

Technical Supplement for  
“Renewing the Field’s Understanding of Personality Data”

*Author names de-identified during the editorial review of the submitted draft article*

**Please Note.** This Technical Supplement is designed to accompany the draft article “Renewing the Field’s Understanding of Personality Data”. For purposes of cross-reference, brief sections of the original article are sometimes included verbatim in the Supplement, but with additional material added before, after, or interspersed.

**Table of Contents**

Introduction to the Technical Supplement..... 3

Supplement to “Prologue: Proliferation of Data Types” (*Renewing*, p. 4)..... 4

Supplement to “Challenge Two: Different Measures Should Converge but They Don’t”  
 (*Renewing*, p. 9) ..... 6

    Excerpt Re. Table 1 from Campbell & Fiske, 1959 (p. 10)..... 6

    Supplement to “After 60 Years, Do Different Measures of the Same Attributes Still Fail to  
 Converge?” (*Renewing*, p. 11) ..... 8

Supplement to “Understanding Why Different Types of Data are Distinct” (*Renewing*, p. 16) .. 10

    Fiske’s First Clue: Different Response Processes (p. 16)..... 10

    Fiske’s Second Clue: The True Score Issue (p. 18) ..... 10

    Further Rationale for the Lack of Convergence (p. 20)..... 11

Supplement to “Integration of Well-Selected Methods” (*Renewing*, p. 27)..... 13

End of Technical Supplement..... 16

### Introduction to the Technical Supplement

This Technical Supplement is intended to accompany the draft article submission *Renewing the Field's Understanding of Personality Data* (henceforth: *Renewing*). Each section of the Supplement is keyed to a heading in *Renewing* itself, and provides additional detail. Note that the Supplement provides additional detail only for some of the *Renewing* sections, not all. The *Table of Contents* that follow also includes this information. Readers may find the most useful supplementary material(s) include:

- a. A table of definitions for the data types discussed throughout the draft article submission.
- b. Calculations we employed to obtain the average convergence correlation across methods in the Campbell and Fiske article
- c. Additional details as to the meta-analyses we reviewed, including the authors, types of data employed, and range of correlations reported, and
- d. More notes on understanding the lack of convergence seen among data

**Supplement to “Prologue: Proliferation of Data Types” (*Renewing*, p. 4)**

**Prologue: Proliferation of Types of Data** discusses the introduction of key types of personality data from the beginnings of scientific psychology until today. That section of *Renewing* provides brief definitions of many data types. Supplement Table 1 indicates the working definitions of many types of data that we employed as we wrote the section. These are arranged in alphabetical order.

Supplement Table 1

A Selected List of Commonly Used Data Types and their Definitions.

<b>Data Type</b>	<b>Other Names</b>	<b>Definition</b>
Act Frequency Data	Act frequency approach	Data that explores the frequency of individual’s behaviors over time to draw conclusions about personality.
Big Data		Currently, data that is characterized by its large size, complexity, and the requirement for special analytical and measurement approaches.
Biographical Data	Biographical data Personal history data L-data (life history) Biodata	Data collected on individual’s personal history or experiences used to predict future behavior.
Implicit Attitude Data		Data that arises from assessments aimed at accessing non-conscious attitudes that influence an individuals’ behavior.
Lifespace Data	Life space Life-space	Data that emerges from the objective observation of individual’s identity, behavior, and the environment, or all the systems that surround and influence an individual’s behavior at a given time. Can be considered a combination of act frequency and biodata.
Mental Ability Data	Ability data Cognitive ability data	Data that assessing individuals cognitive and mental capacities through the use of assessments where they solve specific tasks or problems.
Observer-Judgment Data		Data that tasks individuals with observing and estimating the personality attributes of others.
Performance Data		Data that assesses an individual’s ability to perform specific tasks that assess cognitive abilities. Differs from ability data in that is does not require the use or understanding of language.
Physiological Data		Data collected on bodily functions such as the activity of the nervous system or endocrine system that is then related to psychological phenomenon.
Projective Data	Thematic data	Data that aims to assess projections of non-conscious mental processes that emerge in the attributes and characteristics individual’s assign to ambiguous stimuli.

(Table continues on next page)

Supplement Table 1 (Continued)

A Selected List of Commonly Used Data Types and their Definitions.

<b>Data Type</b>	<b>Other Names</b>	<b>Definition</b>
Self-Report Data	Self-rating Verbal report	Data that tasks individuals with looking inward and reporting information about themselves (i.e., on their own feelings, attitudes, and behaviors).
Self-Judgement Data		Data that emerges from individual's judgement or estimates of their personality attributes.
Think-Aloud Protocols	Protocol analysis	Data that emerges from assessment techniques where individuals are tasked with verbalizing their thoughts while performing specific tasks.

**Supplement to “Challenge Two: Different Measures Should Converge but They Don’t”**

*(Renewing, p. 9)*

**Excerpt Re. Table 1 from Campbell & Fiske, 1959 (p. 10)**

In the article, we remark that Campbell and Fiske set the expectation that the correlation for methods of the same attribute to converge would be  $r = .60$ . This is based on the authors’ Table 1, which contained correlations for three different methods, each applied in turn to traits A, B, and C. The average correlation between methods for the same trait (six correlations) was roughly  $r = .60$ . The actual Table 1 values and exact calculation are:

Supplement Table 2

Cross-Method Correlations for the Same Traits from Table 1 of Campbell & Fiske (1959)
0.57
0.57
0.46
0.67
0.66
0.58
0.56
0.58
0.45
<i>Mean = 0.57</i>

We further aimed to understand the range and mean of the *obtained* validity coefficients across methods for the same personality attribute reviewed by Fiske. That is, what was the average convergence across methods used to assess the same trait? The reported cross-method correlations appeared in Tables 2 through 13 of Campbell & Fiske (1959). We have extracted those from the 12 Tables and copied them into our own Supplement Table 3 below. Overall, the correlations between different methods ranged from as low as  $r = -.10$  to as high as  $.82$ . The average correlation between methods was  $r = .35$ .

Supplement Table 3.

Reported Cross Method Correlations Obtained from Campbell &amp; Fiske's 1959 Review of Data Convergence.

TABLES 1 TO 7		TABLES 8 TO 11		TABLES 12 AND 13	
<i>Table No.</i>	<i>Cross-method rs</i>	<i>Table No.</i>	<i>Cross-Method rs</i>	<i>Table No.</i>	<i>Cross-method rs</i>
Table 2	0.13	Table 8	0.43	Table 12	0.71
	0.12		0.32		0.53
	0.10		0.60		0.43
	0.16		0.74		0.20
Table 3	0.48		0.59		0.47
	0.33		0.20		0.46
	0.44		0.21		0.24
Table 4	0.30		0.13		0.31
	0.38		0.35		0.14
	0.23		0.19		0.35
	0.58		0.17		0.48
	0.31		0.19		0.42
	0.48		0.15		0.22
Table 5	0.31		0.33		0.22
	0.38		0.23		0.26
	0.47	Table 9	0.46	Table 13	0.30
Table 6	0.40		0.40		0.34
	-0.10	Table 10	0.61		0.71
	0.23		0.51		0.53
Table 7	0.19	Table 11	0.57		0.43
	0.08		0.65		0.49
	0.26		0.53		0.20
	0.15		0.32		0.47
	0.66		0.34		0.26
	0.76		0.42		0.30
	0.24		0.39		0.82
	0.12		0.29		0.45
	0.18		0.30		0.45
	0.05		0.56		0.44
	0.66		0.37		0.32
	0.76		0.32		0.41
			0.27		0.26
			0.26		0.43
			0.45		0.42
					0.40
					0.39
<i>Overall Mean and Median Values of Cross Method Correlations</i>					
	Mean	Median	Minimum	Maximum	
	0.35	0.37	-0.10	0.82	

**Supplement to “After 60 Years, Do Different Measures of the Same Attributes Still Fail to Converge?” (*Renewing*, p. 11)**

[For convenience, the following paragraph of the Supplement is copied from *Renewing*, with reference to a further table below]. Looking at six meta-analyses examining the convergence of methods, published between 1992 and the present, the conclusions are resoundingly in keeping with the historical record: self-judgment and lifespace data converge with ability-measured intelligence roughly  $r = .20$  on average (Goff & Ackerman, 1992), cross-informant symptom reports used to assess whether adolescents exhibit internalizing or externalizing symptoms converge roughly  $r = .29$  (De Los Reyes et al., 2015). Self-report and the Implicit Association Test over 126 studies averaged  $r = .24$  (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005), and the association across 24 studies of the Rorschach Mutuality of Autonomy Scale, which measures cooperative versus antagonistic ways of viewing the world averaged  $r = .24$  with a number of scales measuring related constructs (Graceffo, Mihura, & Meyer, 2014). The correlation between self-reported big five and peer observers over 341 studies reviewed yielded a range of obtained  $r$ s for the five factors from  $r = .27$  to  $.41$  ( $N = 3181$  and higher), with strangers’ ratings exhibiting lower convergence and close relatives’ higher convergence (Connolly, Kavanagh, & Viswesvaran, 2007, p. 113, Tables 3 and 4). Finally, a meta-analysis of the “digital footprints” collected from social media, drawing on 14 papers and over 12,000 participants indicated a range from  $r = .29$  to  $.40$  (Azucar, Marengo, & Settanni, 2018).

Supplement Table 4.

Results from Review of Meta-Analyses on Correlations of Cross-Method Assessments

Citation	Brief description	Range of <i>rs</i>
Goff & Ackerman, 1992	Nine areas of attributes including intellectual engagement, hard work items, perfectionism openness absorption, distractibility, extraversion-introversion, energy level; these all were correlated with mental ability measures (p. 539)	$R = -.17$ to $.22$ , with some <i>rs</i> reaching $.42$ —but after adjusting for unreliability.
Azucar, Marengo & Settanni, 2018	Meta-analysis of “digital footprints” (p. 150) collected from social media over Big 5 personality traits. 14 papers, 16 selected studies of Facebook, Twitter, Sina Weibo microblog and Instagram together with Twitter. <i>N</i> varied from 12,210 to 6053 depending on the member of the Big Five	$R = .39$ for openness, $.40$ for conscientiousness, $.40$ for extraversion, $.29$ for agreeableness and $.33$ for neuroticism
De Los Reyes, Augenstein, Wang, Thomas, Drabick, Burgers, and Rabinowitz, (2015), p. 856	Examined 341 studies that reported cross-informant correspondence estimates for internalizing and externalizing psychiatric symptoms of $r = .28$ and $.30$ , respectively	for internalizing and externalizing psychiatric symptoms of $r = .28$ and $.30$ , respectively
Hofmann, Gawronski, Gschwedner, Le, Schmitt (2005)	Examined 126 studies between self-report and the Implicit Association Test (p. 1369)	average effect size of $r = .24$ .
Graceffo, Mihura, Meyer, 2014	Examined 24 samples with an <i>N</i> of 1801 and 91 effect sizes between the Rorschach Mutuality of Autonomy Scale, a measure of the degree to which test-takers view people as autonomous but cooperative versus pathologically destructive, correlated with a variety of other methods including observer-ratings, other clinical measures of the same or similar constructs such as the Thematic Apperception Test and the SCORS, ratings on DSM behavioral-rating scales, and self-report.	the average effect size was $r = .24$ .
Connolly, Kavanaugh & Viswesvaran (2011)	Reviewed 36 studies comparing observer ratings and self-ratings on the big five traits with overall <i>Ns</i> ranging upwards from 5333, depending on the trait.	The obtained <i>rs</i> for the five factors ranged from $r = -.27$ to $.41$ ( $N = 3181$ and higher) for peers, with strangers’ ratings lower and close relatives’ higher.

**Supplement to “Understanding Why Different Types of Data are Distinct” (*Renewing*, p. 16)**

Why, then, the failure? Two clues are offered by Fiske.

**Fiske’s First Clue: Different Response Processes (p. 16)**

The first clue offered by Fiske was a prescient comment that anticipated the introduction today of what goes by the name of “Evidence for Validity Due to Response Processes.” Fiske opined that “Different methods for measuring any construct involve different psychological processes...such as self-reports versus peer ratings...Different items can elicit separate kinds of retrieving and cognizing processes.” (Fiske, 1982, p. 83). And five years later, Fiske (1987, p. 190) opined “We do not know what goes on in the person producing our data. We have little more than our introspections to guide us, and we have done little research on the problem.”

**Fiske’s Second Clue: The True Score Issue (p. 18)**

Both Campbell and Fiske were highly influenced by work in classical test theory of the time, as well as the work on validating tests by Cronbach and Meehl—as were (and are) most psychologists who studied measurement at the time. Fiske (1987, p. 193) later wrote that individual difference and personality psychologists “implicitly assume that there is a correct (true?) value for every person on each trait or disposition;” if so, he implied, their model of convergence across methods *must* be correct. At this stage, twenty-seven years after the first article, and post-Mischel’s (1968) “Personality and Assessment,” Fiske does, however, raise the question of whether true scores might change under some circumstances.

We believe true scores often are stable, but we also believe the field has moved to a position for which it has become accepted that there often are multiple true values (i.e., true scores) for a given trait, depending upon the mental (or digital) process that has produced it.

More specifically, our treatment assumes that true scores for a personality attribute often are different depending upon whether it is measured via a person’s self-judgment or through implicit methods, and also depending upon whether the attribute is measured “inside the person” versus whether it is expressed.

### **Further Rationale for the Lack of Convergence (p. 20)**

To a degree that may not be fully explicit, but is nearly so, this acknowledgment of multiple true scores is hinted at in the recent *Standards of Educational and Psychological Testing*. In both the 1999 and then the 2014 edition of the *Standards of Educational and Psychological Testing* (Joint\_Committee, 2014), the authoring committee introduces and legitimizes under their discussion of evidence for validity, the issue of “response processes”.

They write:

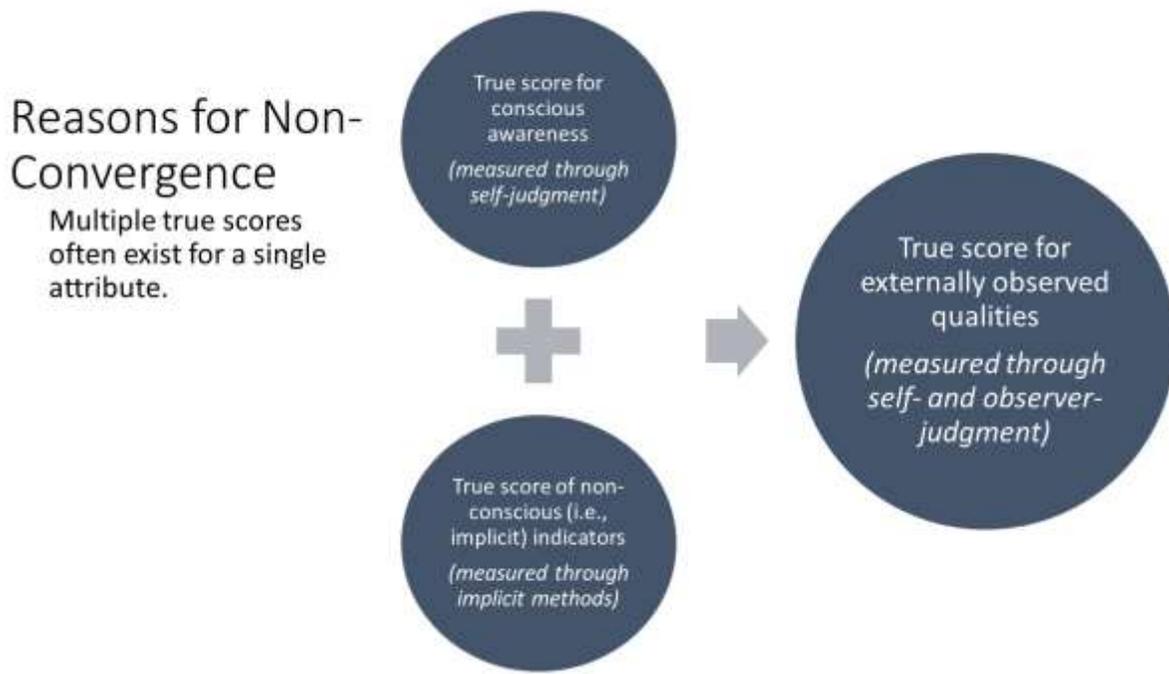
...construct interpretations involve more or less explicit assumptions about the cognitive processes engaged in by test takers. Theoretical and empirical analyses of the response processes of the test takers can provide evidence concerning the fit between the construct and the detailed nature of the performance or response actually engaged in by test takers. (Joint\_Committee, 2014, p. 15)

In more everyday terms: If you want to understand your measurement, you need to know how the person (or measurement instrument) came up with the data you are using. If the measurement process is appropriate to assessing the attribute, you are in good shape; otherwise—watch out!

Different assessment methods represent different information sources: perhaps a person’s self-concept in one instance, a person’s performance, in a second instance, and an independent perceiver’s evaluation, in a third. We would be naïve to believe those would converge—and yet all the measurement methods may all be informative...*about different matters related to the*

*criterion attribute being assessed.* That, we argue, is key to integrating this information that stems from different kinds of data. In essence, different true scores exist for one’s self-conceived attribute, one’s implicit—that is non-conscious processes—related to the same attribute, and for how the attribute is expressed.

This is illustrated in a schematic fashion in Supplement Figure 1.



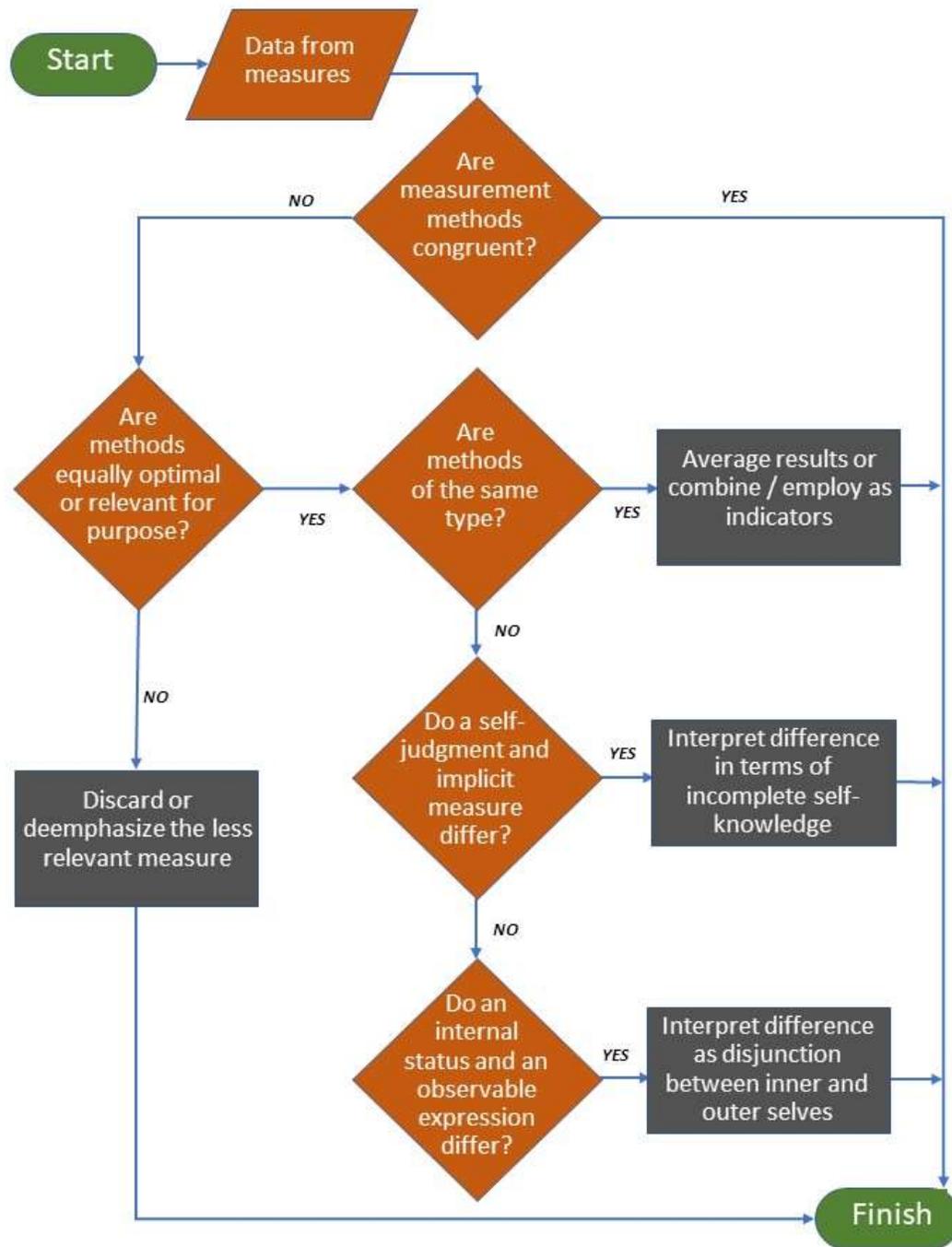
Supplement Figure 1. Personality Attributes May be Manifest in Multiple True Scores. One of our suggestions is that personality attributes—and their true scores—may differ depending upon the mental (or other) process by which they are measured—because, indeed, multiple true scores exist for the same attribute. For example, the true score for one’s self-judged sociability may differ from non-conscious (i.e., implicit) indicators of the same attribute. Similarly, the true score for the same self-judgment may differ from one’s externally-observed qualities. These ideas are indicated in Supplement Figure 1.

**Supplement to “Integration of Well-Selected Methods” (*Renewing*, p. 27)**

The section “Integration of Well-Selected Methods” indicates several guidelines for combining data of different types. Here, we attempt a schematic yet more systematic treatment of the matter using a provisional decision tree, illustrated in Supplement Figure 2, to represent a decision process. The process begins (top left) with collecting different methods. The results from the methods are then evaluated according to whether or not they converge (In practice, this may involve statistical tests utilizing confidence intervals around test scores, etc.).

If the data do not readily converge (central column), the discrepancy may be between different sets of true scores—for example, between self-judgment and implicit measures, or between measures of inner process and outer expressions. Depending upon the specific type of discrepancy, different classes of interpretation are possible.

We believe that such integrations often are made by personality and clinical psychologists dealing with measures of the “same” attribute that indicate different levels of the quality. Figure 2 attempts to capture the commonly applied logic of such interpretations. For example, if a self-judgment and an implicit measure differ, the discrepancy is often attributed to a lack of self-knowledge; if an internal state and observed behavior differ, the discrepancy is often attributed to a disjunction between the individual’s inner self and how the individual personality is expressed be that, again, due to a lack of self-knowledge, or due to an attempt to behave in a socially appropriate (e.g., tactful) manner, or to deceive others.



Supplement Figure 2. A Schematic Overview of Key Types of Integration. Psychologists often interpret instances when different measures diverge. The “Schematic Overview” figure systematizes some of the most common integrations employed in the field. Such integrations often are made by personality and clinical psychologists dealing with measures of the “same” attribute that indicate different levels of the quality. Building on the earlier conception of

multiple true scores for the same attribute, Figure 2 attempts to capture the commonly applied logic of such interpretations. For example, if a self-judgment and an implicit measure differ, the discrepancy is often attributed to a lack of self-knowledge; if an internal state and observed behavior differ, the discrepancy is often attributed to a disjunction between the individual's inner self and how the individual personality is expressed.

## End of Technical Supplement