Assessment Validation Study

An Analysis of the Psychometric Characteristics of the NEW Personality Insights Style Assessment

Study Conducted By Psychometric Statistician
Thomas G. Snider-Lotz, Ph. D.

Personality INSIGHTS
Empowering People to Improve

EXECUTIVE SUMMARY

Research was undertaken to determine the psychometric characteristics of the Personality Insights Style Analysis (NEW), and to compare the characteristics of the PISA with an established instrument from Target Training International. The two instruments were administered to 500 individuals, and were found to have very similar characteristics in all respects. The reliability and validity of the NEW were strong, and the relationships among its scales were similar to those observed for the OLD instrument.

The Personality Insights Style Analysis (NEW) is an instrument that measures an individual's interactional style. It consists of 96 descriptive terms, divided into 24 groups with four terms in each group. Within each group, there is one term for each of the four primary traits: D, I, S, and C. The terms used on the PISA and the traits measured by them are given in Appendices A and B. Table 1 below describes the traits measured by the four primary scales.

Scale	Primary Characteristics	Associated Adjectives
D	Outgoing, task-oriented	Dominant, determined, a doer
I	Outgoing, people-oriented	Inspirational, influential, interesting
S	Reserved, people-oriented	Steady, stable, supportive
C	Reserved, task-oriented	Cautious, conscientious, correct

Table 1. Characteristics measured by the D, I, S, and C scales.

When an individual fills out the Personality Insights Style Analysis, he or she chooses one term from each group that is most descriptive of them, and one term that is least descriptive of them. When the individual is finished, he or she will have chosen 24 terms that are most descriptive of them, and 24 terms that are least descriptive. The numbers of D, I, S, and C terms chosen as "mosts" comprise a profile that reflects the individual's "Environment" style, and the numbers chosen as "leasts" reflect the individual's "Basic" style.

The original PISA, like the OLD instrument, also included a scale labeled the N scale. Terms from the N scale were not scored. The origins of the N scale were not clear, but it would appear that the terms on the N scale may have been written for one of the four basic scales, but were later excluded from those scales because of poor statistical characteristics. Use of the N scale created the situation that some terms might be scored on a primary scale (D, I, S, or C) for the Environment style, but not scored for the Basic style (or vice versa). The numbers of terms on the different scales are given in the table below.

Scale	Environment	Basic Style
D	20	21
I	17	19
S	19	18
C	15	12
N	25	26

Table 2. Number of terms on each scale of the original NEW and the OLD instrument.

Current Project

The research described in this report was undertaken on behalf of Personality Insights, Inc. Its goals was to determine the psychometric characteristics of a new version of the Personality Insights Style Analysis, and to compare its characteristics to an older instrument, that measures the same traits.

The data collection involved administering both the new Personality Insights Style Analysis (NEW) and the (OLD) instruments to a representative sample of 500 individuals at workshops presented by Personality Insights, Inc. Of these 500 individuals, nine were omitted from the analysis because of bad or missing data. The responses of the remaining 491 individuals were used to conduct the analysis described below.

The data were analyzed using SPSS/PC. After invalid or missing data were corrected or removed, frequency and correlational analyses were done to provide descriptive statistics regarding the scales and their interrelationships.

Until this point in the investigation, the NEW had been scored to include an N scale, following the example of the OLD instrument. After the basic statistics had been computed, an effort was undertaken to determine whether the reliability of the NEW scales could be improved by incorporating some or all of the N terms into the four basic scales. Investigation of correlations between responses to the N terms and scores on the D, I, S, and C scales suggested that each of the N terms could be added to one of the four basic scales. The result of incorporating the N terms into the D, I, S, and C scales had several benefits: it eliminated the nonfunctional N scale; it made all of the four basic scales the same length (24 terms); it assigned each term to the same scale for both Environment and Basic Style scoring; and most importantly, it increased the reliability of the D, I, S, and C scales. Unless otherwise indicated, all of the statistics given in this report refer to the updated version of the NEW, from which the N scale has been eliminated.

RELIABILITY AND VALIDITY INFORMATION

Reliability

The reliability of each scale was determined using split-halves (odd-even) reliability. Reliability is measured on a scale of 0.00 to 1.00, with 1.00 representing perfect reliability. The results are

	NEW Environment	NEW Basic Style	OLD Environment	OLD Basic Style
D	0.88	0.88	0.86	0.89
I	0.87	0.84	0.78	0.79
S	0.86	0.84	0.71	0.81
C	0.85	0.84	0.63	0.72

Table 3. Split-halves reliabilities of NEW and OLD scales. N = 491.

given in Table 3. These reliability values for the NEW scales are superior to that of the OLD scales in every case except one, and all values are good values for scales of this length.

Standard errors of measure for the scales are given in Table 4, below. Standard errors of measure (SEMs) are related to the reliability of the scales, in that both are indicators of the precision with which the scales measure. Reliability is measured on an abstract scale ranging from 0.0 to 1.0, while SEMs indicate the number of scale points that should be considered the "margin of error" for each score. For example, an individual's score on the D scale of the NEW, when scored for the Basic Style, should be viewed as accurate within about 2.0 points. All the SEMs are in the range

	NEW Environment	NEW Basic Style	OLD Environment	OLD Basic Style
D	1.7	2.0	1.6	1.8
I	1.8	1.8	1.5	1.6
S	1.9	1.6	2.0	1.4
C	1.9	1.7	1.6	1.6

Table 4. Standard errors of measure of NEW and OLD scales. N = 491.

of 1.4 to 2.0 points, so for practical purposes the margin of error for any score is about 1 ½ to 2 points.

Validity

The appropriate validity evidence for an instrument such as the NEW is construct validity. It has construct validity because all terms were chosen with careful regard for the underlying dimension the four scales were intended to measure. Mathematical evidence for construct validity is provided by correlation of NEW scores with scores on the corresponding scales of the OLD instrument. The table below gives the values of these correlations. All correlations are statistically significant ($p < .001$). All values are positive and moderately high to high, supporting the premise that corresponding scales on the two instruments measure the same constructs.

	Environment	Basic Style
D	0.89	0.90
I	0.84	0.85
S	0.80	0.80
C	0.72	0.74

Table 5. Correlation of corresponding scales on NEW and OLD instrument. All values are significant ($p < .001$).

An additional indicator of construct validity is whether the correlations of scales within instruments follow similar patterns. Table 6 below compares the intercorrelations within the NEW and the OLD instrument. The patterns are similar: for example, the correlation between scores on the D scale and I scale on the Environment choices on the NEW is +.02; the corresponding correlation on the OLD instrument is -.01. Both correlations are essentially equal to zero, indicating that on both the NEW and the OLD instrument an individual's score on the D scale would be of no value in predicting his or her score on the I scale. On the other hand, the D scale and the S scale have a moderate to high negative correlation that holds up for both the NEW and OLD, and for both the Environment scores and the Basic Style scores. This correlation suggests that an individual who scores high on the D scale is likely to score low on the S scale (and vice versa) regardless of which instrument is used, or whether the Environment or Basic Style scores are used.

	Environment		Basic Style	
Scales	NEW	OLD	NEW	OLD
D & I	+.02	-.01	-.21	-.18
D & S	-.64	-.73	-.75	-.82
D & C	-.32	-.42	-.45	-.69
I & S	-.35	-.43	-.18	-.20
I & C	-.65	-.52	-.61	-.40
S & C	-.05	+.29	+.29	+.63

Table 6. Correlations of scales within instruments. Shaded values are statistically significant ($p < .001$).

Throughout the remainder of the table, the corresponding correlations for NEW and the OLD instrument are relatively close in value. This similarity in the pattern of values suggests that the relationships among the scales is the same for both the NEW and the OLD instrument, and is further evidence of construct validity.

DESCRIPTIVE STATISTICS

Scale Means and Standard Deviations

Tables 7 below provides a comparison of the means, medians, and standard deviations of the NEW and OLD scales. The values shown below are for the original NEW scales, before the N-scale items were integrated into the D, I, S, and C scales. The original NEW scales were used for

Environment ("Mosts")

	D		I		S		C	
	NEW	OLD	NEW	OLD	NEW	OLD	NEW	OLD
Number Of Terms	20	20	17	17	19	19	15	15
Mean	3.86	4.33	3.71	3.99	6.13	6.37	4.47	4.53
Standard Deviation	4.13	4.33	3.65	3.23	4.19	3.65	3.14	2.65
Median	3.00	3.00	2.00	3.00	5.00	6.00	4.00	4.00

Correlation Between NEW & OLD	+ .88	+ .84	+ .78	+ .70

Basic Style ("Leasts")

	D		I		S		C	
	NEW	OLD	NEW	OLD	NEW	OLD	NEW	OLD
Number Of Terms	21	21	19	19	18	18	12	12
Mean	8.34	8.52	4.75	3.88	3.75	3.65	3.27	4.55
Standard Deviation	5.23	5.45	4.11	3.41	3.59	3.24	3.29	3.07
Median	8.00	8.00	4.00	3.00	3.00	3.00	2.00	4.00
Correlation Between NEW & OLD	+ .90		+ .86		+ .80		+ .74	

Table 7. Comparison of statistics for Original NEW or OLD instrument. N = 491; shading indicates pairs of means that were statistically different.

this table because they contained the same number of terms as the NEW cales, and thus provided more appropriate comparisons. Means, medians, and standard deviations for the updated version of the NEW, with the former N-scale terms integrated into the D, I, S, and C scales, are given in Table 8. The values for these statistics will tend to be larger than the corresponding values in Table 7, primarily because the revised scales contain more terms than the scales in the original instruments.

	Number of Terms	Environment			Basic Style		
		Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
D	24	4.6	4.8	3.0	9.3	5.9	9.0
I	24	5.5	4.9	4.0	5.7	4.6	5.0
S	24	6.9	5.1	6.0	4.1	4.0	3.0
C	24	7.0	5.0	6.0	4.8	4.3	4.0

Table 8. Means, medians, and standard deviations for updated NEW scales. N = 491.

Sex Differences in NEW Scores

Males and females had different mean scores on several scales on the NEW. Table 9 shows

	Environment		Basic Style	
	Males	Females	Males	Females
D	5.3	3.9	8.8	10.0
I	5.0	6.1	6.2	5.2
S	6.2	7.6	4.8	3.4
C	7.5	6.4	4.2	5.5

Table 9. Mean scores of males and females on PISA (N = 255 males and 235 females). Shading indicates that scores of males and females were significantly different ($p < .01$).

the results of a comparison of scores by males and females. A series of t-tests found the differences in means to be statistically significant in four of the eight comparisons.

If the NEW were a clinical instrument, it would be appropriate to use different norms for males and females. However, as the NEW is used, the focus is on how an individual's style compares to people in general, not just to males or to females. Therefore, it is more appropriate to develop norms based on both males and females.

Norms

Appendices C and D give the norms for the NEW scales.

APPENDIX C
NORMS FOR PERSONALITY INSIGHTS STYLE ANALYSIS SCALES
ENVIRONMENT ("MOSTS") SCORES

D Scale

Score	Percentage[a]	Cumulative Percentage[b]
0	18.4	18.4
1	15.5	33.9
2	9.4	43.3
3	9.2	52.4
4	8.2	60.6
5	8.6	69.2
6	5.1	74.3
7	3.3	77.6
8	4.9	82.4
9	3.1	85.5
10	2.7	88.2
11	1.0	89.2
12	1.8	91.0
13	1.6	92.7
14	1.2	93.9
15	1.2	95.1
16	1.4	96.5
17	1.2	97.8
18	0.4	98.2
19	1.0	99.2
20	0.4	99.6
21	0.0	99.6
22	0.2	99.8
23	0.2	100.0
24	0.0	100.0

I Scale

Score	Percentage	Cumulative Percentage
0	9.6	9.6
1	12.7	22.3
2	13.3	35.6
3	12.0	47.6
4	8.2	55.8
5	6.7	62.5
6	5.1	67.6
7	3.9	71.5
8	3.9	75.4
9	3.5	78.9
10	3.9	82.8
11	2.9	85.7
12	2.2	87.9
13	1.6	89.5
14	2.9	92.4
15	1.6	94.0
16	2.2	96.2
17	1.2	97.4
18	2.0	99.4
19	0.2	99.6
20	0.4	100.0
21	0.0	100.0
22	0.0	100.0
23	0.0	100.0
24	0.0	100.0

[a] Percentage of respondents who achieved this score.
[b] Percentage of respondents who achieved this score or less.

APPENDIX C (Continued)

S Scale			C Scale		
Score	Percentage	Cumulative Percentage	Score	Percentage	Cumulative Percentage
0	7.3	7.3	0	4.7	4.7
1	7.3	14.6	1	6.9	11.6
2	8.8	23.4	2	11.2	22.8
3	7.8	31.2	3	7.3	30.2
4	7.1	38.3	4	8.4	38.6
5	7.6	45.9	5	8.8	47.3
6	7.1	53.0	6	6.3	53.7
7	5.7	58.7	7	5.9	59.6
8	7.6	66.3	8	5.7	65.3
9	5.3	71.6	9	5.9	71.2
10	4.1	75.7	10	5.1	76.3
11	3.9	79.6	11	4.3	80.6
12	3.5	83.1	12	3.5	84.1
13	3.7	86.8	13	1.8	85.9
14	4.1	90.9	14	3.5	89.4
15	2.9	93.8	15	3.1	92.5
16	2.0	95.8	16	2.4	94.9
17	1.4	97.2	17	2.0	96.9
18	1.2	98.4	18	1.6	98.6
19	0.6	99.0	19	0.8	99.4
20	0.2	99.2	20	0.0	99.4
21	0.2	99.4	21	0.2	99.6
22	0.6	100.0	22	0.4	100.0
23	0.0	100.0	23	0.0	100.0
24	0.0	100.0	24	0.0	100.0

APPENDIX D
NORMS FOR PERSONALITY INSIGHTS STYLE ANALYSIS SCALES
BASIC STYLE ("LEASTS") SCORES

D Scale

Score	Percentage[a]	Cumulative Percentage[b]
0	5.1	5.1
1	4.9	10.0
2	6.3	16.3
3	5.1	21.4
4	4.1	25.5
5	4.5	30.0
6	6.5	36.5
7	5.1	41.6
8	5.1	46.7
9	4.5	51.2
10	5.5	56.7
11	5.1	61.8
12	6.1	67.9
13	5.7	73.6
14	3.5	77.1
15	4.7	81.8
16	4.9	86.7
17	2.7	89.4
18	3.5	92.9
19	2.7	95.6
20	1.2	96.8
21	1.8	98.6
22	0.8	99.4
23	0.4	99.8
24	0.2	100.0

I Scale

Score	Percentage	Cumulative Percentage
0	9.8	9.8
1	11.0	20.8
2	10.8	31.6
3	8.8	40.4
4	6.5	46.9
5	9.0	55.9
6	8.4	64.3
7	5.1	69.4
8	3.9	73.3
9	4.9	78.2
10	4.5	82.7
11	4.1	86.7
12	4.3	91.0
13	2.0	93.1
14	2.4	95.5
15	0.4	95.9
16	1.6	97.6
17	1.4	99.0
18	0.6	99.6
19	0.2	99.8
20	0.2	100.0
21	0.0	100.0
22	0.0	100.0
23	0.0	100.0
24	0.0	100.0

[a] Percentage of respondents who achieved this score.
[b] Percentage of respondents who achieved this score or less.

S Scale

Score	Percentage	Cumulative Percentage
0	20.2	20.2
1	13.5	33.7
2	11.4	45.1
3	7.6	52.7
4	9.0	61.6
5	7.1	68.8
6	6.5	75.3
7	6.3	81.6
8	4.5	86.1
9	3.5	89.6
10	2.4	92.0
11	2.0	94.1
12	1.6	95.7
13	0.8	96.5
14	0.6	97.1
15	0.4	97.6
16	1.2	98.8
17	0.2	99.0
18	1.0	100.0
19	0.0	100.0
20	0.0	100.0
21	0.0	100.0
22	0.0	100.0
23	0.0	100.0
24	0.0	100.0

C Scale

Score	Percentage	Cumulative Percentage
0	14.5	14.5
1	12.9	27.3
2	12.2	39.6
3	9.0	48.6
4	8.0	56.5
5	9.4	65.9
6	3.3	69.2
7	4.3	73.5
8	6.1	79.6
9	4.5	84.1
10	4.3	88.4
11	2.9	91.2
12	2.7	93.9
13	1.2	95.1
14	1.4	96.5
15	1.2	97.8
16	1.2	99.0
17	0.8	99.8
18	0.2	100.0
19	0.0	100.0
20	0.0	100.0
21	0.0	100.0
22	0.0	100.0
23	0.0	100.0
24	0.0	100.0

APPENDIX E
GLOSSARY OF PSYCHOMETRIC TERMS

Correlation

- Two measures are said to be correlated if they are related enough that knowing someone's score on one measure allows us to make a good guess about their score on the other measure. The two measures may be very different things (for example, height and weight) or very similar things (for example, scores on the Dominance scale on two different tests).

- Correlations may be positive (for example, the correlation between weight and height) or negative (for example, the correlation between scores on a test of introversion and scores on a test of extraversion).

- A high correlation (either positive or negative) between two measures means that there is a close relationship between scores on the two measures. It does not mean that scores are necessarily the same on the two measures. For example, if people's scores on Test A were always 5 points higher than their scores on Test B, the two sets of test scores would have a "perfect" correlation even though nobody would have equal scores on both tests.

- A low or zero correlation means the two measures are not related, so that knowing someone's score on one of the measures is no help in guessing their score on the other.

- There are numerous ways to calculate correlation coefficients. The most common method is called the Pearson Product Moment Correlation Coefficient, and the number that results from it is represented by an *r*.

- Correlations are measured on a scale from –1.0 to +1.0. A correlation of zero means there is no relationship, $r = +1.0$ indicates a perfect positive relationship, and $r = -1.0$ indicates a perfect negative relationship.

- There are no fixed standards for how large a correlation must be to be considered "high," or even "high enough." There are situations were any correlation higher than 0.20 is considered "high enough." There are other situations where a correlation must be in the .80's or .90's to be "high enough."

- It is possible to calculate the statistical significance of a correlation coefficient. However, there is a danger here, because statistical significance does not guarantee practical significance. It is the nature of statistics that, as the number of people in the sample increases, the minimum value of *r* required for significance drops. In other words, with a large enough sample of people, almost any correlation will be "statistically significant." Responsibility for deciding if such a correlation has practical value belongs to the user of the statistic.

Reliability

- An instrument is reliable to the extent that it gives consistent results. In theory, if a person took a social style instrument several times, he or she should receive the same score each time, unless that person's social style had actually changed in the interim. In practice, it is unlikely that the person would score exactly the same each time, but the scores should be nearly the same. Reliability is a measure of the extent to which scores on an instrument remain constant over repeated administrations.

- On each instrument, each individual has a "true score." The true score is the score that person should get, the score that most accurately describes the person. If a person takes a reliable instrument several times, all of his or her scores should be clustered closely around his or her true score.

- If the instrument was not reliable, his or her scores would be more widely scattered, although the average of all the scores should be close to the true score. Some of the scores might be close to the true score, while others will not be close.

- Since usually a person will take an instrument only once, we need to know how close their actual score is likely to be to their true score. If the instrument is reliable, we can be confident that their actual score is close to their true score. If the instrument is not reliable, we can't have this confidence. Their actual score may be close to their true score, or it might not.

- Another reason reliability is important is that reliability can be an indicator of instrument quality. Many of the causes for low reliability spring from some "problem" with the instrument that causes people to respond inconsistently. For example, if the instrument is ambiguous or confusing, or if it is not relevant to the person's experience, he or she will give inconsistent answers. Low reliability may be an indicator that an instrument needs to be revised.

- Reliability is measured on a scale of 0 to 1.00, where 0 indicates a complete lack of reliability and 1.00 indicates perfect reliability.

- There are no universal standards for what constitutes "good" or "good enough" reliability, but in specific situations it is possible to set criteria for reliability. For instruments like the NEW or the OLD instrument, reliability should be in the upper .70's or above.

Standard Error of Measure

- Suppose a person takes Instrument X three times, and receives scores of 49, 51, and 50. She also takes Instrument Y three times, and receives scores of 44, 60, and 46. Her average score on each instrument was 50, but it is clear that Instrument Y is less consistent than Instrument X.

- One way to describe the consistency of the two instruments is by calculating their reliability, as discussed above. However, one disadvantage of reliability is that it is expressed on an arbitrary scale of 0 to 1, not in terms of the scale normally used to score the instrument.

- It is often more useful to use the standard error of measure (SEM), which measures the consistency of an instrument in terms of the instrument's own point scale. The SEM tells us how close a person's actual score is likely to be to their true score.

- For practical purposes, we can regard the SEM as a "margin of error" to apply to an instrument score. That is, we can say "Her score on Instrument Z was 48, plus or minus *z* points," where *z* is the standard error of measure.

- Unlike reliability, for standard error of measure *smaller is better*, because a smaller standard error of measure means a smaller margin of error.

- Note that one disadvantage of the standard error of measure is that it cannot be used to compare two instruments that have different scoring scales. For example, we could not compare the consistency of the SAT college entrance exam (scored on a scale of 200 to 800) with that of the ACT college entrance exam (scored 0 to 36) using the SEM, because the two instruments generate scores of different magnitudes. We could compare reliabilities, however, since reliability will describe both instruments in terms of the same 0 to 1 scale.

Validity

- There are many ways to define validity, but the most critical aspect is that validity is concerned with whether an instrument is appropriate for the use to which it is put. Validity is not necessarily related to instrument quality, because a high-quality instrument can be invalid if it is used for the wrong purpose. For example, a widely respected intelligence test designed for small children would not be valid for testing the intelligence of normal adults.

- There are several types of evidence for validity. The type of evidence that is used in a particular case depends on how the instrument will be used. For example, an instrument that is used to predict something (for example, an employment test used to predict performance on the job) should be validated by testing its ability to make accurate predictions, preferably in the exact situation where it would be used. On the other hand, an instrument that is used to test a person's knowledge of certain information (for example, an educational test) should be validated by showing that the instrument addresses the information in question.

- The Personality Insights Style Assessment is not used to test knowledge or predict future performance. Rather, it describes and classifies people in terms of an existing set of concepts or constructs -- the D, I, S, and C types. In this case, the instrument should be validated by presenting evidence that it does actually assess those styles. This approach to validation is referred to as demonstrating construct validity.

- There are two primary ways to demonstrate that the NEW successfully measures the D, I, S, and C styles: (1) show that the descriptors used on the NEW are relevant to the styles they are intended to assess; and (2) show that scores produced by the NEW are in agreement with scores produced by an existing instrument that is widely accepted.

- The first requirement, relevant descriptors, is met because the descriptors were chosen by experts with many years' experience with the D, I, S, and C styles. The second requirement, agreement with an existing instrument, is met by demonstrating a high correlation between NEW scores and scores on the OLD instrument.

PERSONALITY INSIGHTS STYLE INVENTORY
FACT SHEET

THE INVENTORY

96 descriptive adjectives or phrases, arranged in 24 groups of 4 descriptors

DESCRIPTIVE STATISTICS

Note: All statistics are based on a representative sample of 500 people.

	Number of Terms	Environment			Basic Style		
		Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
D	24	4.6	4.8	3.0	9.3	5.9	9.0
I	24	5.5	4.9	4.0	5.7	4.6	5.0
S	24	6.9	5.1	6.0	4.1	4.0	3.0
C	24	7.0	5.0	6.0	4.8	4.3	4.0

PSYCHOMETRIC CHARACTERISTICS

Reliability

The table below shows the split-halves reliabilities of individual scales (1.00 is the maximum possible value).

	Environment	Basic Style
D	0.88	0.88
I	0.87	0.84
S	0.86	0.84
C	0.85	0.84

Validity

Construct validity is measured as correlation with scores on a widely accepted measure of social style (1.00 is the maximum possible value). All values are statistically significant ($p < .001$).

	Environment	Basic Style
D	0.89	0.90
I	0.84	0.85
S	0.80	0.80
C	0.72	0.74

Standard Errors of Measure

	Environment	Basic Style
D	1.7	2.0
I	1.8	1.8
S	1.9	1.6
C	1.9	1.7