high as .84. The highest random-data value observed was **.69**.

5.3 CPP Levels of Work

Table 17 reports the Gower agreement indices for the current (cLOW) and potential (pLOW) Levels of Work attributes. Given the rather small range of possible values for current and potential Levels of Work, we also computed Goodman-Kruskal Gamma ordinal correlation coefficients between the two occasions' data. Gamma, like the Pearson r, is primarily a measure of monotonicity rather than absolute agreement.

Table 17: Study 2: Retest Reliability CPP Levels of Work {cLOW range 1-4, pLOW range 1-5}

Duration	<= 6 months, n=236			> 6 months but <= 1 year n=219			> 1 year n=2,265		
Level of Work(LOW)	Gower	MAD	Gamma	Gower	MAD	Gamma	Gower	MAD	Gamma
CPP Current (cLOW)	.84	.48	.92	.83	.50	.81	.83	.51	.82
CPP Potential (pLOW)	.86	.56	.82	.86	.55	.76	.86	.54	.78

Note: **MAD** = Mean Absolute deviation between the paired class-categories

Both Gower and Gamma coefficients are statistically significant at $p < 0.0001$

For the cLOW data with an integer measurement range between 1 and 4, a bootstrap analysis was undertaken, generating 20,000 random samples of for each duration number of cases, forming empirical sampling distributions of possible Gower indices.

- The median expected random-data Gower varied between **.65** and **.66.**
- The lowest observed Gower index in Table 17, any duration is **.83**.
- None of the 20,000 random samples produced a Gower as high as .83. The highest random-data value observed was **.73**.

For the pLOW data with an integer measurement range between 1 and 5, a bootstrap analysis was undertaken, generating 20,000 random samples of for each duration number of cases, forming empirical sampling distributions of possible Gower indices.

- The median expected random-data Gower index was **.66**.
- The lowest observed Gower index in Table 17 is **.86**.
- As for the cLOW index, none of the 20,000 random samples produced a Gower as high as .86. The highest random-data value observed was **.73**.

Overall, the CPP ranked styles, Levels of Work class-categories, and processing scores in Study 2 show reasonable to excellent retest reliability/magnitudes agreement over three durations; compatible with and in most cases exceeding such reliabilities found using conventional psychometric tests over this duration. Clearly, the attributes assessed here seem to be relatively stable over time.

6. Results – Study 2, long-term retest-duration > 5 years

Of interest perhaps is evaluating the CPP retest reliability of the same attributes reported for Study 2 in section 3.2 above, but where we only include cases whose 2nd retest occasion was more than 5 years since their 1st test occasion. Within the Study 2 dataset, we have **475** cases whose durations extend from **261 weeks** (5 years, 1 week) through to **613 weeks** (11 years, 41 weeks).

6.1 CPP Ranked Styles

For every individual, the scores for each cognitive style are rank-ordered in their report; in this dataset *(unlike in Study 1)* the rank orders range from 14 = most preferred, 1 = least-preferred. Table 18 provides the Gower agreement indices (reliabilities) for this long-duration group.

Table 18: Study 2: Long-Duration (> 5 years) Retest Reliability CPP Ranked Styles

Ranked Style	Gower	MAD
CPP - Explorative	.78	3
CPP-Analytical	.73	4
CPP Structured	.82	2
CPP Holistic	.82	2
CPP Intuitive	.87	2
CPP Memory	.73	4
CPP Logical	.75	3
CPP Impulsive	.83	2
CPP Random	.81	3
CPP Integrative	.85	2
CPP Learning	.80	3
CPP Quick insight	.81	2
CPP Reflective	.81	3
CPP Metaphoric	.74	3

Notes: **MAD** = Mean Absolute deviation between the paired observations.
The MAD values are rounded integers as the possible 'score' ranges are integer ranks {1…14}

For these data with an integer measurement range between 1 and 14, a bootstrap analysis was undertaken, generating 20,000 random samples of n=475 cases forming an empirical sampling distribution of possible Gower indices.

- The median expected random-data Gower index was **.67**.
- The lowest observed Gower index in Table 18 is **.73**.
- None of the 20,000 random samples produced a Gower as high as .73. The highest random-data value observed was **.71**.

For two styles, Analytical and Memory, a ±4 average rank-discrepancy between occasions is substantive.

6.2 CPP Information-Processing Competencies

Within this retest dataset, all 14 process scores were expressed on a 0 to 100 integer scale. Table 19 provides the Gower agreement indices between the two-occasion CPP assessments for the three duration groups.

Table 19: Study 2: Long-Duration (> 5 years) Retest Reliability CPP Information Processing Competencies

Process Score	Gower	MAD
CPP-Pragmatic	.88	12
CPP-Exploration	.94	6
CPP-Analytical	.84	16
CPP-Rule-Oriented	.91	9
CPP-Categorisation	.93	7
CPP-Integration	.93	7
CPP-Complexity	.91	9
CPP-Logical Reasoning	.86	14
CPP-Verbal Conceptualisation	.87	13
CPP-Use of Memory	.91	9
CPP-Memory Strategies	.92	8
CPP-Judgement	.90	10
CPP-Quick Insight Learning	.93	7
CPP-Gradual Improvement Learning	.89	11

Note: MAD = Mean Absolute deviation between the paired class-categories

For these data with an integer measurement range between 0 and 100, a bootstrap analysis was undertaken, generating 20,000 random samples of **n=475** cases forming an empirical sampling distribution of possible Gower indices.

- The median expected random-data Gower index was **.67**.
- The lowest observed Gower index in Table 19 is **.84**.
- None of the 20,000 random samples produced a Gower as high as .84. The highest random-data value observed was **.71**

Some of the processing competency scores remain remarkably similar over a 5 year or longer retest duration, especially Exploration, Categorisation, Integration, and Quick Insight Learning. The Analytical competency shows the greatest discrepancy – with a ±16 average score-discrepancy between occasions, although relative to a 0-100 integer possible score-range.

6.3 CPP Levels of Work

Table 20 reports the Gower agreement indices for the current (cLOW) and potential (pLOW) Levels of Work attributes. Given the rather small range of possible values for current and potential Levels of Work, we also computed Goodman-Kruskal Gamma ordinal correlation coefficients between the two occasions' data. Gamma, like the Pearson r, is primarily a measure of monotonicity rather than absolute agreement.

Table 20: Study 2: Long-Duration (> 5 years) Retest Reliability CPP Levels of Work

Level of Work (LOW)	Gower	MAD	Gamma
CPP Current LOW	.84	.48	.81
CPP Potential LOW	.87	.53	.78

Note: **MAD** = Mean Absolute deviation between the paired class-categories

Both Gower and Gamma coefficients are statistically significant at $p < 0.0001$

For the cLOW data with an integer measurement range between 1 and 4, a bootstrap analysis was undertaken, generating 20,000 random samples of for each duration number of cases, forming empirical sampling distributions of possible Gower indices.

- The median expected random-data Gower was **.65.**
- The cLOW Gower index in Table 20 is **.84**.
- None of the 20,000 random samples produced a Gower as high as .84. The highest random-data value observed was **.70**.

For the pLOW data with an integer measurement range between 1 and 5, a bootstrap analysis was undertaken, generating 20,000 random samples of for each duration number of cases, forming empirical sampling distributions of possible Gower indices.

- The median expected random-data Gower index was **.66**.
- The observed pLOW Gower index in Table 20 is **.87**.
- As for the cLOW index, none of the 20,000 random samples produced a Gower as high as .87. The highest random-data value observed was **.70**.

Overall, the CPP ranked styles, Levels of Work class-categories, and processing scores for the longer-duration subsample of Study 2's data, show reasonable to excellent retest reliability/magnitudes agreement over three durations; exceeding such reliabilities found using conventional psychometric tests over this long-duration. Clearly, many of the attributes assessed here seem to be relatively stable over time, reflecting their substantive, but not identical, relationship with cognitive ability.

7. The Logic of CPP retest assessment: An Advisory

Because the CPP capitalizes on a person's cognitive response to new and unfamiliar information, the first CPP is always the most valid – particularly if the person's performance has not been affected by extreme performance anxiety or demotivation. Note that a manageable degree of performance anxiety may even improve concentration.

CPP re-assessment, especially where the first CPP can be regarded as valid, should therefore be postponed by at least 4 to 5 years or more, if possible.

However, at times it is useful to evaluate the impact of developmental initiatives, work exposure, maturity, changes in attitude and interest on cognition, or to reassess those with invalid reports. The second set of CPP results then has to be interpreted qualitatively. Cognadev can assist consultants in doing so.

Higher CPP scores are often obtained with the second assessment, especially in the case of operational profiles. Certain processing dimensions also tend to improve with a second assessment, such as the Analytical skills. Other dimensions are more resistant to change. These include the Potential level of work indication, the Units of information or Complexity preferences, Integration and Judgement skills. In the case of Strategic profiles, the second set of results may, however, be somewhat lower than the first, as the candidate is likely to approach a familiar problem somewhat differently from an unfamiliar problem. This may include taking short cuts based on what is already known and capitalizing on memory.

Seeing that the CPP capitalises on measuring thinking processes in response to unfamiliar problems, the most valid results are obtained with a first assessment. Test administrators should therefore take care that test candidates are calm, rested and motivated to complete their first CPP assessments. The results may remain a valid reflection of the person's approach for a long period which exceeds the 5 year cut-off point specified for a valid reassessment. However, developmental initiatives, personal motivation and work experience may improve a person's cognitive approach and this can be assessed for by means of the CPP after a certain period of time.

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. The Gower coefficient varies between 0 and 1 (or 0% and 100%).

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