

An Executive Summary of the Development, Validity and Reliability of the English Version 3.0 of the Insights Discovery Evaluator

Dr. Stephen Benton

Director of The Business Psychology Centre (bpc),
The University of Westminster, 309 Regent Street, London, W1B 2UW, UK

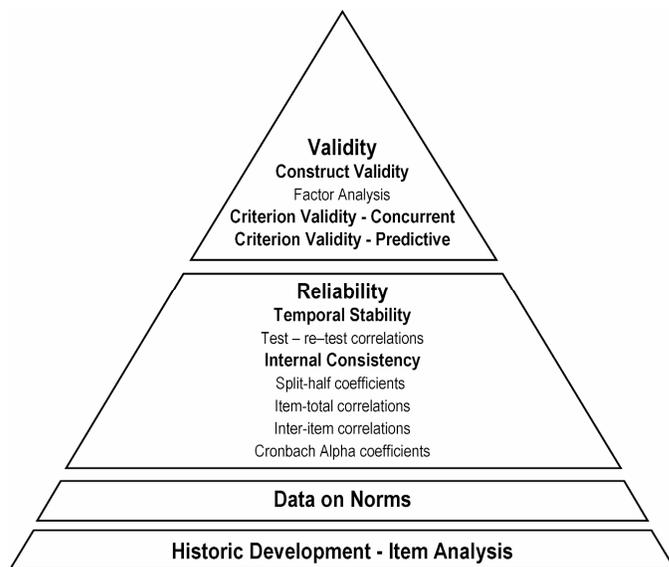
Dr. Corine van Erkom Schurink

Director of The Analytical Research Bureau (Pty) Ltd,
16 Central Avenue, Pinelands 7405, South Africa

Stewart Desson

Head of Learning, Insights Learning & Development Ltd,
Jack Martin Way, Dundee, DD4 9FF, Scotland, UK

This executive summary presents key data on the psychometric properties of the Insights Discovery Evaluator (IDE). It draws upon an extensive research and development programme undertaken between Insights Learning & Development Ltd. and the University of Westminster, aimed at the development of a psychometrically robust evaluator. For a more comprehensive review of the IDE’s properties, please read ‘An Overview of the Development, Validity and Reliability of the English Version 3.0 of the Insights Discovery Evaluator’ produced at the University of Westminster’s Business Psychology Centre (bpc).



This summary presents information covering ‘item analysis’, ‘norms data’, ‘reliability’ and ‘validity’.

Key statistics have been computed for each of these four areas and they have been benchmarked against international standards.

To convey these standards in easily understandable terms, the statistics are presented in a four segment pyramid.

Figure 1 – Pyramid of key psychometric statistics

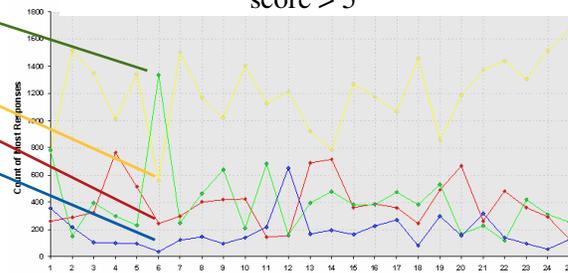
Historic Development – Item Analysis

There are 100 colour ‘items’ spread over the 25 frames in the IDE. Item analysis involves empirically testing the quality of these 100 items and replacing weaker items with better ones. One example of an item analysis on the 25 ‘Sunshine Yellow’ items is show in Figure 2. On the horizontal axis are the 25 frames. On the vertical axis is the number of respondents that highlighted ‘Sunshine Yellow’, ‘Fiery Red’, ‘Earth Green’ or ‘Cool Blue’ as the ‘most’ in the evaluator. However, this sample of respondents is comprised of those who, on average across all 25 frames, have scored 5 or more (out of 6) for ‘Sunshine Yellow’. The 6th frame of the evaluator is weak and was therefore subjected to systematic re-evaluation as new word pairs were empirically tested. The best results were found when the ‘Earth Green’ item was changed from ‘Accommodating and Helping’ to ‘Relating and Amenable’, combined with a change in the ‘Sunshine Yellow’ item from ‘Upbeat and Hopeful’ to ‘Expressive and Hopeful’.

Insights Discovery Preference Evaluator Name: _____

6. Accommodating and helping	L 1 2 3 4 5 M
Upbeat and hopeful	L 1 2 3 4 5 M
Powerful and assertive	L 1 2 3 4 5 M
Thinking and distant	L 1 2 3 4 5 M
7. Demonstrative and ready	L 1 2 3 4 5 M
Painstaking and discerning	L 1 2 3 4 5 M
Tough and initiating	L 1 2 3 4 5 M
Settled and reflective	L 1 2 3 4 5 M
8. Determined and resolute	L 1 2 3 4 5 M
Social and cheerful	L 1 2 3 4 5 M
Relating and amenable	L 1 2 3 4 5 M
Consistent and correct	L 1 2 3 4 5 M
9. Sensitive and diplomatic	L 1 2 3 4 5 M
Precise and deliberate	L 1 2 3 4 5 M
Encouraging and valuing	L 1 2 3 4 5 M
Results-oriented and fast	L 1 2 3 4 5 M

Count of ‘Most’ responses for people with Total Sunshine Yellow score > 5

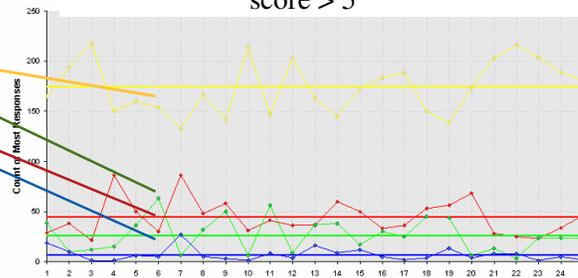


Question Numbers 1-25
Before Changes

Insights Discovery Preference Evaluator Name: _____

6. Relating and amenable	L 1 2 3 4 5 M
Expressive and hopeful	L 1 2 3 4 5 M
Powerful and assertive	L 1 2 3 4 5 M
Thinking and self-contained	L 1 2 3 4 5 M
7. Demonstrative and persuasive	L 1 2 3 4 5 M
Questioning and reflective	L 1 2 3 4 5 M
Immediate and initiating	L 1 2 3 4 5 M
Stable and concerned	L 1 2 3 4 5 M
8. Resolute and confident	L 1 2 3 4 5 M
Social and cheerful	L 1 2 3 4 5 M
Faithful and helping	L 1 2 3 4 5 M
Consistent and correct	L 1 2 3 4 5 M
9. Sensitive and diplomatic	L 1 2 3 4 5 M
Precise and deliberate	L 1 2 3 4 5 M
Encouraging and valuing	L 1 2 3 4 5 M
Results-oriented and fast	L 1 2 3 4 5 M

Count of ‘Most’ responses for people with Total Sunshine Yellow score > 5



Question Numbers 1-25
After Changes

Figure 2 – Example of results of item analysis - graphs show ‘before’ and ‘after’ item changes

Data on Norms

The norms data for the IDE is of good quality, being available segmented by; the language of the evaluator completed; the country a respondent is based in; age (in ten-year bands) and occupation (with over 100 different occupations analysed). Here is an example of norms data for CEOs and CFOs. While this data indicates that people in certain roles tend to have a preference for certain colours, it does not correlate or necessarily relate to how well they are doing they're job or how capable they are in fulfilling that role.

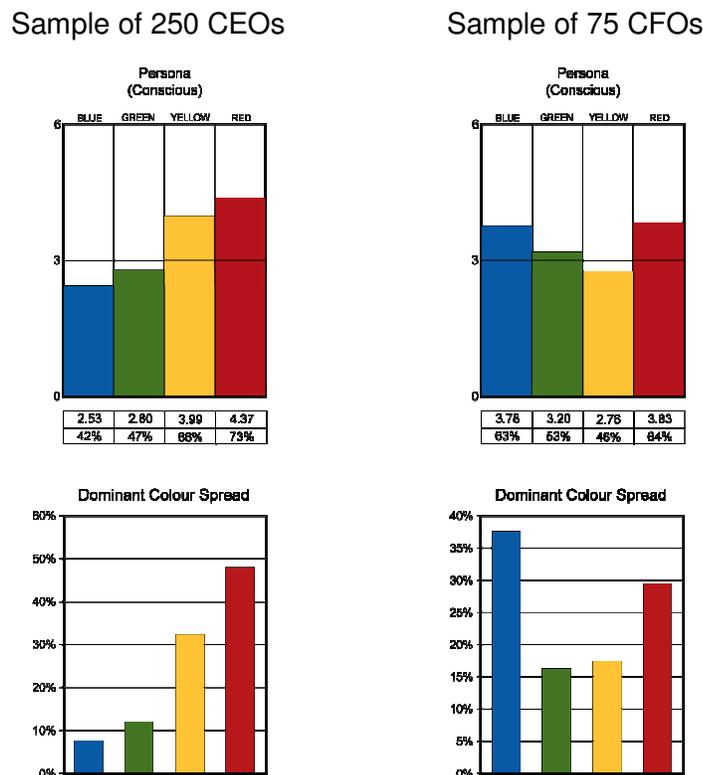


Figure 3 - A graphical view of CEOs vs. CFOs, in support of the argument for predictive validity

Reliability: 'Inter-item' and 'Item-total' Correlations

'Inter-item' correlation coefficients have been calculated using the Pearson Product-Moment Correlation. This involved creating four colour based '25 by 25' matrices showing the correlation between the 25 colour items. In 1991, Robinson et al.¹ concluded that the mean 'inter-item' correlation coefficient should equal or exceed 0.30. The analysis of the 24,224 evaluators (completed between 31/11/2003 and 31/7/2004 by people in the UK), shows that, for each of the four colours in the evaluator, the average 'inter-item' correlation coefficient is significantly above 0.3 providing strong evidence of the case for reliability.

¹ Robinson, J.P., Shaver, P.R., Wrightman, L. S (1991) 'Criteria for Scale Selection and Evaluation In Measure of Personality and Social Psychological Attitudes' Calif: Academic Press, San Diego

Table 1 – Inter-item correlations

N = 24,224	Colour preference			
	Cool Blue	Earth Green	Sunshine Yellow	Fiery Red
Mean	0.33	0.32	0.31	0.35
Minimum	0.10	0.12	0.06	0.13
Maximum	0.58	0.60	0.60	0.56

Reliability: Cronbach-Alpha Coefficients

This coefficient measures the error variance on the average inter-item correlation. When the error variance is low, which is desirable, the alpha coefficient approaches 1.0. A value of 0.70 is the commonly accepted inferior limit². Analysing the same 24,224 completed evaluators shows the four colours to have very high Cronbach-Alpha coefficients, providing further evidence of excellent reliability.

Table 2 – Cronbach-Alpha coefficients

N = 24,224	Colour preference			
	Cool Blue	Earth Green	Sunshine Yellow	Fiery Red
Cronbach-Alpha Coefficients	0.92	0.92	0.93	0.92

Reliability: Split-Half Coefficients

The final measure of internal consistency that supports the case for reliability is the ‘split-half’ measure. In split-half reliability we randomly divide all items that are thought to measure the same construct into two sets e.g. we create two sets of Fiery Red items. We test the evaluator on a sample of people and compute the total score for each randomly divided half. The split-half assessment of reliability is based on how well these two total scores correlate.

² DeVellis, R. F. (1991) ‘Scale Development: Theory and Applications’, Sage Publications, Newbury Park, CA
 Robinson, J.P., Shaver, P.R. (1973) ‘Measure of Psychological Attitudes’ MI: Survey Research Centre Institute for Social Research, University of Michigan
 Robinson, J.P., Shaver, P.R., Wrightsman, L. S (1991) ‘Criteria for Scale Selection and Evaluation In Measure of Personality and Social Psychological Attitudes’ California: Academic Press, San Diego
 Swailes, S., & McIntyre-Bhatty, T. (2002) ‘The “Belbin” team role inventory: reinterpreting reliability estimates’, Journal of Managerial Psychology, 17, 6, 529 – 536

The split-half measures for the IDE were achieved by splitting the 25 frames into two groups of 12 and 13. The colour results are computed for each of the two groups and then correlated. A high correlation suggests high reliability i.e. the higher the association (correlation coefficient) between the two data sub-sets, the higher the internal consistency of the scale. The analysis shows high coefficients for the IDE, with a key summary being:

- Cronbach-Alpha Coefficients above 0.8 for each half
- Pearson Correlation Coefficients above 0.7 i.e. the 2 halves correlate highly

Reliability: Temporal Stability – Test / Re-test Correlations

‘Temporal stability’ or ‘test/re-test’ reliability is determined through the administration of the same evaluator across time and it helps us gauge how robust the items are. A convenient sample of 1,435 people, who needed to complete the evaluator twice, had their original and re-tested colour scores assessed through a Pearson correlation analysis. Reliability is expressed as correlation coefficients, ranging from 1 to 0. Temporal stability tests are generally expected to yield reliability coefficients ranging between 0.70 and 0.90.

Table 3 – Test re-test correlation

N = 1,435	RETEST Cool Blue	RETEST Earth Green	RETEST Sunshine Yellow	RETEST Fiery Red
TEST Cool Blue	0.85	0.12	-0.72	-0.30
TEST Earth Green	0.14	0.81	-0.17	-0.66
TEST Sunshine Yellow	-0.74	-0.16	0.86	0.15
TEST Fiery Red	-0.29	-0.65	0.13	0.82

All correlations in this table are significant at the 0.01 level (2-tailed).

The results of the test/re-test analysis performed on the four colour scores show a very high reliability, translating into coefficients ranging from:

- 0.81 to 0.86 for the Pearsons correlation coefficients and
- 0.89 to 0.92 for the Cronbach-Alpha reliability coefficients for the same data.

Validity - Confirmatory Factor Analysis

Confirmatory Factor Analysis was used to test the hypothesized factor structure of the Insights Discovery model. Specifically, it is hypothesized that the four sets of 25 colour based items in the IDE, should load onto the factors such that:

- The polar opposite nature of the ‘Fiery Red’ vs. ‘Earth Green’ items is apparent
- The polar opposite nature of the ‘Sunshine Yellow’ vs. ‘Cool Blue’ items is apparent
- ‘Fiery Red’ items should not load significantly onto any factor that ‘Cool Blue’ and/or ‘Sunshine Yellow’ items load onto
- ‘Earth Green’ items should not load significantly onto a factor that ‘Cool Blue’ and/or ‘Sunshine Yellow’ items load onto
- ‘Sunshine Yellow’ items should not load significantly onto any factor that ‘Fiery Red’ and/or ‘Earth Green’ items load onto
- ‘Cool Blue’ items should not load significantly onto a factor that ‘Fiery Red’ and/or ‘Earth Green’ items load onto

The results that follow successfully confirm this hypothesized structure and offer evidence for the construct validity of the IDE. These two factors account for 34% of the variance. Generally factor loadings greater than 0.30 or below -0.30 are considered to meet the minimal level for significance required³. Using these criteria the statistically significant factor loadings have been highlighted in a larger bold font in the table.

Table 4 – English S3.0 IDE - factor loadings summary table

The ‘Fiery Red’ and ‘Earth Green’ items load strongly on Factor 1, with average factor loadings of *minus* 0.59 and *plus* 0.56

The ‘Cool Blue’ and ‘Sunshine Yellow’ items load strongly on Factor 2, with average factor loadings of *minus* 0.57 and *plus* 0.32

The opposite signs of these loadings support the theoretical construct that ‘Fiery Red’ and ‘Earth Green’ are polar opposites. The same holds true for ‘Cool Blue’ and ‘Sunshine Yellow’.

N=20,948 IDE English version S3.0 Completed in UK 31/11/03 to 2/7/04	Average Factor Loadings	
	Factor 1	Factor 2
Earth Green	0.56	0.06
Sunshine Yellow	-0.05	0.32
Cool Blue	-0.09	-0.57
Fiery Red	-0.59	0.19

This table is an average of the factor loadings. However, it is also possible to analyze the factor loadings for each of the 100 items in the IDE. Figure 4 is a scatter plot of the 100 items. It shows the top two factor’s loadings onto all 100 items. Both diagonal’s scales range from 0.8 to minus 0.8, with the axis crossing at zero.

³ Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., (1998) ‘Multivariate Data Analysis’, 5th ed, Prentice-Hall, Inc.

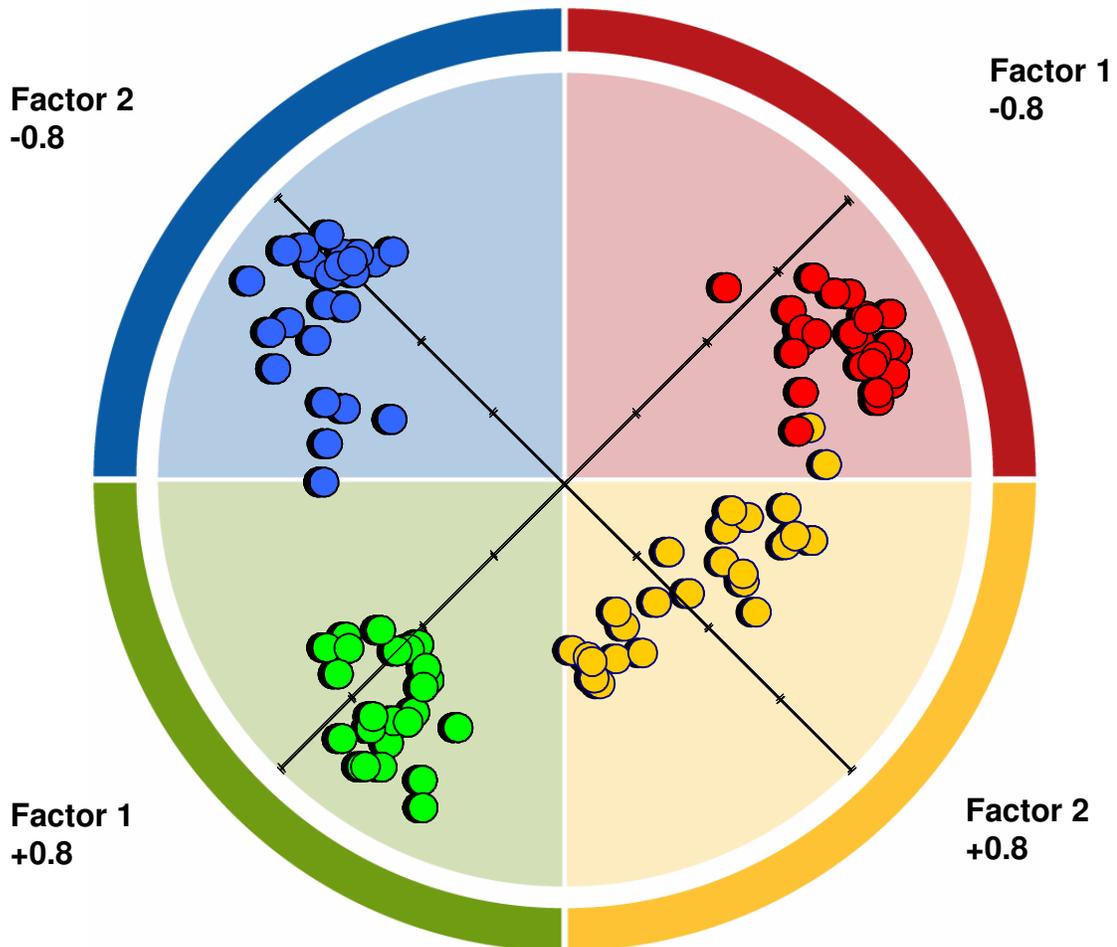


Figure 4 – English S3.0 IDE – Graph of the 100 items (25 x 4 colours) plotted against the factors

The graph in Figure 4 has been superimposed onto the Insights Discovery Wheel. It shows the relationship between each of the 100 items (four colours multiplied by 25 frames) and the top two factors. It can be seen that 97 out of 100 items appear in the ‘correct’ quadrant. Two yellow items appear just inside the red quadrant and one ‘Cool Blue’ item appears on the border of the ‘Cool Blue’/‘Earth Green’ quadrants.

This item level data provides further evidence of the bi-polar nature of the colour scores and the construct validity of the model.

Testimonials and professional feedback on the full paper: ‘An Overview of the Development, Validity and Reliability of the English Version 3.0 of the Insights Discovery Evaluator’

“This is a fascinating paper that provides the hard evidence required to support the use of the Insights Discovery Evaluator in both my counselling practice as well as for organisational development in the corporate world. I often find that organisations need and demand confirmation of validity before they will invest their time and money in using a psychometric. I am pleased to say that this paper provides the answers to their questions and, furthermore, gives a sense of the brilliance behind the instrument itself”.

Professor Peter Smyth, M.Sc.Ed., M.S.W., Ph.D., (C) OACCPP,
Lecturer at York University Toronto and Faculty at the Centre for Excellence in Critical Care
Medicine, Mount Sinai Hospital, University of Toronto, Canada.

“The integrity of the Insights Discovery Evaluator (IDE) rests solely upon its thorough and rigorous measures of validity/reliability and its congruence with the colour scores measuring personality preferences that embody Jung’s concept of opposites. Instruments similar to the IDE have come and gone as each fails to meet one or more of these vigorous measures. Only the best survive. This present landmark effort assures the entire professional community of researchers, developers, practitioners and consumers alike, that their belief in the integrity in the IDE is well placed and positively confirmed. If one was ever skittish about pressing questions of the reliability and validity of the IDE, one can now move onto other concerns”.

Professor B. Bradley West, B.S., A.M., Ph.D.
Research Specializations in MBTI, Psychometrics and Projective Techniques
Michigan State University, East Lansing, Michigan, 48864, U.S.A.

“Having used the Insights Discovery Evaluator extensively in workshops, research and client work at the University of Dundee; I read the recent publication by the University of Westminster on the psychometric properties of the Insights model with great interest. In presenting a thoughtful, in-depth and clear analysis, the authors have provided a persuasive argument for pursuing typological research with confidence. They vindicate my long-held opinion that the instrument has solid statistical merit, and an underlying robustness that compares very favourably with similar systems.”

Nick R. Halpin, PhD., C. Psychol, Dip. Couns., MBACP (Accred)
Head of Counselling, University of Dundee, UK.

“This is a robust body of work that documents the development of a novel psychometric tool and reports the theoretical principles of reliability and validity with factual accuracy”

Professor Tony Towell Ph.D.
Senior Research Fellow, Research Methods, University of Westminster
Department of Clinical Neurophysiology, Hospital for Sick Children, Great Ormond Street,
London, UK.