ImpactAI 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 ImpactAI, 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 ImpactAI

How trustworthy the results are

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

IMPACT AI

BY THINK ANDERSON

The last several years has impacted consumers' lives in so many ways—socially, technologically, physically, environmentally, financially, and emotionally.To overcome these barriers, brands must demonstrate that they not only have a clear purpose and deliver a relevant value, but also make a desirable impact on the things that matter to their customers, prospects, and employees. People just aren't receptive to "messages without substance" or "business as usual." The Anderson Group helps brands assess opportunities to deepen their purpose and increase their value and impact on their customers and prospects, their employees, and their local/global communities. By maximizing the impact a brand has enables brands to achieve greater engagement, usage of products/services, and favorability from consumers, employees, and prospects. This paper specifically focuses on brands' impact on consumers.

ImpactAI is a new platform that uses patent-pending systems for AI-powered projective tests. This platform helps brands identify their brand impact quotient (TAG-IQ) specifically on consumers' lives. In other words, the extent to which a brand has an impact on consumer connections to culture, category, and consumer psychology. The ImpactAI platform offers a true barometer of a consumer's relationship with the brand.

How do we do it? ImpactAI 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:

- Brands have an impact on more than one area of a consumer's life
- Tapping into and identifying this impact can help brands form and sustain long-term profitability
- ImpactAI surveys consumers on their perception of the impact a brand has on key areas of their life: social, psychological, family, work, environment, physical, technological, purpose, and financial
- The ImpactAI platform offers brands solutions and detailed suggestions to enhance their relationship with consumers

What can you learn? By using the ImpactAI platoform, a brand can get access to insights about:

1. Scope of a brand's impact quotient score
2. What the scores mean
3. Deep dive into their impact quotient score
4. Recommendations for changing a brand's impact quotient score
5. Simulations to showcase the implications of increasing their impact quotient score

Key Benefits of I-Factor

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

Automated Predictive AI. The ImpactAI platform automatically creates predictive algorithms unique to each brand. The predictive AI not only provides brands with their brand impact quotient score and recommends solutions for improvements, but it also predicts how changing a brands impact quotient score will impact (1) consumer engagement, (2) product/service usage, (3) likelihood of consumers to recommend the brand, and (4) brand favorability. The predictive AI grows better each time it is used.

Less Questions, More Insights. Less Questions, More Insights. With the ImpactAI platform, brands answer less questions and get more insights. For years market researchers have been saying surveys in our industry are too long, leading to poor data quality from burnt out survey respondents. So at Inkblot Analytics, we wanted to create a solution that allows researchers to get the same amount of data by asking less questions.

Both Quantitative and Qualitative. Both Quantitative and Qualitative. ImpactAI studies both quantifiable data and emotional insight using a visual library to uncover secret sentiments that consumers harbor towards a brand. This enables brands to uncover consumers' deep-seated thoughts and feelings above and beyond a typical survey or interview. Concurrently, brands have access to quantitative data with a tangible brand impact quotient score through the platform. The combination of quantitative and qualitative insights offers a 360 visualization of their target audience.

How Does I-Factor Work?

This paper is focusing specifically on the TAG-IQ scale. Obtaining brand impact quotient scores involves a four step process:

The Testing Step
Taking the TAG-IQ scale

The Scoring Step
Scoring responses to the secret sentiments test

The Profiling Step
Identifying which profile is predominant for the individual

The Predicting Step
Predicting outcomes and solutions for brands

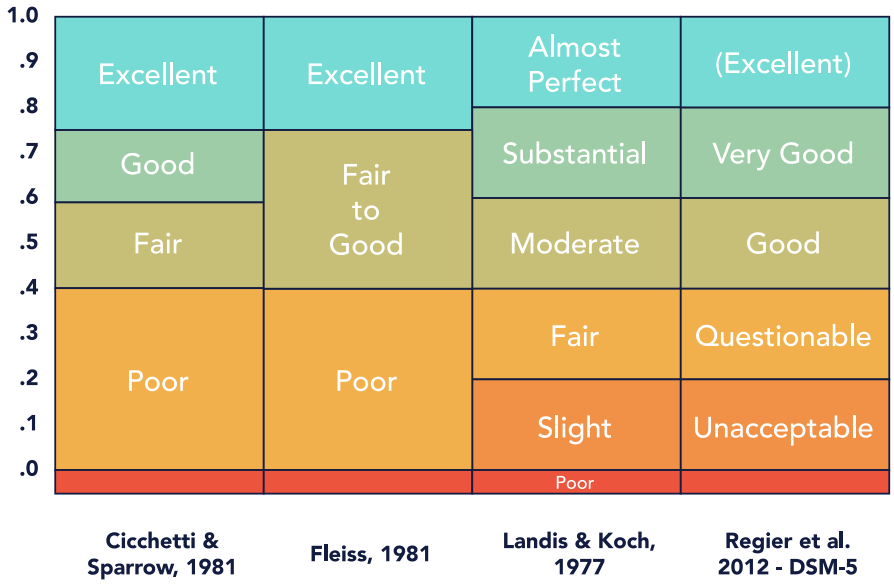
Each one of these steps has a scientific process built into them. For the **testing step** (i.e., when the participant takes the TAG-IQ scale), we want to make sure the data are good quality. So we use an algorithm that measures the extent to which a respondent is intentionally trying to deceive the 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 Brand Impact

Once the test data is collected, we are able to use our proprietary algorithms to help build brand impact quotient 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" the construct of brand impact on social issues relative to the consumer (issue involvement), but in reality we might be measuring the "general impact a brand has on social issues."

To measure brand impact or the extent to which a consumer perceives the impact a brand has on them, we created the TAG-IQ survey. A brand's impact on consumers can be best measured by Culture, Category, and Consumer Psychology. Culture refers to the extent to which consumers feel that a brand impacts their community and current social/cultural issues that matter to the consumer. This can be further divided into two facets: Issue Involvement and Community Involvement. Category refers to the extent to which consumers feel that a brand impacts their perception of an ideal and unique experience relative to competitor brands in the same categories.

The Profiling Step: Psychometrics of Brand Impact

The factor Category can be broken down into two facets: Ideal Experience and Unique Experience. Finally, Consumer Psychology refers to the extent to which consumers feel that a brand impacts various aspects of their life, specifically social, psychological, family, work, environment, physical, technological, purpose, and financial aspects. These aspects can be grouped into two distinct facets: Psychosocial and Socio-Cultural. In this section, we walk you through the scientific process of how we evaluated the psychometric properties of the TAG-IQ, using the construct Culture as an example.

TAG-IQ – Brand Impact Quotient

We determined that a brand's impact on consumers can be best measured by the three C's: Culture, Category, and Consumer Psychology.

For the Culture construct:

Culture - Issue Involvement

A high score indicates that consumers feel that they are more involved or engaged in current social and cultural issues because of a brand. In other words, the brand has an impact on the consumer's relationship with issue involvement.

Culture - Community Involvement

A high score indicates that consumers feel that they are more involved or connected to their community because of a brand. In other words, the brand has an impact on the consumer's relationship with community involvement.

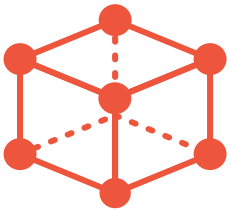
We determine the extent to which individuals feel a brand is high on Culture adding up scores on Culture - Issue Involvement and Culture - Community Involvement. We repeat this process for the remaining TAG-IQ constructs. Together, the three C's from the TAG-IQ. Brands can use this information to target specific constructs within the three C's to improve how consumers relate and feel towards their brand.

I-Factor Part 1: Scale Validity

For the TAG-IQ Scale 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:

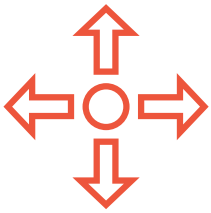


Structural Validity

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

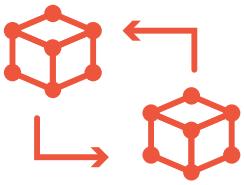
Convergent Validity

Does the construct, Culture, relate to other constructs it should be theoretically related to?



Divergent Validity

Is the construct, Culture, unrelated to constructs it shouldn't be related to?



Nomological Validity

Does the network of constructs around the construct, Culture, show relationships that are expected?

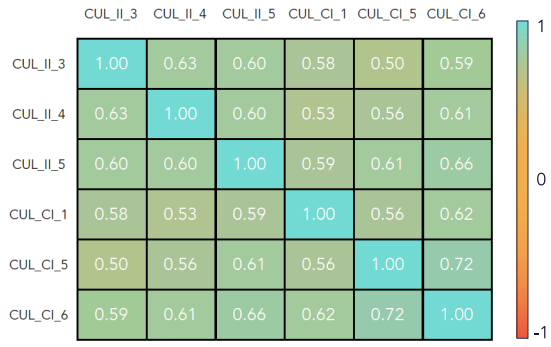
Construct Validity: Structural Validity.

For the Culture construct, we want to make sure that the items for Culture - Issue Involvement are measuring the extent to which a brand impacts a consumers relationship with social/cultural issues and items for Culture - Community Involvement are measuring the extent to which a brand impact the consumers relationship with community, and all items together are measuring the Culture construct. To do so, we assess structural validity by using both exploratory factor analysis and confirmatory factor analysis.

Exploratory Factor Analysis.

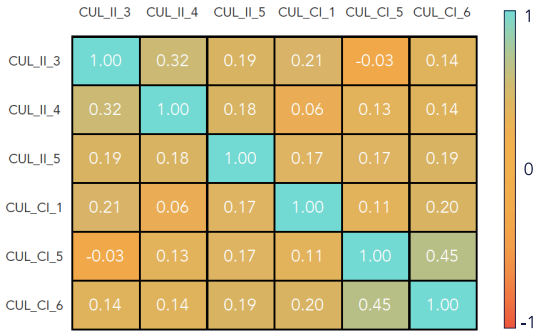
- Step 1: Correlation Check

» 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). Items with correlations below this threshold As you can see in the example below, the three items included in Empathy Identification all have correlations, on average, around .50 with each other. Similarly, all three items Empathy Emotion have correlations around .40 with each other. Together, the items have correlations above .40 with each other. Therefore, all items for the Empathy construct were retained.

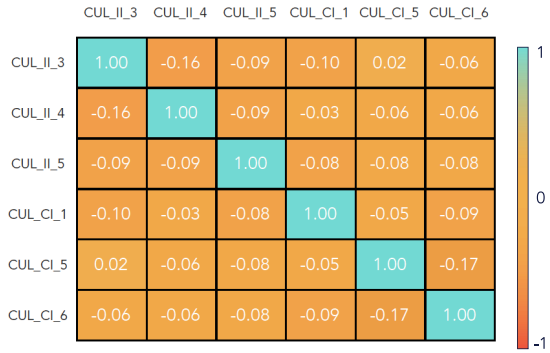


» Traditional bivariate correlations only provide a part of the picture, so we also examined **partial correlations.** Partial correlations refer to the correlation between two items after controlling for the effect of all other items. In other words, partial correlations are the correlations that are 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, the three items included in Issue Involvement have partial correlations below .7 with each other. Similarly, all three items in Community Involvement have partial correlations below .7 with each other.

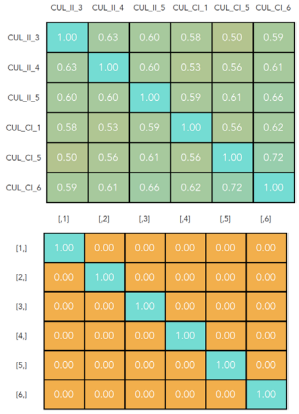
Together, all items have partial correlations below .7 with each other. Therefore, all items for the Culture construct were retained.



» 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.



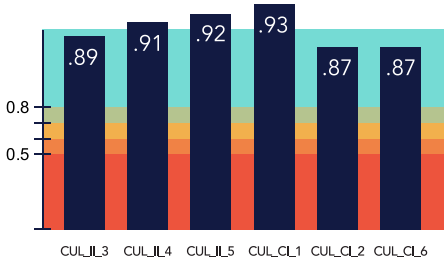
» **Bartlet test of sphericity** compares the correlation matrix to the identity matrix, checking to see if there is any redundancy between the variables. High redundancy is indicative that the variables have common variance and therefore can be loaded on similar factors. If there is high redundancy, 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 Culture construct, the correlation matrix was significantly different from the identity matrix.



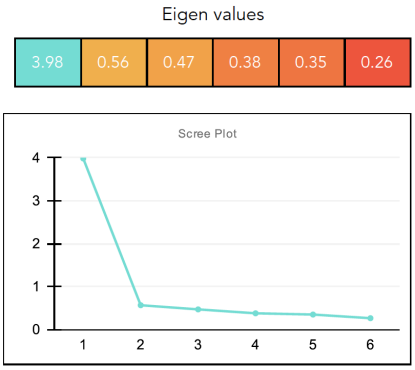
» Lastly, the **Kaiser-Meyer-Olkin Measure of sampling adequacy** measures the extent to which the variance of the items might be caused by an underlying factor. The higher proportion of variance caused by underlying factors, the better your factor solution might be. Consequently, the following is what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

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

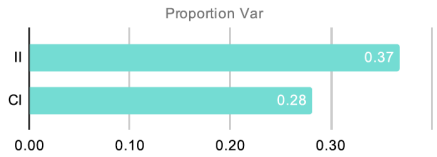
There is cause for concern, if the KMO drops below .60. All items for the Culture construct have values above .87, indicating that they are meritorious.



- 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



» 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 Culture, the proportion of variance for the two facets indicate that items measuring the Culture can also be further divided and represented with the Issue Involvement facet and Culture Involvement facet.

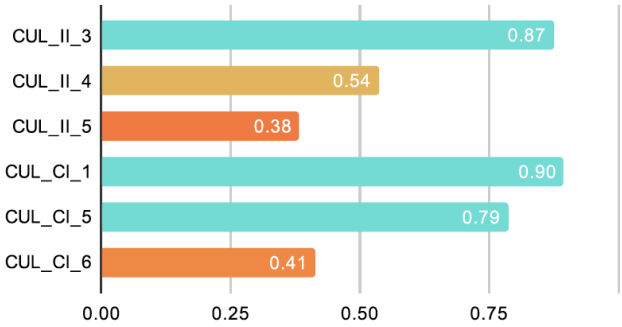
- Step 3: Item Check. Once we've decided on the number of factors that should be retained, the question becomes what items are associated with the factors (and which items are not).
  - » For practical significance of factor loadings, we follow the below approach:

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

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

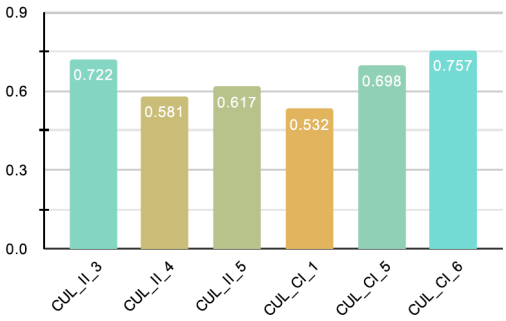
*Significance is based on a .05 significance level (α), a power level of 80 percent, and standard errors assumed to be twice those of conventional correlation coefficiencies

You can see the following example:



- » The lowest factor loading for the Issue Involvement Facet is .38, indicating that the item is minimally viable. The item relates the least to the construct compared to the other items. Similarly, the lowest factor loading for Community Involvement is .41.
- » 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.
- » With a sample size of 284 participants we can safely conclude that factor loadings for Issue Involvement and Community Involvement 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).



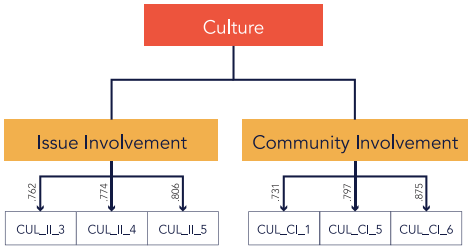
- » All items have communalities above .5, indicating that at least half of the variance in the common factor can be explained by the items.

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 Issue Involvement and Community Involvement load highly and ideally on their respective factors.

> .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 both Issue Involvement and Community Involvement 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

	CUL_II_3	CUL_II_4	CUL_II_5
CUL_II_3	0.000	0.054	-0.018
CUL_II_4	0.054	0.000	-0.027
CUL_II_5	-0.018	-0.027	0.000

	CUL_CI_1	CUL_CI_5	CUL_CI_6
CUL_CI_1	0.000	-0.038	-0.025
CUL_CI_5	-0.038	0.000	0.036
CUL_CI_6	-0.025	0.036	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. One pathway recommended the addition of a correlation path between items of Issue Involvement and Community Involvement. 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 second pathway was the correlation between the two facets. This is also not surprising, as both facets together create a higher construct of Culture. 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	0.987
TLI	0.975
RMSEA	0.074
SRMR	0.027

  - » 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. For both, Issue Involvement and Community Involvement, 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.

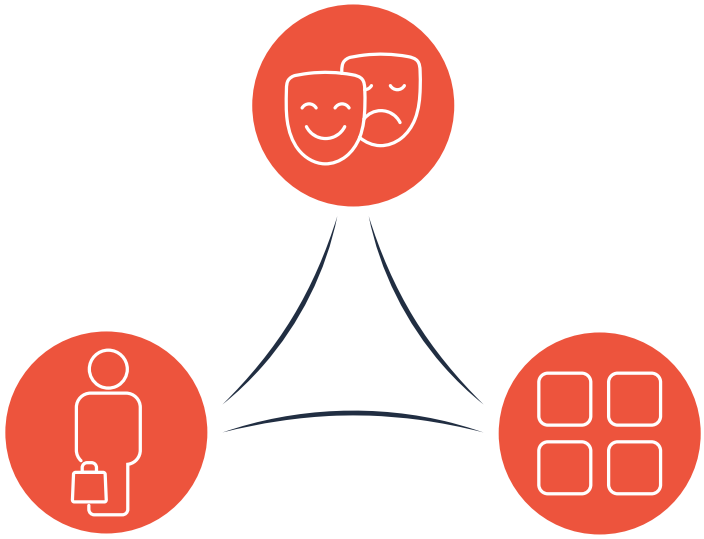
With Bifactor Modeling.

- 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:



- The higher a consumer rates scores on the Culture construct, the higher they score on the Category construct.
- The higher a consumer rates scores on the Culture construct, the higher they score on the Consumer Psychology construct.
- The higher a consumer rates scores on the Consumer Psychology construct, the higher they score on the Category construct.

These relationships between constructs make sense, as a brand that makes an impact on a consumer's relationship with their community is likely to also impact consumer psychology.

TAG-IQ 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 Issue Involvement, we get the following:

	Culture
CUL_II_3	0.7940
CUL_II_4	0.8000
CUL_II_5	0.8300

Notice all items are above the .70 threshold. Similarly, when looking at the scores for Community Involvement, we get the following:

	Culture
CUL_CI_1	0.7890
CUL_CI_5	0.8110
CUL_CI_6	0.8620

Again all items are above the .70 threshold.

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 Issue Involvement meet the .70 threshold for reliability. Items for Community Involvement meet the threshold for acceptable reliability. Additionally, the general Culture construct meets the reliability threshold.

	II	CI	Culture
Omega	0.84	0.85	0.91

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

	II	CI	Culture
Alpha	0.83	0.83	0.90

When looking at Issue Involvement and Community Involvement, scale reliability is .83, indicating good internal consistency. The general Culture construct has a reliability of .90 indicating excellent internal consistency. The higher reliability for the general construct further adds to the evidence that all items together measure the impact brands have on the consumers relationship with Culture.

Summary

At this point, I hope you can see just how much rigor goes into the platform, Ex-Score Scale.

From the perspective of scale construction and use, scales must have adequate psychometric properties to be used. Both example scales reported on in this paper--Empathy Identification and Empathy Emotion--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.