

**TRACOM<sup>®</sup> GROUP**

THE SOCIAL INTELLIGENCE COMPANY<sup>®</sup>



# RETEST RELIABILITY

*Documenting the Reliability of SOCIAL STYLE<sup>®</sup>*

## OVERVIEW

The reliability of TRACOM®'s SOCIAL STYLE Model™ and SOCIAL STYLE® assessments is the focus of this whitepaper. It specifically looks at retest reliability over time including each of the three components of SOCIAL STYLE: Assertiveness, Responsiveness and Versatility.

## WHAT IS RETEST RELIABILITY?

An important aspect of assessment instruments is their stability across time, often called

*"We judge ourselves by what we feel capable of doing,  
while others judge us by what we have already done."*

-HENRY WADSWORTH LONGFELLOW

retest reliability. Retest reliability indicates the likelihood that a person's profile results will remain the same or similar when profiled more than once over time. Multiple factors can affect a person's responses to the same questionnaire when taken more than once. For example, if I'm in a very good mood during the first administration, but in a very bad mood during the second administration a month later, I might respond differently. This type of unreliability is due to the individual.

Reliability can also be affected by environmental factors. I might be in a hurry during the second administration and feel unusually stressed, or there might be loud construction noises from outside my office window that make it difficult for me to concentrate. A more important environmental factor can affect multi-rater profiles: a different group of people might rate a person at two different administrations. This can affect results, though TRACOM's research on inter-rater reliability and agreement indicates that raters tend to evaluate individuals very similarly (see TRACOM's Technical Report for information on these studies).

Unreliability due to the individual and due to the environment is outside of our control and can affect any given individual at any time. The study presented here focuses on the third source of retest reliability, the reliability of the instrument itself. For this type of study it is important to understand that the unit of analysis is not any given individual, but rather a large group of individuals. Statistical research virtually always applies to groups and not individuals. An instrument might demonstrate high reliability, but some individuals will still score differently at different times. A good analogy for this is the use of polls during elections. Based on a sample of only several thousand people, pollsters can predict the outcome of elections for entire nations within a certain level of confidence. But of course these polls don't, or shouldn't, affect how any given individual votes.



## STUDY DESCRIPTION AND RESULTS

To determine retest reliability, TRACOM analyzed data from 814 individuals who were profiled across time on our multi-rater SOCIAL STYLE & Enhanced Versatility Profile. These individuals came from a variety of organizations, multiple occupations, and more than 25 industries. Seven percent of the group was from outside of North America, while the rest were from the U.S. or Canada. The time between administrations ranged from less than one month to more than four years, with an average of 15.6 months.

Reliability was calculated based on individuals' multi-rater scores from co-workers, not on their own self evaluations. We based our analysis on others' ratings because the perception of others is integral to TRACOM's profiles and the lessons we teach in our materials and courses. This research design is unique; in fact, in a literature review we found just one unpublished study that examined personality retest reliability based on other-ratings. Related to this, research has shown that others' perceptions of an individual are not only more accurate than self-perception, but are also better predictors of job performance. In a meta-analysis (an analysis of multiple research studies), researchers found that when personality profiles were based on others' perception, the relationship between personality and job performance was much greater than when personality profiles were based on self-perception. In fact, using just one "other" rater made a significant difference, and the effect was magnified with multiple raters. The authors of this study concluded that the validity of personality for predicting job performance is much greater than previously believed, but this can only be shown when personality is evaluated by others who know the person.

In the current study, we utilized a sample of convenience that included all individuals who had reprofiled in our database; therefore, we had no way to control whether the raters at time two were the same people who rated at time one. It is almost certain that many or most of the raters were different between the two administrations. As mentioned previously, the inability to empirically control for differences in rater groups can increase the amount of statistical "error" in ratings across time periods.

Like other forms of reliability, retest reliability is analyzed using a coefficient statistic. In general, correlations above 0.70 are considered reliable. The table to the right shows the correlations between the two time periods for Assertiveness, Responsiveness, and Versatility. The results show good consistency across time for the two scales that comprise SOCIAL STYLE

SCALE	CORRELATION BETWEEN TIMES 1 AND 2
Assertiveness	0.73
Responsiveness	0.76
Versatility	0.55

– Assertiveness and Responsiveness. The correlation for Versatility is lower, which is to be expected since Versatility is less stable and can change across time and circumstances. In fact, this is one of the central principles of TRACOM’s teachings and programs.

Because the time lapse between administrations varied widely among individuals, we ran partial correlations to statistically control for this effect. A partial correlation “partials out” the effects of a third variable that could be responsible for the initial correlation, ensuring that the correlation between the two variables of interest is accurate and is not due to an uncontrolled variable. In this case, the third variable is the amount of time between the two surveys. Controlling for time lapse did not change the correlations for any of the scales. This means that people who re-profiled years after their first profile were just as likely to maintain consistent scores as people who re-profiled only a few weeks after their first profiles.

## RETEST RELIABILITY OF SIMILAR INSTRUMENTS

To provide a baseline for these results, we reviewed retest reliability studies conducted on other personality and behavioral style measures.

**Myers-Briggs Type Indicator®** — The Myers-Briggs Type Indicator® (MBTI®) is an assessment of psychological type based on Carl Jung’s theory of personality and is sold by CPP, Inc. Its typology is composed of four pairs of opposite preferences, called dichotomies:

- Extraversion (E) or Introversion (I)
- Sensing (S) or Intuition (N)
- Thinking (T) or Feeling (F)
- Judging (J) or Perceiving (P)

In a report released by CPP, retest reliabilities on these four scales for the Form M assessment were calculated for time intervals ranging from less than three weeks to greater than a year. The reliabilities ranged from 0.67 to 0.73 (all time intervals combined).

**DiSC® Model** — The DiSC model of human behavior was developed in the 1920s by William Moulton Marston. The DiSC profile is considered a public domain and many variables exist. The profile measures four dimensions of behavior: Dominance (D), Influence (i), Steadiness (S), and Conscientiousness (C). In a technical report, one-year retest reliabilities ranged from 0.71 to 0.80 on the four scales.

**Big Five Personality Model** — The Big Five personality model is one of the most popular and well-researched personality models in use today. It consists of five personality dimensions: Emotional Stability, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. A meta-analysis of multiple studies that examined retest reliability on the Big Five model found reliability coefficients that ranged from .69 to .76 across the five personality dimensions.

A separate meta-analysis looked at personality trait retest reliability for people of different age groups. This study found that the consistency of personality traits increased from 0.31





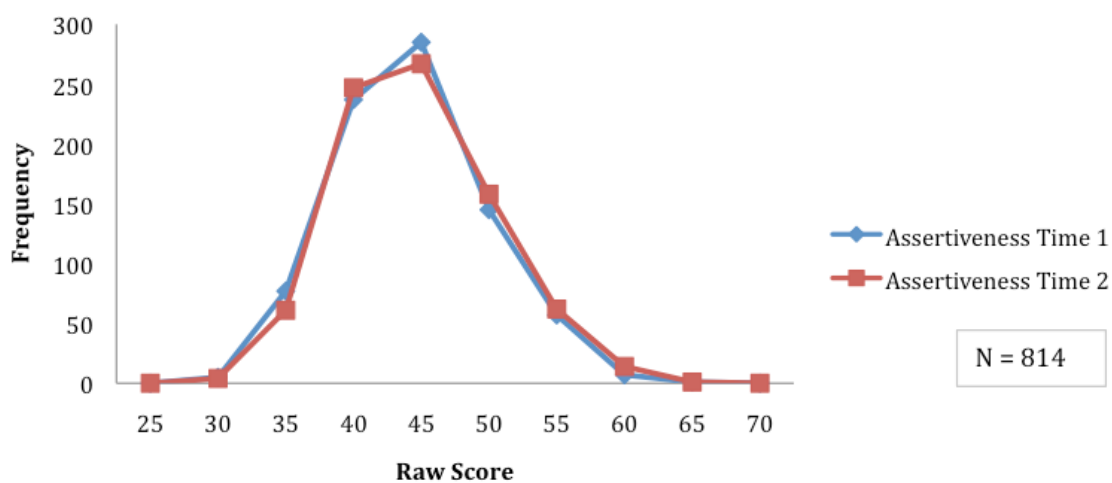
in childhood to .54 during the college years, to 0.64 at age 30, and then reached a plateau around 0.74 between ages 50 and 70. Our research was not able to examine differences across age groups, but our findings are consistent with the highest range of reliability for personality that the meta-analysis found throughout the life span.

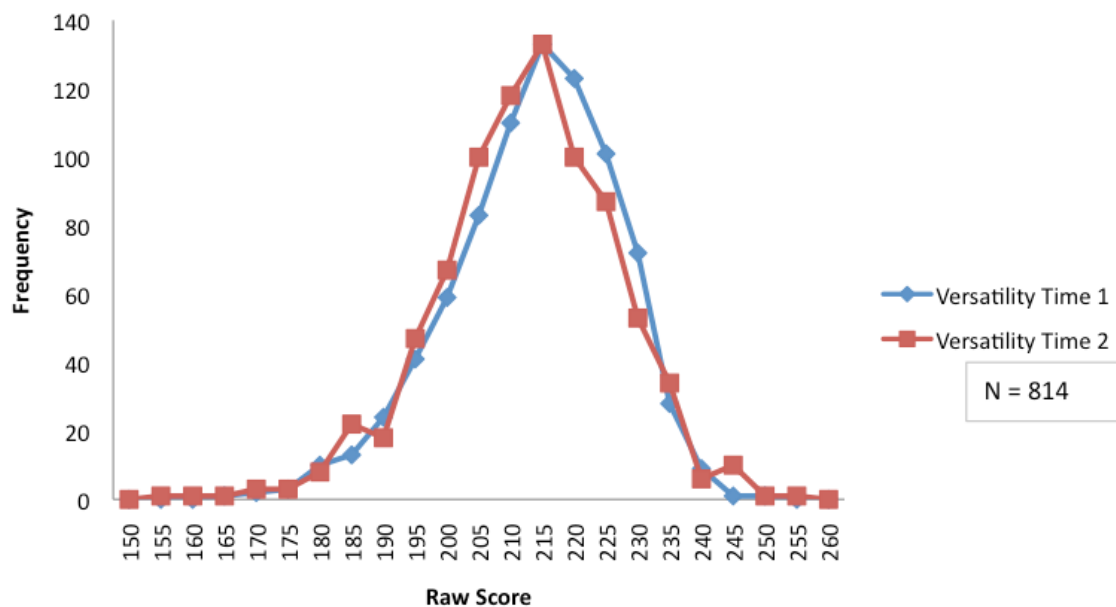
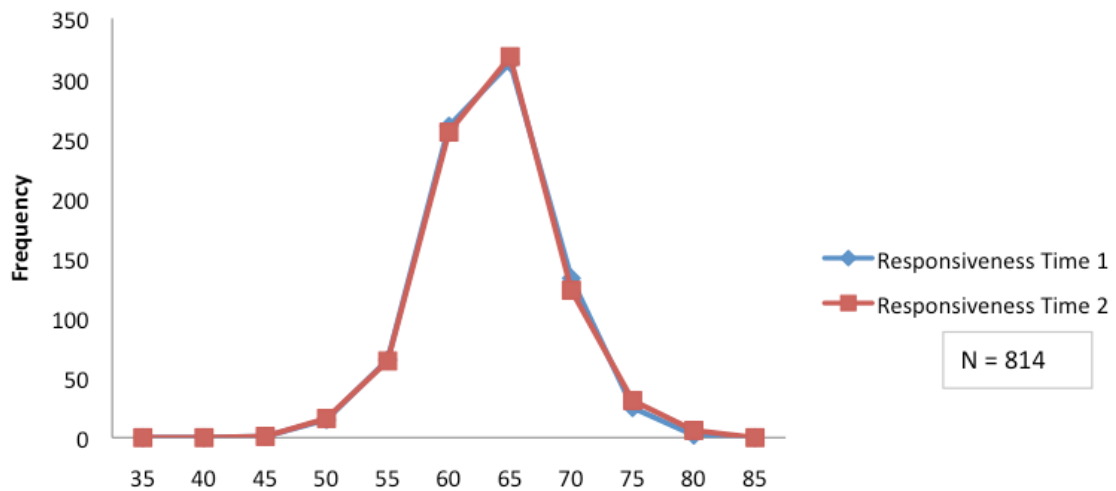
Finally, Connelly studied other-ratings of Big Five personality traits. In a meta-analysis he found that other-ratings of personality are measured at least as reliably as self-ratings. He concluded that for other-ratings to be accurate, however, they must have adequate opportunity to observe the target person. This accuracy is enhanced when other-raters have access to internal aspects of the target person's personality (thoughts, emotions, values, etc.) as a result of interpersonal intimacy.

**Summary of Comparisons with Other Instruments** — This review of personality and behavioral style measures shows that the Multi-Rater SOCIAL STYLE & Enhanced Versatility Profile compares favorably with other personality and behavioral style measures, with the exception of the Versatility scale which was designed to be more transient.

## DISTRIBUTION OF SCORES ACROSS TIME

A helpful way to understand the consistency of scale scores across time is to visually plot scores from the two time periods against one another. Below are frequency distributions for each scale across the two time periods. These graphs show the research group's distribution of scores on each scale for the two time periods. While these graphs do not directly plot each individual's scores across the two time periods, the consistency of the group's scores is clearly visible.





## SUMMARY

This research study shows that TRACOM's Multi-Rater SOCIAL STYLE & Enhanced Versatility Profile has good retest reliability, specifically for the ratings of "others." While any given individual's profile results can change across time due to a variety of reasons, the measure itself has reliability that is comparable or better than other personality and behavioral style measures. Critically, this retest reliability information is based on the ratings of others, typically co-workers, showing that the behavior measured by the Profile is observable to others and remains reliably consistent over time. Other research from TRACOM has established the high degree of reliability that groups of raters have with one another when rating an individual at one point in time (i.e., inter-rater reliability and agreement). Also noteworthy is that Versatility showed lower retest reliability than the other scales. This corroborates the philosophy and design of this scale; Versatility is changeable across time and circumstances, whereas Assertiveness and Responsiveness are more stable.



*About the Author*

**CASEY MULQUEEN, PH.D.**

*Senior Director of Learning & Development*

Casey Mulqueen oversees the research and development of TRACOM's various assessment instruments and products. He has experience developing a wide variety of assessments such as personality inventories, 360-degree feedback programs, performance appraisal systems, and employee engagement programs. His expertise in cross-cultural assessment and norming has helped ensure that TRACOM's global surveys are valid and reliable throughout the world. He is a writer who has authored a variety of materials including books, book chapters, and peer-reviewed journal articles. Casey has an M.S. in clinical psychology and a Ph.D. in industrial/organizational psychology.

# ABOUT **TRACOM<sup>®</sup> GROUP**

---

THE SOCIAL INTELLIGENCE COMPANY<sup>®</sup>

## **[WHY we do]**

We believe that improving peoples' understanding of themselves and others makes the world a better place.

## **[WHAT we do]**

We synthesize our discoveries into actionable learning and resources that improve an individual's performance in all parts of their lives. We call this Social Intelligence.

## **[HOW we do it]**

Through research and experience we uncover the hidden barriers to individuals achieving their maximum potential and identify how to help overcome them.

For more information, visit [WWW.TRACOMCORP.COM](http://WWW.TRACOMCORP.COM)  
or call (303) 470-4900 — (800) 221-2321 (U.S. only)