

The Birkman Method[®]

**Reliabilities
and Validities Summary**

March 2003

Entire contents copyrighted 1992...2003
by Birkman International, Inc.
Reproduction in whole or in part
is strictly forbidden
without the prior written consent
of the publisher.

Reliabilities and Validities Summary
Frank Reed Larkey, Ph.D.

0303A

Birkman International, Inc.
3040 Post Oak Blvd., Suite 1425
Houston, TX 77056
(713) 623-2760

The Birkman Method[®]

Reliabilities and Validities Summary

The purpose of this summary is to provide an introduction to the reliabilities and validities of The Birkman Method to the non-statistician. Unfortunately, it is impossible to discuss the topic of reliabilities and validities without using some numerical tables. The fact is that while reliabilities and validities can be talked about conceptually, the usual “definition” is defined statistically (mathematically).

Therefore, the goal of this short summary is to blend the conceptual with the numerical. For a comprehensive presentation see **The Birkman Method[®]: 2000 Reliabilities and Validities** (Sadler & Mefferd, 2001) and **The Birkman Method[®]: Reliabilities and Validities for Business and Industry** (Mefferd, 1975).

The following topics will be discussed:

- Scientific and organizational context within which The Birkman[®] was developed
- Overview of item development and the four major areas of characteristics measured
- Reliability of Birkman scales as demonstrated in a test-retest study
- Reliability of Birkman items as demonstrated by internal consistency analysis
- Validity of Birkman: face validity, construct validity, and criterion-related validity
- Validity of career families: criterion referenced scales and cross validity analysis
- Birkman constructs and cultural comparisons
- Translation reliability and validity protocol
- Research references

Birkman International, Inc.
March 2003

A Context of Scientific Rigor

The Birkman Method® was conceptualized, developed, and refined by Roger W. Birkman, Ph.D., as part of his dissertation under the guidance, direction, and tutelage of the highly skilled professional psychologists, statisticians, and psychometricians at the University of Texas. The Method was created under a rigorous, academic, scientific, industry-focused process.

The Birkman Method was born in 1950 while Roger W. Birkman was working with a group of scientists at the University of Texas surveying potentially useful psychological instruments for pilot selection for the US Air Force.

Birkman was pursuing a Ph.D. in Educational Psychology under the direction of Benjamin Fruchter, his dissertation chairperson. Fruchter was himself a student of the renowned psychometrician J. P. Guilford, whose expertise was both personality assessment and multivariate statistical analysis.

Birkman believed he could create a much-needed instrument that would measure social expectations, self-concepts, interests, and stress behavior in a single assessment tool. He further believed that such an instrument would be of great value to both organizations and individuals.

The Professionals Who Helped Make It Possible

While Birkman was pursuing his research interests, he met Roy B. Mefferd, Jr. Ph.D., who was completing his dissertation in Bacteriology-Biochemistry as a participant in an interdisciplinary program under the rubric of “Gene Research”. He was to become Dr. Birkman’s longtime research partner.

Mefferd went on to publish over 200 scientific publications over a wide range of topics with a particular emphasis in complex multivariate statistical techniques. He held the positions of adjunct Professor of Psychology, Graduate School, and Clinical Professor of Psychology, Department of Psychology, University of Houston; Director of Psychiatric and Psychosomatic Research Laboratory, Veterans Administration Hospital in Houston; Professor of Physiology, Departments of Psychiatry and Physiology, Baylor College of Medicine, Houston.

Timothy G. Sadler, Ph.D., joined Birkman International in 1970 and, while actively participating in ongoing research, also took the lead in developing new applications, training, and business consulting. Together with other licensed professionals, Sadler co-developed many innovative reports to support the constant push by Dr. Birkman to remain on the cutting edge of industrial-organizational research, psychometrics and technology.

Overview of Item Development and Scales

“Following the administration of the seventh revision to several thousand subjects ranging from professional sales personnel to non-professional employees, the form used in this study was introduced in 1957. Generally speaking, the items which had been developed through the eight years of work and statistical analysis were judged to be highly polished and sensitive to personality differences”.

(Birkman, 1961, p.29)

Item Development for a Test of Social Comprehension

Dr. Birkman’s first questionnaire was called a **Test of Social Comprehension**. Before form construction, an extensive survey of the literature on personality questionnaires was conducted. Virtually every item that existed in widely used questionnaires at that time was studied carefully. Thousands of items were examined. The reduction of items was based on whether the items related to deeper attitudes and experiences not easily identified as true or false by casual observation.

Seven successive revisions of the form were performed to improve the accuracy, reliability and validity. **Each** revision was administered to more than 1000 participants selected from a business or industrial setting (N = 7000+). After each testing, responses were recorded and analyzed for compatibility. Each new form represented an improvement and variation of the previous form. Both reliability and validity of this instrument was demonstrated by predicting successful and non-successful employee performance within a nationwide service organization.

The Birkman Method[®] Scales

While many current Birkman[®] reports are based on a subset of over 55 scales, the instrument is essentially based on four primary constructs of human characteristics and thus four *combination* scales each measured on two dimensions (the frequency of particular scores for a construct, e.g. Empathy, is technically called a “scale”). They are:

Usual Behavior – an individual’s effective style of dealing with tasks and relationships

Underlying Needs – how an individual wants or needs others to act toward them

Stress Behavior – an individual’s ineffective way of dealing with tasks and relationships

Interests – an individual’s expressed motivation based on job title preferences

Each of the four types of scales above contains specific items that measure constructs relevant to the particular grouping of scales. The four scales are interpreted on two dimensions using a matrix format (basically an X dimension and Y dimension).

Reliability: Test-Retest

Current and past research results both support the idea that The Birkman Method® is a very reliable instrument. **What does the term reliability mean?** When a person thinks of reliability, many things may come to mind – my friend is very reliable, my car is very reliable, my internet bill-paying process is very reliable, and so forth. However, in each of these examples one does not expect perfect performance. Reliability by nature assumes variability – your car might fail to start once or twice a year. Likewise, individual responses will vary over time. The research question is “how much do individual responses vary when taking the Birkman® twice in a scientifically controlled situation?”

In order to determine if Birkman scores are reliable, we must measure them. Many people associate the Birkman’s reliability with their personal scores, “If I take the Birkman again, will my scores change”? Indeed, clients talk about reliability typically only in regards to their own individual scores. However, the scientific method is applied not to an individual’s score but to a group of scores. So when instrument reliability is discussed, the unit of analysis is not the individual but a group. Therefore, an instrument might demonstrate a very high reliability but some individual scores will certainly vary. One way to measure this reliability is through the application of the test-retest method. Table 1 below shows the scores for a test-retest conducted in 2002.

Table 1
Two Weeks Test-Retest 2002

Components	Usual	Need	Interests		
Esteem	0.81	0.70	Persuasive	0.93	
Acceptance	0.85	0.76	SocialService	0.88	
Structure	0.75	0.72	Scientific	0.92	
Authority	0.82	0.59	Mechanical	0.96	
Advantage	0.80	0.77	Outdoor	0.94	
Activity	0.88	0.72	Numerical	0.89	
Challenge	0.72	0.72	Clerical	0.86	
Empathy	0.88	0.78	Artistic	0.89	
Change	0.80	0.74	Literary	0.90	
Freedom	0.77	0.78	Musical	0.72	
Thought	0.78	0.77			
	ave	0.81	0.73	ave	0.89
	stdev	0.05	0.05	stdev	0.07
	Range	Range		Range	.72 to .96
	.72 to .88	.59 to .78			
All significant at the .0001 level					

The table displays how little the scores changed for one group (N = 77) when they took the Birkman twice (two-weeks between tests), i.e. the scores are reliable. Numbers in the three columns are “correlations”. A correlation measures the relationship strength of two variables (reliability). The correlation for Esteem (U) is circled. The statistic for Esteem ($r = .81$) reflects a strong relationship between responses in the first session and responses in the second session. All correlations above .60 are considered reliable.

Reliability: Internal Consistency

A second measure of reliability analyzes items. Birkman® “items” are the individual statements and job choices that require a response. The Birkman has 442 items used in scale construction. The issue of internal consistency is seldom raised by non-statisticians. Nonetheless, the analysis is important and easy to understand conceptually.

A scale, like Esteem Usual, is internally consistent if the items used to create the scale are more related to the characteristic of Esteem Usual rather than any other component (e.g., Acceptance Usual). Thus, the items related to Esteem should be homogeneous and internally consistent. Researchers Kuder and Richardson in 1937 devised a formula for estimating the reliability of a test, known as Kuder-Richardson 20 or KR20. The value of KR20 will vary between 0 and 1, and scales with values over 0.60 are generally taken to be internally consistent.

Table 2 displays the KR20 values for each component. The value (number) in the two columns represents the Kuder-Richardson coefficient and is a measure of inter-item consistency. Assuming a homogenous scale (one measuring a single attribute), this value indicates the extent to which the items, taken together, are consistently measuring that attribute.

Reliability coefficients (KR20) of the Birkman components:

Table 2

	Needs	Usual
Esteem	.749	.779
Acceptance	.758	.801
Structure	.727	.723
Authority	.614	.662
Advantage	.843	.646
Activity	.744	.720
Empathy	.817	.803
Change	.618	.657
Thought	.720	.676
Challenge	.935	.891
Freedom	.825	.715

The reliability coefficients found through analyses of the items currently comprising the Birkman components, both their Needs and Usual response aspects, are all in the high range (i.e. over .60). Note that these “coefficients” compare favorably with the “coefficients” in Table 1 in the previous section. Taken together, the two reliability analyses demonstrate that the Birkman scales are statistically reliable.

Validity

A general definition of validity is: a questionnaire, test, or psychological instrument that measures what it claims to measure. This is indeed a very broad definition. Specific examples of how validity is measured helps to blend the concepts with the measurement. Four types of validity will be discussed in this section.

Face Validity

Case example:

While you are on a flight from LA to New York, you begin to read the airline journal article and notice a questionnaire on “Leadership”. This seems to be very opportune because you are moving to New York as a newly promoted VP. After completing your questionnaire you compute your score and you are classified as “9” on a 10 point leadership scale. “Wow, you say to yourself, this simple test seems to accurately measure my leadership characteristics!”

This is an example of Face Validity. Many consultants rely on this definition of validity. Face validity is perhaps the simplest scientific definition of validity since it is defined as the mere appearance that the results are relevant, important, or make sense to the test-taker. This validity is important since it is a measure of the perceived accuracy of the test.

The Birkman[®] has always received positive feedback on its face validity by both individuals and groups.

Individual Face Validity Test. A large measure of the success consultants experience during feedback sessions using the Birkman is attributed to face validity. An example of a face validity test that any consultant can apply is the last page of the Life Style Grid[®] report. On this last page, the verbal description of a person with a particular combination of Interest, Usual, Needs and Stress scores is provided. Typically, the client is asked to read the text at the bottom of this last Grid and, more often than not, the client remarks something like: “Wow, this description appears to be right on target for me!”

Organization Face Validity Test. Perhaps the strongest case for face validity is the large number of organizations that have “perceived” that the instrument is relevant, important and makes sense. For over fifty years The Birkman Method[®] has been perceived as a major tool in organizational analysis, promotion, team building, selection, manager development, coaching, merger and acquisitions, and decision making.

Table 3 on the next page lists current organizations and groups using the Birkman managed either by external or internal certified Birkman Consultants.

The Birkman Method clearly demonstrates Face Validity.

Face Validity: Perceived Meaningfulness in Organizations

Table 3

3D/International, Inc.	Marathon Oil Company
ABB Vetco Gray	McLane Company
Acxiom Corporation	MD Anderson Hospital
Alabama Power Company	Merrill Lynch
American Red Cross	Money Mailer, Inc.
Ameritas Life Insurance	NASA
Assn. of Junior Leagues International	OGE Energy Corporation
Austin Energy	Outback Steakhouse
Baker Hughes	Owens Corning
Baylor College of Medicine	Pentasafe, Inc.
Beckman Coulter	Pentax Precision Instrument Corp.
BMC Software	Philips Oral Healthcare, Inc.
BP Amoco	Port of Houston Authority
Campus Crusade for Christ	PowerFood, Inc.
Cardinal Health, Inc.	Prentiss Properties Ltd.
Chevron Texaco	Reliant Energy
Citizens Bank of Canada	Right Management Consultants
Conectiv	St. Luke's Episcopal Hospital
Cooper Cameron	Seattle University
Cox Communications	Shell Oil Company
Deloitte & Touche	Siemens Management Learning
Discovery Communications	Snap-On, Inc.
Drake Beam Morin	Spherion
Dynergy, Inc.	State of Washington
Eli Lilly	Steelcase, Inc.
Emory University	TEC Laboratories
Exxon Mobil Chemical Company	Telus Corporation
Federal Aviation Administration	UN World Food Program
Fluor Daniel, Canada	University of Nebraska at Omaha
Frank Russell Company	University of Colorado at Boulder
Frito Lay	University of South Carolina
Georgia Tech, Dupree College of Management	University of Texas Medical Branch
Hawaiian Electric Company, Inc.	University of Texas Career Center
Hines Interests Ltd. Partnership	Universal Underwriters Group
Houston Astros Baseball Club	US Department of Agriculture
Hughes Christensen, Inc.	US Postal Service
Intel Corporation	Wal-Mart Stores, Inc.
Invesco Funds Group	Waterpik Technologies, Inc.
Iowa State University	Wells Fargo Home Mortgage
Lee Hecht Harrison	Weyerhaeuser Company
	World Bank
	World Vision

Validity: Construct Validity

Construct Validity is defined as the extent to which a test measures the “construct” (attribute, characteristic, trait, etc.) that it is designed to measure.

Case Example:

Suppose that you are applying for a job as an art director and you are given an assessment to measure your creativity. The assessment is comprised of one coloring book and a box of crayons and you are asked to color in all the pictures. You are disappointed when you receive a very low score in this “creativity” assessment. It seems that you colored outside of the lines on 87% of the drawings and were given a low score of 13 (100 – 87). Very disconcerted, you ask the HR professional, “What does staying within the lines of a coloring book have to do with creativity! I have just graduated as an Art major with an A average. This assessment is not valid!”

In a sense, the individual is claiming that high grades in Art school were a valid measure of creativity and, therefore, a score on the creativity test should have been highly, positively related. In other words, the two measures should be highly correlated – by knowing the individual’s scores in Art school, the person should be confident in predicting a high score on any test that claimed to measure the construct of creativity. In the case above, the person would be correct in questioning the validity of the second creativity test.

The person’s question deals with the strength of relationship of two separate measurement instruments. As in the reliability measures, the strength of the relationship between two scores is a measure of “correlation”. The question of construct validity is a very practical question – if I get a high score on creativity on one instrument I assume that I would get a high score on creativity on a second instrument.

When comparing personality, interest, and motivational instruments, construct validity deals with the strength of the relationship between the various personality-related instruments. The question of construct validity is a very practical question. Birkman® researchers have established construct validity with directly comparable instruments such as the 16 PF® (16-Personality Factors), the MMPI® (The Minnesota Multiphasic Personality Inventory), the CPI® (California Psychological Inventory), the EPI® (Eysenck Personality Inventory) and the MBTI® (The Myers-Briggs Type Indicator).

Since this is a comprehensive discussion that would require hundreds of pages, one example has been selected to clarify the understanding on Birkman construct validity. On the next page in Table 4 is a comparison between the component Structure to related 16 PF constructs of Perfectionism, Rule Consciousness, and Openness to Change.

Validity: Construct Validity (continued)

According to the 2001 R&V report, the Structure scales measure:

“an orderliness-based construct that includes following plans, finishing tasks, dealing with detailed tasks, using a systematic approach and their bipolar opposites. Related to conscientiousness, this construct measures the manner of dealing with systems and procedures. High scores reflect orderly, systematic and detail oriented behavior, a preference for planned and controlled tasks and a tendency to become overly constrained by existing plans, procedures or way of doing things when stressed by rapid change of approach, lack of predictability or feelings that tasks are out of control...”

In order to establish construct validity, the same group of individuals must complete each instrument. The table below compares Birkman® and 16 PF® factors. Correlations are the values represented in each cell. Note the intersection of the Structure column and the rows listing Perfectionism, Rule-Consciousness, and Openness to Change.

Table 4: Comparison of Cattell's 16PF Factors with the Birkman Self scales
(all correlations are significant, p = .01)

16PF Scale			Birkman Self Scores									
Scales	Original Labels	5 th Ed. Labels	Emp	Tho	Est	Chg	Aut	Adv	Acc	Act	Str	Fre
O	Timidity	Apprehension	.45	.42	.39	.36	.17	.21	-.18	-.33	na	na
Q ⁴	Ergic Tension	Tension	.50	.37	.34	.42	.30	.17	-.17	-.32	na	na
L	Protension	Vigilance	.31	na	.23	.18	.46	.35	na	na	na	na
H	Parmia	Social Boldness	na	-.32	-.33	na	.17	.64	.31	na	na	na
C	Ego Strength	Emotional Stability	-.43	-.45	-.33	-.37	-.24	-.19	.27	-.39	na	na
A	Cyclothemia	Warmth	na	na	na	na	na	na	.33	na	na	na
F	Surgency	Liveliness	na	-.19	na	na	.27	na	.53	.20	na	na
Q ²	Self-Sufficiency	Self-Reliance	na	na	na	na	na	na	-.39	na	na	na
E	Dominance	Dominance	na	na	na	na	.34	.17	.23	na	na	na
Q ³	Self-Sentiment	Perfectionism	-.19	na	na	na	na	na	na	na	.42	na
G	Superego Strength	Rule-Consciousness	na	na	na	na	na	-.18	na	na	.48	-.21
Q ¹	Radicalism	Openness to Change	na	na	na	na	na	na	na	na	-.20	na
I	Premisa	Sensitivity	na	na	na	-.10	na	na	na	na	na	na

The following correlations had no correlations at p.01 level

M	Alexia	Abstractiveness	emp=empathy	aut=authority	str=structure
N	Shrewdness	Privateness	tho=thpught	adv=advance	fre=freedom
B	Intelligence	Reasoning	est=esteem	acc=acceptance	
			chg=change	act=activity	

In each cell, there is either a correlation or a “na” which means there was not a statistically strong relationship between the component and the particular factor. Note the Structure scale is positively related to Perfectionism and Rule-Consciousness and negatively related to Openness to Change. These correlations would suggest that an individual with a high score on Structure would probably also get high scores on Perfectionism and Rule-Consciousness and a low score on Openness to Change (.42 is a moderate relationship, typical of factor correlations, and it is statistically significant – not due to chance).

Other construct validity analyses have shown that Structure is also positively related to Conscientiousness (“Big Five” - NEO-PI-R™) and Judging (MBTI®). Clearly, Structure is measuring the same underlying construct as these other instruments. The other nine components can be compared in the same manner.

Validity: Criterion Validity

The third definition of validity has two subtypes: Predictive and Concurrent Validity. Criterion validity is based on how well a test, assessment, or inventory predicts a particular outcome (criterion) such as job success, job satisfaction, or retention rates.

Predictive Validity is demonstrated when the same individuals who complete an assessment are followed into their careers and perform at the level predicted over a period of time. This process usually requires two to five years.

Concurrent Validity is demonstrated when one group of individuals are used to create the prediction formula (model) and “concurrently” the performance of a second group of individuals are evaluated based on the scores of the original group. Since individuals do not have to be followed for years, the results can be analyzed within months.

Concurrent and predictive validity has been demonstrated for a variety of outcome measures during the fifty years that Birkman[®] has conducted research in business, educational, industrial, educational, and community settings.

Example Longitudinal Study

The Birkman research team entered a three-year longitudinal study with Combined Insurance to study the effectiveness of organization-wide application of The Method to specific organizational development issues. The four objectives and results were:

- *Objective:* To reduce hiring costs, recruiting and training time spent by managers
- *Results:* The average reduction of turnover in a three-year period was 53% - a savings in hiring alone of around \$1.6 million
- *Objective:* To demonstrate successful selection improves tenure
- *Results:* With improved tenure in the selected sample 54,000 individual employee weeks were added
- *Objective:* To replace the bottom 48% of the existing sales force with individuals similar to the individuals in the upper 52% of the sales force
- *Results:* The percentage of average and good performers increased from 69% to 90% of the sale force
- *Objective:* To find short-term predictors of success for hiring at the management level
- *Results:* All predictors showed important increases based on the new managers selected.

Validity: Criterion Validity for Interview Guide Scores

The Birkman Interview Guide scores are a product of research using regression analysis. While the term may seem a bit daunting at first, the concepts behind it are easy to grasp. A line-by-line explanation is provided using Table 5 below.

Table 5

Predictor: Directive Management Style

<i>Cross Validations</i>		<i>r</i>	<i>N</i>
Based on criterion		0.41	867

Sample Descriptions		
Code	N	Descriptions of Contrasted Groups
1	414	Operations managers, engineering managers, project managers, construction managers, oil and gas managers, utility company managers, police management
0	568	Public school counselors, psychologists, psychiatrists, physicians, college professors, medical researchers, therapists, lawyers
0	438	Property managers, hotel managers, restaurant managers, insurance sales managers, sales managers, bank managers, account managers

<i>Characteristics of Multiple Regression Equations</i>						
Equation	N	R	R²	F	(d.f.)	Signif F
Validation 1	1420	0.54	0.29	48.77	12/1407	0.0001

The Directive Management score, generated in the Interview Guide, is a measure of how closely the individual's answering pattern resembles that of a well-established group previously identified as demonstrating a directive management style. Thus, the score is a prediction of success. The higher the score, the more similar the response pattern is to the criterion. However, since groups were used to predict group behavior, it is not appropriate to overemphasize a single individual score. Rather, a pattern of scores should always be analyzed. This management "style" is created by adding 12 to 16 Birkman[®] scores together. When combined, they predict a portion of the characteristics required to be a successful directive style of manager.

Validity was demonstrated in two cross validation analyses. First, the table shows a correlation of .41 between a Criterion Group and a Contrast Group (total sample size of 867). This information is provided in the first three lines of the table.

Second, a Criterion of "Red" managers (N = 414) and two Contrast Groups (Blue Job Titles, N = 568; Green and Yellow Job Titles, N = 438) were analyzed through regression analysis. As the last two lines of the table indicate, the regression equation was able to discriminate between the 414 Directive Style managers and the 1,006 mixed Green and Blue managers (total N = 1420).

This second validity analysis yielded a correlation of .54, and the equation accounted for 29% ($R^2 = .29$) of the variance. The results were highly statistically significant ($p = .0001$) which means that these results were clearly not due to chance. In simple terms, 29% of the successful Directive Management Style can be accounted for by the predictive model (R^2 is simply the correlation squared: $.54 * .54 = .29$).

Validity of Career Families

Validity for Career Families is provided through the application of cross-validities research. Cross validity is a two-stage process. Following is the methodology used by researchers at Birkman®.

In the first stage, Birkman scores, collected between 1980 and 1995, were used to develop scales of career orientation (Job Titles, Job Families, General Groups). Birkman scores (a set of predictors) were used to separate a large group of people who had taken the Birkman into two distinct groups. One group is the Criterion Group (25,000 people of known job title) and the second group is the Contrast Group (36,000 people in general). Numerical scales were developed that could classify individual scores into one of the Criterion Groups (e.g., Educational Career Scores, Accounting Career Scores) as opposed to the Contrast Group (average reference group). The scales developed are represented in scores reported in the Career Management Report.

In the second stage, Birkman scores collected between 1995 and 2001 (17,166 people with adequate career data) were used to independently replicate the Job Titles, Job Families, General Groups scores. Correlations between the Job Families scores collected in the first stage and Job Families scores in the second stage provide the measure of validity.

Table 6 below shows one example of the cross-validity results for the comparison.

Table 6

Cross Validities of OrganizationalSM Strengths		
Scale	r₁	n
Planning Group (Blue)		
Artistic Careers	0.52	799
Educational Careers	0.45	801
Social Service / Counseling	0.44	553
Employee Relations / Training	0.40	1603
Medical Professions	0.33	727

1. All r values significant beyond the .01 level

The far left column contains the list of Job Families (Blue). The middle column (r) displays the correlation values. The last column (n) gives the number of people in each group. Therefore the Artistic Careers scores for the first sample collected from 1980 to 1995 and scores on Artistic Careers for the second sample collected from 1995 to 2001 are related by a correlation of .52.

The cross validity correlations for across all job families range from .29 to .58 and are all significant at the .01 level. These correlations are very conservative given the use of two independent raters and the time period (20 years) covered by the study.

Cultural Comparisons

The Birkman Method® is intended for use with a wide range of people of various ages, walks of life, as well as different languages. Table 6 below lists the average scores for non-English translations collected over the past year (2002). The data reveal similarities and differences in average scores (the English language population is the sample (N=107539) used in the 2000 R&V report).

	Table 6				
	US	Spanish	German	French (Canadian)	Dutch
Usual					
Esteem	23.26	19.55	22.73	34.98	19.15
Acceptance	76.71	65.20	72.23	57.45	69.67
Structure	69.31	85.76	62.02	77.85	66.14
Authority	53.69	65.77	73.91	62.50	68.62
Advantage	19.59	22.94	25.42	21.70	25.74
Activity	79.73	83.61	73.82	75.21	80.45
Challenge	50.03	57.37	37.79	52.51	54.73
Empathy	39.60	44.22	57.65	51.93	46.13
Change	55.06	55.70	49.55	55.14	52.33
Freedom	35.70	35.90	42.78	35.81	42.47
Thought	37.44	35.70	39.75	47.69	35.82
Need					
Esteem	54.45	64.49	49.27	66.00	33.63
Acceptance	54.93	35.67	44.66	31.99	57.89
Structure	52.33	58.83	52.45	45.82	59.22
Authority	57.79	62.57	61.51	71.48	68.95
Advantage	54.49	61.23	57.89	75.75	60.07
Activity	53.86	47.20	58.59	39.91	62.05
Challenge	50.03	52.51	54.73	37.79	57.37
Empathy	56.11	64.89	58.26	79.73	64.16
Change	56.26	60.17	54.46	68.47	62.97
Freedom	54.91	65.05	57.16	79.24	65.97
Thought	57.28	63.75	56.19	71.42	61.38
Interest					
Persuasion	54.02	59.82	66.57	50.19	58.96
Social Service	48.40	54.57	65.63	53.70	55.93
Scientific	51.81	39.00	40.72	52.96	45.67
Mechanical	52.23	57.15	40.70	56.33	52.12
Outdoor	49.85	49.12	44.43	48.95	47.56
Numerical	53.19	63.97	42.48	53.92	42.25
Clerical	49.89	66.85	62.68	61.35	52.01
Artistic	48.00	40.70	53.53	52.71	58.56
Literary	54.59	43.69	57.44	52.16	53.03
Musical	54.18	49.50	48.80	48.68	49.42

Continuous reliability and validity research is being conducted by Birkman® researchers to determine what differences can be attributed to sampling differences and what differences could point to cultural differences. The protocol for establishing reliability and validity across languages is provided on the next page. Conclusions based on all non-English data collected up to this point in time is insufficient to warrant cultural generalizations.

Non-English Research and Validity Protocol

1. The Translation

For the Birkman[®] Questionnaire initial translation, BI selects translators who have a high degree of facility in both English and the target language.

2. Back-Translation

A back-translation from the target language to English is completed by an independent bilingual translator.

3. Comparison of the English and Back-Translated Versions

The back-translation is compared with the original English version to determine if the item meanings have been correctly translated.

4. Study of Equivalency

- The accuracy of the Questionnaire translation for the target language will be evaluated through a preliminary pilot study.
- Revision of translation released as pilot study version (PV1).

5. Study to Determine the Adequacy of the American Norms

- A study is conducted of the norm scores for the Questionnaire in the target language.
- A research study is conducted where a sample is selected based on the demographic group targeted as customer base.
- Decision to create separate norms is made based on empirical results.

6. Development of Norms for the Target Language

If the norms for the translated Questionnaire are significantly different from the American norms, a representative sample of the results from the translated Questionnaire will be collected to serve as a basis for new norms.

7. Reliability Studies Research

- Identify research partner to begin long-term reliability studies.
- Series of test-retests and internal consistency analyses to be conducted.

8. Validity Studies Research

- Identify research partner to begin long-term validity studies.
- Face, construct, and criterion-validity studies to be conducted.

As a general rule, it will take approximately 3-5 years to complete all the steps listed above for each new Questionnaire translation.

References for The Birkman Method®

- Bernstein, A.B. and Schaffzin, N. R. (1998). The Princeton Review: Guide to Your Career (3rd Ed.), New York, Random House, Inc.
- Birkman, R. W. (1961). Development of a personality test using social and self-perception. Unpublished Doctoral Dissertation. University of Texas, Austin.
- Birkman, R. W. (1967). A New Frontier for the Church in Mental Health, in O. Hobart Mowrer (Ed). Morality and Mental Health, pp. 304-305, Chicago: Rand McNally & Company.
- Birkman, R. W. (1995). True Colors. Nashville, Tennessee: Thomas Nelson, Inc.
- Birkman, R. W. (2001). The Birkman Method® Reliability and Validity Study. Houston, TX: Birkman International, Inc.
- Broussard, W. J., Sadler, T. G., & Mefferd, R. B. Jr. (1976). The Birkman Method®: An industrial psychometric tool applied to vocational counseling and career planning for an ex-offender population. Sociological Abstracts, 66, 244. Texas Criminal Justice System and the Board of Pardons and Paroles, 1972-1980.
- Cattell, Raymond B., & Eber, Herbert, & Tatsuoka, Maurice. (1970). *Handbook for the Sixteen Personality Questionnaire*. Champaign, Illinois: Institute for Personality and Ability Testing, Inc.
- Cruz, P. (2003). Evaluation of Others as a Method of Personality Assessment. White paper.
- Dowling, S. S. and Larkey, F. R. (2000). Personality Factors as Predictors for Success in Supervisory Training. Presentation session, American Speech-Language-Hearing Association Annual Convention. Washington, D.C.
- Eissler, V. C., Sadler, T. G., Kirkpatrick, D. E., & Mefferd, R. B. Jr. (1976). Vocational potential: A comparison of inmates and free world workers via the Birkman Method®. Sociological Abstracts, 66, 244. Texas Criminal Justice System and the Board of Pardons and Paroles, 1972-1980.
- Eysenck, H. J. & Eysenck, S. G. B. (1975). *The Eysenck Personality Questionnaire*. Seven Oaks, England: Hodder & Stroughton.
- Gough, H. G. and Bradley, P. (1996). *California Personality Inventory Manual (3rd ed)*. Palo Alto, CA: Consulting Psychologists Press.
- Hathaway, S. R., and McKinley, J. C. (1983). *Minnesota Multiphasic Personality Inventory Manual*. New York, NY: The Psychological Corporation.

- Justice, B., & Birkman, R.W. (1972). An effort to distinguish the violent from the non-violent. *Journal of Southern Medical Association*, 65, 703-706.
- Krausz, M. J. (1979). Psychological correlates of the criminal personality: a quantification of the Yochelson-Samenow zero-state concept. Unpublished dissertation, University of Houston, Houston, Texas.
- Larkey, F. R. (2001). Personality Guidebook: An Introduction to Personality Assessment Applications. Published in Houston, Texas, by The Reed Larkey Group.
- Larkey, F. R. (2002). The Birkman Method® Developmental Heritage. White paper, <http://birkman.com/Researchppf.html>.
- Larkey, F. R. (2002). How to Understand the Myers-Briggs Type Indicator® from The Birkman Method® Perspective. White paper, <http://birkman.com/Researchppf.html>.
- Larkey, F. R. (2002). The Birkman Method® Developmental Heritage: Reliability and Validity (1972). White paper, <http://birkman.com/Researchppf.html>.
- Larkey, F. R. and Dowling, S.S. (2000). Speech-Language Pathology: Personality Types, Team Participation. Poster presentation, American Speech-Language-Hearing Association Annual Convention. Washington, D.C.
- Larkey, F. R. and Knight, J. L. (2002). Reliabilities/Validities: The Birkman Method® and the Myers-Briggs Type Indicator®. White paper, <http://birkman.com/Researchppf.html>.
- Larkey, F. R. and Knight, J. L. (2002). Test-Retest Reliability and The Birkman Method®. White paper, <http://birkman.com/Researchppf.html>.
- Larkey, F. R. (2003). Master the Mystery: Investigate Yourself. InterChange, Schlumberger Technology Corporation, Sugar Land, Texas.
- The Birkman Method® Developmental Heritage: Reliability and Validity (1972). White paper, <http://birkman.com/Researchppf.html>.
- McCall, G. A. (1970). The Birkman Method®: a resource for small group experience in the local congregation, unpublished dissertation, Southern Methodist University, Dallas, Texas.
- Mefferd, R. B., Jr. (1966). Structuring physiological correlates of mental processes and states: the study of biological correlates and mental processes (Chapter 20). In R. B. Cattell (Ed.) Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally, pp. 684-710.
- Mefferd, R. B., Jr. (1968). Technique for minimizing the instrumental factor. Multivariate Behavior Research, Vol. 3, pp. 339-354.

- Mefferd, R. B., Jr. (1971). Use of individual differences in personnel selection. In U.S. Army Advanced Material Concepts Agency Report, AMCA 71-010, 26-27.
- Mefferd, R. B., Jr. (1972). The Birkman Method[®] for manpower selection, classification, assessment, motivation counseling, and training: Its reliabilities and validities as of March 1972. Houston: Birkman & Associates, Inc.
- Mefferd, R. B., Jr. (1974). The scores on the six board examinations of medical school graduates: Discrimination by personality variables of psychiatry and public health versus the other four speciality areas. Houston: Baylor College of Medicine.
- Mefferd, R. B., Jr. (1975). The Birkman Method[®]: Reliabilities and validities for business and industry. Houston: Birkman & Associates, Inc.
- Mefferd, R. B., Jr. (1975). Experimentation and research in the Birkman Program. In Changing tasks and roles in higher education, Austin College Total Institutional Project. National Science Foundation and National Endowment for the Humanities, 1972-1975. Sherman, TX: Austin College.
- Mefferd, R. B., Jr. (1977). Comments of the comparability of the Birkman Method[®] with other standard tests and inventories. Houston: Birkman & Associates, Inc.
- Mefferd, R. B., Jr. (1978). The development of special aids in determining dispositions (AIDD) of adult offenders. Texas Board of Pardons and Paroles.
- Mefferd, R. B., Jr. (1983; revised 1988). Validity Report: The Birkman Method[®] interview guide. Houston: Birkman & Associates, Inc.
- Mefferd, R. B., Jr. (1984). Interaction of teachers' judgments of students: Personality and school performance. Psychological Reports, 54, 921-922. Sheldon Independent School District.
- Mefferd, R. B., Jr. (1984). Performance assessment and personalities of teachers and students. Psychological Reports, 54, 317-318. Sheldon Independent School District.
- Mefferd, R. B., Jr. (1984). Prediction of susceptibility to substance abuse. Psychological Reports, 54, 317-318.
- Mefferd, R. B., Jr. (1987). A content-valid test (with an alternate form required for difficulty level) for level-3 customer service representatives of the public service of Oklahoma Utility Company. Houston: Birkman & Associates, Inc.
- Mefferd, R. B., Jr. (1988). Introduction to the perceptions of job requirements[®] (PJR). Houston: Birkman & Associates, Inc.
- Mefferd, R. B., Jr. (1989). Work related attitudes of employees of the Compressor

Engineering Corporation. Houston: Birkman & Associates, Inc.

- Mefferd, R. B., Jr. (1992). Cross-cultural validity of the Birkman Method[®]: Comparison of Native American and Native Oriental graduate students in schools of business. Houston: Birkman & Associates, Inc. (funded by the William M. Mercer Group for a study at the University of Chicago).
- Mefferd, R. B., Jr. (1993). Manual for the Birkman Method[®]. Houston: Birkman & Associates, Inc.
- Mefferd, R. B., Jr., Hale, H. B., Shannon, I. L., & Prigmore, J. R. (1970). Stress responses as criteria for personnel selection: A baseline study. Aerospace Medicine, *42*, 42-51.
- Mefferd, R. B., Jr., & Sadler, T. G. (1976). Job analysis and validation of a performance evaluation procedure of county-level and area-level agents of the Mississippi Cooperative Extension Service. Starkville, MS: Mississippi Cooperative Extension Service.
- Mefferd, R. B., Jr., & Sadler, T. G. (1984). Review of development of a performance evaluation procedure of a performance evaluation instrument by the Mississippi Cooperative Extension Service: 1974-1984. Starkville, MS: Mississippi Cooperative Extension Service.
- Myers, I. B., McCauley, M. H., Quenk, N. L. & Hammer, A. L. (1998). *MBTI manual (3rd ed.)*. Palo Alto, CA: Consulting Psychologists Press).
- Price, J. R, Sadler, T. G., & Mefferd, R. B. J. (1971). Set, satiation, and confusion: A reply to Sadler and Mefferd. Perceptual & Motor Skills, *33(3, Pt. 1)*, 703-707, 999-1005.
- Sadler, T. G., & Mefferd, R. B. J. (1971). The interaction of extraversion and neuroticism in human operant behavior. Journal of Experimental Research in Personality, *5(4)*, 278-285.
- Sadler, T. G., Mefferd, R. B. J., & Houck, R. L. (1971). The interaction of extraversion and neuroticism in orienting response habituation. Psychophysiology, *8(3)*, 312-318.
- Schweiger, D. M. (2002). M&A Integration: A Framework for Executives and Managers. New York: McGraw-Hill.