Reliability & Validity

of the Real Colors® Instrument

Holly O'Rourke, PhD.

Reliability & Validity of the Real Colors® Instrument

INTRODUCTION

The value of a temperament assessment lies in its reliability and validity. Does it measure what it intends to measure and does it do so consistently?

For centuries, people have searched for a convenient way to peek into the human mind. The Greek philosopher, scientist, and physician Hippocrates wrote over twenty-three hundred years ago that all human beings could be divided into four distinct personality types. Since the time of Hippocrates, researchers have continued to build on the body of research furthering the desire to better understand themselves and each other.

It is from the innate desire to understand human behavior and the motivations behind it that the Real Colors® instrument was developed. Since 1993, Real Colors® has sought to be a provider of simple, easy to apply, research-based tools and training that will allow people to better understand themselves and others. Because of our commitment to ongoing research and development, Real Colors® finds it necessary to periodically assess the reliability and validity of its own assessment. When assessing reliability, one must ask, "Does the instrument give consistent results?". And when assessing validity, one must ask, "Does the instrument measure what it intends to measure?". On both these criteria, the Real Colors® instrument performs well.

TYPES OF RELIABILITY AND VALIDITY

Reliability and validity are two ways researchers can evaluate the quality of assessment tools. They indicate how well a method, technique, or test measures what it intends to measure. Reliability provides information on the consistency of a measure, and validity provides information about the accuracy of a measure.

Reliability definition. Reliability is the extent to which results can be reproduced when research is repeated under the same conditions. It is also known as consistency. The more consistent results are when achieved under the same conditions, the higher the reliability of a test. Two important types of reliability are individual test reliability and test-retest reliability.

> **Individual test reliability.** Individual test reliability (also known as internal consistency) provides information on the consistency of items within a single test. That is, it is a measure of how closely related a set of items are as a group within a test. Individual test reliability is measured using a statistic called Cronbach's Alpha. If this statistic Alpha is larger than .7, this indicates high individual test reliability.

Reliability & Validity of the Real Colors® Instrument

Types of Reliability and Validity Continued

Test–retest reliability. Test-retest reliability provides information on how consistent the same test is over time. This is important for getting results to replicate with a single test under similar conditions. This can be measured by administering the same test twice over a period of time to the same group of individuals. The scores from Time 1 and Time 2 can then be correlated in order to evaluate the test for test–retest reliability using a correlation statistic. A correlation larger than .7 with a statistical significance of $p < .05$ is evidence of test–retest reliability.

Validity definition. Validity is the extent to which a test that is designed to measure a specific concept accurately measures that concept. It is the extent to which results of a test truly measure what they are supposed to measure. This study focuses on one type of validity called convergent validity.

Convergent validity. Convergent validity measures the degree to which two distinct but related theories agree with each other, or how they converge. Our study was designed to measure the convergent validity of Real Colors® with respect to aspects of the Myers-Briggs Type Indicator (MBTI), specifically how each color relates to specific aspects of the MBTI. Convergent validity is measured with correlations between items that are thought to be related. Significant correlations ($p < .05$) between items that are hypothesized to be related are evidence of convergent validity.

The next section describes this study's findings for these types of reliability and validity.

RESEARCH FINDINGS

Study Design

The Real Colors® instrument is a personality and temperament tool similar to the Myers–Briggs Type Indicator (MBTI). In order to ensure that Real Colors® is a high-quality instrument – meaning it is both reliable and valid – a study with exploratory analyses was conducted to assess reliability and validity of the instrument.

For this study, it was hypothesized the following categories from the Real Colors® instrument correlate to the MBTI as follows:

- Individuals who scored as a Primary **GOLD** on the Real Colors® instrument should score as SJ's (Sensing/Judging) on MBTI: ESTJ, ISTJ, ESFJ, and ISFJ.

- Individuals who scored as a Primary **BLUE** on the Real Colors® instrument should score as NF's (Intuiting/Feeling) on MBTI: ENFJ, INFJ, ENFP, and INFP.

Reliability & Validity of the Real Colors® Instrument

Research Findings Continued

- Individuals who scored as a Primary **GREEN** on the Real Colors® instrument should score as NT's (Intuiting/Thinking) on MBTI: ENTJ, INTJ, ENTP, and INTP.

- Individuals who scored as a Primary **ORANGE** on the Real Colors® instrument should score as SP's (Sensing /Perceiving) on MBTI: ESTP, ISTP, ESFP, and ISFP.

From June to October of 2019, data was collected from participants to test for reliability and validity of the Real Colors® instrument. Eighty-eight participants were enrolled to take paper versions of Real Colors® and MBTI from seven sites across the United States. Of the eighty-eight participants, eighty-four participants completed the initial Real Colors® instrument (RC Time 1) and MBTI. To achieve reliability, seventy-three of the participants who took the initial Real Colors® instrument returned thirty days later to take an identical version of the Real Colors® instrument (RC Time 2). The environments were controlled to be as similar as possible. The study participants were diverse in age, gender, race and ethnicity, as well as educational level.

Individual test reliability.

Individual test reliability was measured by grouping together item responses by color from the Real Colors® survey instrument and using Cronbach's Alpha coefficients to determine reliability. A Cronbach's Alpha of larger than .7 indicated acceptable individual test reliability. Table 1 shows Cronbach's Alpha values for each color from the Real Colors® survey instrument at RC Time 1. The Cronbach's Alpha statistics ranged from .854-.888, indicating high individual test reliability for all colors at RC Time 1.

Table 1			
Color	**Sample Size**	**Questions on RC (1–10)**	**Cronbach's Alpha**
ORANGE	81	a	.885
BLUE	84	c	.870
GOLD	84	d	.854
GREEN	81	b	.888

Real Colors®

Reliability & Validity of the Real Colors® Instrument

Research Findings Continued

Table 2 shows Cronbach's Alpha values for each color from the Real Colors® survey instrument at RC Time 2. The Cronbach's Alpha statistics ranged from .857–.890, indicating high individual test reliability for all colors at RC Time 2.

Table 2			
Color	**Sample Size**	**Questions on RC (1–10)**	**Cronbach's Alpha**
ORANGE	69	a	.879
BLUE	72	c	.890
GOLD	72	d	.876
GREEN	69	b	.857

Test–retest reliability.

Test–retest reliability was measured by examining significance tests of correlations between each Real Colors® overall color score (determined by summing scores on picture, card, and survey portions of the instrument) measured at RC Time 1 and RC Time 2. A correlation larger than .7 with a statistical significance of $p < .05$ indicated acceptable test–retest reliability. Table 3 shows correlations between time points for each overall color score from the Real Colors® instrument. The correlations ranged from .861–.924 (well above the threshold of .7 for acceptable reliability) and p values were all less than .05, indicating high test–retest reliability for all colors.

Table 3			
Color	**Sample Size**	**Correlation Between T1 & T2**	***p***
ORANGE	71	.896	<.001
BLUE	73	.861	<.001
GOLD	73	.924	<.001
GREEN	70	.898	<.001

Reliability & Validity of the Real Colors® Instrument

Research Findings Continued

Convergent validity.

Convergent validity was measured by examining significance tests of correlations between each Real Colors® overall score (determined by summing scores on picture, card, and survey portions of the instrument) measured at RC Time 1, and each MBTI score corresponding to Real Colors® colors also measured at RC Time 1. Significant correlations with a statistical significance of $p < .05$ indicated acceptable convergent validity. Table 4 shows correlations between RC Time 1 overall color scores and RC Time 1 MBTI corresponding scores. The correlations ranged from .369–.801 with $p < .05$ for all correlations, indicating acceptable convergent validity for all colors. It should be noted that the correlation between Real Colors® Primary **ORANGE** total score and MBTI SP total score was lower than other correlations; correlations ranged from .675–.801 for other colors. This may be due to characteristics of the sample collected for this study. Additional research is being conducted to determine if this pair is consistently correlated lower than the other Real Colors®/MBTI scale pairs.

Table 4				
Color	**MBTI**	**Sample Size**	**Correlation Between RC & MBTI**	**p**
ORANGE	SP	79	.369	<.001
BLUE	NF	83	.685	<.001
GOLD	SJ	83	.801	<.001
GREEN	NT	80	.675	<.001

SUMMARY

This study was undertaken to investigate the reliability and validity of the Real Colors® instrument. Results indicated that the Real Colors® instrument has high individual test reliability and high test–retest reliability for all colors. Using the MBTI instrument as a basis for validity, the Real Colors® instrument has acceptable validity for all colors as well. Although our sample showed a weaker relationship for Primary **ORANGE**/SP MBTI pair than for the other colors and their corresponding MBTI pairs, the convergent validity correlations are still medium–large by conventional standards for behavioral data and well within the acceptable range to indicate convergent validity (Cohen, 1988).

Real Colors®

Reliability & Validity of the Real Colors® Instrument

REFERENCES

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

ABOUT THE RESEARCHER

Holly O'Rourke is an assistant professor with the Measurement and Statistical Analysis (MASA) group at the T. Denny Sanford School of Social and Family Dynamics at Arizona State University. She received her PhD in Quantitative Psychology from Arizona State University in 2016, and completed a postdoctoral fellowship at the REACH Institute before entering her current position. Her research focuses on the intersection of quantitative methodology and prevention research. Her statistical work is grounded in real-world questions and issues in prevention, particularly alcohol and substance use. She is currently engaged in research that draws on several research areas, investigating longitudinal mediation models that are most appropriate for alcohol use outcomes in treatment and prevention studies.