An Empirical Study of Test/Retest Reliability of the Kiersey Bates Temperament Sorter

Jerome Lavelle, Dennis Krumwiede, and Duane Brown
Department of Industrial Engineering, Department of Statistics
Kansas State University

ABSTRACT

This paper discusses the results of an empirical study to investigate the test/retest reliability characteristic of the Kiersey Bates Temperament Sorter (KBTS) personality type indicator. The study was conducted during the fall semester of 1995. Test subjects were undergraduate students in the business, engineering and sociology curricula at Kansas State University. Statistical measures used to provide an indication of reliability included: a percentage agreement comparison, test versus retest correlations, and a correlations comparison. The experimental results indicate that in general the KBTS proved very reliable in terms of test/retest as a personality type indicator. The results of this study are of potential importance to those interested in using the KBTS for personality typing in lieu of, or as a surrogate for, the more popular and widely tested Myers Briggs Type Indicator.

MOTIVATION FOR RESEARCH

The motivation for establishing the test/retest reliability characteristic of the Kiersey Bates Temperament Sorter (KBTS) involves research that the authors are undertaking on leadership personality and effectiveness in Total Quality Management (TQM) implementations. There are many applications of the use of personality indicators in the context of TQM. Companies interested in TQM (or Continuous Improvement (CI)) are very interested in the proper use of team centered skills. Increasingly, the Myers Briggs Type Indicator (MBTI) is being utilized to properly orient and understand people within this team-based environment. Also, as Walton[1] explains, W. Edwards Deming specified that management and company leadership ultimately establish success factors for long term sustenance of a CI philosophy. The authors' motivation was to ask: Is it legitimate to utilize the KBTS in lieu of the MBTI in conducting research involving TQM and personality? Two issues surface as one asks this question: (1) In the KBTS reliable? and (2) Is the KBTS valid? This paper reports on the first of these two questions.

KBTS AS A PERSONALITY MEASURER

The KBTS uses much of the same construct as the MBTI. With the KBTS and MBTI, personality types are derived from four preference scales. These dimensional scales are: Extroversion - Introversion (E-I), Sensation - Intuition (S-N) , Thinking - Feeling (T-F), and Judgment - Perception (J-P). The KBTS uses 70 questions (less than the MBTI) written to test preferences with respect to the four preference scales. There are sixteen unique "personality types" formed from the four personality preference scales ($2^4=16$). Example personality types would be: ENTJ,

ISFP, and ESFJ. Temperament can be inferred from a subject's personality type (the combination of the values of these four scales). Four unique temperaments are derived from the sixteen different personality types.

EXPERIMENTAL PROCEDURE
Task: Investigate the test/retest reliability of the KBTS.

Subjects: The subjects consisted of 209 volunteers from undergraduate business, engineering, and sociology classes at Kansas State University (KSU) in Manhattan, Kansas, USA. The subject pool consisted of 40.8% females and 59.1% males. All subjects signed an informed consent statement per KSU policies involving the use of human subjects in research. All data collection sessions involving the subjects were conducted on campus in a classroom setting.

Experimental Procedure: The following list delineates the process used to collect KBTS test/retest data from the subject pool. This data is analyzed for reliability.

> 1. Conducted Session #1 — Collection of "Test" Data
>   a. Orientation to experiment, signed consent form
>   b. Subjects provided answers to 70 questions on KBTS. Subjects were given 15 minutes to complete the 70 questions
>
> 2. A Re-test interval of 6 Weeks Elapsed
>
> 3. Conducted Session #2 — Collection of "Re-test" Data
>   a. Re-orientation to experiment
>   b. Subjects provided answers to same 70 questions in 1b. above within 15 minutes.

STATISTICAL RESULTS
A statistical analysis of the empirical data collected as part of this research was performed with the objective of investigating the test/retest reliability of the KBTS. Test/retest reliability is determined by comparing a subject's KBTS score on the "test" phase with the score from the "retest" phase. This test/retest relationship was compared to reliability data for the MBTI[2]. Specifically, the KSU data was compared with MBTI data taken from a student group at Mississippi State University (MSU). This data set was chosen for comparison because of its similarity to the KSU data, the MSU data used: a student pool, a similar test/re-test time interval, both male and female subjects, and a sample size greater than one-hundred. The following statistical tests provided insight into the test/retest reliability of the KBTS.

Given in Table 1 are the percentage agreement and percentage unchanged data for each of the KBTS typing categories for the KSU data set.

Table 1: Test/Retest Agreement and Preference Category Changes for KSU Data

Preference Category	E-I	S-N	T-F	J-P
Percent Agreement Between Test and Retest	78	82	75	84

Number of Preferences Unchanged	4	3	2	1
Percent of Preferences Unchanged From Test to Retest	41	43	13	3

To compare the KSU data set with the MBTI MSU data set a Chi-Squared test on the proportion (p) for each preference category was used. This test compares the observed cell and expected cell counts for each personality preference category at an alpha level of 0.05. Table 2 illustrates the tested data.

Table 2: Observed and Expected Cell Counts for Each Preference Category for both KSU and MSU data sets

Preference Category	E-I	S-N	T-F	J-P	TOTAL
KSU Observed	164	171	158	176	669
MSU Observed	143	158	147	149	597
Total Observed	307	329	305	325	1,266
KSU Expected	162.23	173.86	161.17	171.74	
MSU Expected	144.77	155.14	143.83	153.26	

The hypothesis tested is:

H_0 : $p_{11} = p_{21}$; $p_{12} = p_{22}$; $p_{13} = p_{23}$; $p_{14} = p_{24}$

H_1 : not H_0

For this test:

Test Statistic is k = sum [(expected-observed)^2/expected]

Critical Value = x^2 (0.05,3) = 7.81

Conclusion: There is insufficient evidence to show a significant difference between the KSU and MSU data sets in terms of percent agreement from test to retest.

Test 2: Test Versus Retest Correlations

Several correlation values were calculated on the test versus retest values where personality type for each category is taken as a continuous score. Both the KBTS and MBTI produce such continuous scores as part of quantifying preferences for each category. Table 3 gives the

test/retest correlation coefficients of these overall continuous scores for the Pearson, Kendall and Spearman correlation tests.

Table 3: Test/Retest Correlations on Continuous Scores for KSU Data

Preference Category	PEARSON	KENDALL	SPEARMAN
E-I	0.7883	0.6551	0.7862
S-N	0.7953	0.6373	0.7813
T-F	0.7807	0.5984	0.7413
J-P	0.8327	0.6763	0.8307

Test 3: Correlations Comparison

Pearson correlations from the KSU data were compared with the MSU MBTI correlations for the test and retest continuous scores. For this comparison Fisher's transformation was used to convert the correlations into standard normal deviates. To determine if there is a difference between KSU's Pearson correlations and the MSU data taking into account all of the categories, binomial probabilities were calculated assuming n = (# of categories being used) and p=0.05 (this is the probability of rejecting a true null hypothesis). Table 4 shows the Fisher Z-score for both the KSU and the difference data (between KSU and MSU). The binomial probability calculated between the KSU and MSU data was 0.1855.

Table 4: Pearson Correlations Comparison

Preference Category	E-I	S-N	T-F	J-P
KSU Pearson Correlations	0.7883	0.7953	0.787	0.8327
KSU Z-scores	1.0669	1.0857	1.0472	1.1969
KSU σ^2	0.0049	0.0049	0.0049	0.0049
KSU Sample Size	209	209	209	209
MSU Pearson Correlations	0.8200	0.8700	0.7800	0.8100
MSU Z-scores	1.1568	1.3331	1.0454	1.1270
MSU σ^2	0.0088	0.0088	0.0088	0.0088
MSU Sample Size	117	117	117	117
KSU/MSU Diff Z-score	-0.7689	-2.1158	0.0154	0.5978
KSU/MSU Diff p-value	0.4420	0.0344	0.9877	0.5400

The hypothesis tested is:

$H_0:$ p_{KSU} - $p_{MSU} = 0$

$H_1:$ not H_0

For this test:

Fisher's Transformation: $Z = (1/2)(\ln(1+r)-\ln(1-r))$ and $\sigma^2 = (1/(n-3))$

Critical value: 1.96

Conclusion: KSU correlations are not significantly different from correlations from the MSU data set in terms of continuous scores for personality preference scores.

CONCLUSIONS

The results of this study indicated that the KBTS had a comparable level of test/retest reliability to the MBTI. Data from Test 1 indicated that the KBTS had a high level of percent agreement both in terms of agreement at the personality-preference level and overall percentages unchanged. When compared to the MSU MBTI a difference could not be shown in the data from a percentage agreement perspective. Test 2 listed various correlation coefficients for the KSU data in terms of test versus retest continuous scores. These data indicated a strong linear relationship. Test 3 sought to compare the correlation values calculated in Test 2 with those from the comparison data set (MSU MBTI data). No statistical differences were found in the correlations for the two tested data sets.

The results from this empirical study are clear — in terms of test/retest reliability, the KBTS is reliable and comparable with published data on the reliability of the MBTI for a similar data set. If the results of this experiment can be generalized, one could believe that in terms of test/retest reliability the KBTS is a suitable substitute for the MBTI for quantifying subjects' personality preferences. To provide a more clear general picture on KBTS test/retest reliability it may be useful to study other comparable published MBTI data. It would also be of interest to address the question of the validity of the KBTS to provide a more complete commentary of the substitutability of the KBTS for the MBTI.

REFERENCES

[1] Walton, M. The Deming Management Method, New York: Putnam, 1986.
[2] Briggs, I. and McCaulley, M.H., Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator, Consulting Psychologist Press, Inc. 1985, pp. 172-173.

BIOGRAPHICAL INFORMATION

JEROME LAVELLE Ph.D. is an Assistant Professor in the Industrial Engineering department at Kansas State University. His PhD in IE is from North Carolina State University. Dr. Lavelle's teaching and research at KSU are in engineering economics and total quality management — he is the coordinator of engineering management at KSU. Currently he is an officer in both the engineering economy and engineering management divisions of ASEE. Dr. Lavelle is also a member of ASEM and IIE. He will be spending summer-1997 at the NASA Kennedy Center as part of the NASA/ASEE Summer Faculty Fellowship Program.

DENNIS KRUMWIEDE has 15 years of experience as an industrial engineer at companies such as AT&T, Rockwell International, Ball Aerospace and Vaisila OY. He is Ph.D. candidate in the Industrial and Manufacturing Systems Engineering Department at Kansas State University. Mr. Krumwiede teaches in the Management Department at KSU in the areas of POM and TQM. His research interests are with the international community of businesses in the areas of engineering management and operations management.

DUANE BROWN is a PhD candidate in the Statistics department at Kansas State University.