



**Birkman International conducted a test-retest reliability analysis in April-May 2018. This document highlights the procedures followed and reports the results obtained.**

Birkman International (BI), founded by Dr. Roger Birkman, is a behavioral and occupational assessment company with a global reach. The Birkman Method® (TBM), our industry-leading personality assessment, is widely accepted and critically acclaimed. The prototype of The Birkman Method was created in 1951 and attained scientific form in 1965. TBM has been constantly reviewed, updated, and improved by qualified psychometricians and organizational psychologists. TBM demonstrates strong psychometric and statistical properties.

The Birkman Method is unique within the behavioral assessment arena in that it compares a person's own perceptions about *self* with his/her perceptions about *most* people. In effect, it reports a person's social worldview. Another unique contribution of TBM is that it reports each respondent's *social desirability* bias. It is well-known that individuals have a tendency to present themselves in a socially desirable light when answering questions about themselves. TBM integrates social desirability into its scoring processes and effectively deals with it, instead of ignoring the concept or applying "correction techniques" like many personality assessments that are available today.

**Scales:** With more than forty scales, TBM reports characteristics of a respondent, characteristics of the interaction between a respondent and other people, and characteristics of the interaction between a respondent and situations, all with one single assessment. To conduct test-retest reliability analysis, the nine Components (Social Energy, Physical Energy, Emotional Energy, Self-Consciousness, Assertiveness, Insistence, Incentives, Restlessness, and Thought), the six Perspectives (Distinctiveness, Alignment, Image Management, Social Acuity, Self-Affirming, and Others-Affirming), and the ten Occupational Interests (Musical, Scientific, Artistic, Literary, Social Service, Numerical, Technical, Administrative, Outdoor, and Persuasive) are chosen as scales.

**Test-Retest Reliability:** An important element of the reliability testing for any psychometric instrument is the degree of agreement between the results of successive measurements, separated in time, of the same scale carried out under the same conditions of measurement [1,2]. The test-retest reliability coefficient measures the intensity and direction of such association between two successive values of scales. It also indicates the overall consistency of scales over time. To conduct test-retest reliability, the assessment is administered to a group of people twice, with a fixed amount of time between each administration. The results from each session are then compared to evaluate the relationship between each set of scores.

**Procedure:** BI's overriding goal for this effort was to minimize or eliminate any inherent bias in either the form or the process. Toward this end, BI hired a third-party market research firm to recruit potential respondents with no existing knowledge of TBM, and who could reasonably be presumed to be willing to complete the questionnaire twice (although that was not part of the initial request). Keeping in mind that TBM is a behavioral and occupational assessment tool, BI also stipulated that the sample population should consist only of working-age people and that the demographic makeup of the sample should – as closely as possible – match the demographics of the working population in the United States. A link to

TBM assessment was forwarded to the selected respondents (583), of which 422 successfully completed the questionnaire. Please note that in this study the test takers were financially incentivized to participate. It is also worth noting that at this point the test takers did not know that they would be asked to complete the assessment a second time. After 15 days, a new link to the assessment was forwarded to those who had completed it the first time. Successful completion yielded higher incentive for this second iteration. Out of 422 respondents, 414 successfully completed the task.

**Analysis:** Following successful completion of the data collection phase, BI analyzed the data for initial insights. Almost immediately an anomaly was discovered – the amount of time used by some of the respondents to complete the questionnaire was either extremely small or extremely large in comparison to the time taken by the average Birkman respondent. Historically, most respondents took anywhere between 15 to 45 minutes to complete, with an average time of around 30 minutes. Clearly many (if not all) of the respondents with very small time-to-complete figures were not fully engaged in the process, likely not even fully reading the questions. The data associated with those instances are spurious and serve to arbitrarily and negatively impact the results. Out of those concerns, BI imposed filtering criteria based only on the total time taken to complete the assessment. Considering only those respondents who took anywhere between 15 to 45 minutes to complete the assessment yielded 120 respondents. It was deemed necessary to take such an action in order to ensure a reasonable level of data quality. It was presumed that the final sample of 120 respondents represented people who were “serious” about their charge, and who were honest in their answers to the questionnaire. The actual analysis of test-retest reliability was conducted on this sample. Tables 1-6 highlight the demographics of the final set of respondents.

**Results:** The range of possible values for the test-retest reliability coefficient is 0 to 1. In practice, however, scores of “0” and “1” do not occur in the domain of psychometrics. Generally speaking, higher values indicate higher levels of reliability, and lower values are associated with lower levels of reliability. Essentially, we are concerned with answering the question, “Does TBM consistently and reliably measure the same thing(s) over time?” Or, in other words, is the Method stable? Scales with a reliability coefficient closer to 1 are considered more stable than scales with a reliability coefficient closer to 0. Table 7 provides a general guideline about the interpretation of the test-retest reliability coefficient(s). Tables A-C highlights the reliability coefficients of the scales under scrutiny. One can observe that most of the TBM scales exhibit good reliability – Among Components, Emotional Energy and Physical Energy; among Perspectives, Self-Affirming and Alignment; and among Interests, Musical and Outdoor top the list.

**Conclusions:** From the scores observed for each independent TBM scale, the reliability coefficients indicate that scale values result from systemic rather than chance or random factors [3] and that the scales are stable over time. These results are very consistent with historical analyses of test-retest reliability (even though in this case the test conditions were less than ideal) and provide us with high confidence that TBM is quite stable and consistent. In summary, these results indicate that a given respondent would get essentially the same results from one administration to another, given similar testing conditions and minimal knowledge acquisition between the administrations.

**Demographics of The Respondents:**

Gender	Percentage
Female	52.43
Male	47.57

Age Bin	Percentage
20 – 29	6.32
30 – 39	16.84
40 – 49	16.84
50 – 59	25.26
60 – 69	32.63
70 – 79	2.11

Table 1 and 2: The distribution of respondents by Gender and Age, respectively, is depicted in the above tables. One can note that most of the respondents are Female, and are between ages 30 and 69.

Ethnicity	Percentage
African American	9.30
Asian or Pacific Islander	3.49
Caucasian	77.91
Hispanic	4.65
Native American or Alaskan	1.16
Native Hawaiian	1.16
Mixed	1.16

Education Level	Percentage
Middle School	0.81
High School	29.27
Associate	18.70
Bachelors	26.83
Masters	13.82
Doctoral	4.07
Other	6.50

Table 3 and 4: The distribution of respondents by Ethnicity and Education Level, respectively, is portrayed in the above tables. It is evident that most of the respondents are Caucasian, and finished High School.

Employment Status	Percentage
Never Been Employed	0.82
Not Currently Employed	3.28
Part Time Employed	4.92
Full Time Employed	90.98

Number of Years in Current Job	Percentage
00 – 09	40.83
10 – 19	35.00
20 – 29	18.33
30 – 39	5.83

Table 5 and 6: The distribution of respondents by Employment Status and Number of Years in Current Job, respectively, is presented in the above tables. One can note that most of the respondents are Full Time Employed, and have less than 10 years work experience.

**Test-Retest Reliability Coefficient and Stability:**

Reliability Coefficient	Stability
≥ 0.9 & < 1.0	Excellent
≥ 0.8 & < 0.9	Good
≥ 0.7 & < 0.8	Acceptable
≥ 0.6 & < 0.7	Questionable
≥ 0.5 & < 0.6	Poor
≥ 0.0 & < 0.5	Unacceptable

Table 7: A general guideline about the interpretation of test-retest reliability coefficient and stability.

**Test-Retest Reliability Coefficients of The Birkman Method:**

<b>Components</b>	<b>Reliability Coefficient</b>
Social Energy	0.84
Physical Energy	0.85
Emotional Energy	0.86
Self-Consciousness	0.83
Assertiveness	0.79
Insistence	0.70
Incentives	0.77
Restlessness	0.72
Thought	0.80

*Table A: Reliability coefficients of the Components.*

<b>Perspectives</b>	<b>Reliability Coefficient</b>
Distinctiveness	0.72
Alignment	0.85
Image Management	0.82
Social Acuity	0.76
Self-Affirming	0.86
Others-Affirming	0.80

*Table B: Reliability Coefficients of the Perspectives.*

<b>Interests</b>	<b>Reliability Coefficient</b>
Artistic	0.82
Administrative	0.84
Literary	0.83
Technical	0.83
Musical	0.89
Numerical	0.83
Outdoor	0.88
Persuasive	0.78
Scientific	0.72
Social Service	0.85

*Table C: Reliability coefficients of the Interests.*

References:

- 1) *Types of Reliability*. Social Research Methods Knowledge Base. Last Revised: 2006 October 10
- 2) Joint Committee for Guides in Metrology (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement*, doi JCGM 100:2008
- 3) Salkind, N. J. (2010). *Encyclopedia of research design*, doi: 10.4135/9781412961288