# Measuring Personality Constructs: The Advantages and Disadvantages of Self-Reports, Informant Reports and Behavioural Assessments

Enquire 1(1): 75-94



Jennifer Dodorico McDonald, University of Cambridge

## Abstract

*Achieving construct validity, or using measures that accurately represent particular theoretical constructs, is an important goal of social science research. This article reviews arguments regarding the strengths and limitations of personality assessment methods in terms of methodological feasibility and whether they are accurate in measuring personality constructs. Specifically, it considers the advantages and disadvantages of assessing personality with self-report questionnaires, informant reports, and behavioural measures. Acknowledging that no method is perfect, the discussion then focuses on the value of incorporating multiple methods. In light of the reliance on the self-report method in personality psychology, it will be argued that researchers could maximize the validity of the measurement of personality constructs by combining the questionnaire approach with other methods.*

## Introduction

According to Cronbach and Meehl (1955), psychological and personality constructs are “postulated” or inferred characteristics or traits of a person. There are many constructs, or concepts, in psychology that are not tangible; if we cannot physically see personality traits, for instance, there is the potential to question whether they are really there. Is it possible to measure such an abstraction as ‘Conscientiousness’ in the way that physical attributes, like height, are assessed? Social scientists have been making inferences about what people are like and thus measuring these sorts of hypothetical constructs for years, which is necessary in order to more fully understand how people behave (Smith, 2005). Smith points out, though, that it is important to ensure that these theoretical constructs are measured “in a convincing, valid way” (p. 396). In general, validity of a measurement device refers to the extent to which it actually measures what it intends to measure. Construct validity, then, refers to the accuracy of a measurement of the theoretical concept (e.g. John & Soto, 2007; Messick, 1995; Ozer, 1999). Therefore, abstract personality constructs can only be accepted and made more concrete if the validity of the methods used to measure them can be ascertained.

This paper will focus on evaluating the usefulness and accuracy of different methods of measuring personality constructs. According to Pervin (1999), the determining of the best methods of measuring personality constructs is one of the most prominent issues in personality psychology. The central debate involves the accuracy of self-report data in obtaining information about an individual's personality, in comparison with ratings from others or the use of alternative methods of assessment. This is especially important in light of recent questionnaire findings that suggest that personality psychologists primarily rely on self-report measures, yet belong to a methodologically diverse field overall (Robins, Tracy, & Sherman, 2007). Here, I will discuss the value of considering self-reports, informant or observer reports, and behavioural measures, by presenting the advantages and disadvantages of each of these methods. I will assert that it is not sufficient to simply assume the accuracy of any one measure in fully representing the desired personality trait. I will demonstrate this by briefly addressing issues related to construct validity. Overall, this paper will suggest that a variety of methods should be employed in assessing personality constructs.

## Self-Reports

### *Advantages and disadvantages of self-reports*

In order to later propose that more than one method is required to obtain more accurate understandings of personality constructs, I will discuss the value of the most common measure: self-reports. Using objective self-reports, or asking people directly for information relating to a particular construct, is extremely prevalent in most areas of the social sciences, including personality psychology (Schwarz, 1999). In the field of personality psychology, asking people to respond to questions or statements about what they are like or how they behave seems to be the most preferred method. According to an analysis conducted by Vazire (2006), 98% of the studies assessing personality traits published in the Journal of Research in Personality in 2003 used self-report measures. Moreover, in 70% of these studies, the self-report was the only measure that was used. More than 95% of studies reported in the Journal of Personality in 2006 used self-report questionnaires (Kagan, 2007). Research by Robins and colleagues (2007) similarly found that, though a variety of methods are accepted by the personality psychologists that they polled, self-reports are "by far" the most frequently used (p. 677). Clearly, the questionnaire is perceived as central to measuring constructs.

On the surface, the fact that obtaining self-reported data is so popular makes complete sense – if I want to learn more about somebody, why would I

not go directly to that person? One would expect that the individual possessing the particular personality traits should be able to provide the most informative and accurate information about these constructs. In accordance with the basic foundation of such models as the five-factor theory of personality, people can convey a vast amount of information about themselves through the expression of certain "relatively enduring patterns of thoughts, feelings, and actions" (McCrae & Costa, 1999, p. 140). Correspondingly, these 'Big Five' traits (namely Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness) can be directly assessed through carefully constructed personality questionnaires, such as John and Srivastava's (1999) Big Five Inventory or Costa and McCrae's (1992) NEO-PIR, which measures the five factors as well as more specific facets of the traits. Paulhus and Vazire (2007) remind us that "no one else has access to more information" than oneself, and that this information is rich with motivational and other introspective details that others might not be aware of (p. 227). Self-reported questionnaires are also advantageous in that the respondents are likely to be more motivated to talk about themselves than others, and they identify with the questions in ways that others do not (Paulhus & Vazire, 2007). It seems that the most accurate information is that which comes straight from the horse's mouth, so to speak. Therefore, it appears that one valid way to shed light on the personality traits that an individual has is to measure them through self-reports.

According to Kline (1993), administering personality inventories directly to the person is advantageous because scoring the results is very straightforward. However, marked evidence of the validity and the reliability, or consistency, of the actual assessment device must first be established, which does require a lot of effort (John & Benet-Martinez, 2000). In addition to being easy to interpret, self-reports are also used because they are an inexpensive and relatively quick way to collect a lot of data (Kline, 1993). Paulhus and Vazire (2007) emphasize the practicality and efficiency of self-report measures in obtaining data from a large number of participants, even at one time, such as in a university lecture hall or even via the Internet. Though the pros and cons of obtaining data using online survey tools or websites is another topic altogether, internet data collection is becoming increasingly popular and research has shown that some of the fears about the method (e.g. lack of diverse samples) might just be myths (e.g. Gosling, Vazire, Srivastava, & John, 2005; Sue & Ritter, 2007). In terms of quickness of administration, there are even some questionnaires (e.g., the Ten-Item Personality Inventory) that have the potential to assess personality traits in as little as one minute, although limitations of measures this short have been acknowledged (Gosling, Rentfrow, & Swann, 2003). Nevertheless,

convenience and ease are vital for researchers who want to have a high number of cases in their study to improve on the statistical strength of their results (Westen & Rosenthal, 2005). There are clearly practical and meaningful reasons for using self-report measures; however, I will go on to suggest that are also a number of reasons to be cautious of using them, especially when on their own.

Though there are many strengths of using self-reports to measure psychological constructs, there are also potentially a number of weaknesses. First, the structure of the questions affects whether the reported information accurately measures the construct under consideration. According to Schwarz (1999), "self-reports are a fallible source of data, and minor changes in question wording, question format, or question context can result in major changes in the obtained results" (p. 93). There are many potential problems with errors on the part of the respondent. Moskowitz (1986) recognized that self-reports leave a lot of room for 'response biases', which according to Paulhus (1991), involve "a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content (i.e., what the items were designed to measure)" (p. 17). For example, people often respond in such a way that presents them in a more favourable light, even if these responses do not reflect how they actually think or behave ('Socially Desirable Responding'; Paulhus, 1991). 'Acquiescent Responding', in which individuals agree with responses without considering what the question is asking, and 'Extreme Responding', or giving extreme ratings on scales, are other common response tendencies (Paulhus & Vazire, 2007). Lack of credibility due to biased responding is a major issue because it could impede the validity of the self-report as a measure. Although this is a concern, there are actions that can be taken to try to reduce these biases in terms of improving on the questionnaire construction and instructions given to the participants (Moskowitz, 1986) For example, balancing the scoring key of the questionnaire when constructing the measure can help to reduce the effects of the acquiescence bias (Paulhus & Vazire, 2007). Additionally, response biases can be measured and controlled for with specially-designed scales. For example, the MMPI Lie Scale presents attitudinal and behavioural statements that are common, yet not favourable to admit (e.g. 'I get angry sometimes'). If a particular number of statements are selected as 'false', this shows that the respondent is engaging in socially desirable responding (Paulhus, 1991).

Another related concern regarding credibility of respondents is that individuals do not just respond in a socially desirable manner because they want to present themselves in a certain way. Instead, there is a theory that we respond more positively because we have a distorted outlook about who

we are (John & Robins, 1994). In this view, it is the self-perceptions themselves that provoke the biased responding in the individual because people are predisposed toward self-enhancement, or trying to maintain positivity about the self at the expense of being unrealistic (Fiske & Taylor, 1999). This causes different problems compared with when one consciously engages in impression management; if there is a distortion in how people perceive themselves beyond the typical response biases like social desirability, it cannot be corrected for through lie scales or other such controls.

An intriguing query and potential limitation is whether people know enough about themselves to be able to accurately portray what the self-report is attempting to determine. In terms of personality constructs, are we actually capable of reporting on our own personality or do we have a lack of selfknowledge? Can we accurately provide information about what we are like? If not, there would be serious repercussions for the methodologies employed in attempting to measure these constructs. Self-reports are built on the assumption that the individuals "have access to the psychological property that the researcher wishes to measure" and that the participants "are willing to report that property" (Judd and McClelland, 1998, p. 202). However, according to Kagan (1988) people do not possess the kind of self-awareness about why they do the things they do, that self-report measures take for granted that they have. Kagan (2007) reminisces that before the emphasis of factor-derived traits emerged in personality psychology, projective or indirect tests (e.g. Rorschach ink blot tests) were used rather than self-reports, partially because of the notion that individuals could not consciously provide accurate self-information. Kagan (2005; 2007) finds some absurdity in the fact that the most accepted notions of personality are based on questionnaire answers, or interpretation of words, and equated this idea to biologists obtaining verbal reports to determine diseases. He argues that there may be some variation in interpretations of the questions because the "naked predicates" often provided are not context-specific and thus ambiguous (Kagan, 2007, p. 369). The other side of the debate regarding self-knowledge and distorted self-perception is evidenced by the vast amount of self-report measures that have been carefully constructed and frequently used. Though Kagan (2007) does not believe that questionnaires are as objectively accurate as behavioural measures, he does point out that self-reports are the only way to get at personal conceptions of the self and should thus be combined with more direct methods.

Other potential issues with self-reports noted by Paulhus and Vazire (2007) were that test construction is a time consuming process because of the need to show construct validity of the measure (though this can be

avoided by using an established inventory), as well as cultural differences in questionnaire response style. Since there are a number of concerns that arise as a result of using self-report methods, it is clear that other approaches to assessing psychological constructs should be considered. The advantages and disadvantages of some of the alternatives to self-reports will now be presented, and the need for the use of multiple methods will become more evident.

## Informant Reports

### *Advantages and disadvantages of informant reports*

It is worth considering the value of utilizing other ways of measuring personality traits, as self-reports may not be absolutely effective in accurately representing these constructs. Perhaps then, as Hofstee (1994) recommends, it might be better to ask other people about what someone they know is *really* like in order to obtain a more complete description of his or her traits. This would involve such methods as obtaining peer reports in which a number of other informants provide ratings about the individual. Though personality might be perceived as pertaining to internal attributes that can only be accessed by the individual him- or herself, we do get to know people based on their overt behaviours and actions. Therefore, others who have observed the target in their everyday life, even very informally, have the ability to judge that individual's personality (Funder, 1989). Craik considered judgments of this sort to be "an indispensable methodological tool for researchers in the social sciences" (as cited by John & Robins, 1993, p. 522). Based on a number of studies that have been conducted, informant reports could be defined as inventories on which a target's friends, acquaintances, spouses, and the like, provide ratings that are based on their overall conception of the individual. The other form of informant reports could involve judgments by peer observers. These methods are built upon the notion that other people can offer a unique perspective on an individual's personality (Hogan, 1998), and this is considered one of the key reasons why reports from others could be obtained when attempting to measure individual differences. Essentially, the objectivity that might come with personality ratings from others is presumably seen as an improvement on normally subjective self-reported assessment devices.

According to Vazire (2006), although data from informants are such a "rich source of information" (p. 2), peer reports are often avoided because many costs are perceived, in comparison with self-report measures. For example, most researchers consider collecting reports from people other than the target to be expensive in terms of time and money. As well, this method is

commonly viewed as being difficult and invalid, in terms of attempting to recruit uncooperative informants and ensuring that they do not give dishonest answers. However, Vazire dismissed these grievances and provided strategies to overcome any potential problems, such as obtaining the data via the Internet. Therefore, the burden of asking others for information about a target could be reduced. The title of Vazire's (2006) paper, "Informant reports: A cheap, fast, and easy method for personality assessment" summed up her sentiments about this means of measuring personality constructs.

From Hofstee's (1994) point of view, one of the largest advantages of collecting information from informants rather than the individual target is that in receiving a number of judgments from many people, reliability of the results is increased. He called this the 'principle of aggregation', and stated that it is this principle that would most strongly justify the argument that asking for other people's opinions is the best way to assess personality. However, since this aggregation principle could refer to the idea that "the sum of a set of multiple measurements is a more stable and unbiased estimator than any single measurement from the set" (Rushton, Brainerd, & Pressley, 1983, p. 19), Hofstee's praise of the value of using multiple informant judgments on their own is paradoxical. Although more than one informant report obviously yields more than one measurement, the principle of aggregation also implies that more than one method should be used in order to obtain a more accurate result. Rushton et al. (1983) interpreted this principle as evidence that a multimethod approach to measurement, along the lines of Campbell and Fiske's (1959) conception of construct validity, would be most helpful in enhancing the definition of a hypothesized trait. Though Cronbach and Meehl (1955) originally described construct validation as a way of investigating how a psychological construct relates to other theoretical concepts in the absence of a criterion to which to compare it, Campbell and Fiske (1959) declared that constructs must not only correlate with theoretically *related* variables ('convergent construct validity'), but they also should be uncorrelated with theoretically *unrelated* concepts ('discriminant construct validity'). More recent notions of construct validity treat it as a 'unitarian' concept, encompassing a variety of types of validities including 'face validity', 'content validity', 'predictive and concurrent criterion validity', and 'convergent and discriminant construct validity' (Messick, 1995). The potential benefits of incorporating multiple measures in general, for obtaining overall validity of the construct in question, will later be discussed. So far, it is noticeable that self-reports are not the only potentially valuable option for personality measurement.

It might be useful, however, to compare the informant judgments with the target's self-reports in order to determine if there is congruence between the

two. In an extensive review of research in interpersonal perception, Kenny (1994) proclaimed that there are indeed correlations between self and observer ratings, otherwise known as 'self-peer or self-other agreement'. This type of agreement has been shown, for example, in research conducted by McCrae and his colleagues over a number of decades using versions of the NEO Inventory (McCrae and Weiss, 2007). Although one might infer accuracy as a result of the tendency toward agreement, it should be noted that accuracy and congruence are not the same (Funder & Colvin, 1997). More valuable information could be provided by instead comparing the ratings made by the different observers, which refers to what is called 'peer-peer agreement' or 'interjudge agreement' in the literature. This is important to determine, because a lack of consensus could mean that "personality is in the eye of the beholder" (Kenrick & Funder, 1988, p. 25). Although there is often disagreement among judges, an empirical review by Kenny, Albright, Malloy, and Kashy (1994), showed that there is also consensus among peer ratings, and the level of agreement even has the potential to increase under certain conditions. For example, it is widely accepted in the literature that consensus is greater when the informants know the target well and when the traits being rated are more overt, like Extraversion. Once again, however, consensus does not necessarily equal accuracy.

Even still, it is interesting to consider if there is more agreement among peer-peer or self-peer ratings, due to the potential implications for endorsing particular methodologies. A study by John and Robins (1993) found that self-other agreement was not as high as interjudge agreement when 'evaluative' traits, or characteristics that carry a valence of desirability or undesirability, were what was being rated. This result was attributed to a "difference in the process of self-and-other perception" (p. 548). However, John and Robins indicated that this does not necessarily mean that the self possesses biases that would undermine the use of self-report measures. On the other hand, Kenny (1994) suggested that peer observers provide more insightful information than the target because these judgments are based more on "currently observable reality" (p. 191). He thus argued that "the orientation that views personality as the study of self-report inventories" (p. 194) is flawed, implying that judgements from knowledgeable others are better predictors of behaviour and should be considered.

Another advantage of asking peers for information is that "informants have often had the opportunity to observe the individual engaging in many different behaviors; consequently, the report can include many attributes that are supposed to reflect the characteristic" (Moskowitz, 1986, p. 305). Thus, informants can provide useful information about the individual's personality in general, across situations.

As mentioned, obtaining peer or informant data is considered less practical and efficient, and more costly than obtaining data directly from the individual under study. Though Vazire (2006) presented convincing arguments to dispel some of the myths, it cannot be disputed that it takes more effort to get two or three people to participate in research than to get one person, the target, to do so. Also, according to McCrae and Weiss (2007), because informant reports are in questionnaire format, there is the potential for many of the same response biases found in self-report respondents, like acquiescence and extreme ratings, to occur. One bias that would not arise in informant reports is social desirability because self-presentation practices cannot occur if the respondent is not reporting on him- or herself. However, informants might instead be motivated to show the target in a favourable light – especially if they are a spouse or close friend, or if there is an important reason for the research, like selecting prospective job candidates (McCrae & Weiss, 2007). Therefore, enhancement biases (or even diminishing biases, perhaps out of spite) could occur. Also, drawing on theories of attribution style, responses from others may be affected by new biases such as the fundamental attribution error. This involves the tendency to emphasize dispositions of others over situational factors in explaining behaviour (e.g. Ross & Nisbett, 1991). Additionally, a major limitation of informant reports is that, regardless of whether target individuals can accurately report on their own personalities, informants will never have as much information to draw from as the target (Paulhus & Vazire, 2007). They just do not have access to another person's thoughts, feelings, and motives, which is a weakness.

Furthermore, though it is advantageous that peers are able to form their conceptions based on an aggregate of behaviours across situations, an issue arises: other people would not be able to report on how their peers would react in a more specific situation (Moskowitz, 1986). In that case, it might be better to consider other types of methods such as behavioural observation, which will now be addressed.

## Behavioural Measures

### *Advantages and disadvantages of behavioural measures*

Another form of measurement that could be employed, collecting behavioural data, is very important because in personality psychology, traits are overtly expressed through behaviour. The most seemingly obvious and necessary way to gauge an individual's personality is to "see how they act" (Furr & Funder, 2007, p. 273), yet the use of the behavioural observation method is extremely rare in personality research (Robins et al., 2007). This would usually involve having external judges view and code an individual's

actions, either in a laboratory or naturalistic (real-world) setting. According to Furr and Funder (2007), the method is beginning to gain popularity; but they described the amount of research studying links between personality and behaviour as "embarrassingly short" (p. 275). The benefit of directly measuring something so inherent and central to the construct of personality—behaviour—is clear. Furr and Funder also point out that the importance of behaviour is the main reason why observational methods should be used in studying personality, even though there are a number of costs and considerations when conducting this type of research.

In terms of overall costs of observational methodology, and reasons why behaviour is not more commonly measured, Furr and Funder (2007) suggest that: 1. It is much more practical and convenient to use other methods like questionnaires, since there are many ethical issues in viewing others' behaviours or manipulating situational circumstances; 2. Developing and assessing coding schemes that are required involves a great deal of effort; and 3. The time and money involved in carrying out the research is a major drawback.

The fact that there is an option to conduct behavioural observations in either an artificial or natural setting is one strength of the method. However, there are different pros and cons associated with each approach. For instance, in a laboratory setting, the researcher could simulate a particular situation and observe how the person responds. This has a clear advantage over informant reports in that it can assess how a person demonstrates situation-specific traits, which is often difficult to determine when behaviour is rated by peers. It also bypasses the problems associated with assessing behaviours retrospectively, as often occurs with self- and peer reports (Henry, Moffitt, Caspi, Langley, & Silva, 1994). Then again, this laboratory approach has many weaknesses as well. These include artificiality, or the lack of representativeness of how the individual generally acts; the potential for social desirability to be exhibited by the target as a result of being watched; susceptibility to demand characteristics, which involves attempting to appease the observer (Moskowitz, 1986); and various ethical concerns. Also, because just one situation is often measured in this type of research, this limits the behaviours being assessed (Kagan, 2007). According to Furr and Funder (2007), observational data collected on one occasion might reflect specific situational factors rather than dispositional factors.

In order to get around some of these limitations, behavioural assessment can also be conducted in naturalistic settings. This would likely involve something along the lines of participant observation, or setting up equipment like cameras or microphones in order to record the individual's reactions,

perhaps on a number of occasions. These approaches could nevertheless be considered unattractive because they are often expensive and time-consuming for the researcher, and may be obtrusive to the participant. However, if reviewing long tapes of data is of concern, it is possible to implement more modern methods such as the 'EAR' ('Electronically Activated Recorder'), which captures short auditory snapshots of what the individual is doing at numerous times over a couple of days (Mehl & Pennebaker, 2003). This was developed to minimize the problems that come with artificial observations as well as with behaviour-oriented self-reports or even typical questionnaire-based experience sampling methods (e.g., Reiss & Gable, 2000).

The EAR has recently been used in a large-scale study of behavioural manifestations of personality, which will be briefly described here due to its novel methodology. Mehl, Gosling, and Pennebaker (2006) collected auditory EAR clips of individuals' lives over the span of a couple of days, and had research assistants code the sounds on a number of variables including interaction types (e.g. phone calls, group conversations), mood (e.g. laughing, arguing), activities (e.g. socializing, working), language use (e.g. use of long words or swear words), and so on. The coded variables were then subjected to correlation analysis with personality data obtained through a well -established personality measure. Results showed that there were patterns of correlations between self-reported personality traits and the 'observed' (heard) behaviours (e.g. Extraversion was significantly positively related to percentage of time spent talking), and this was interpreted as support for the idea that personality is manifested in everyday life behaviour. Mehl and colleagues (2006) also discussed the implications of their naturalistic methodology for expansion of the study of personality using behavioural measures, which included the potential for increasing the ecological validity (true-to-life nature of personality research), as this is not often found using other methods.

A final consideration related to behavioural measures is that it might be conceptually difficult to capture a particular construct by observing behaviour because "the link between a specific behaviour and a specific personality characteristic may not be direct" (Furr & Funder, 2007, p. 275). If assessing idiographic constructs was of interest, it would be necessary to combine methods such as these with other methods like questionnaires. However, implementing more than one measure might be valuable to do in any case. This proposition will now be investigated.

## Multiple Method Approaches

It has been demonstrated that there are a number of possible alternatives to using the common self-report method in measurement of personality constructs. However, in reviewing these methods it has become evident that there are a number of considerations involved in applying any measure of personality. A summary of the advantages, disadvantages, and debates surrounding the methods reviewed can be found in the Appendix to this document. Now there is the potential concern of how to wade between the pros and cons and go about choosing between the measures. Or is it better to use several methods?

In a review of self-reports, peer reports, and behavioural studies regarding their ability to measure persistent or 'global' personality characteristics, Moskowitz (1986) declared that there was no method that stood out as the best one to use. However, it was then suggested that different methods should be applied for different purposes. Although this point is well-founded, perhaps Moskowitz's argument would have been more compelling if he had considered the option of using multiple methods, rather than merely selecting one that appears to have a good "fit between operation and construct" (p. 309). This is important because utilizing more than one method to measure a concept can help to demonstrate construct validity (Campbell & Fiske, 1959). This does not mean that every research study must create a correlation matrix that compares multiple traits assessed by one measure with one construct assessed by many measures, which is the essence of Campbell and Fiske's multitrait-multimethod approach (West & Finch, 1997). Instead, it is necessary to simply realize that many of the methods that have been traditionally used to measure constructs like personality traits do have faults, so accuracy would be increased if these methods were applied in combination with one another.

An additional reason to consider multiple measures of personality is the issue of measuring certain aspects of a construct that could not be addressed to the right extent if only one method is used. Personality is a complex and multi-faceted concept because two potential perspectives are involved: 'personality from the actor's perspective and personality from the observer's perspective' (Hogan, 1998, p.6). Therefore, other people who are familiar with a particular individual might be able to provide some valuable insight about what that person is like, which would contribute to enhancing the ability to accurately measure personality constructs. However, since there are two underlying perspectives, it would not make sense to choose one method over the other; we would need to consider both the self-reported and informant-reported information in order to get a balanced and more complete depiction

of the construct. The soundness of Hogan's (1998) conception of personality could potentially be rebutted, but it nevertheless provides another illustration of why it is important to move away from the mentality that self-reports should be used on their own. In accordance with this argument, Vazire (2006), proclaimed that informant reports should be employed in conjunction with self -reports not only 'for the sake of multi-method assessment' (p. 2) but also because these alternative measures "allow researchers to address new questions that cannot be examined with self-reports alone" (p. 8). Paulhus and Vazire (2007) made a similar suggestion when it comes to self reports: "As with the use of any method, self-reports should be corroborated with alternative assessment methods' (p. 235).

A number of other prominent names in the personality literature have suggested the need for multiple measures and construct validity (e.g. Funder, 1989; 2002; John & Soto, 2007; Pervin, 1999). For example, Funder (2002) called for the use of 'innovative techniques that go beyond, without replacing, self-report measures…', which he named 'The new personality psychology' (p. 639). In his earlier discussion of peer observation measures, Funder (1989) asserted that construct validity must only be established vis-à-vis the 'relationships between the peer judgment and other sources of data about the target of judgment' (p. 220), thus implying the notion of the multi-method approach to construct validity. Even researchers who obviously favour one approach over the others, such as Hofstee (1994) or Kagan (2007), recognize the value in adopting multiple methods. Hofstee, for example, was careful to point out that he was not entirely rejecting the use of self-reports in his endorsement of informant reports. He suggested that 'the study of personality should employ a construct-validation strategy...pitting methods against each other (self and peer reports, tests, behavioural observations, physiological measures), rather than taking one approach as its Archimedic point' (Hofstee, 1994, p. 150, citing a personal communication with Funder). Clearly, more is better when it comes to measuring psychological constructs.

To summarize, the main strength of multiple methods is to improve construct validity, or accuracy of the data in measuring what is intended. In addition to improving validity, using multiple methods can lead to obtaining richer data and can answer more research questions. It is worth mentioning that multiple method research is not the same as what Creswell and Plano Clark (2007) refer to as 'mixed methods research'. This increasingly popular mixed approach "focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of studies" (Creswell & Plano Clark, 2007, p. 5). For example, a study could involve running focus groups to collect qualitative data that can then be used to inform questionnaire items for a quantitative inventory. Conversely, in-

depth interviews could be conducted following a quantitative study in order to get richer data. Alternatively, both quantitative and qualitative data could be given equal weight. The main purpose of this combination of approaches is to gain more complete insight about a research problem than could be done with quantitative or qualitative methods alone. In this way, mixed methods are related to 'methodological triangulation' in which the purpose is "to obtain different but complementary data on the same topic" (Morse, 1991, as cited by Creswell & Plano Clark, 2007) to increase understanding of the research question. Though slightly different in premise, both mixed methods research and studies using multiple methods suffer from similar limitations. Namely, using different methodologies or data is not easy; this requires extra time commitment, money, and resources, and training to implement. This explain why, for example, some personality tests are constructed without properly assessing validity, or why some informant research is conducted without assessing self-other agreement.

Although the value of multiple methods is known, the accuracy of the methods being used to measure the psychological constructs is too often taken for granted. In the era of wanting "fast data" (John & Benet-Martinez, 2000, p. 365), "research using multiple methods to measure the same construct has not been very frequent" (p. 356). Pervin (1999) argued that the value of self-reports "must be established rather than assumed" and suggested that the way to do this would be to consider the use of other approaches (p. 691). Conversely, just as the validity of self-reports needs to be demonstrated, so does the accuracy of the other methods. For example, it should not be assumed that aggregating the results of a number of observer reports indicates validity; this only really provides evidence of reliability (Kenny, 1994). Furthermore, validity of all methods should be formally established because *all* measurement in psychology is susceptible to biases and errors (John & Benet-Martinez, 2000). It is imperative that researchers understand the importance of accurate measurement, because: "Regardless of a measure's other properties, if there are no valid inferences from the measure's scores, assessment has been pointless" (Ozer, 1999, p. 678). Once again, in personality research, more than one method should be used in order to get a better portrayal of the construct.

## Conclusion

In assessing the strengths and limitations of a variety of methods to measure personality constructs, it has been argued that no single measurement device is better than any other. Therefore, a better way to measure these constructs might be to integrate a variety of methods in order

to improve the accuracy of the measurement. Most prominently, it is necessary to decrease the overwhelming reliance of self-reports on their own.

Essentially, personality psychologists should not presuppose the value of self-reports or any of the other measures; and incorporating more than one approach to measure personality constructs will help to increase the accuracy of measuring the intended theoretical concept. With that being said, it is important to state that the measures that are employed should always be driven by the specific research question that the researcher wishes to address. Sometimes the frequency of certain behaviours is what is under study, in which case asking for either self- or peer reports would not contribute to the investigation of the hypothesis. Other times, perhaps using only informant ratings would be sufficient to address a certain research question. However, when there is some doubt about whether the construct under investigation can be represented to its fullest extent, the use of multiple methods should definitely be considered. As emphasized by Judd and McClelland (1998), the bottom line is simply that we need to take the time to consider measurement when conducting research.

With proper measurement it is possible to shift latent and unobservable constructs, such as personality, more into the observable realm. The existence of these constructs is made more concrete by the methods that purport to measure them. Therefore, the establishment of the validity of these measures through incorporating a variety of methods has the potential to bring the constructs to life.

## Bibliography

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. London: Sage.

Costa, P.T.,Jr.,& McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEOPI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-301.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw Hill.

Furr, R. M., & Funder, D. C. (2007). Behavioral observation. In R. W. Robins,

R. C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 273-291). London: The Guilford Press.

Funder, D. C. (1989). Accuracy in personality judgment and the dancing bear. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 210-223). New York: Springer-Verlag.

Funder, D. C. (2002). Personality psychology: Current status and some issues for the future. *Journal of Research in Personality, 36*, 638-639.

Funder, D. C., & Colvin, C. R. (1997). Congruence of others' and self-judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617-647). San Diego: Academic Press.

Gosling, S. D., Rentfrow, P .J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504-528.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93-104.

Henry, B., Moffitt, T.E., Caspi, A., Langley, J., & Silva, P. (1994). On the "remembrance of things past": A longitudinal evaluation of the retrospective method. *Psychological Assessment, 6*, 92-101.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality, 8*, 149-162.

Hogan, R. (1998). Reinventing personality. *Journal of Social and Clinical Psychology, 17*, 1-10.

John, O. P., & Benet-Martinez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 339-369). Cambridge: Cambridge University Press.

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206-219.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*, 521-551.

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley & R. F.

Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461-494). London: The Guilford Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). London: Guilford Press.

Judd, C. M., & McClelland, G. H. (1998). Measurement. *Handbook of social psychology* (Vol. 1, pp. 180-232). Boston: McGraw Hill.

Kagan, J. (2007). A trio of concerns. *Perspectives on Psychological Science, 2*, 361-376.

Kagan, J. (2005). A time for specificity. *Journal of Personality Assessment, 85*, 125-127.

Kagan, J. (1988). The meanings of personality predicates. *American Psychologist, 43*, 614-620.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford Press.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the Big Five. *Psychological Bulletin, 116*, 245-258.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112*, 165-172.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43*, 23-34.

Kline, P. (1993). *Personality: The psychometric view*. London: Routledge.

McCrae, R. R., & Costa, P. T. Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139-153). New York/London: Guilford Press.

McCrae, R. R., & Weiss, A. (2007). Observer ratings of personality. In R. W. Robins, R. C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 259-272). London: The Guilford Press.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90*, 862-877.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural

conversations. *Journal of Personality and Social Psychology, 84*, 857-870.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50*, 741-749.

Moskowitz, D. S. (1986). Comparison of self-reports, reports by knowledgeable informants, and behavioral observation data. *Journal of Personality, 54*, 294-317.

Ozer, D. J. (1999). Four principles for personality assessment. In L.A. Pervin & O.P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 671-686). London: Guilford Press.

Paulhus, D. P. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego: Academic Press.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R.C. Fraley & R.F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). London: The Guilford Press.

Pervin, L. A. (1999). Epilogue: Constancy and change in personality theory and research. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 689-704). London: Guilford Press.

Reis, H. T., & Gable, S .L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190-222). Cambridge: Cambridge University Press.

Robins, R. W., Tracy, J. L., & Sherman, J. W. (2007). What kinds of methods do personality psychologists use? A survey of journal editors and editorial board members. In R. W. Robins, R. C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 673-678). London: The Guilford Press.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94*, 18-38.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.

Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment, 17*, 396-408.

Sue, V. M., & Ritter, L. A. (2007). *Conducting online surveys*. London: Sage.

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality, 40*, 472-481.

West, S. G., & Finch, J. F. (1997). Personality measurement: Reliability and validity issues. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 143-164). San Diego: Academic Press.

Westen, D., & Rosenthal, R. (2005). Improving construct validity: Cronbach, Meehl, and Neurath's Ship. *Psychological Assessment, 17*, 409-412.

# Appendices

| Method | Advantages | Disadvantages |
|---|---|---|
| Self-Reports | Practical and efficient<br>Convenient and easy to administer<br>Inexpensive<br>Direct insight into unique personal information<br>Individual motivation to respond<br>Can control most response biases<br>Many readily available psychometrically-tested inventories<br>Most commonly used method | Potential issues with credibility of responses (due to response biases):<br>Socially Desirable Responding<br>Acquiescent Responding<br>Extreme Responding<br>Assumes that respondents are self-knowledgeable and do not have distorted self-perceptions (debates involved with using)<br>Issues with non-context-specific language use of questions<br>Cultural limitations |
| Informant Reports | Can provide objective information about a target<br>Potential to be practical, inexpensive and convenient (e.g. Internet)<br>Multiple raters and aggregation of data can lead to reliability of results<br>Others can provide insights on behaviour, especially across-situations<br>No socially desirable response bias | Less efficient and more effort required to obtain third-party data than to simply ask the target<br>Similar response biases to self- Reports (e.g. Acquiescent and Extreme Responding)<br>Issues with choosing informants – possibility for biases based on relationship or research aims<br>Others cannot access certain personal information about target<br>Difficult to assess situation-specific behaviour |
| Behavioural Methods | Directly examines behaviour, which is central to examining personality<br>Can get situation-specific information<br>Fewer response biases than with questionnaires<br>Two potential settings: lab or field (which carry different pros and cons) | Least practical and convenient method<br>Ethical concerns<br>Involves complex coding schemes<br>Expensive in time and money<br>Possible disconnection between behaviour and specific traits |
| Multiple Methods | Same strengths as above<br>Added advantage of the improvement of construct validity<br>Can answer more research questions and lead to richness of data, especially if part of a triangulation or 'mixed methods' strategy | Same weaknesses as above<br>Requires more effort, money, resources, time and training to implement |

***Summary of Advantages and Disadvantages of Methods of Personality Measurement***