StandOut Assessment Development and Validation

WHITE PAPER SERIES »

tmbc.com

Introduction

» The StandOut assessment measures personal strengths to identify the assessment taker's competitive advantage in the workplace.

The StandOut strengths assessment was developed using a mixed-method research approach, combining quantitative analysis with qualitative techniques such as interviews, case studies, and focus groups to provide a deeper dive into strengths. Over the last fifteen years, we have administered a talent inventory to 435,564 participants in order to uncover the talents that most reliably predict multi-industry job performance for six job families: Leader, Manager, Professional, Sales, Service, and Support. From this research emerged the 18 themes that are included within the assessment's nine strength Roles.

TMBC's senior researchers continue to scrutinize the StandOut assessment to deepen our understanding of its psychometric properties. This paper describes the development, psychometric support to date, and ongoing research efforts surrounding the assessment.

Validity and Purpose

The primary application of the StandOut assessment is to evaluate an individual's strength Roles (combinations of dominant patterns of thought, feeling, and behavior that can be productively applied) and then provide targeted development strategies for using those strength Roles at work.

The StandOut assessment was designed to:

- Help leaders pinpoint their own strengths to leverage them with their team members.
- Help team members identify their own strengths to use them in their work.

The StandOut assessment is not:

- A selection instrument.
- A performance management tool.
- A comparison tool.
- An omnibus measurement of the full spectrum of personality.

It is important to understand what the StandOut assessment is and is not, because the validity of an assessment must be evaluated through its intended purpose and consequences.

Messick (1989) serves as the beginning of the modern view of validity. The modern approach demands long-term commitment, which means that the validation process can never be entirely complete. TMBC is continuously studying the consequences of assessment use, evaluating and correcting measurement error issues, and understanding the evolving context of use. This white paper discusses the extent of our research surrounding the validity evidence.

Scoring

The precise scoring of the StandOut assessment is proprietary to TMBC. Nevertheless, a general analysis of the scoring process will allow for the establishment of validity and reliability.

The StandOut assessment measurement is completed using a timed Situational Judgment test (SJT). Each question gives participants 35 seconds to choose one of four fixed options as a response to a hypothetical real-world situation. Each situation was carefully chosen to represent an impactful moment in life in which a person's pre-existing skills or technical knowledge would have minimal reference to response options.

Most of the response options are associated with a specific talent theme — an innate pattern of thought, feeling and behavior. Eighteen talent themes served as the fundamental building blocks of the nine strength Roles measured in the StandOut assessment. A proprietary formula assigns a value to each item response, which helps to derive a strength Role score. Results are presented to the respondent as a rank ordering of the nine strength Roles.

The relation of talent themes to strength Roles is derived from research with high-performing individuals and the ways that their talents tend to cluster into identifiable patterns.

As part of our research in the validation of the StandOut assessment, we sought to understand excellence in practice through the lens of our nine strength Roles. With each sample, we brought prior expectations from our historic experience. These recurring patterns of powerful talent combinations provide a basis for analyzing the keys to success for individuals who share the same talents.

Theory

Seligman and Csikszentmihalyi (2000) charged psychology as a field to understand and document "what work settings support the greatest satisfaction among workers." In response to this challenge, the discipline of positive psychology emerged as an attempt to change the preoccupation with "repairing the worst things in life" into "building positive qualities" (Seligman & Csikszentmihalyi). Positive psychology studies the strengths and virtues that enable individuals and communities to thrive (Bakker & Schaufeli, 2008). This approach underpins the StandOut assessment.

Research

The research process to develop and validate the StandOut assessment has continued for over 15 years, combining qualitative and quantitative data. Following are more detailed descriptions of a selection of the studies that have informed the assessment's development.

Study 1

The first study set out to begin construct definition and development. Between 2000 and 2010, we conducted over 200 focus groups with top-performing leaders, managers, professionals, sales, and service specialists in multiple industries.

Analysis of the qualitative data revealed a pattern of 18 talent themes.

Based on these talent themes, we developed interviews to validate and refine the profiles of top performers (which would become the basis for the nine strength Roles). The discovery interviews were conducted as part of 73 validation studies, which included over 1,000 participants.

The 18 themes were administered as part of a talent inventory to 435,564 participants in order to calibrate the items and uncover the most reliable talents for prediction. To ensure that the talent themes were a valid predictor of success, we conducted several predictive validity studies with top performers from different industries and organizations. In addition, each of the themes was analyzed for internal consistency using Cronbach's Alpha. Across the samples, alphas for the 18 talents ranged from .64 to .93, demonstrating an acceptable high internal consistency to the industry standard of alpha = .6.

Table 1: Reliability of Predictive Talents by Strength Role

Talent	# of Items	Cronbach's Alpha Reliability Coefficient (n=94756)	Strength Role
Problem Solver	6	.76	Advisor
Common Sense	8	.69	Advisor
Initiator	6	.64	Connector
Team	9	.76	Connector
Adaptability	7	.76	Creator
Analytical	5	.76	Creator
Responsibility	8	.68	Equalizer
Structure	8	.79	Equalizer
Persistence	5	.69	Influencer
Courage	7	.70	Influencer
Achiever	5	.75	Pioneer
Belief	7	.68	Pioneer
Relator	8	.74	Provider
Service	7	.84	Provider
Positivity	5	.83	Stimulator
Intensity	5	.70	Stimulator
Developer	9	.76	Teacher
Individualization	5	.77	Teacher

***p < .05, **p < .01**

While the goal of the StandOut assessment is not to predict job performance, this preliminary research helped to establish the validity of the talents that leaders leverage to achieve success.

Study 2

In 2009, we completed a second study to test the Alpha version of the StandOut assessment. We created hundreds of situations and talent-based response options, then trialed them in multiple test environments. The assessment was administered to 232,000 participants across jobs and industries globally. We then analyzed the psychometric properties of the assessment to refine StandOut's 146 situations and response combinations.

We then completed differential item functioning (DIF) for gender and age. DIF occurs when respondents from different groups show differing probabilities of endorsing an item while controlling for ability. Content experts reviewed the scenarios that showed DIF to determine if any adverse impact existed with the situation or the response options. Some of the scenarios were removed from the StandOut assessment to prevent any item bias or adverse impact.

Study 3

In the fall of 2010, to test the stability and reliability of responses to the StandOut assessment, we conducted a third study using a test-retest analysis. A random sample of N=269 participants from the US workforce completed the assessment online. They received no specific strengths feedback or information about the purpose of the assessment. Six months later, the same participants completed the assessment a second time.

We conducted a statistical analysis called the Chi-Square Test of Independence using a dichotomous variable labeled "Strength Role Match from Time One to Time Two." A review of the top three strength Roles computed from the analysis found that 90% of the sample had their leading Role in the top three from time one to time two, and 47% had two matches from time one to time two among their top three. All of the nine strength Roles had significant Chi-Square results, indicating that their presence in the top three Roles on the initial administration of the assessment was significantly related ($p<.05$) to their presence in the top three Roles during the second administration of the assessment.

Table 2: Chi-Square Test of Independence Results (N=269)

Theme	Chi-Squared	Significance
Advisor	5.039	.025*
Connector	6.335	.012*
Creator	11.245	.001*
Equalizer	29.911	.000*
Influencer	9.444	.002*
Pioneer	25.460	.000*
Provider	6.822	.009*
Stimulator	28.906	.000*
Teacher	20.530	.000*

**Correlation is significant at the .05 level (1-tailed)*

To scrutinize further the consistency of the results, the test-retest analysis involved correlating theme means from time one and time two to examine theme comparisons outside of the respondent's top three Roles. The test-retest correlational values of the dimensions range from .384 for the Creator profile to .600 for the Pioneer profile. The average correlation is .484, shown in Table 3.

Table 3 : Six to Seven Month Test-Retest Reliability Estimate N = 269

Theme	Test-Retest Reliability Estimate
Advisor	.539*
Connector	.435*
Creator	.384*
Equalizer	.517*
Influencer	.389*
Pioneer	.600*
Provider	.429*
Stimulator	.535*
Teacher	.531*

**Correlation is significant at the .001 level (2-tailed)*

A dependent sample t-test was completed to determine if there were significant differences in the strength Roles from time one to time two. We found no significant differences, lending more evidence for the stability of an individual's strength Roles over time.

Table 4: Test-Retest Paired Comparison (N=269)

Theme	t-test	Significance
Advisor	.488	.626
Connector	1.025	.306
Creator	1.944	.053
Equalizer	-.353	.724
Influencer	1.816	.070
Pioneer	.767	.443
Provider	.200	.842
Stimulator	-.038	.970
Teacher	.074	.941

The overall results show an acceptable amount of test-retest reliability for this population. TMBC plans to continue examining the stability and reliability of the strength Roles over time, also involving international participants.

Reliability

According to Classical Test Theory, the reliability of a score is an estimate of its stability (true score), or the portion of the score not due to error (Nunnally & Bernstein,1994). For instruments like the StandOut assessment, there are two generally used types of reliability estimates: 1) Internal Consistency and 2) Test-Retest Reliability.

- *Internal consistency* is a measure based on correlations between different items on the same assessment. It looks at how well the proposed items measure the same general construct and produce similar scores. Cronbach's Alpha is often used to measure this type of reliability.
- *Test-retest reliability* is completed by using the same group of respondents who take the assessment at two administrations separated by time. The results are compared from time one to time two for consistency in responding. A paired sample t-test is used to capture the stability of responses over time.

Other methods such as Exploratory or Confirmatory Factor Analysis are often used to establish the validity and reliability of instruments. However, given how the StandOut assessment is scored, it is not appropriate to conduct an Exploratory or Confirmative Factor Analysis, because the comparison of individual responses is ipsative or recursive in nature, which would introduce multicollinearity into the model and be problematic.

Another method, Cluster Analysis, is often proposed for understanding groupings of items; however, the method is not clearly established. A common criticism is that statisticians can manipulate the data to fit a proximity matrix and link groups until they "discover" the structure that they proposed to begin with. For this reason, cluster analyses are not appropriate for the StandOut assessment and have not been completed.

Study 1 and Study 3 both provide support for the reliability and stability of the StandOut assessment, and constitute sufficient evidence of the consistency and stability of the StandOut assessment for its given purpose.

Validity

Validity evidence collection for the StandOut assessment started with focus groups, interviews and literature reviews of relevant constructs. The traditional view of validity contains three types of validity evidence: construct, content, and criterion-related. While there is nothing wrong with this view of validity, other considerations are important when examining the validity evidence. Validity has to be considered based on the inferences or consequences that come from the assessment. Again, the purpose of the StandOut assessment is to classify an individual's strength Roles to increase targeted strengths-based personal development. Because the assessment is not a selection tool or recruitment tool, we have not collected any evidence that one profile is better for any job or function. While we have examined the criterion relationship to the underlying talents for the assessment, we have not completed any criterion-related studies using the current assessment.

Construct Validity

The validation process began with the construct definition stage before any of the items was written. We explored the data patterns from half a million talent assessments and listened through thousands of interviews to understand what explains success. It became apparent that the 18 core predictive talents combine into common clusters with dominant peaks of frequency and intensity that best explain how someone will tend to think, feel, and behave. StandOut measures how these talents converge into nine strength Roles, which are powerful frequently recurring patterns that emerged through our research. Figure 1 on page 16 shows the hypothesized measurement model for the StandOut assessment. Each of the nine strength Roles is tied to two sub-dimensions that are measured by a varied number of questions.

Content Validity

Once the constructs were well defined, the instrument was built and validated against known questions that measured the dimensions hypothesized to be contained in the construct. Individuals who were well versed in the strengths literature participated in the qualitative item analysis. These content experts examined the responses to the SJT to determine if each of the item responses grouped into the construct of interest. Items were classified as representative of one strength Role or another. Each of the items achieved an acceptable expert classification.

Future Research

TMBC continues to evaluate the items as well as relationships for the StandOut assessment as data becomes available from clients and research partners. We will continue to conduct:

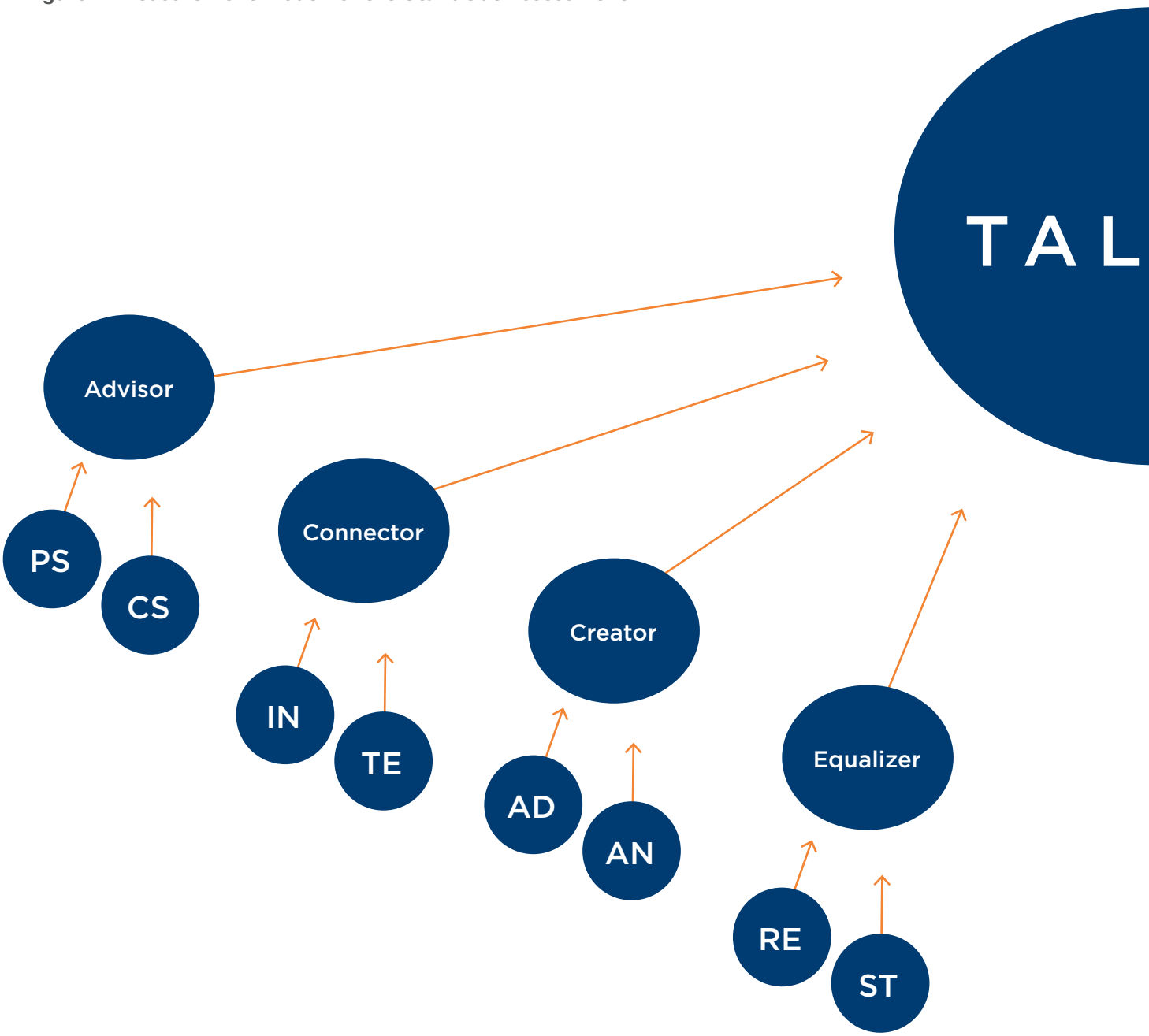
- Continuous item calibration studies using data from current users (n=229,701) of the tool. This data is used to solidify the relationships within the tool and the nine strength Roles.
- Further test-retest reliability studies in the US population as well as international populations.
- Demographic studies to understand any generational differences or gender differences that affect results of the StandOut assessment.
- Content validation studies with new response options to refine further the StandOut assessment.
- Ongoing research to understand the relationships between team leaders' strength Roles and their team members' strength Roles.

TMBC's senior researchers also continue to examine the validity, stability, and reliability of the StandOut assessment and modify it based on research findings when appropriate.

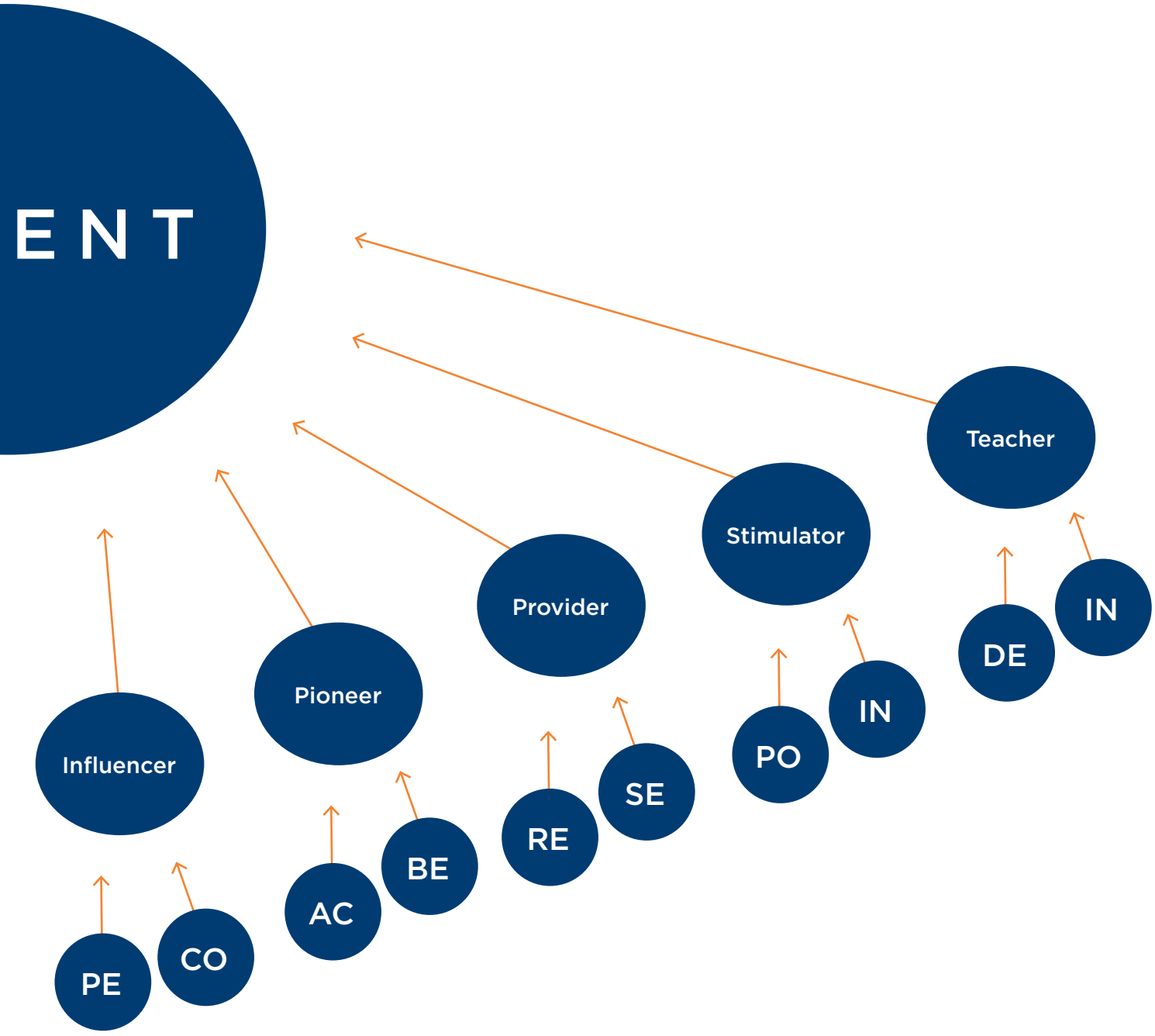
Closing

Half a million individuals have begun or continued their strengths-based journey using the StandOut assessment. Strengths facilitators, using this assessment, have helped to assist with the development of individuals across multiple industries and job roles. Research for this tool is provisional, as TMBC's researchers will continue to learn and modify the tool based on evidence collected from ongoing studies. To date, research with top executives and team leaders in multiple fields has led us to a greater understanding of how individuals turn their talents into success. Our understanding of strengths continues to grow from each study that we conduct with our clients and research partners.

Figure 1: Measurement Model for the StandOut Assessment



See Table 1 on page 7 for a full list of the 18 talent themes.


ENT
Teacher
Stimulator
Provider
Pioneer
Influencer
IN
DE
IN
PO
SE
RE
BE
AC
CO
PE

References

Bakker, A.B., & Schaufeli, W.B. (2008). Positive organizational behavior: Engaged employees in flourishing organizations. Journal of Organizational Behavior, 29, 147–154.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational Measurement (3rd ed., pp. 13–103). New York: Macmillan.

Nunnally, J.C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill.

Seligman, M.E. & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. American Psychologist, 55, 5–14.