MatchAI Technical Report

Scale Validation and Prediction Accuracy

WHITEPAPER

Psychology is trendy in market research these days.

There are a number of AI-powered psychology-based platforms on the market. The question is, which one of those platforms are substantive vs just for show? Or, which one will provide you with insights that give you a real business advantage? In the following paper, we will review the psychometrics behind the MatchAI, highlighting its validity and reliability. Our hope is that by the end of this paper, you will be able to see:

How much scientific rigor was put into the construction of the platform MatchAI

How trustworthy the results are

How to compare the rigor of our platform with other platforms of the market

MATCHAI

BY BRANDTRUTH

In the past few years, people's approach to employment has dramatically changed. For example, the Great Resignation era has seen employees leave work because they don't believe in the company. Employees have become highly selective about who they choose to work with and more mindful about how compatible they are with their employer. Brand Truth helps consumers save time, money, and frustration associated with the job search process by calculating their compatibility with employers based on key psychographic indicators to bring about better employer-employee alignment.

MatchAI is a new platform that uses patent-pending systems for AI-powered projective tests. This platform helps brands identify the C-Score or the rating of employer-employee alignment. In other words, the extent to which an individual's personality profile is compatible with a potential employer.

How do we do it? MatchAI uses a unique combination of psychological science, data science, and machine learning algorithms to produce intelligent AI-powered projective test technology built on the following underlying conceptual principles:

- Employees are selective about who they choose to work with
- The job search process can be time consuming, costly, and often frustrating
- Identifying an employee's personality profile and an employer's personality profile can help facilitate the job search process
- MatchAI survey individuals and employers on their personality profiles and computes a C-Score or compatibility score for specific potential employers

What can you learn? By using the MatchAI platform, a brand can:

1. Take Workplace Characteristics Scale to get personality profiles
2. Deep dive into personality profiles
3. Get insights about their C-Score
4. Access simulations showcasing predicted C-Score based on external factors

Why use our platform? While there are a number of ways the MatchAI platform is beneficial, there are three primary benefits:

Automated Predictive AI. The MatchAI platform automatically creates predictive algorithms unique to each individual and employer. The predictive AI offers individuals and employers their personality profile scores and can simulate a predicted C-Score considering factors such as paid-time-off or salary. The predictive AI grows better each time it is used.

Key Benefits of the MatchAI Platform

No More Guesswork. With the MatchAI platform, individuals can see their C-Score with various employers, indicating the extent to which their personality profile aligns with the employer. This can help employees and employers alike save time, money, and frustration associated with the job search process and take some of the guesswork about how well an employee will fit in out of the process.

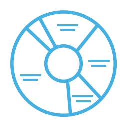
Both Quantitative and Qualitative. MatchAI studies both quantifiable data and emotional insight using a visual library to uncover and enhance our understanding of personality profiles of individuals. This enables individuals and employers to uncover their deep-seated thoughts and feelings above and beyond a typical survey or interview. Concurrently, brands have access to quantitative data with a tangible C-Score through the platform. The combination of quantitative and qualitative insights offers a 360 visualization of each individual.

How Does the Workplace Characteristics Scale Work?

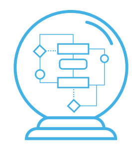
In order to compute a C-Score, we first have to assess the personality profiles for the employee and the employer. This paper is focusing specifically on the employee personality measure, referred to as the Workplace Characteristics scale. Obtaining the personality scores and eventually the C-Score involves a four step process:

The Testing Step
Taking the Workplace Characteristics scale

The Scoring Step
Scoring reponses to the Truth Detector test



The Profiling Step
Identifying which profile is predominant for the individual



The Predicting Step
Predicting C-Score for different employers

Each one of these steps has a scientific process built into them. For the **testing step** (i.e., when the participant takes the Workplace Characteristics scale), we want to make sure the data is good quality. So we use an algorithm that measures the extent to which a respondent is intentionally trying to deceive the test, not take it seriously, or enter in bad quality data. For the **scoring step** we use measures of inter-rater reliability. For the **profiling step**, we use classic psychometric measures of validity and reliability to know the traits we're measuring are trustworthy. For the **predicting step** we use the model's error (the difference from the predicted score and actual score) to know how accurate/precise the model's predictions are. Over the course of the rest of this paper, we'll go in depth on each of these aspects so that you can see just how science-based this tool is.

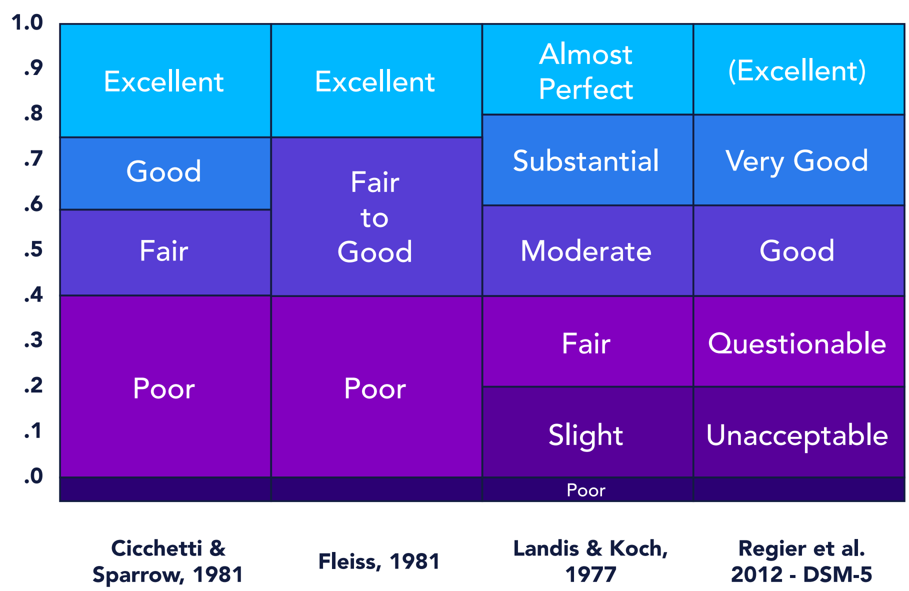
The Scoring Step: Measuring Inter-rater Reliability

Due to the high velocity of data we sometimes receive, we have multiple coders who apply a specific scoring scheme to the secret sentiments portion. However, as you may suspect, everyone has a slightly different way of interpreting ambiguous data. As a result, all coders are put through a training program for how to score the secret sentiments portion. Once the coders have sufficiently passed a scoring test, they are allowed to work on scoring project data. For any given project, we have 2 coders score the responses separately. No coder is able to see how any other coder has scored the responses, keeping all parties independent of possible scoring influences. However, to continually check that all coders are scoring the responses similarly, we calculate inter-rater reliability on all projects, and overall, on an ongoing basis.

Inter-rater reliability (IRR) is a statistic that measures the consistency of our coding methods. Basically, it's a check to see if our trained coders are applying the same codes to the same responses.

Historically, there are a few different approaches as to what is considered a "good" versus "bad" reliability score. You can see these approaches, and their references, in the accompanying chart. At Inkblot Analytics, we traditionally follow the inter-rater reliability approach outlined by Regier et al (2012), shooting for .80 reliability or above. This means that we always expect our coders to agree on a minimum of 80% of the scoring they do.

This section is an add-on service.



The Profiling Step: Psychometrics of Personality Profiles

Once the test data is collected, we are able to use our proprietary algorithms to help build personality profile scores. First, however, we have to make sure that our prosperity scales accurately and consistently measures each aspect or construct of brand irresistibility. In other words, we have to make sure that our scales have strong psychometric properties. Without assessing the psychometric properties of constructs, we can't be certain if we are "tapping into" the construct we are interested in. For example, we may think we are "tapping into" an individual's sociable personality traits, but in reality we might be measuring how friendly a person is in specific situations.

C-Score – Personality Profiles

We measure five key personality profiles in the Workplace Characteristics scale for each individual: Sociable, Agreeable, Resolute, Mindful, and Creative.

Sociable

High score indicates that the individual is outgoing, gains energy in social situations, and helps others feel energized and enthusiastic. A sociable personality profile is best suited for **people-oriented careers such as sales, public relations, human resources, and education.**

Creative

High score indicates that the individual is a curious and artistic person, who enjoys new experiences and has many interests. A creative personality profile is best suited for **artistic careers (i.e., arts, writing, designing, writing) and even challenging careers (i.e., lawyer, pilot, entrepreneur).**

Agreeable

Someone who scores high on this trait is someone who is kind, easily trusts people, and loves helping others. An agreeable personality profile is best suited for **careers that prioritize empathy and care for others, such as nursing, teaching, and counseling.**

Mindful

Someone who scores high on this trait is a thoughtful, goal-directed, and organized individual. This individual plans ahead and adheres to a schedule. A mindful personality profile is best suited working as **scientists, doctors, politicians, or entrepreneurs.**

Resolute

Someone who scores high on this trait is resilient, calm in stressful situations, and rarely makes impulsive decisions. A resolute personality profile is best suited for **high stress jobs such as surgeon, firefighter, or dentist.**

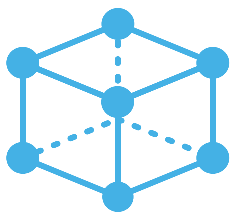
We added scores for each item for their respective personality profile in the Workplace Characteristics scale to obtain overall personality profile scores. We then determine which personality profile is most dominant for an individual. A similar personality test is administered to various employers. Together, the two tests are used to compute a C-Score specific to a potential and/or current employer. The C-Score represents the extent to which the individual or employee's personality is compatible with the employer's personality, providing key insights that help facilitate the job search process and take the guesswork out of it. For the current paper, we will focus on the Sociable personality profile, which taps into the Sociablility construct.

C-Score Part 1: Scale Validity

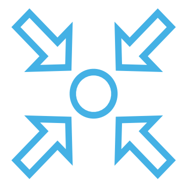
For the Workplace Characteristics s cale to work, we had to train and test how responses to the scale were related to scores on each of the constructs and if the scale had acceptable psychometric properties. The first psychometric property we looked at was construct validity.

Construct Validity

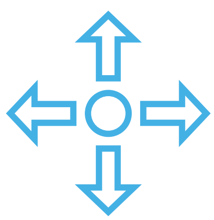
Validity corresponds to the extent to which the scale accurately measures reality. Construct validity is an assessment as to whether or not the measure we created is measuring what we want it to measure. For example, is our measure of Empathy truly assessing the extent to which a brand understands their consumers? Or is it measuring something else? To test construct validity, we look at four areas:



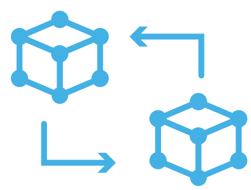
Structural Validity
Does the factor structure support that items are all measuring the same construct?



Convergent Validity
Does the construct, Sociability, relate to other constructs it should be theoretically related to?



Divergent Validity
Is the construct, Sociability, unrelated to constructs it shouldn't be related to?



Nomological Validity
Does the network of constructs around the construct, Sociability, show relationships that are expected?

Construct Validity: Structural Validity.

For the Sociability construct, we want to make sure that each of its items are measuring various aspects of the construct including: friendly, assertive, excitement, cheerful, talkative, and outgoing and all the items together measure the Sociable Personality Profile. To do so, we assess structural validity by using both exploratory factor analysis and confirmatory factor analysis.

Exploratory Factor Analysis.

- Step 1: Correlation Check
  - To determine which items to include or exclude in factor analysis, we first examined the **bivariate correlations** to identify any items with small bivariate correlations (r <.30). As you can see in the example below, all items have correlations, on average of .35 with each other. This indicates that all items are retained.

	SO_1	SO_2	SO_3	SO_4	SO_5	SO_6
SO_1	1.00	0.56	0.43	0.43	0.37	0.43
SO_2	0.56	1.00	0.51	0.51	0.44	0.46
SO_3	0.43	0.51	1.00	0.41	0.41	0.41
SO_4	0.43	0.51	0.41	1.00	0.37	0.50
SO_5	0.37	0.44	0.41	0.37	1.00	0.42
SO_6	0.43	0.46	0.41	0.50	0.42	1.00

  - Traditional bivariate correlations only provide a part of the picture, so we also examined **partial correlations**. Partial correlations refer to the correlation between two items after controlling for the effect of all other items. In other words, partial correlations are the correlations that are left over after the common variance is extracted. As a rule of thumb, we include items with a partial correlation <.70 in the analysis and exclude items that exceed this threshold. As you can see in the example, all items have partial correlations below .70 with each other. Therefore, we retain all items for the Sociability construct.

	SO_1	SO_2	SO_3	SO_4	SO_5	SO_6
SO_1	1.00	0.33	0.12	0.11	0.07	0.14
SO_2	0.33	1.00	0.23	0.22	0.16	0.09
SO_3	0.12	0.23	1.00	0.11	0.17	0.11
SO_4	0.11	0.22	0.11	1.00	0.07	0.27
SO_5	0.07	0.16	0.17	0.07	1.00	0.19
SO_6	0.14	0.09	0.11	0.27	0.19	1.00

  - We also look at the **anti-image correlation matrix**, which contains the negatives of the partial correlation coefficients. Consequently, these values are the magnitude of the variable that can't be regressed on, or predicted by, the other variables. If variables can't be regressed on, or predicted by, the other variables, then the variables are not likely related. If variables aren't related, then they will not likely load on the same factor. Consequently, large magnitudes indicate the possibility of a poor factor solution. However, as you can tell from the light-mid colors in the corrogram heat map, majority correlations in the anti-image correlation matrix are close to 0. This means all items on both constructs are retained.

	SO_1	SO_2	SO_3	SO_4	SO_5	SO_6
SO_1	1.00	-0.19	-0.08	-0.07	-0.04	-0.09
SO_2	-0.19	1.00	-0.14	-0.13	-0.10	-0.05
SO_3	-0.08	-0.14	1.00	-0.07	-0.11	-0.07
SO_4	-0.07	-0.13	-0.07	1.00	-0.05	-0.17
SO_5	-0.04	-0.10	-0.11	-0.05	1.00	-0.13
SO_6	-0.09	-0.05	-0.07	-0.17	-0.13	1.00

    - **Bartlet test of sphericity** compares the correlation matrix to the identity matrix, checking to see if there is any redundancy between the variables. High redundancy is indicative that the variables have common variance and therefore can be loaded on similar factors. If there is high redundancy, then the correlations in the correlation matrix should be higher in magnitude. Therefore, when it's compared to the identity matrix (where values are mainly 0), the two matrices will not be similar. If there is little redundancy, then the correlations in the correlation matrix should be close to zero. This means when it is compared to the identity matrix, the two matrices will be similar, indicating the possibility of a poor factor solution. In the case of the Sociability construct, the correlation matrix was significantly different from the identity matrix.

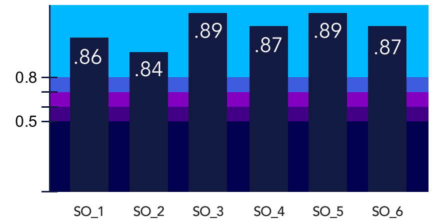
	SO_1	SO_2	SO_3	SO_4	SO_5	SO_6
SO_1	1.00	0.56	0.43	0.43	0.37	0.43
SO_2	0.56	1.00	0.51	0.51	0.44	0.46
SO_3	0.43	0.51	1.00	0.41	0.41	0.41
SO_4	0.43	0.51	0.41	1.00	0.37	0.50
SO_5	0.37	0.44	0.41	0.37	1.00	0.42
SO_6	0.43	0.46	0.41	0.50	0.42	1.00

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1.00	0.00	0.00	0.00	0.00	0.00
[2,]	0.00	1.00	0.00	0.00	0.00	0.00
[3,]	0.00	0.00	1.00	0.00	0.00	0.00
[4,]	0.00	0.00	0.00	1.00	0.00	0.00
[5,]	0.00	0.00	0.00	0.00	1.00	0.00
[6,]	0.00	0.00	0.00	0.00	0.00	1.00

» Lastly, the **Kaiser-Meyer-Olkin Measure** of sampling adequacy measures the extent to which the variance of the items might be caused by an underlying factor. The higher proportion of variance caused by underlying factors, the better your factor solution might be. Consequently, the following is what we use to determine whether or not to continue with the factor analysis.

> .80	Meritorious
> .70	Middling
> .60	Mediocre
> .50	Miserable
< .49	Unacceptable

There is cause for concern, if the KMO drops below .60. All items for the Sociability construct have values above .84, indicating that they are meritorious.

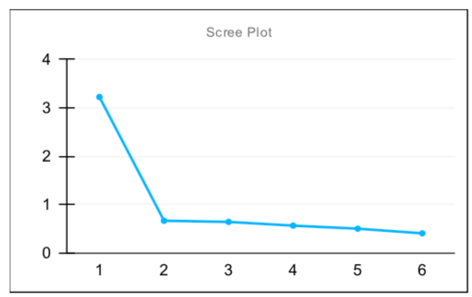


- Step 2: Factor Check. Once the correlations check out for each construct, and a final list of items are retained, we then run the factor analysis. The first thing to consider in this process is how many factors to retain in the solution. To determine this, we use two general principles:

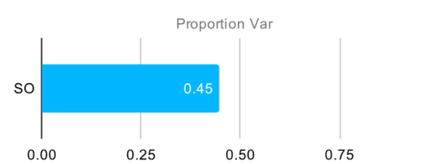
  » Only retain factors with eigenvalues > 1

Eigen values

3.22	0.67	0.64	0.56	0.50	0.41



  » Only retain factors with variance > 5% OR factors whose variance sum to 60% or more

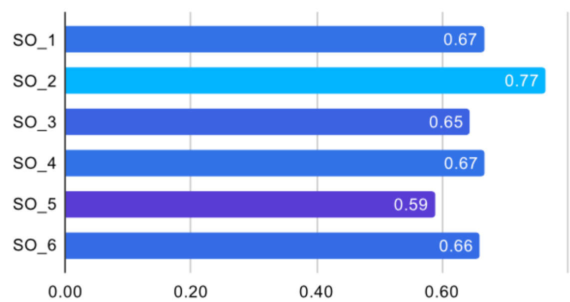


  » While the eigenvalues present evidence for a one-factor solution, in that all six items come together to represent the Comprehend construct, the proportion of variance for the two facets indicate that items measuring the Comprehend construct can also be further divided and represented with the Experience facet and Knowledge facet.

- Step 3: Item Check. Once we've decided on the number of factors that should be retained, the question becomes what items are associated with the factors (and which items are not).
  - For practical significance of factor loadings, we follow the below approach

> .70	Indicative of a well-defined structure
.50 - .69	Practically significant
.30 - .49	Minimally viable for a factor structure
< .30	Unrelated

You can see the following example:



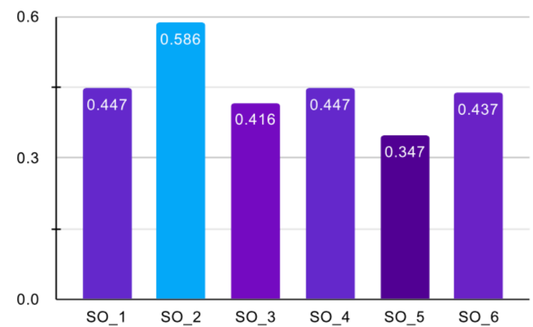
  - In this case, all items are at least practically significant for the factor structure, meaning there is no cause for concern thus far.
  - For statistical significance of factor loadings, there are a few different approaches that researchers can take. However, factor loadings significance changes as a function of sample size. Consequently, we generally adhere to the following significance of factor loadings given the sample size.

Factor Loading	Sample Size Needed for Significance*
.30	350
.35	250
.40	200
.45	150
.50	120
.55	100
.60	85
.65	70
.70	60
.75	50

*Significance is based on a .05 significance level (α), a power level of 80 percent, and standard errors assumed to be twice those of conventional correlation coefficiencies

  - With a sample size of 401 participants we can safely conclude that factor loadings for the Sociability construct are statistically significant.

- Lastly, when determining what items to retain, we look at communalities. Communalities are the proportion of each variable's variance that can be explained or accounted for by the factors. As a general rule of thumb, we shoot for Communalities > .5 (i.e., retaining items in which a half of the variance of each variable should be accounted for by the factor solution).



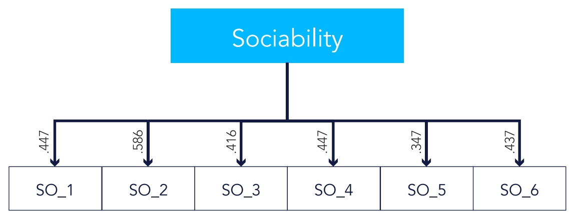
  - Some items have communalities that are below .50. However, since the items represent various aspects of the Sociability construct it is not surprising that people in the sample endorsed some traits over others. We retain all items anyway.

Confirmatory Factor Analysis.

Exploratory Factor Analysis is only half of the equation. At Inkblot Analytics, we also use Confirmatory Factor Analysis to help with structural validity. While exploratory factor analysis was a data-driven approach, confirmatory factor analysis is a theory-based approach that helps us "confirm" if our theory matches the data. There are four things we look for in a confirmatory factor analysis that supports structural validity:

- **Standardized loading estimates should be high.** Standardized loading estimates are the same as standardized regression coefficients—they quantify the magnitude and direction of the relationship between the item and the factor. We want the relationship between the item and the factor to be high, therefore standardized loadings estimates must be high to be retained for adequate structural validity. More specifically, we use the accompanying rule. Notice, items for the Sociability construct load highly and ideally on the factor.

> .50	Ideal
.30 - .49	Minimally Viable
< .30	Unacceptable



- **Standardized residuals should be small.** Standardized residuals are a calculation of the error in a model. Basically, it is a calculation of the magnitude of difference between observed and expected values. If our factor structure is not valid, then there is likely to be more error. Consequently, for items to be retained, we look for low values.
  - Notice that the values for all items are less than .20. This means that the expected values are a close match to the observed values. Very little error was produced when we estimated our theoretical model.

< .20	No problem
.21 - .39	"Red flag"
> .40	Unacceptable

	SO_1	SO_2	SO_3	SO_4	SO_5	SO_6
SO_1	0.000	0.055	-0.014	-0.027	-0.027	-0.012
SO_2	0.055	0.000	0.011	-0.004	-0.008	-0.042
SO_3	-0.014	0.011	0.000	-0.016	0.024	-0.011
SO_4	-0.027	-0.004	-0.016	0.000	-0.019	0.059
SO_5	-0.035	-0.008	0.024	-0.019	0.000	0.032
SO_6	-0.012	-0.042	-0.011	0.059	0.032	0.000

- **Model Indices should be small.** Modification indices represent the improvement a model would see (that is, improvement in units of chi-square values) if a particular relationship was added or deleted to the model. For a factor structure to be structurally valid, we want to minimize the number of modification indices and their values. Our current rule of thumb is that modification indices > 4 suggests improvements can be made to the model and therefore represent a poor factor structure.
  - CFA is a theoretically guided analysis. So the researcher must be selective in what modification indices to use. The algorithm will give any/all modifications that can be made to your model, not just the ones that are theoretically relevant. In this case, two modification indices were flagged. Both flags recommended the addition of a correlation path between two items. This is not surprising, as all items are related to each other and adding any of these paths would not change the interpretation of the model. The modification suggestions are theoretically trivial. To be through, we tested each modification suggestion and did not find a significant improvement in model fit for all. In other words, adding any of the four recommended paths had a minimal (non-significant) impact on our final conclusions, so we retained our hypothesized model.

- Model Fit Indices should indicate a good fit. Lastly, there are a number of model fit values that provide an overall assessment of how well the model fits the data. We use many of these to assess model performance and overall structural validity. The table below will show you what values we use for our cutoff.

Factor Loading	Standard for Acceptable Fit
CFI	> .90 (marginal fit) ; > .95 (good fit)
TLI	> .90 (marginal fit) ; > .95 (good fit)
RMSEA	< .08
SRMR	> .05 (i.e., not statistically significant)
PClose	< .08
CD	The closer to 1, the better the fit
AIC	When comparing models, the lower the better
BIC	When comparing models, the lower the better

CFI	0.988
TLI	0.98
RMSEA	0.049
SRMR	0.025

  » Notice that our model fits the data very well. Since all other empirical evidence points to a good fit, we move forward.

- AVE > .5. With CFA, the average variance extracted is calculated by the average of the variance explained by the factor for each item that loads on it. Said differently, it's the sum of the squared standardized loadings of all items on a factor, divided by the number of items on that factor. If an AVE < .50, then it suggests that error explains more about the item's variance than is explained by the factor structure. In this case, the average variance extracted was greater than .50.

Construct Validity: Convergent & Divergent Validity.

Other forms of construct validity are known as convergent and divergent validity. Convergent validity refers to the relationship between variables that should be theoretically related. Divergent validity refers to the relationship between variables that should not be theoretically related.

With Regular Bivariate Correlations.

When looking to support convergent and divergent validity, the use of bivariate correlations can show us just how related different measures are. At Inkblot Analytics, we use the accompanying rules of thumb.

> .90	Indicates the same construct
.70 - .89	Convergent validity for highly related constructs
.50 - .69	Convergent validity for somewhat related constructs
.40 - .49	No man's land
.20 - .39	Divergent validity for somewhat unrelated constructs
.10 - .19	Divergent validity for highly unrelated constructs
0 - .09	Indicates no relationship

With Confirmatory Factor Analysis.

» AVE > Correlation. Convergent validity is supported by finding two constructs are related, but are NOT the same construct. For this to be shown, the variance extracted by a factor should be GREATER than the variance explained by the related construct. So when doing a CFA, we're looking for the AVE for two factors to be greater than the correlation between the two factors.

» A model with cross-loadings should be a poorer fitting model. When performing a CFA, if construct validity is to be theoretically supported, there should not be any cross-loaded items. If there were to be cross loaded items, removing them should make the model better. To test this out, we force some items to cross-load (that is, load on to the original construct and the related construct). By doing this, your model should get worse. If it gets better, then you know both constructs might be measuring the same thing.

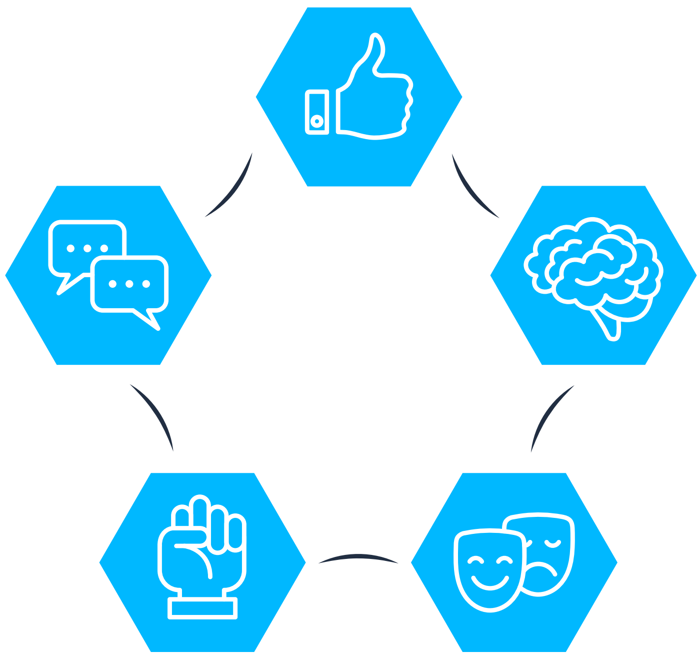
With Bifactor Modeling.

» Test a bifactor model and see if it gets worse. A bifactor model is usually used when you want to test the presence of a general factor that all items load onto. This approach helps identify the plausibility of a scale having multiple factors that are theoretically uncorrelated.

Convergent and divergent validity analysis are add-on features.

Construct Validity: Nomological Validity.

Typically, at Inkblot Analytics, we use other construct types for convergent and divergent validity, while using variables from the same construct type for nomological variability. For nomological validity we look at a correlation matrix and identify the biggest correlations. In theory these relationships should correspond to how you would theoretically think variables within the same construct type would be related. For example we found the following correlations:



- People who score highly on the Sociable personality profile are likely to score highly on the Agreeable personality profile.
- People who score highly on the Sociable personality profile are likely to score highly on the Resolute personality profile.
- People who score highly on the Mindful personality profile are likely to score highly on the Resolute personality profile.

These relationships between constructs make sense, because an individual can manifest many other personality traits that are closely related.

C-Score Part 2: Scale Reliability

Item-to-total correlations > .5.

One of the first things we look at is to what extent each scale item correlates with a composite score of the scale (i.e., with all items for the scale scored properly). Generally speaking, we look for an item-to-total correlation of at least .50. When looking at the scores for the Sociability construct, we get the following:

	Emp
SO_1	0.7720
SO_2	0.7960
SO_3	0.7050
SO_4	0.7340
SO_5	0.6580
SO_6	0.7180

Notice all items are above the .50 threshold, indicating that each item correlates strongly with the composite score that represents the Sociable personality profile.

CFA's Composite Reliability >.70.

We calculated the composite reliability of the CFA models. This includes both Alpha and Omega values of reliability. Generally speaking, we use the following criteria:

> .70	Suggests good reliability
.60 - .69	Acceptable

As you can see below, items for Sociability construct meet the .70 threshold for reliability. Taking it one step further, items measuring each personality trait, collectively meet the .70 threshold for reliability. This means that all items in the Workplace Characteristics scale reliability measure personality.

	Sociable	Total
Omega	0.86	0.93

Chronbach's Alpha > .70.

One of the most prolific ways of checking scale reliability is by calculating Chronbach's alpha. When calculating scale reliability at Inkblot Analytics, we use the following standards:

Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	Excellent
$0.9 \geq \alpha \geq 0.8$	Good
$0.8 \geq \alpha \geq 0.7$	Acceptable
$0.7 \geq \alpha \geq 0.6$	Questionable
$0.6 \geq \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

	Sociable	Total
Alpha	0.83	0.92

Items for the Sociable personality profile and altogether items of the Workplace Characteristics scale have good internal consistency.

Summary

At this point, I hope you can see just how much rigor goes into the platform, MatchAI.

From the perspective of scale construction and use, scales must have adequate psychometric properties to be used. Items measuring the Sociable personality profile in the Workplace Characteristics scale have good to excellent psychometric properties.

No matter what part of the tool you're looking at, our results are backed by a rigorous vetting process.