

The Birkman Method® Developmental Heritage: Reliability and Validity (1972)

This paper is a summary of the first comprehensive reliability and validity report published by the President and Director of Research, Roy B. Mefferd. The report is based on in-depth research studies and the application of the highest level of statistical analysis and experimental design. The primary purpose of this summary is to clearly demonstrate the continual research integrity of the Birkman organization as The Method® evolved. The first white paper in this series emphasized the test development guidance of Dr. Birkman by leading psychometricians (e.g. Benjamin Fructer, the author of Introduction to Factor Analysis in 1954). To reinforce the quality of the professionals who further developed Dr. Birkman's ideas, a complete bibliography of Dr. Mefferd is included on the web site. The original report is over forty-eight pages and is available by request.

The Birkman Method® for Manpower Selection, Classification, Assessment, Motivation Counseling, and Training: Its Reliabilities and Validities as of March 1972

Roy B. Mefferd, Jr., Ph.D.
President and Director of Research
The Birkman-Mefferd Research Foundation

Abstract

“This report examines the professional personnel, psychological instruments, methodology and analytical practices used by the personnel management consulting firm, Birkman and Associates, and the reliabilities and validities of each aspect of the components of the Birkman Method—an integrated package of selection, classification, training, counseling, and motivation procedures. The firm has been in business for two decades, and associated with it is a non-profit research organization that provides public service functions—The Birkman-Mefferd Research Foundation. One function of the Foundation is to do research on the instruments and procedures of the firm to increase their cultural fairness, reliabilities and validities, and to report regularly on these. The senior staffs of both organizations are fully qualified and experienced professionals.

The instruments are low-pressure, un-timed, and non-threatening. They require about an hour to complete. They are extremely difficult to “fake” and there is a simplified “child’s” form. Intelligence is ordinarily not measured per se—rather the emphasis is on interests, motives, attitudes, and the like. Standardization norms are available for various groups (e.g., 11 categories of managers, 12 categories of non-managers, professional engineers, teachers and ministers, prisoners; children and adults; males and females; whites, blacks and Spanish surnamed), and the categories are being extended rapidly. A group of special equations are available for predicting job classification, work performance, strengths and work problems, absenteeism, violence-proneness and the like.

Multivariate techniques (e.g., multiple regressions, multiple discriminate function analysis, factor analysis, etc.) are used almost exclusively. A large number of data points are obtained, and these are shown to be very powerful in discriminating the pertinent attributes. Close controls on the statistical methodology itself are maintained. Individuals and employees are offered a pattern of options (i.e., predicted classification and performance) for 23 broad job classifications. These were developed by means of hierarchical categorization and factor analytic techniques.

Birkman & Associates emphasize the importance of objective, unbiased criterion measures (e.g., or work performance, absenteeism), the homogeneity of the job categories, and an accurate delineation of these categories. Using these various criteria, five basic equations are developed, in sharp contrast to the usual one or two: job classification, job performance, support based on the breadth and variety of jobs for which an individual is qualified, self-control and self-discipline, and life style-success profile. All enter the final decision. Additionally, equations are developed for predicting separately Good, Average, and Weak-Poor work performance, and these are then combined into a single prediction.

A number of reliabilities having to do with the characteristics of the instruments per se are computed: (1) Inter-rater Reliability, $r = .88$ to $.96$; (2) Internal Consistency, $r = .78$ to $.95$ (odd-even reliabilities, corrected for test length for six different samples including children); and (3) Test-Retest for the individual item scores (mean $r = .74$), and 48 basic scores derived from the 336 items, ($r = .69$ to $.96$). A split-sample reliability for factor structure was determined from a basic sample of 1,008 managers and good employees. The three sets of factors (viz. for 1,008; 672 and 336 subjects) were virtually identical.

Seven large groups of subjects were factor analyzed and the factor structures for the primary factors were virtually identical for managers, non-managers, school teachers, insurance salesmen, item salesmen, children and prisoners. However, there also were quite specific factors for most of these.

A simplified form suitable for children and poorly educated adults was developed and standardized against the adult form: $r = .88$ for the 48 basic scores for a group of prisoners.

A variety of validities (i.e., does the instrument or procedure do what it is intended to do?) are presented for a variety of groups and uses: of the Seminars (is there increased self-understanding?); of the Predictive Equations (empirical, statistical)—(concurrent and cross validities); of the Birkman Patterns of “hang-ups” and risk factors, and of the Life Style graphical procedure. The statistical parameters of the current multiple regression and discriminant function equations are presented. All these validities and values are highly significant and successfully discriminate the pertinent groups.

The statistical dependencies within the total data points (337) of the basic instruments were examined in several ways and they were found to exercise only a minor negative effect. At the same time the variety of scores did increase accuracy of predictions by tapping more specifically the interaction variance.

The cultural fairness of the instruments and procedures were examined in several ways. It was concluded that, to date, there is no evidence that any of these led to unfair discrimination of the major ethnic groups or of the sexes. A positive program of affirmative action to reduce unfair discrimination of minority groups is in effect at Birkman & Associates.

Correlates of the basic scores from the Birkman instruments are presented (e.g., age, education, intelligence, sex).

A list of current clients who may be consulted for information about the practical success of the Birkman Method is provided.”

(Quoted from Mefferd, 1972, p. ii-iii)

Validity Analyses

A series of analyses were conducted to establish validity of the Birkman Method® in order to ensure that it was measuring what it purported to measure.

Criterion-Related Validity

A major focus of this original research conducted at Birkman was the use of newly developed multivariate inferential statistics to establish criterion-related validity through the use of a concurrent methodology. This concurrent validity was established through a multiple step process. First, performance appraisals were obtained for a minimum of 100 employees in each company that was analyzed. Second, each organization's idiosyncratic rating system was distilled into four global categories: 1) "Definitely keep even during personnel reduction," 2) "Keep as long as possible," 3) "Transfer or let go during any personnel reduction," or 4) "Will discharge or transfer as soon as possible." Third, these ratings were then used to sort employees into performance groups (e.g., Poor, Weak, Average, Good, Very Good, Excellent) along a ranking scale. Fourth, after the Birkman Method data was collected from employees, a series of performance regression equations were produced for three contrasted groups: Good vs. All Others, Average vs. All Others, and Weak/Poor vs. All Others. Furthermore, classification equations were also created comparing employees in different job families in order to create profiles across different job clusters (23 clusters based on analyses of 3,000 employees at the time of writing). Finally, after the equations had been produced, Birkman used the equations and compared the *predicted* job performance of 1,371 employees (across 12 different job classifications) to their *actual* job performance. The equations produced extremely high accuracy "hit" rates ranging from 71% to 94% of employees correctly classified as "Good" or "Other." After additional data had been collected 18 months later, the equations were recalculated, and the prediction was markedly improved for 2,141 employees across 4 different job classifications.

These equations (and three others to be described later) were then used in a predictive fashion during organizations' selection procedures. Specifically, applicants were evaluated on five criteria: "1) job classification equation, 2) performance equation, 3) a support predictor based on the breadth and variety of jobs for which he is qualified, 4) a predictor of self-control and self-discipline, and 5) a life style, success profile based on a graphically derived equation" (p. 11). These predictions were summarized through the acronym *MASEL*, or Motivation, Aptitude, Supporting Strengths, Equilibrium, and Life Style. To establish predictive and cross-validity, Birkman & Associates worked with a service technician firm (N = 29) and a sales firm (N = 23) to re-apply their original prediction equation to employees who had been hired (either using the equation or the company's older selection procedures) in order to determine how well it had worked in predicting job performance. In both instances, the Birkman had a hit rate at least 20 percentage points higher than the older method. Notably, this number would likely have been even higher if the poor performers who were selected using the old method (but who had subsequently been fired or who had quit) were included in the sample. A different sample using 76 prisoners and 76 non-criminals revealed that the Birkman prediction equation was 95% effective at classifying the two different groups into their respective categories.

Seminar Validity

Although there were no empirical, quantitative data collected as to the effectiveness of the Birkman Seminars, qualitative data revealed that over 75% of seminar participants believed

that their reports and their group members' reports were accurate, that the reports were useful, and that others in their organizations would benefit from using the report.

Validity of "Patterns"

The Birkman Method® originally measured "Patterns," or "estimates of the degree of balance a person has between his various personality traits. Whatever the personality traits may be, are they intensive or mild, do they harmonize or blend, do they contrast properly with each other, or do they clash and interfere with his judgment in making decisions or in making constructive use of his own talents?" (p. 30). Data collected from 200 salesmen whose performance was classified as "Good" or "Poor" revealed that good performers had far fewer "hang-ups" or disturbances in patterns than poor performers. A follow-up study using 250 good and 250 poor insurance salesmen indicated that, although the univariate pattern criterion was not able to distinguish well between the groups, the multivariate risk criterion was. This same finding was true in an additional sample of 1,200 employees.

Validity of the Graphical Procedure

Birkman uses a graphical procedure to plot people's Most People, Self, and Interest scores within a quadrant. "The vertical bipolar dimension extends from Controlling at the bottom to Commanding at the top, and the horizontal bipolar dimension extends from Active on the left to Verbal on the right. The resulting quadrants are characterized clockwise from the upper left as Organizing (Production-centered [Red]), Supervising (People-Centered [Green]), Planning (Idea-Centered [Blue]), and Controlling (Procedure-centered [Yellow])" (p. 32-33). A factor analysis of 164 salesmen, 100 Good and 100 Poor employees, and 80 public school teachers revealed that the stability of the graphical points was sound and that the graphical data have a valid and reliable relationship with the basic data.

Reliability Analyses

The consistency and stability of the Birkman Method® Questionnaire was assessed through several tests of reliability, including internal consistency analyses, two different levels of test-retest reliability, and factor structure comparisons across gender, age, race, test forms, and job clusters.

Internal Consistency

Odd-even reliability is calculated by correlating the summed score on the odd-numbered items with the summed score of the even-numbered items (and then typically using a correction for test length to increase the validity coefficient). Data from 483 people (109 male salesmen, 100 good performers, 100 poor performers, 42 male prisoners, 132 children) were used to calculate this split-half reliability with a Spearman-Brown Formula correction. Results indicated exceptionally good reliabilities, ranging from .88 to .95.

Test-Retest Reliability

An immediate test-retest reliability analysis was conducted on a sample of 42 prisoners and revealed correlations from the first to the second test ranging from .69 to .96. Furthermore, a two-week test-retest reliability analysis was conducted on 132 children (a group thought to have large systematic changes and fluctuations in their self-concept) in order to establish a set of

minimum correlations. As expected, these component reliabilities were lower, but still high, ranging from .40 to .99.

Factor Structure

To determine if and to what extent the factor structure of the Birkman Method® could be maintained across different groups, a series of principle components factor analyses with a varimax rotation was used. Specifically, analyses were conducted comparing across 954 managers (ranked good, average, and poor), 92 public school teachers, 59 public school administrators, 106 prisoners, 79 insurance salesmen, 109 item salesmen, 1051 employees, and 132 children. With the diversity of age, race, gender, occupation, and work performance in the sample, it was concluded that the inherent factor structure was stable across these different groups and that the Birkman Method® did not cause any potential adverse impact towards any group. To further demonstrate consistency, a split-sample reliability for factor structure was calculated, and the consistent results indicated a stable pattern of inter-relationships among the large samples.

Additional Material

The remainder of the volume is devoted to providing answers to commonly asked questions. The questions, along with the parsimonious Yes or No answer, are below. More complete and thorough answers can be found in the complete report.

- Does the Birkman provide employees with a range of options for job classification? **Yes**
- Is there a negative effect of using both basic and derived scores in multivariate equations in the Birkman? **No**
- Can the Birkman validly be used across different cultural groups for making predictions? **Yes**
- Can the Birkman make meaningfully significant (as opposed to just statistically significant) distinctions among people? **Yes**

Additionally, some preliminary concurrent validity comparisons are made between the Birkman and the Otis Quick Scoring Test (N = 132 children), but other studies were not conducted using different instruments because of “the large expense/small benefit ratio involved” (p. 44).

Bibliography

- Berdie, R. F. "Validities of the Strong Vocational Interest Blank." In The Strong Vocational Interest Blank: Research and Uses. Layton, W. L. (Ed.), Minneapolis: University of Minnesota Press, 1960.
- Cattell, R. B. "The Scree Test for the Number of Factors." Multivariate Behavioral Research, 1966, Vol. 1, pp. 245-276.
- Cattell, R. B. "Theory of Situational, Instrument, Second Order, and Refraction Factors in Personality Structure Research." Psychological Bulletin, 1961, Vol. 58, pp. 160-174.
- Cronbach, L. J. H. & Glesner, G. C. Psychological Testing and Personnel Decisions. Urbana: University of Illinois Press, 1965.
- Department of Labor, "Validation of Employment Tests by Contractors and Sub-Contractors Subject to Provision of Executive Order 11246." Federal Register, 1968, Vol. 33, No. 186.
- Einhorn, H. J., & Bass, A. R. "Methodological Considerations Relevant to Discrimination in Employment Testing." Psychological Bulletin, 1971, Vol. 75, pp. 261-269.
- Ghiselli, E. E. The Validity of Occupational Aptitude Tests. N. Y.: John Wiley & Sons, 1966.
- Guilford, J. P. Psychometric Methods. N. Y.: McGraw-Hill, 1954.
- Harman, H. H. Modern Factor Analysis. Chicago: Univ. Chicago Press, 1960.
- Hiestand, D. L. "Discrimination in Employment. An Appraisal of the Research." Policy Paper No. 16. Ann Arbor: The Institute of Labor and Industrial Relations, University of Michigan, 1970.
- Horst, P. "The Logic of Personnel Selection and Classification." In Psychological Principles in System Development. R. M. Gagne (Ed.). N. Y.: Holt Rinehart and Winston, 1963, pp. 231-271.
- Justice, B., & Birkman, R. W. "An Effort to Distinguish the Violent from the Non-Violent." Journal Southern Medical Association (in press).
- Kirkpatrick, J. J., Ewen, R. B., Barrett, R. S., & Katzell, R. A. Testing and Fair Employment. New York: New York University Press, 1968.
- Mefferd, R. B., Jr. "Structuring Physiological Correlates of Mental Processes and States: The Study of Biological Correlates and Mental Processes" (Chapter 22) in R. B. Cattell (Ed.) Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally, 1966, pp. 684-710.
- Mefferd, R. B., Jr. "Technique for Minimizing the Instrumental Factor." Multivariate Behavioral Research, 1968, Vol. 3, pp. 339-354.
- Mefferd, R. B., Jr. "Use of Individual Differences in Personnel Selection, and Discussion" in J. R. Hennessy (Ed.). The Prediction and Understanding of Human Behavior, U. S. Army Advanced Material Concepts Agency, Report AMCA 71-010, May, 1971.
- Mefferd, R. B., Jr., Hale, H. B., Shannon, I. L., & Prigmore, J. R. "Stress Responses as Criteria for Personnel Selection: Baseline Study." Aerospace Medicine, 1970, Vol. 42, pp. 42-51.
- Mefferd, R. B., Jr. & Wieland, B. A. "Extraction and Utilization of Components of Variance from Psychophysiological Variables," Society for Psychophysiological Research, Washington, D. C., October 1964.
- Mefferd, R. B., Jr. & Wieland, B. A. "Comparison of Responses to Anticipated States: The Study of Biological Correlates and Mental Processes." (Chapter 22) in R. B. Cattell

- (Ed.). Handbook of Multivariate Experimental Psychology. Chicago: Rand, 1966, pp. 684-710.
- Nunnally, J. C. Psychometric Theory, N. Y.: McGraw-Hill, 1967.
- Pokorny, A. D., & Mefferd, R. B., Jr. "Geomagnetic Fluctuations and Disturbed Behavior." Journal of Nervous and Mental Diseases. 1966, Vol. 143, pp. 140-151.
- Roessler, R., Burch, N. R., & Mefferd, R. B. Jr. "Personality Correlates of Catecholamine Excretion under Stress." Journal of Psychosomatic Research, 1965, Vol. 11, pp. 181-185.
- Rulon, P. J., Tiedeman, D. V., Tatsuoka, M. M., & Langmuir, C. R. Multivariate Statistics for Personnel Classification. N. Y.: Wiley, 1967.
- Sadler, T. G., Mefferd, R. B. Jr., & Houck, R. L. "The Interaction of Extraversion and Neuroticism in Orienting Response Habituation." Psychophysiology, 1971, Vol. 8, pp. 312-318.
- Sadler, T. G., & Mefferd, R. B. Jr. "The Interaction of Extraversion and Neuroticism in Human Operant Behavior." Journal of Experimental Research in Personality, 1971, Vol. 5, pp. 278-285.
- Sanders, E. M., Mefferd, R. B. Jr., & Brown, O. H. "Verbal-Quantitative Ability and Certain Personality and Metabolic Characteristics of Male College Students." Educational and Psychological Measurement, 1960, Vol. 20, pp. 491-503.
- Snedacor, G. W., & Cochran, W. G. Statistical Methods. Ames: Iowa State Univ. Press, 1967, p. 407.
- Stewart, N. "A.G.C.T. Scores of Army Personnel Grouped by Occupation." Occupations, 1945, Vol. 22, pp. 5-38.
- Tatsuoka, M. M. Multivariate Analysis in Educational and Psychological Research, N. Y.: Wiley, 1971.
- Tatsuoka, M. M. Discriminant Analysis. Champaign: Institute for Personality and Ability Testing, 1970.
- Wechsler, D. The Measurement and Appraisal of Adult Intelligence. Baltimore: Williams & Wilkins Co., 1958.
- Wittreich, W. J., & Miner, J. B. "People: The Most Mismanaged Asset: Rational Selection and Promotion." Business Horizons, 1971, pp. 69-77.