

Claremont Colleges

Scholarship @ Claremont

CMC Senior Theses

CMC Student Scholarship

2019

The Use of Personality Testing in Personnel Selection

Riya Kumar

Follow this and additional works at: https://scholarship.claremont.edu/cmc_theses



Part of the [Dance Commons](#), [Industrial and Organizational Psychology Commons](#), and the [Organization Development Commons](#)

Recommended Citation

Kumar, Riya, "The Use of Personality Testing in Personnel Selection" (2019). *CMC Senior Theses*. 2038.
https://scholarship.claremont.edu/cmc_theses/2038

This Open Access Senior Thesis is brought to you by Scholarship@Claremont. It has been accepted for inclusion in this collection by an authorized administrator. For more information, please contact scholarship@cuc.claremont.edu.

Claremont McKenna College

The Use of Personality Testing in Personnel Selection

submitted to
Professor Jay Conger

by
Riya Kumar

for
Senior Thesis
Fall 2018
December 10, 2018

TABLE OF CONTENTS

Acknowledgements.....	1
Abstract.....	2
Chapter 1: Assessing Job Fit in Personnel Selection	3
Chapter 2: A History of Personality Testing	20
Chapter 3: The Limitations of Personality Tests in Predicting Job Performance.....	36
Chapter 4: The Big Five Model & Derivative Tests.....	50
Chapter 5: The Future of Personality Testing.....	68
References:.....	80
Appendices:.....	100

ACKNOWLEDGMENTS

I would like to first and foremost acknowledge my advisor, Professor Conger, who provided an enormous amount of guidance and insight throughout the writing process. His detailed feedback on each one of my drafts and his willingness to answer all questions made this process substantially easier. I would also like to thank the professors from CMC's Psychology department who provided me with additional resources and research materials. Lastly, I would like to thank my family and friends for their unparalleled support in this journey.

ABSTRACT

Research has shown that more than 45% of American companies are opting to integrate personality tests in their recruitment processes. Given this surge in personality testing, this thesis examines whether personality testing is a valid predictor of job-fit and performance in the context of personnel selection. A large proportion of this paper is focused upon the Big-Five factor model, its limitations, and derivative tests of the model. The impact of technology upon personality testing is also discussed as an emerging field. By tracing and examining the history of personality testing to current day, I have found that personality tests are best administered when they provide incremental validity over other tools and are matched to specific job-criteria.

Keywords: personality testing, job performance, Big Five

Chapter 1

Assessing Job Fit in Personnel Selection

Introduction

Companies can choose from a variety of assessment methods in personnel selection. Prior to recruitment, the organization's human resources or recruiting department usually decide the criterion for selecting workers and the tools to be employed in the process. In hiring a candidate, companies focus upon two concepts: person-job (P-J) fit and person-organization (P-O) fit. P-J fit assesses whether a person has the skill-set and ability required by the job, while P-O fit refers to how closely a candidate matches a given organizational environment with respect to the needs and characteristics of both the candidate and the organization (Sekiguchi, 2004). Both P-O and P-J fit are examined in selection, however, there is greater concentration upon the latter in employee recruitment.

Methods and tools in personnel selection, such as interviews or cognitive tests, have developed with changes in organizational needs, complexity, and societal shifts. The following sections of this thesis will describe the evolution of some of these tools and their utility to person-job fit, namely intelligence and aptitude testing, interviews, integrity testing, assessment centers, simulations, and personality tests. Ultimately, the central focus of this thesis will be on personality testing, given its resurgence in the last decade and its increased integration in the hiring process.

Intelligence and Aptitude Testing

During the 1920's, psychology garnered widespread acceptance as a tool that could address social problems in a growing industrial society (Dennis 1984). In particular, the use of IQ tests flourished as a means to test the knowledge and intelligence of candidates and the industries they were suited to. The United States army, for example, administered the Army Alpha test, an evaluation of intelligence based on the Stanford Binet test. The primary motivation of the Army Alpha was to eliminate intellectually deficient individuals who were deemed unfit to work in the army. The Army Alpha test propelled mental testing into mainstream society despite concerns with its validity (Dennis 1984). Intelligence tests were then integrated into school curriculums as a method to differentiate grade levels. Moreover, the tests were also administered for college applications as well as application processes for potential employees (Dennis, 1984). In the realm of personnel selection, intelligence testing appeared to be predominately used to screen out candidates who did not have the perceived minimum required intelligence for the job.

However, intelligence testing fell out of favor between 1922 and 1925. Many companies chose to give up on the use of these tests as a result of poor turnover and disappointment with what these tests were expected to deliver (Vinchur & Bryan, 2012). Furthermore, the Alpha Army test also became a part of a larger controversy on differences in intelligence between racial groups, calling into question its validity. Some psychologists suggested that the test measured innate intelligence, independent from social experiences and education. As a consequence, it was used to bolster notions of

white supremacy and intelligence over other minority groups, as Caucasians tended to have higher scores (Dennis, 1984). Thus, controversies in the interpretation of test results and problems with validity undermined the usefulness of cognitive testing as a component of personnel selection.

During World War II, however, personality testing resurfaced at the forefront due to military needs in the recruitment of aviators and soldiers (Vinchur & Bryan, 2012). Moreover, increasing organizational complexity elicited the need for more scientific methods of testing. Consequently, psychologists and researchers showed more concern to validity and reliability and the need for sound statistical techniques (Scroggins, Thomas, & Morris, 2008). For example, in WWII, there was a need to recruit candidates for highly specialized tasks. It was realized that the job demands and skills for a pilot, as compared to a flight engineer or a navigator, were extremely different. Moreover, it was found that intelligence tests were not sufficient in selecting candidates for flight school, as they did not provide an assessment of a candidate's hard skills. The army thus replaced the Alpha test, which was used to screen out recruits, with the Army General Classification test, which was designed to measure whether a candidate was suited to the role of a soldier. The Army General Classification test reduced the focus on verbal abilities, instead concentrating on quantitative and spatial abilities, which were deemed more important to the role of soldiers (Harrell, 1992). Specialized aptitude tests were also developed for technical and mechanical jobs in the army. Moreover, the United States Employment Service developed the General Aptitude Test Battery which consisted of 12 tests measuring 9 aptitude abilities (Appendix A). The test was used for new entrants into the labor force, individuals who desired an occupational change, and those who wanted

vocational training (Dvorak, 1956). Thus, there was a shift in the scope of testing; organizations, including the military, did not solely rely on intelligence testing. To counter this drawback, aptitude tests were developed to measure specific abilities of the candidate. Companies also began to administer job-specific aptitude tests, realizing different roles within the same job required a unique skill-set (Lunenburg, 2011).

Moreover, we have seen a revision in what we classify as intelligence. In the early 1900's, employers assessed intelligence in the context of a candidate's knowledge about the world, such as factual information and current events. In the last 10 to 20 years, psychologists have, however, expanded the concept of intelligence to include emotional intelligence. Emotional intelligence is defined as the "ability to monitor one's and other's emotion, to discriminate among them, and to use to information to guide one's thinking and action" (as cited in Lievens & Chan, 2017). Goleman argues that our current society is characterized by emotional ineptitude and that we have witnessed a rise in the number of individuals suffering from depression and other mental health issues. As a result, there has been increased scientific study on emotions, particularly emotional intelligence (as cited in Cadman, 2001). In a meta-analytic study, researchers found that self-reported emotional intelligence had more incremental validity than cognitive ability tests and personality measures in predicting job performance (Oboyle, Humphrey, Pollack, Hawver, & Story, 2010). Studies investigating the link between emotional intelligence and job performance have found a relationship with leadership effectiveness. This is because awareness of one's and other's emotional states enables a leader to foster and build close relationships with co-workers. Moreover, knowledge about another's emotions allows leaders to predict emotional reactions in different situations and enables

better management of team members. Furthermore, leaders with effective emotion regulation have a positive effect on team performance (Prati, Douglas, Ferris, Ammeter, & Buckley, 2003). Thus, some companies are now opting to test for emotional intelligence, especially in leadership roles such as managers and executives.

Today, many organization use a battery of aptitude tests that more holistically reflect the candidate's abilities, knowledge, and character. Intelligence tests are still regarded as good predictors for jobs which demand problem solving, thinking, and reasoning (as cited in Lunenburg, 2011). For example, some organizations require applicants to submit their scores on standardized tests such as the SAT and ACT; however, companies are now opting to supplement intelligence tests with personality tests and aptitude tests. Furthermore, the term "intelligence" has now broadened to encompass other kinds of intelligences, such as emotional intelligence. As research has shown, companies are now opting to integrate emotional intelligence in their recruitment processes. Thus, over time, we have witnessed a rapid expansion in the types of test used to assess job fit as well as the importance of job specificity in testing.

Interviews

The interview is one of the most common procedures used in personnel selection. An interview is defined as an instrument used to predict a candidate's future job performance and compatibility with a role on the basis of his/her oral responses. The birth of the interview is traced to Thomas Edison, who devised the Edison questionnaire, comprised of 48 questions, for the position of an industrial chemist (Dennis, 1984) (Appendix B). Although some questions were connected to the job that Edison was hiring

for, a large proportion were focused upon the candidate's general knowledge, which were irrelevant to the job at hand. For example, questions like "Who assassinated President Lincoln?" and "Who was the Roman Emperor when Jesus was born" (Berlinger, 2018) tested a participant's general knowledge about world events rather than his/her knowledge about chemistry or job-specific tasks. However, Edison did revise his test to include items that assessed an individual's character such as "When do you consider a lie permissible?" and "If you were to inherit \$1,000,000 within the next year, what would you do with it?" (Dennis, 1984), yet the primary focus of his interview remained upon testing general knowledge.

Today, interviews have shifted from focusing solely on intelligence testing and general knowledge, to examining the candidate's aptitude, knowledge of and interest in the company, past experiences, and confidence. Interviews are typically differentiated by their degree of structure and are thus classified into structured, semi-structured, and unstructured interviews (Azarpazhooh, Ryding, & Leake, 2008). Unstructured interviews are the most common type of interview and are characterized to be informal, conversational, and non-directive. The interviewer does not prepare questions beforehand and answers are not evaluated using a formal scoring guide (McDaniel, Wherzel, Schmidt, & Maurer, 1994). On the other hand, structured interviews consist of job-related questions that are evaluated using a predetermined scale (Azarpazhooh et al., 2008). Although structured interviews are used less frequently, research has shown that they result in superior employee performance (Zee, Bakker, & Bakker, 2002). This is because structured interviews provide a more consistent standard for interviewees to assess whether a candidate meets the criteria of the job and has the relevant skills. In an

unstructured interview, the interviewee or recruiter is forced to rely on subjective opinion and intuition which are more prone to biases.

Content of Interviews

The content of these varied forms of interviews are further subdivided into situational interviews, behavioral interviews, and technical interviews. In a situational interview, candidates are presented with work-related dilemmas and then asked what actions they would take if confronted with these kinds of situations (Oostrom, Melchers, Ingold, & Kleinmann, 2016). This allows a potential employer to assess a candidate's capabilities and problem-solving skills in a situation that is similar to what they might encounter in that organization. For example, if you were testing for interpersonal skills, a possible scenario could be: "A very angry client walks up to your desk. She says she was told your office sent her an overdue check 5 days ago. She claims she has not received the check. She says she has bills to pay, and no one will help her. How would you handle this situation?" ("Structured interviews", n.d.). This tests the candidate's ability to handle a difficult work-related situation as well as his/her interpersonal skills in a client-facing task. Research conducted on situational interviews has shown that these types of interviews can predict job performance (Latham & Sue-Chan, 1999). During a behavioral interview, the interviewer asks the candidate a set of standardized questions regarding behavior in past situations. The situations are pre-determined and job-related (Alonso, Moscoso, & Salgado, 2017). Questions for a behavioral interview include: "Tell me about a mistake you made during a project or at a certain point in your career" or "Describe a time you had a conflict with a co-worker, and how you resolved it" (Becker,

n.d.). Similar to the situational interview, the behavioral interview permits the employer to assess a candidate's personality, ability to handle work situations, and their skill-set. Lastly, the technical interview is when a candidate is asked technical questions pertaining specifically to the job role. For example, many software companies administer a programming test to web developers to assess their coding abilities, a skill that is pivotal to the job description. (Ford, Barik, Parnin, 2016). Technical interviews are predominantly used in jobs relating to engineering and computer science, as these professions require hard skills. Thus, all three types of interviews are intended to assess a candidate's abilities, skills, and future performance in work-related scenarios.

Companies use a variety of interview techniques in their recruiting process to assess job fit. For example, consulting firms, like the Boston Consulting Group, use both behavioral and technical interviews. The first round, a behavioral interview, assesses the candidate's background and experiences and whether he./she has the required experience for the job .The technical interview, which is in the form of a case study, is used to evaluate the candidate's thinking process and strategic skills, two components that are essential to consulting. Lastly, during the course of the interview, they also look at the candidate's curiosity, ability to communicate effectively, and positivity. (“Interview Process & Preparation”, n.d.).

Despite the prevalence of interviews in personnel selection, researchers have shown that they have considerable limitations. One of the biggest drawbacks is related to the role of the interviewer. Interviewers are often untrained and therefore are more likely to ask irrelevant questions and make inconsistent judgements (Niece, 1983). For example,

interviewers who have little knowledge about the position will be unlikely to pose questions that capture the essence of the job. Moreover, the problem of personal bias can also negatively impact the candidate interviewing for the job. Interviewers who have preconceived notions regarding an applicant, race or ethnicity, may be more likely to unfavorably rate the participant. Research has also shown that some interviewers are impressed with surface attributes of the candidate such as physical appearance and signs of composure (as cited in Lunenburg, 2011). These drawbacks are particularly magnified in unstructured interviews where the interviewer is not given a set of predetermined questions or an answer key. Thus, companies should opt for interviews that have a formal structure, and moreover, should also provide formal training to personnel conducting interviews (Niece, 1983).

Thus, the interview process has become more interactive and expansive. Rather than simply administering an intelligence test, interviewers now also assess a candidate's aptitude, familiarity with the company, past experiences, and confidence. Moreover, interviewers are also attuned to the candidate's non-verbal cues such as confidence, composure, and eye-contact. Research has shown substantial drawbacks to the use of unstructured interviews for personnel selection and job performance, therefore organizations should take this into consideration when recruiting.

Integrity Testing

Integrity tests, predominantly paper and pencil tests, are used in the pre-hire process to determine whether a candidate is inclined towards dishonest behaviors. The earliest form of integrity testing for pre-employment was the polygraph, a lie-detector

test, which measured deception via physiological responses (Saxe, 1994). Polygraphs typically measure a participant's response to a set of questions using physical indicators such as heart rate, respiration and skin conductivity. Based on a physiological response, the administrator of the polygraph ascertains whether the participant is lying or telling the truth (Saxe, 1994). However, evidence has shown low validity of physiological indicators as predictors for deceptive behavior in the context of personnel selection (Sackett, Burris, and Callahan, 1989). As a result, federal law banned the use of polygraphs as a screening tool in personnel selection with the establishment of the Polygraph Protection Act of 1988 (Faust, 1996). Polygraphs were also banned due to privacy issues, as some candidates found the test intrusive (Faust, 1996).

With restrictions on the use of the polygraph in 1988, integrity-specific instruments began to gain prominence. Unlike polygraphs, integrity tests are not intended to reveal whether a person is lying; these tests reflect an individual's sense of morality and honesty by assessing attitudes. While it is difficult to ascertain the exact number of companies that use integrity testing, it is estimated that around 5000 companies utilize integrity tests in personnel selection (Cullen & Sackett, 2004). Integrity testing is now divided into two types of tests: overt-integrity tests and personality based measures. Overt-integrity tests generally measure of attitudes towards dishonesty and theft, in addition to admissions of past theft or participation in illicit activities. On the other hand, personality-based measures or covert integrity tests do not directly measure honesty; these tests are designed to measure counterproductive work behaviors. Examples of this include the Personnel Decision Employment inventory which measures employee deviance and the Reliability Scale of the Hogan Personnel Selection series

which measures organizational delinquency. The Hogan Reliability scale consists of items that deal with conscientiousness, authority, thrill-seeking, and social insensitivity (Sackett, Burris, and Callahan, 1989). Meta-analytic investigations have found that integrity tests have good criterion validity in the prediction of job performance and counterproductive work behaviors (as cited in Cullen & Sackett, 2004). Both the polygraph and the integrity tests are designed to screen out candidates who are unsuitable for the job due to their inclination towards deceit and dishonesty. However, the integrity tests utilized today differ as they also measure traits and attitudes that are counterproductive to a job, thus screening out prospective hires who are likely to engage in disruptive behavior or take risky actions on the job.

Both overt and covert integrity tests, however, have been shown to be susceptible to faking and coaching. A study has shown that individuals were able to fake being more honest when instructed to do so (as cited in Alliger, Lilienfield, & Mitchell, 1996). Moreover, integrity tests are also susceptible to coaching. The content and domains measured by integrity test tend to be similar, thus participants can anticipate the kind of questions they will be asked during the test. Moreover, there are coaching tutorials available that candidates can use to increase the probability that they will pass an integrity test (Alliger et al., 1996).

There are also other objections to the use of integrity tests. Some argue that integrity tests should be culturally dependent, as culture can influence an individual's perception of honesty versus dishonesty. For example, a gift in one culture may be viewed as a bribe in another culture. Moreover, another objection is that more honest

people maybe more likely to admit to wrong-doings in the past than dishonest individuals. However, despite these limitations, integrity tests have shown incremental validity in relation to job performance (Alliger et al., 1996).. Therefore, some companies continue to integrate them in their recruitment processes

Assessment Center

An assessment center is defined as a system of global assessment which includes interviews, aptitude tests, personal history tests, and situation tests. The precepts of the assessment center were seen in World War II in the evaluation of candidates for sensitive assignments. The center comprised of traditional interviews, psychological tests, and situational exercises. Candidate's performance was evaluated by trained professionals in individual and group settings (Alder, 1987). In the 1950s, AT&T became the first private company to use the assessment center as means to predict performance for managers (Vinchur & Bryan, 2008). The fruition of the assessment center was the result of an AT&T Management Progress Study which examined the performance of managers in a longitudinal study. The authors of the study, Dr. Bay, Richard Campbell, and Donald Grant found that assessments conducted early in a manager's career were valid predictors of job performance. As a result of the study, AT&T established the assessment center, where candidates spent three and a half days partaking in various tests and simulations. Based on their performance, assessors evaluated whether the candidate could potentially achieve a middle management role at AT&T. Consequently, more than 1000 companies established assessment centers by the 1970s. However, despite studies that showed good predictive validity of assessment centers in evaluating job performance, there are

limitations to the use of assessment centers. The cost of building and implementing an assessment center is substantial. It is labor intensive as teams of three to seven professionals are needed to assess a group of 12 candidates. Moreover, the testing process is also time consuming (Alder, 1987). Thus, cost can be a limiting factor for companies

Simulations

Simulations generally comprise of three categories of testing techniques: assessment centers, situational judgement tests, and work samples. The first examples of simulations, in the context of personnel selection, were seen in assessment centers established by the military in the early 1940s. Since that time, assessment centers have predominantly been used in the selection and assessment of managers. They are also used in the selection of salespersons and public safety officials. The exercises employed in assessment centers comprise of in-basket exercises, leaderless group discussions, and role playing (Tuzinski, 2013). For example, in an in-basket exercise, candidates are given a scenario and then asked to write out responses to how they would tackle the situation (“Personnel selection: Methods”, n.d.). In a leaderless group discussion, a cohort of applicants are asked to discuss a job-related problem where examiners would assess the leadership and communication skills of each candidate. Lastly in a role-playing exercise, candidates interact with another individual in a job-related situation. The latter is generally a trained role player who responds to the candidate. (“Personnel selection: Methods, n.d.)

Situational judgement tests (SJTs) also fall under the bracket of simulations. In this type of test, candidates are presented with a situation they might encounter at the job

for which they are applying to (Tuzinski, 2013). SJTs were used in the army to assess the judgement of soldiers. In the 1950's they were also administered for managerial selection. Lastly, work samples, which test candidates on the actual kinds of tasks they will perform on the job, have been shown to have high criterion related validity for job performance (Tuzinski, 2013). An example of a work sample for the position of an administrative assistant could involve transcribing a memo using a word processor as this would allow the employer to evaluate the candidate's task proficiency for the position. ("Assessment and Selection", n.d.).

There are a number of advantages to using simulations in personnel selection and job performance. Simulations are less susceptible to faking and social desirability as compared to self-reports. In a simulation, a participant is made to behaviorally demonstrate how they would react in a job-related scenario using their technical knowledge, personality, and cognitive abilities (Boyce, Corbet, & Adler, 2013). Thus, the simulation is able to measure a participant's actual behavior, whereas other selection techniques can only measure behavior in a non-behavioral format. Moreover, simulations allow for greater flexibility as they can be used to measure a wide range of constructs from personality to cognitive ability. They can also be customized and altered to suit a wide range of occupations. Despite its perceived benefits, the cost of implementing simulations can be a drawback for some companies (Boyce et al., 2013). However, developments in technology have made simulations more accessible and affordable, which has resulted in a greater number of organizations adopting them into the recruitment process.

Personality Tests

The history of personality testing can be broadly classified into five waves. The first wave involved an obsession with maladjustment. Researchers developed uni-dimensional measures such as the Woodworth Personal Data Sheet which was designed to screen out candidates who demonstrated psychoneurosis. The second wave of personality testing, from the 1940s to the 1960s, saw the establishment of personality inventories comprised of a wide variety of personality dimensions. The majority of tests developed during this period had weak ties to maladjustment. In the third wave from 1960-1980, the use of personality testing actually declined due to a meta-analytic study which showed poor criterion validity of personality to job performance. The fourth wave, from 1980 to 2000, witnessed a resurgence in personality testing as a result of the five-factor model, an agreed upon taxonomy of personality. Moreover, meta-analytic investigations investigating the role of personality in predicting job performance were optimistic regarding the use of personality tests in job fit. Finally, in today's context, personality tests have transitioned from paper-and-pencil assessments to online, computerized assessments. Moreover, technology, especially artificial intelligence, have transformed personality testing by automating the process of feedback on tests as well the assessment of individuals personality through the use of algorithms .The developmental history of personality testing will be examined and analyzed in greater detail in the next chapter.

Personality tests are generally broken down into objective and projective tests (Meyer & Kurtz, 2006). Objective tests require candidates to describe how accurately a

proposition describes their personality. The response options to these questions are limited. For example, participants may be asked true and false questions, or they might respond using a Likert scale. In contrast, projective tests consists of tasks or activities where the candidate is required to produce a response without restrictions on the nature of the response. Objective tests are predominantly used in the realm of job performance as there is no judgement required by the test-giver; answers are scored according to a key (Meyer & Kurtz, 2006). Thus, my analysis will focus on objective personality tests given their relevance to job performance.

Conclusion

As discussed, organizations no longer rely on a single assessment in recruitment; candidates may be subjected to a battery of aptitude and personality tests in the first few rounds, and if successful, they proceed to the interview stage. Moreover, the assessment center also allowed for the integration of a wide number of assessment tools, thus increasing criterion validity in the prediction of job performance. In the early 1900's, there was also a reliance on self-report measures, but with the advent of simulations, employers are now able to observe a candidate's true behavior in a work-related situation. By screening candidates with a wide number of assessment techniques, organizations now have a more holistic view of a candidate's profile.

Given the range of evaluative tools from which organizations can choose, I will focus my discussion on personality testing and its impact on job-fit and performance. There has been a renewed interest in this field over the last two decades, attributed to meta-analyses that have examined the relationship between the Big Five Personality traits

and job performance (Barrick & Mount, 1991; Salgado, 1997). A number of companies have chosen to integrate personality tests as a component in their hiring processes, making it important for researchers to study whether this is an effective tool in predicting job performance and fit. Thus, my thesis will trace the evolution of personality tests, their role in job fit and performance, common tests and the Big-Five model, and explore whether these tests are both reliable and valid measures of job performance. Ultimately, I will deliberate whether personality tests should constitute an important part in the recruiting landscape and how companies can use and interpret the results of these tests in making personnel decisions.

Chapter 2

A History of Personality Testing

Introduction

The first chapter of my paper provided a brief background to the different waves of personality testing. In this section, I will provide a detailed description of the history of personality testing, the different types of tests that were developed, and their use in evaluating job fit and performance.

A History of Personality Testing

The origins of personality testing stemmed from military needs. From 1910-1930, personality testing focused on maladjustment and largely disregarded other constructs of personality that were relevant to predicting job performance and fit, as the military wanted to eliminate candidates with behavioral and mental health issues. This obsession with maladjustment continued into the early 1940's; however, focus on maladjustment slowly decreased as broader personality inventories came into the forefront of testing.

The Woodworth's Personal Data Sheet (WPDS), developed in 1919 as a response to military requirements during World War I, was the first formal objective personality test identified by researchers (Gibby & Zickar, 2008). As a result of enemy bombardment, many American soldiers during WW1 experienced shell-shock¹, now termed PTSD, which rendered them unable to continue fighting in the war. When the war

¹ Symptoms include uncontrollable nausea, night shakes, heart palpitations

concluded, shell-shock victims were estimated at 800,000 French and British soldiers, and 15,000 American soldiers. Consequently, to address this concern, the American military commissioned the scale of Psychoneurotic Tendencies, developed by Robert S. Woodworth, which was designed to screen soldiers who demonstrated emotional instability. Psychoneurosis is classified as a mild mental illness not caused by an organic disease. Symptoms include anxiety, obsessive behavior, stress, and hypochondria (“APA dictionary of Psychology”, n.d.). Woodworth developed questions based on interviews and case studies with individuals who showed psychological disturbance. If candidates exhibited a number of neurotic symptoms on the test, they were then interviewed by a military psychologist. Thus, the military used the test to identify and eliminate candidates who showed traits that impeded their ability to perform their duties (Gibby & Zickar, 2008). Given the high risk nature of the military and active duty, soldiers are required to remain calm and collected when making important decisions. Individuals who are inclined towards instability are more likely to make poor decisions when on the battlefield and are also more prone to shell-shock, thereby making them poor candidates for the army.

In 1919, Woodworth adapted the Psychoneurotic Tendencies scale to the Woodworth Personal Data Sheet, which was used in industrial research (Gibby & Zickar, 2008). The WPDS consisted of 75 yes/no questions. Sample questions included: “Do you ever get so angry that you see red?” and “Are you troubled with dreams about your work?” (Gibby & Zickar, 2008). Again, as highlighted by the nature of the questions, the central focus of this test was to identify maladjustment and psychoticism, a personality type defined by aggressive and hostile behavior, in the test-taker.

Throughout the 1920's and the 1930's, a number of tests focusing on maladaptive facets of personality came into being, namely the Colgate Mental Hygiene Test (1925), and the Personality Schedule (1930) (as cited in Gibby & Zickar, 2008). In part, the management literature accounted for this obsession with maladjustment. Elton Mayo, a professor at the Harvard Business School, attributed workplace issues to the "mental disintegration" and lack of emotional control that workers experienced (Vinchur & Bryan, 2012). He reasoned that identifying and eliminating individuals who exhibited these behaviors would increase productivity and reduce group conflict in work settings. Mayo's work was heavily influential in the psychology community, thus the backbone of personality tests at this time were centered on maladjustment. Moreover, the rise of psychiatry elicited the creation of a mental hygiene movement (Vinchur, 2018). Mental hygiene refers to the prevention and care of individuals with mental illnesses. Thus, management literature and mental hygiene reiterated an emphasis on measuring maladjustment.

These aforementioned tests borrowed heavily from the WPDS, focusing on emotional instability and control. The Colgate Mental Hygiene Test was intended to identify and quantify to what degree individuals deviated from normal behavior on a trait. It also identified individuals "in need of mental hygiene" (Vinchur, 2018). Questions included: "Have you felt a strong desire to take things that were not yours?" and "Has it made you uneasy to sit in a small room with the door shut?" (Laird, 1925). Thus, items on the test measured an individual's propensity to show psychoneurotic traits. However, researchers later added an extraversion-introversion dimension to the test. Moreover, the Thurstone Personality Schedule incorporated items from both the WPDS and the Colgate

Inventory (Vinchur, 2018). The test included 223 yes/no items that aimed to measure a participant's psychoneurotic tendencies (Harvey, 1932). Sample questions include: "Have you ever had a nervous breakdown?" and "Did you ever have a strong desire to run away from home?" (Conklin, 1937). Researchers used a predetermined key that comprised of how neurotic individuals would answer each item and compared those scores of the respondent. Although they were not used widely in industry, the Colgate Mental Hygiene Test was administered to female nursing and female liberal arts college students to determine the kinds of personality traits exhibited by nurses for research purposes. The results from the test found that women nursing students were more emotionally stable and extroverted than the college students (Elwood, 1927). The Thurstone Personality Schedule was used to determine adjustment difficulties faced by women teachers. Therefore, these tests were primarily used to identify, and also diagnose individuals who demonstrated psycho neuroticism.

In the 1930's, there was a transition from unidimensional to multidimensional tests with the establishment of the Bernreuter Personality Inventory (BPI) and the Humm-Wadsworth Temperament Scale (HWTS) (Gibby & Zickar, 2008). While these tests borrowed items from the Colgate Mental Hygiene test and the Thurstone Personality Schedule, they did not focus solely on measuring neuroticism, but also measured a wider range of personality constructs. The BPI, developed by Robert Bernreuter, consisted of 125 items that measured four dimensions of personality: Neurotic Tendency, Self-Sufficiency, Introversion-Extraversion, and Dominance-Submission. In 1935, the dimensions of Self-Consciousness and Solitariness were added to the test. Although the BPI measured multiple dimensions of personality, it retained a focus on maladjustment,

as measurement “was concentrated upon the maladaptive continuum” of each dimension (Gibby & Zickar, 2008). However, the BPI’s popularity as one of the first multi-dimensional tests of personality ensured its wide application to industry and personnel selection. Employers administered the test to dance instructors and salesmen to assess extraversion, as these types of occupations rely on comfort with interpersonal communication. It too was employed as a means of screening out maladjusted individuals and identifying healthy individuals in a number of occupations such as salespersons, given the addition of extraversion-introversion, and engineers (Gibby & Zickar, 2008). Thus, while the BPI did consist of a wider set of personality constructs, it was still employed, to a certain extent, as a means to eliminate maladjusted individuals.

The HWTS, which measured seven dimensions², was developed as a response to workplace violence in which an employee murdered a supervisor (Gibby & Zickar, 2008). It was based on a theory of personality which posited that all individuals possessed some degree of abnormal traits, namely, antisocial, cyclothymic³, autistic, and epileptic traits. However, those with normal personalities were more successful in inhibiting these abnormal traits (Gibby & Zickar, 2008). Sample questions on the test include “Would you mind work which would require being alone for long periods?” and “Does a snappy pep talk make you want to try harder?” Responses to the items were then statistically compared to responses from normal respondents (Wiggins, Behrends, Ben-Porath, and Blatt, 2003). Thus, as indicated by the items on the scale, the test sought to

² hysteroid, manic, depressive, autistic, paranoid, epileptoid, self-mastery, and a response bias scale

³ Cyclothymic means manic-depressive

measure the emotional disposition and social dimensions of a subject's personality to determine suitability to a job (Lussier, 2018). Like the BPI, the HWTS also enjoyed immense popularity in industry given due to marketing efforts on the part of the tests makers. It was used in the selection of individuals for a wide range of jobs, particularly jobs where temperament and emotional stability were important, such as those of policeman or pilots. Lockheed Martin, an aerospace and defense-security company, implemented the HWTS for all potential employees to screen out those with emotional instability (Gibby & Zickar, 2008). Thus, in the realm of job fit, the HWTS was essentially to identify applicants who were well-adjusted, and eliminate those with temperamental issues and emotional maladjustment.

Tests assessing maladjustment continued in the 1940's with the development of the Minnesota Multiphasic Personality Inventory (MMPI). Yet, this was also a period where personality inventories assessed a wide variety of dimensions, with relatively weak ties to maladjustment. The MMPI, formulated by Hathaway and McKinley in 1939, was focused on identifying individuals with adjustment problems (Gibby & Zickar, 2008). Sample true and false questions on the MMPI include: "At times I have fits of laughing and crying that I cannot control" and "I see things or animals or people around me that others do not see" (Seedhouse, 2013). In developing the MMPI, Hathaway and McKinley took an approach that differed from previous tests, such as the HWTS, which were based on pre-existing theories. Rather, Hathaway and McKinley developed a number of items that they administered to psychiatric patients and then compared responses to normal samples. The items that discriminated psychiatric patients from normal patients were then used to develop the scales of the MMPI. Before the MMPI, personality tests such as the

HWTS focused upon the maladjustment of otherwise “psychologically healthy individuals” (Gibby & Zickar, 2008). However, the MMPI was targeted towards individuals who demonstrated more severe mental illnesses. For example, the MMPI initially targeted mental hospital patients (Gibby & Zickar, 2008). The MMPI was used predominantly in clinical settings to measure psychopathology or mental disorders such as schizophrenia and hysteria (Gibby & Zickar, 2008). It was also used with war veterans who faced difficulties reentering the workforce. (Christiansen & Tett, 2014) Researchers believed this was due to psychological problems, such as shell-shock, and therefore administered the MMPI to identify veterans with mental disorders (Zickar & Kostek, 2014).

Although the MMPI was designed for use in clinical settings, it was also used in industry. In 2001, the MMPI was eventually revised to the MMPI-2. Today, high risk jobs, such as those in security and defense (policemen, nuclear plant personnel, fire-fighters), require applicants to take the MMPI-2 to ensure that potential candidates do not suffer from psychological disorders (“MMPI-2 Overview”, n.d.). However, the American Disabilities Act (ADA), instituted in 1990, also prohibited the use of personality tests that screened out applications with disabilities. This had an impact on the use of the MMPI, as the ADA Act did not allow for medical examinations to be used in the pre-employment process. Since the MMPI was used in the diagnosis of mental disorders it was considered a medical examination (Goldstein & Epstein, 2008). Yet, the MMPI could be administered once an employee begins work to determine if he/she poses a direct threat

or has a medical condition that hinders performance in the job⁴. Thus, the MMPI, similar to the HWTS, is generally administered in “sensitive” jobs to eliminate and screen out candidates who demonstrate some form of psychopathology (Riggio, 2018); however, its primary application remains in “inpatient and outpatient” psychiatry settings (as cited in Gibby & Zickar, 2008).

Apart from the MMPI, Cattell’s Sixteen Personality Factor Questionnaire (16PF) and the Myers Briggs Indicator (MBTI) also rose to prominence as tests that moved away from maladjustment, incorporating a broader range of personality dimensions. Cattell’s 16PF consists of 16 personality factors, which Cattell believed to be the building blocks of personality. Cattell’s approach to his model was lexical, an approach that posits that the most salient personality traits are encoded within languages (as cited in John & Srivastava, 2011). Thus, he drew upon Allport’s list of 17,935 trait words gathered from an English dictionary. He then used questionnaires and applied factor analysis to reduce and group Allport’s traits into a 16 Personality factor model (Cattell 1943; Cattell, 1957). Cattell further summarized these 16 factors into five primary scales (Appendix C). Only three of the sixteen factors had some relation to adjustment. These include Emotional Stability (Factor C), Apprehension (Factor O), and Tension (Factor Q4). While Cattell’s test could predict adjustment-related criteria, it was also able to successfully predict criteria that was not related to adjustment. This is because the 16PF provided an in-depth assessment of the individual as a whole using a wide range of dimensions (Cattell &

⁴ The extent to which companies follow these rules are unknown.

Mead, 2008). For example, the 16 PF was used to select salespersons, as well as predict job performance for police officers.

The MBTI was developed by Isobel and Katherine Briggs in 1943. The founders of the test were motivated to develop a measure that would allow provide insight into individual preferences in relation to the war effort. They both believed that a greater understanding of preference would help women determine war-time jobs they were suited for when entering the workforce (Myers & Myers, 1995). It was not until the 1960s that companies began employing the MBTI for job assessment in the United States (Hosie, 2017). The MBTI was based on Carl Jung's type theory which posited that every individual has a personality type (Zickar & Kostek, 2014). It consisted of four personality types - Extroversion/Introversion, Sensing/Intuition, Thinking /Feeling, and Judgement/Perception (Zickar & Kostek, 2014). On completion of the test, each individual receives a personality type that comprised of one of each preference (ex: ESFJ). The underlying aim of the test is to measure an individual's personality type and determine which roles they are most suited to. Today, the Myers Briggs Type Indicator (MBTI) is one of the most widely used tests, employed by more than 88% of Fortune 500 companies ("Myers-Briggs Type Indicator", n.d.).

From the 1960s to the 1980s, however, researchers paid little attention to the use of personality measures. This was largely attributed to the results of Guion and Gottier's study which found poor validity for personality measures and their utilization in job performance. Moreover, another research study conducted by Mischel (1968) suggested that situations, rather than traits, predicted and determined behavior (Johnson, 2003).

Although Tupes and Christal released their paper on the five-factor model of personality in 1964, their paper did not receive any attention as it was published in an obscure journal. As a result of these factors, personality testing faded into the background.

A congressional hearing in 1965 on the use of paper and pencil personality tests, namely the MMPI, was a pivotal reason for the decline of these types of personality tests. The denigration of the MMPI “tainted” all other personality tests, thus leading to the reduction in their use in pre-employment screening (Haney, 1982). Furthermore, the Civil Rights Act of 1964, which prohibited employers from discriminating based on age, sex and race, also attributed to the decline in personality testing. Although Title VII of the act did not preclude the use of personality testing, it propagated the use of personality tests only as a measure of job performance. Personality tests that resulted in a racial imbalance, for example, tests that excluded more blacks than white, were deemed unlawful. Thus, the fall of the MMPI, legal concerns with the use of personality tests, and a greater concern for the protection of minority groups contributed to the decline of personality tests (Haney, 1982).

A notable resurgence in the use of personality tests was observed beginning in the 1980’s. This could be attributed to the findings of supportive research studies (Barrick & Mount, 1991; Salgado, 1997) and the establishment of the Big Five Factor Model, an agreed upon taxonomy for classifying personality traits. Although studies yielded low correlations between the five factor dimensions of personality and job performance (Barrick & Mount, 1991; Salgado, 1997), they were generally optimistic about their findings suggesting that personality had a place in the assessment of job performance

(Morgenson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt, 2007). As a result of these studies, a number of companies have now adopted or are beginning to adopt personality tests as a component of their hiring process. Moreover, the institution of the Employee Polygraph Protection Act in 1988 banned private employers from using polygraph tests. Thus, many companies opted to switch to personality tests (Stabile, 2008). Some of the most popular personality tests, used today, based on the Five-factor model, include the Hogan Personality Inventory, the NEO-PI-3, the Caliper Profile, and the Occupational Personality Inventory.

Big Five Tests

The Hogan Personality Inventory, formulated in 1980, is an inventory of normal or bright-side personality based on Five Factor Model of Personality designed specifically for use by the business community (Appendix D). Bright side personality encompasses positive personality traits such as conscientiousness and sociability. For example, individuals with bright side personalities do not exhibit psychoticism (Ferguson, Semper, Yates, Fitzgerald, Skatova, & James, 2014). Furthermore, the HPI was also grounded in socio-analytic theory which states that individuals are biologically motivated to engage in social interaction (Hogan & Hogan, 2007). The HPI has seven primary scales: Adjustment, Ambition, Sociability, Interpersonal Sensitivity, Prudence, Inquisitive, and Learning Approach (Hogan & Hogan, 2007). Hogan developed his scale using the FFM, but he believed that the Big Five model consisted of seven components. Thus, he split extraversion into sociability and ambition, and openness to experience into inquisitive and learning approach (Appendix E). Moreover, each of the seven

components were further broken down into groups of “related sub themes” or homogenous item composites (HICs) (Salgado, Moscoso, Alonso, 2013). The scales can be found in Appendix F. These HIC scales facilitated the development of criterion-related occupational scales. The latest version of the HPI comprises of six occupational scales: Service Orientation, Stress Tolerance, Reliability, Clerical Potential, Sales Potential, and Managerial potential.

While adjustment occupies one component on the HPI’s scale, a large proportion of the test evaluates an individual's ability to work with others in social settings, their leadership capabilities, and dependability. Moreover, the HPI is not intended to be used in psychiatric or clinical settings, or for the evaluation of mental illnesses and disorders (Hogan & Hogan, 2007). With respect to job-fit, the HPI is used across a wide variety of employers in personnel selection from HR specialists to consultants. A number of companies also employ the HPI as a pre-hire tool for managerial positions given that managers are expected to be able to work collaboratively and demonstrate leadership skills, traits that are included in the scales and subscales of the HPI (Hogan & Hogan, 2007).

The Neuroticism Extraversion Openness Five Factor Inventory (NEO) was designed specifically to test for the FFM. Its first version only assessed three factors (neuroticism, extraversion, and openness to experience); however, in subsequent revisions the test encompassed all five factors of the FFM by including conscientiousness and agreeableness. The second version, NEO-PI-R, is a 240 item questionnaire that examines 6 traits for each of the 5 personality dimensions (Appendix G) (McCrae, Costa,

& Martin, 2005). The NEO-PI-R also comes in two forms- a self-report and an observer report. In the observer report, questions are framed in third person thus allowing third party assessment of personality. However, the NEO-PI-R was revised to the NEO-PI-3 to make it more accessible to adolescents, who were unable to comprehend the meanings of some of the words in the items. For example, words such as fastidious, adhere, and lackadaisical, were removed (McCrae et al., 2005). Studies investigating the NEO-PI-R also revealed a low loading factor (correlation) for items in Openness, Extraversion, and Agreeableness (Aluja, Garcia, Rossier, & Garcia 2004), thus threatening the construct validity of the test. As a result, McCrae developed the NEO-PI- 3 which was more readable and also had a slightly higher internal consistency (Appendix G) (McCrae et al., 2005).

Unlike the HPI, the NEO-PI-R and its subsequent versions are used in vocational assessments to assess the strengths and weakness of a candidate and determine the occupational area to which they are most suited to (Costa, 1996). For examples, openness was related to artistic and investigative interests, and extraversion was found to be related to social and enterprising vocations. One study found correlations between NEO-PI-R scores and supervisor ratings of job performance. Results from the study highlighted that extraversion was positively correlated to oral expression, agreeableness was related to social interactions, and conscientiousness was related to a wide variety of performance dimensions including quality of work, oral expression, adaptability, and adherence to rules (as cited in Costa, 1996). Moreover, the NEO-PI-R is also used in clinical settings to identify adjustment issues and assist in diagnosis and counselling (Costa & McCrae,

2008). Thus, given the NEO-PI-R's validity to job performance, it is also used for employee placement and selection.

The Caliper Profile consists of 5 sections of a 112 questions. The test measures 22 personal attributes which are grouped into four competency categories: Leadership/Persuasiveness, Interpersonal/ Service Orientation, Problem Solving and Decision Making, and Personal Organization and Time Management ("Caliper Profile Overview", n.d.). The specific traits or personal attributes examine how an individual approaches each competency category (Appendix H). The Caliper Profile was partially formulated from constructs in the Bernreuter Personality Inventory (Prewett, Tett, & Christiansen, 2013). According to its website, the Caliper Profile is an instrument designed to measure an individual's job performance potential and determine whether the individual is suited to the job. It can be used in hiring, employee development, and team development ("Caliper Profile Overview", n.d.).

The Occupational Personality Inventory (OPQ), developed in 1984, is widely used both in personnel selection and staff development. The OPQ combines a range of personality theories and management style constructs formulated by researchers Cattell and Eysenck (Saville, Sik, Nyfield, Hackston, & MacIver, 1996). The OPQ measures personality at three level: a five-factor level (based on the Big Five Personality model), a 16 factor solution, and a concept model consisting of 30 scales (Appendix I). The OPQ was further developed and revised to the OPQ 32 which was more reliable and also focused more on job-related traits such as sociability, creativity, and analytical skills (Appendix J). The OPQ 32 measures 32 facets of personality relevant to selection,

promotion, counselling, team building, and training. The 32 facets are grouped into three domains: Relationships with People, Thinking Style, and Feelings and Emotions (“Occupational Personality Questionnaire, n.d.). The OPQ 32 is used in assessing the behavioral and thinking styles of individuals to determine which job they are best suited to. Moreover it is also used in personnel selection, counselling, development, promotion, and team-building (“Occupational Personality Questionnaire, n.d.).

Similarities and Differences between the FFM tests

All four personality tests can be used in personnel selection, however the degree to which they are used differs. Although the NEO-PI-R can be used in the job context as it assesses normal personality, it is primarily employed in clinical settings. The HPI, on the other hand, is heavily marketed towards the business community. The makers of the HPI, however, acknowledge that it has potential clinical value. The HPI’s School Success Scale correlated with the MMPI’s Psychopathic scale and its Adjustment scale can potentially measure a participant’s inclination towards neuroticism. Further research would need to be conducted to assess whether it should be deployed in clinical settings (Hogan & Hogan, 1995). The OPQ is also more popular in vocational settings as compared to personnel selection since it provides test-takers with an overview of their strengths and weaknesses and the jobs that most align with their skill-set.

Moreover, both the HPI and the Caliper Profile emphasize interpersonal communication and social interaction as reflected in their scales. The HPI has an interpersonal sensitivity and sociability scale (Appendix F) while the Caliper Profile has an Interpersonal Service Orientation scale. However, unlike the HPI, NEO-PI-R and the

OPQ, the Caliper Profile also examines the candidate's problem solving skills with abstract and numerical reasoning tests (Appendix H).

Conclusion

As the history of personality testing stemmed from military recruitment, tests developed during WWI, such as the HWTS, focused upon screening out individuals with maladjustment. The changing needs of the military in WWII, establishment of legislation that prohibited medical examinations in pre-employment, and the incorporation of personality assessment in industry, however, led to the development of inventories that tested a wide variety of personality constructs. Both researchers and employers shifted from a 'screen out' to a 'screen in' approach, formulating assessments that holistically assessed a candidate rather than his/her sole propensity towards counterproductive behaviors. Moreover, greater research directed into the realm of personality facilitated the establishment of the Five-factor model, giving researchers a structured and pre-determined taxonomy to investigate relationships between personality and job performance. Tests based upon the FFM, such as the HPI and the OPQ, also expanded upon the dimensions of job performance, measuring constructs like leadership effectiveness and interpersonal skills.

Chapter 3

The Limitations of Personality Tests in Predicting Job Performance

Introduction

The previous chapter outlined the history and evolution of personality tests from the 20th -21st century. In this chapter, I will examine some of the concerns with validity and reliability, methodological issues, as well as the ability of these tests to predict job performance and other work-related dimensions. By doing so, I will be able to assess inherent limitations within each test as well as limitations common to a number of personality tests, namely, the problem of faking and response distortion. Derivative tests of the Five Factor Model, such as Hogan's Personality Test and the NEO-PI-R Personality test, will be discussed in the next chapter, given the breadth of research on the five factor model in the context of its use in predicting job fit and performance.

Unidimensional Psychometric Tests

As mentioned earlier, early personality tests predominantly emphasized maladjustment, namely tests like the Woodworth's Personal Data Sheet, the Colgate Mental Hygiene Test, and the Thurstone Personality Schedule. Despite their use in industry and clinical settings, all three tests suffered from concerns with validity.

The concerns with the reliability and validity of test measures began to surface with the development of the WPDS in 1917. Studies examining the WPDS found the test to have poor construct validity. For example, many of the items on the WPDS appeared to measure different constructs. Items such as "Have you ever lost your memory for a

time?” or “Can you sit still without fidgeting” appear to be measuring a diverse set of constructs given current knowledge of clinical psychology (Strauss & Smith, 2009). For example, the first item could be assessing dementia or Alzheimer’s while the second item could be potentially getting at individuals with anxiety disorders or individuals with Attention Deficit Hyperactivity Disorder. As a result, researchers who attempted to ascertain the validity of this test found that it was unable to differentiate between college students and individuals with psycho-neuroticism (Garrett & Schneck, 1928). Moreover, another study found that scores on the WPDS did not correlate with teachers ratings of emotional instability in students (Fleming & Fleming, 1929). The limitations to the WPDS largely stemmed from the fact that the study and understanding of psychopathology as a field was limited in itself. In the construction of the WPDS, Woodworth relied upon case studies of neurotic individuals; however, the authors of these case studies did not possess sufficient knowledge of psychopathology (Strauss & Smith, 2009). Thus, items on the WPDS had poor construct validity as they did not accurately measure what they intended to measure, i.e., psychopathology and emotional instability in the individual. The Colgate Mental Hygiene test and Thurstone Personality schedule likely suffered the same dilemma, given that psychopathology was still an emerging field.

One of the limitations of the Thurstone Personality Schedule is the inability of the test to distinguish between psychotic or psychopathic individual and normal individuals. Moreover, even if an individual has a high neurotic score, this does not necessitate that he/she will “crack” under the pressure of a neurotic breakdown, especially if he/she has “strong mental constitution” (Thurstone & Thurstone, n.d.). On the other hand, an

individual with few neurotic tendencies might be more affected by a neurotic breakdowns (Thurstone & Thurstone, n.d.). Thus, the test can only show us which individuals experience more neurotic difficulties and tendencies, not which individuals are more affected by them.

The unidimensional nature of these three tests were arguably the most significant limitations to their use in the realm of job assessments. All three tests focused on maladjustment which did not allow employers to assess whether the candidate has the desired traits and characteristics required by the job. Consequently, researchers forayed into developing multi-dimensional tests to address this drawback.

Multidimensional Psychometric Testing

Building on the Woodworth Personal Data Sheet (WPDS), the Bernreuter Personality Inventory (BPI) and Humm-Wadsworth Temperament Scale (HWTS) were some of the first multidimensional personality tests that were used for employee selection in the job market. As noted in the prior chapter, these tests were immensely popular as they were able to assess a wide variety of personality constructs as compared to unidimensional tests which generally focused upon neuroticism.

Despite the BPI's use in personnel selection for a wide variety of occupations, its validity for predicting performance and behavior was lacking (Kanfer, Ackerman, Murtha, & Goff, 2011). In Super's review of research conducted on the BPI, it was noted that the test was unable to differentiate between abnormal and normal individuals. For example, it was found that some normal individuals reported a greater inclination to daydream and "have ideas run through their head" than neurotic individuals (Super,

1942). These unexpected results were attributable to the methodology used to formulate the items on the test. Items were selected based on behaviors that were believed to have psychological significance in clinical studies; this approach was logical rather than psychological (Super, 1942). Moreover, studies found that the BPI was poor at predicting job performance. Research found that there was little correlation between scores on the BPI and work-related criteria for salesmen, grocers, and cotton-mill supervisors (as cited in Gibby & Zickar, 2008). Thus methodological issues and concerns with validity hindered the applicability of the BPI to accurately predict job performance in personnel selection.

The HWTS has serious concerns with validity. While the developers of the test claim a validity coefficient of 0.85, subsequent studies have produced validities of 0.65 and 0.40. The ability of the HWTS to predict job performance has also been called into question. In one study, 405 employees were made to take the HWTS. Ten years later, researchers found that 191 individuals were employed at the same company, 75 employees had been dismissed, and 139 individuals had terminated employment without any “unfavorable service record”. Of the 191 still employed only 9.4 % had good profile ratings on the HWTS. From the 75 who had been dismissed, 12% had good ratings (Humm & Wadsworth, n.d).

Although the BPI and the HWTS measured a larger set of personality constructs than preceding tests, the overarching focus, again, remained on maladjustment. The focus on negative aspects of personality is likely the underlying cause of the BPI’s poor prediction of sales performance. This emphasis on maladjustment coupled with the

inability of these tests to accurately distinguish between normal and abnormal individuals limited their use in industry and clinical settings.

Minnesota Multiphasic Personality Inventory (MMPI)

The MMPI fell out of favor in the 1970s as a consequence of the American Disabilities Act. Moreover, some of the test concepts of the MMPI became outdated with advancements in the realm of psychopathology. The MMPI was thus revised in 1989 with the development of the MMPI-2 which used a new standardized sample of representative individuals than the MMPI. Although a large proportion of the MMPI scales are retained on the MMPI-2 scales (Butcher, Graham, and Ben-Porath, 1995), researchers expanded the scope of measurement by including sub-scales like Shyness/Self-Consciousness, Social Avoidance, and Self/Other Alienation (Ben-Porath, Hostetler, Butcher, & Graham, 1989).

The limitations of the MMPI and the MMPI-2 in the prediction of job performance and fit are similar to both the BPI and the HWTS. In the realm of job performance and fit, the MMPI is most heavily used in the selection of police officers (Lough, & Treuer, 2013). However, a large number of studies found either negative or inconclusive results for the MMPI as a good predictor of performance issues or job performance in police officers (Daniels & King, 2002). For example, a study comparing a group of problem police performers with a control group found that the MMPI could not distinguish between the MMPI profiles of both groups. The MMPI was designed to predict psychopathology and was not intended to be used as an identifier for behaviors that do not fall under the breadth of psychopathology. Researchers have thus suggested

that scales which measure desirable and positive traits are more beneficial and suited in determining whether a candidate has the required characteristics for the job. The use of the MMPI is recommended only when it is supplemented with other instruments that are able to measure more than just psychopathology (Daniels & King, 2002).

Methodologically, scores on the MMPI were compared to scores from a group of 725 Minnesotan normal individuals. These individuals set the behavioral norms which were used for baseline comparisons. However, this sample of “normals” were comprised of solely white individuals where the average member had an 8th grade education. Researchers deemed the sample to be too normal and unrepresentative of the population in the United States. Thus, in the revised MMPI-2, the normal sample was made to be more representative consisting of individuals from varied ethnic races. Furthermore, the revised MMPI-2 also dropped any sexist and offensive questions in the original MMPI (Brataas, 1989). Moreover, the MMPI-2 was found to have better predictive validity for disruptive behaviors in patients with personality disorders (Scholte, Tiemens, Verheul, Meerman, Egger, Hutschemaekers, 2012). With respect to job performance, the MMPI-2 is useful in predicting counterproductive work behaviors. For example, scores on the Hypochondriasis, Schizophrenia, and Social Introversion scale were positive related to police counter productivity (Porath, n.d.).

Cattell's 16PF and the MBTI

One of the most significant limitations to Cattell's 16PF questionnaire has to do with the replication of his 16 Factor Model of Personality. In 1971, a study conducted by Howarth and Brown found that 10 of the 16 factors failed to relate to items in his model,

suggesting that the 16PF did not accurately measure the factors it intended to (as cited in Fehring, 2001) . Furthermore, multiple studies have found that Cattell's primary factor did not replicate in their factor analysis (as cited in Eysenck, 1977). Methodologically, Cattell's reliance on a lexical approach also has its drawbacks. One limitation is that language differs across communities and individuals, therefore there is limited generalizability if a personality model is developed from within one language. For example, in China, researchers have identified a tradition factor that is not included in the Big Five model of personality (Gurven, Rueden, Massenkoff, Kaplan, & Vie, 2013) A second limitation is that some of these trait terms are ambiguous and loosely defined. For example, the adjective “outgoing” can take on varied meanings in different contexts. Outgoing can refer to behavior, such as socializing, and it can also refer to a mental state, such as an individual who feels comfortable around strangers (Saucier & Srivastava, 2008). This has been acknowledged by researchers, such as Allport, who stated that he was dissatisfied with using “natural language” as a basis to compile personality traits (as cited in John, Angleitner, & Ostendorf, 1988). Thus, issues with replication and methodology have called into question whether the 16PF is a valid measure of personality.

With regards to job performance, Cattell’s 16PF questionnaire has largely been successful in predicting job-related dimensions such as creativity, social skills, leadership styles, and team roles. Moreover, the questionnaire has also been used to determine the personality traits of successful supervisors, executives, and managers. It has also been shown to be effective in predicting the personality profiles for salespersons and police officers. For example, successful salespersons tend to score on Extraversion and the traits

of Warmth, Social Boldness, Liveliness, and Group Orientation. Moreover, salespersons tend to be low on Anxiety and the sub-traits of Apprehension and Emotional Stability (Cattell & Mead, 2008). Thus, the 16PF questionnaire is regarded as an effective tool in predicting the profiles of individuals in a variety of occupations.

The Myers-Briggs Type Indicator (MBTI), in particular, has been subject to widespread criticism despite revisions to the test and its continued popularity in industrial and clinical settings. Research has shown that the MBTI suffers from poor reliability and validity. Studies investigating the MBTI have highlighted its low test-retest ability coefficient (Pittenger, 1993). Thus, a person's personality type, as determined by the MBTI, is susceptible to change even during a short time frame. This makes it difficult to ascertain whether one's personality preference on the first attempt of the MBTI is actually one's personality type. With respect to validity, researchers have conducted a factor analysis to determine whether the four factors in the MBTI exist, and whether an individual's MBTI type allows us to predict how he/she will perform in different situations. However, the results of study assessing the MBTI's use of factor analysis, with college students as samples, found six factors instead of four. The study also highlighted that 83% of the differences among the colleges students could not be accounted for by the MBTI, suggesting that the factors found in the statistical analysis were not consistent with MBTI theory (as cited in Pittenger, 1983)

Despite the popularity and proliferation of the MBTI, researchers have highlighted a number of issues and limitations within the test. Psychologists claim that the MBTI is based on a very simplistic type theory which assumes that individuals fall

explicitly on one of the two polar dimensions. According to the MBTI, a person is either extroverted or introverted; there is no middle ground. Psychologists view personality on a continuum where individuals can fall on a range of multiple dimensions. (Zickar & Kostek, 2014). The MBTI, however, does not allow for this as individuals are stereotyped into one of the two dimensions of each personality type.

With regards to job selection, the MBTI tests individuals in a specific job to determine the personality profile for that job. Thus, in hiring decisions, some companies administer the MBTI to examine whether the potential candidate fits the personality profile of the role that they are interviewing for. This approach, however, has limitations. There is the possibility that a person employed in a particular profession was chosen for reasons unrelated with their ability to do the work. For example, employers may hire individuals they believe to be authoritarian for police work, but this does not necessitate that the individual will perform well in the job. Furthermore, there is also the argument that one's personality changes, as success in an occupation may demand that individuals adapt behaviorally and cognitively. Therefore, when measuring individuals who are currently in a profession, there is the confounding bias of the environment that produces a change in their personality. Thus, the MBTI possesses limitation in its prediction of job performance given its emphasis on determining the job profile of candidates (Pittenger, 1993). Moreover, the MBTI is not intended to be used as a measure for selecting employees. While it appears intuitive that certain personality types would perform well in certain jobs, the use of the MBTI in excluding a candidate for selection is not recommended (Coe, 1992). This is because the MBTI classifies individuals into strict categories (introvert vs extroverted); however, some individuals may exhibit facets from

all eight dimensions of the MBTI. Lastly, the MBTI is also not an indicator of how well a candidate can perform tasks as it only measures an individual's personality type (Coe, 1992).

Literature on faking

The literature on faking, or response distortion, and the consequences of faking on personality tests has been subject of serious debate among psychologists in the academic community. Some researchers argue that faking can fundamentally threaten the validity of personality testing, while others claim that it has limited impact upon criterion related validity. Faking itself can manifest in three forms: 1) random faking, where candidates deliberately sabotage results in a random fashion, 2) motivational distortion, where respondents intentionally fake to achieve a certain profile, and 3) sheer ignorance, where respondents lack self-understanding and are therefore unable to accurately respond to items on a test (Furnham, 1990). The second form of faking, motivational or response distortion, is the most commonplace in personality testing (Furnham, 1990). In motivated response distortion, candidates may be motivated to present an image that reflects a positive self-concept, matches what the candidate believes the job demands, and exhibits traits of what they believe is the ideal employee (Rosse, Stecher, Miller, & Levin, 1998).

Response distortion consists of two separate constructs: self-deception and impression management. Self-deception is defined as a “tendency to think of oneself in a favorable light,” while impression management is an intentional attempt to distort responses “in order to create a favorable impression with others” (Barrick & Mount, 1996). For example, some personality tests include items like: “I am a sociable person” or

“I am a hard-working person” (Barrick & Mount, 1996). It is thus easy for applicants to distort responses on such items by making themselves appear more favorable if they choose to do so (Barrick & Mount, 1996).

Research has shown that respondents have the ability to fake in a socially desirable manner. A meta-analytic study found that participants were able to elevate their scores on a personality measure, when instructed to do so (Viswesvaran & Ones, 1999). However, the research on whether response distortion affects the validity of personality assessments for job performance is mixed. Some research studies have shown that faking has little impact on the validity of a personality test (Barrick & Mount, 1996). In Barrick and Mount’s study, it was found that although participants engaged in self-deception and impression management, neither type of distortion decreased the predictive validities of the Big Five factors (1996). Another study revealed that less than one third of comparisons in validity between accurate and faked responses were significantly different (as cited in Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007). Lastly, another study examined whether faking reduced the validity of personality measures by statistically controlling for social desirability (Ones, Viswesvaran, & Reiss, 1996). Results showed no effect on criterion validity of the Big Five personality traits on job performance even when social desirability was singled out. Thus, the researchers concluded that social desirability or response distortion had minimal impact upon personality testing and thus did not threaten personnel selection (as cited in Griffith, Chmielowski, & Yoshita, 2005).

While some argue that faking is a non-issue when it comes to the use of personality tests in selection, other researchers assert that faking can impact rank-order decisions of candidates. If candidates are ranked based on their scores on a personality tests, an applicant that fakes may have an advantage over a candidate who does not fake (Mogensen et al., 2007). Thus, the debate on the impact of faking in recruitment and personnel decisions is still in contention. For example, one study examining the extent to which Big Five personality traits could be faked, found that faking can change the rank order of high scoring participants (Hartman & Grubb, 2011)

The Benefits of Personality Testing in Industry

Given the limitations to personality testing, why do companies still choose to employ them as a tool in personnel selection?

One of the most important reasons that companies employ personality tests, is to assess whether a candidate is a good fit for a specific job (Stabile, 2002). In some jobs, individuals that possess a specific trait are more likely to perform better than individuals who do not. For example, recruiters looking to hire for a sales position might veer towards candidates who are extroverted, due to the high degree of interpersonal and customer contact that sales require (Barrick & Mount, 1991). Thus, a personality test can help determine if an individual has traits that are suited to a job, or conversely, traits that will inhibit successful performance. However, this claimed benefit of personality is still in contention given evidence that contradicts a relationship between personality and job performance.

While the use of personality tests is still contested, researchers have identified both financial and legal advantages to employing these tests as a tool for personnel selection. Firms are now continually opting to use personality tests to reduce the costs and legal risks of making poor hiring decisions. The cost of replacing a bad hire can be significant to companies; it is estimated to be 1.5 times the workers' salaries and benefits (Stabile, 2002). Companies also look to avoid the financial cost of high employee attrition and turnover. Therefore, personality tests can screen for candidates who will likely be successful at the job, and thus less likely to quit.

Moreover, workers who engage in theft or substance abuse can cause a company significant economic losses. In these cases, a personality tests can predict the likelihood of an employee partaking in disruptive behaviors that are counterproductive to the workplace. There are also legal risks that a company can incur if employees engage in criminal or hostile behavior. The Respondeat Superior Doctrine states that an employer can be liable for criminal acts committed by an employee if the acts falls under the job's scope (Cavico, Mujtaba, Lawrence, and Muffler, 2015). Moreover, background checks and referrals sometimes fail to provide critical information about employees, due to the fear of employee defamation suits (Stabile, 2002). Managers or supervisors are sometimes unwilling to fully disclose details of their employees, often only providing neutral information. Although there have not been a great number of defamation suits, the cost of potentially defending one likely deters individuals from disclosing negative information about employees. Thus, it is in the best interest of the company to screen employees for violent or dishonest traits and tendencies to reduce costs and avoid any legal consequences (Stabile, 2002).

Conclusion

This chapter has addressed inherent limitations of personality tests from the early 1900's to the early 2000's. Until the 1950's, personality tests were largely centered upon screening out or eliminating candidates who exhibited signs of maladjustment. Today, however, companies use broader personality inventories that measure a wider range of personality constructs. Thus, companies are now able to screen in candidates that demonstrate the personality traits they are looking for. Yet, despite the transition to more holistic personality tests, some tests, particularly the MBTI, lack the required validity to be used as effective measures of job performance and fit. The issue of faking also plagues a number of tests, but research is still to determine whether it has a substantial impact upon personnel selection. Given this research, we are left with important questions pertaining to the use of personality tests in job contexts. Do the drawbacks of personality tests outweigh the potential benefits that they can provide to employers when making decisions regarding personnel selection? These questions will be deliberated upon and examined in the subsequent chapter.

Chapter 4

The Big Five Model and Derivative Tests

Introduction

This chapter will briefly trace the origins of the Big Five Model, studies on the Big Five and its ability to predict job performance, as well as limitations of this model of personality. Furthermore, derivative tests, such as The Hogan Personality Inventory, the NEO-PI-3, the Caliper Profile, and the Occupational Personality Questionnaire will also be discussed with respect to their reliability, validity, and limitations.

A brief history of the Big Five Model

The move towards a hierarchical-structure of personality, where each primary dimension has a subset of components traits, can be traced to the work of two German psychologists - Klages (1926) and Baumgarten (1933). Both psychologists adopted a lexical approach to personality, i.e., the hypothesis that the most salient personality traits are encoded within languages (as cited in John & Srivastava, 2011). According to the lexical hypothesis, individual differences that are most relevant to people will be found in the vocabulary of language (as cited in John, Angleitner, & Ostendorf, 1988). Therefore, by looking through the dictionaries of languages, one can find and group words that describe personality traits and differences. Klage developed a list of 4000 words in the German language that related to personality and inner states. Building on Klage's work, Baumgarten gathered a list of 941 trait-descriptive adjectives and 688 nouns from German dictionaries and publications (as cited in John et al., 1988).

Klages and Baumgarten's work provided a foundation for psychologists to continue developing a taxonomy for personality (as cited in John & Srivastava, 2011). Allport and Odbert's study identified 17,953 trait-words from an English dictionary and assembled them into four mutually exclusive categories: personality traits, temporary states, evaluative judgements, and physical characteristics (as cited in John & Srivastava, 1999). Their findings influenced British-American psychologist, Raymond Cattell, whose systematic approach in classifying personality traits is now identified as the cornerstone for the emergence of the five-factor model (Digman, 1990). While Cattell employed a lexical approach, his work was more grounded in objective forms of research, using questionnaires and factor analysis to reduce and group Allport's list of personality traits, eventually arriving at a 16 Personality Factor model (John & Srivastava, 2011). Cattell referred to these 16 factors as surface traits which formed the root base for personality. Cattell was thus able to eliminate more than 99% of terms on Allport's list using factor analyses (John & Srivastava, 2011).

The emergence of a five-factor model for personality (FFM) can be attributed to Donald Fiske's study on personality. Building upon Cattell's work, Fiske analyzed 22 of Cattell's variables and found that only five factors replicated in self-ratings, peer ratings, and observer ratings (as cited in Digman, 1990). Fiske used self and peer ratings of 128 male graduate students and performed a factor analysis, arriving at a five-factor solution for personality. He termed his dimensions Confident Self-Expression, Social Adaptability, Conformity, Emotional Control, and Inquiring Intellect (Fiske 1949). In 1957, Tupes and Christal conducted a study to predict officer effectiveness in the Air Force using rating questionnaires that incorporated Cattell's bipolar scales (as cited in

Digman, 1990). The studies corroborated Fiske's findings of a five factor model of personality; however, the naming of dimensions differed. In their analysis of the correlations between traits, they suggested the following five factors: Surgency, Agreeableness, Dependability, Emotional Stability, and Culture.

Following this, many researchers began replicating and re-examining Tupes and Christal's findings (Borgatta, 1964; Norman, 1967; Smith, 1967). Their studies replicated a five-factor model; however, there were variations in the naming of the dimensions. Norman's identification of the five dimensions- Extraversion, Emotional Stability, Agreeableness, Conscientiousness, and Culture- more closely align to the current traits that we see today (as cited in John & Srivastava, 2011). Ultimately, it was Costa and McCrae's labels for the dimensions of the FFM - Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness- that are most often used by researchers in academic papers (Costa & McCrae 1991).

Definitions of the Big Five Dimension

The first dimension, extraversion, is comprised of facets such as activity, assertiveness, and self-confidence (Tonetti, 2011). Individuals who are extroverted tend to be sociable, assertive, talkative, and active (Barrick & Mount, 1991). The second dimension, neuroticism, is defined as the inability to control one's anxiety, anger and emotionality (Tonetti, 2011). Common traits in this dimension include being angry, embarrassed, and insecure (Barrick & Mount, 1991). The third dimension, agreeableness, refers to "a concern and sensitiveness towards others and their needs" (Tonetti, 2011). Individuals who are agreeable tend to be courteous, flexible, and co-operative (Barrick & Mount,

1991). Conscientiousness, the fourth factor, is defined as the ability to self-regulate in both proactive and inhibitory modes (Tonetti, 2011). Traits associated with conscientiousness include being responsible, organized, and dependable (Barrick & Mount, 1991). Lastly, openness is defined as the “propensity to novelty, the tolerance of different values” and an interest for “different habits and lifestyles (Tonetti, 2011). Individuals who demonstrate openness are inclined towards being imaginative, cultured, curious, and artistically-sensitive (Barrick & Mount, 1991).

The Big Five Model in Job Performance and Fit - Meta Analytic Investigations

In the 1990's, researchers undertook extensive meta-analytic investigations to examine whether the Big Five Model of personality predicted job performance and which traits more closely correlated with dimensions of job performance. Researchers generally divide job performance into two dimensions: task performance and contextual performance. Task performance is defined as the extent to which individuals demonstrate proficiency with technical skills and activities that are required for that specific job (Borman, Bryant, & Dorio, 2010). Contextual performance is defined as the contributions of an employee, beyond technical performance, that impact the organizational, social, and psychological environment of an organization (Borman & Motowidlo, 1997). Contextual performance includes interpersonal facets such as co-operating and maintaining good relationships with other individuals. It also includes motivational aspects, such as “persisting in the face of adversity” and assisting with additional tasks in the organization, which is referred to as job dedication (Scotter & Motowidlo, 1996).

From all three meta-analytic investigations, conscientiousness appeared to be the sole factor that predicts job performance across all occupational groups. A number of meta-analytic investigations have supported this claim. Barrick and Mount's study, which compared the Big Five dimensions to performance in five occupational groups (professionals, police, managers, sales, and skilled/semi-skilled labors), revealed that conscientiousness was a valid predictor for performance across all occupation groups (Appendix K) (1991). Although Barrick and Mount's study used only American samples, these findings were replicated across other meta-analytic studies which used samples from other countries (Aarde, Meiring, Wiernik, 2017; Salgado 1997). Thus, employees who are conscientious tend to be organized and meticulous allowing for greater efficiency in work tasks and meeting performance goals and deadlines.

To a certain extent, emotional stability, which is the opposite of neuroticism, is also regarded as a valid predictor of job performance in a large number of occupational groups; however, the correlation between emotional stability and job performance is weaker than that of conscientiousness. In Salgado's study, which used European samples, he found that emotional stability was a valid predictor across performance criteria and occupations. This finding was supported in other studies (Tett, Jackson, & Rothstein, 2006), but was not supported in Barrick and Mount's study (Appendix K). Barrick and Mount, however, suggest that their unexpected results could be attributed to range-restriction, based on a selecting-out process, where individuals with low emotional stability were already excluded from the labor force (1991). In general, individuals who are not emotionally stable and exhibit neurotic traits will have poor job performance as these characteristics inhibit work performance, while individuals with high levels of

emotional stability are likely to be successful across occupations. Therefore, both conscientiousness and emotional stability are regarded as valid predictors of job performance in all occupational groups.

The remaining three factors extraversion, agreeableness, and openness to experience appear to be valid predictors of job performance only for specific occupations and specific criteria's of performance. Extraversion is a valid predictor of task and contextual performance in jobs which involve a high degree of sociability and interpersonal contact. The results from Barrick and Mount's investigation revealed that extraversion was a valid predictor for jobs with reliance on interpersonal skills, namely, managers and sales-representatives ($\rho = .18$ and $.15$) (1991). As these types of jobs are based on interactions with other individuals, traits such as being talkative, assertive, and gregarious are beneficial to performing well. In contrast, these traits are less important to semi-skilled/skilled jobs which do not emphasize client facing interactions and require more administrative work. Moreover, Salgado's findings support Barrick and Mount's results (1997). The results from the study indicated a positive correlation between extraversion and two occupations: managers and police, (Salgado 1997) as both occupations involve high levels and interactions and co-operation with others (Barrick & Mount, 1991). These studies suggest that extraversion does not generalize to all occupational groups. Rather, extraversion is a good predictor of job performance in careers that require high levels of interpersonal contact and communication.

With respect to agreeableness, Barrick and Mount's study found that it was not a valid predictor for any criterion of job performance. However, another study, conducted

by Hurtz and Donovan, contradicted Barrick and Mount's findings (2000). Their study addressed some of the methodological limitations of previous meta-analytic investigations, as they only included studies with inventories that were designed to measure the Big Five personality constructs. They found that while Agreeableness had no influence on task performance, it was related to the interpersonal component of contextual performance (Hurtz & Donovan, 2000) (Appendix L). Individuals who are agreeable tend to be likeable, flexible, and co-operative and thus perform well in interpersonal interactions. Furthermore, a study in 2002 found an interactive effect between conscientiousness and agreeableness on job performance (Witt, Burke, Barrick, & Mount, 2002). The researchers suggested that individuals who are high on conscientiousness but low on agreeableness may have lower contextual performance. Result from the study found that among highly conscientious workers, those who scored high on agreeableness had higher ratings on job performance than those who scored lower on agreeableness. However, this finding was not held for jobs with little interpersonal interactions and jobs where interactions were characterized by "leading, supervising, and delegating" to others as seen in managerial situations (Witt et al., 2000). Thus, agreeableness appears to be related to the interpersonal dimension of contextual performance.

Lastly, openness to experience is the least predictive factor of the Big Five when it comes to predicting job performance. Barrick and Mount's study found that it was correlated to only one dimension of job performance: training proficiency. This is because individuals who score high on this factor are broad minded and curious, and thus they tend to approach learning experiences with positive attitudes (Barrick & Mount,

1991). Moreover, there are studies which have shown a linkage between openness to experience and creativity (Rothmann & Coetzer, 2003). Moreover, a study conducted by Schilpzand, Herold, and Dhalley using undergraduate students found that teams that were diverse on openness to experience had the highest level of team creativity (2010).

Researchers suggest that individuals who score high on this dimension tend to have more active imaginations and a preference for variety and are thus more able to apply themselves creatively.

Limitations and Methodological issues of Meta-analytic Investigations

A considerable proportion of research dedicated to investigating the relationship between the FFM and job performance is meta-analytic in nature (Barrick & Mount, 1991; Salgado, 1997). A key limitation in these articles, is the methodology used in deriving validity coefficients from past studies (Hurtz & Donovan, 2000). The validity coefficients were predominantly taken from other studies that contained measures which did not explicitly assess the Big Five dimensions. The data obtained from these studies were categorized into Big Five measures post-hoc, thus threatening construct validity. Therefore, it is unclear whether these measures truly mapped onto the Big-Five dimensions. Furthermore, these correlation coefficients are only estimates of validities of the Big Five measures as they were not based on the FFM (Hurtz & Donovan, 2000). Thus, they are not accurate estimates of the true validity of the Big Five dimensions.

Moreover, issues raised with interrater agreement also presented a threat to construct validity. In Barrick and Mount's study, personality scales from prior studies were translated into a Big Five Scale by six trained raters. Raters were given the

definition of the Big Five dimensions and the personality scales with a definition of each inventory. They then proceeded to categorize each inventory into one of the Big Five dimensions. If the inventory did not fit into any of the five dimensions, it was assigned to a sixth miscellaneous dimension. Their study reported a 83 % or better agreement on 68% of the classifications, which Hurtz and Donovan view as “less than desirable interrater agreement” (2000). Thus, given disagreements in classifications, it is plausible that errors were made in classifying scales into the Big Five dimensions. Salgado notes that there is inherent ambiguity in how scales map onto the Big Five dimensions, and this ambiguity makes it difficult for researchers to adequately determine which dimension best fits to a scale (1997).

Inherent Limitations of the Five Factor Model

While there appears to be a consensus on the number of factors, debate regarding the operational definition of each factor continues. Researchers generally agree on Extraversion and Emotional Stability/Neuroticism as the first and second dimension (Digman 1990). However, Hogan’s study suggests that Extraversion should be divided into Sociability and Surgency, thus accounting for a six-factor model of personality (Hogan & Hogan, 1992). Hogan defines sociability as “the degree to which a person seems to need and/or enjoy interacting” (Hogan & Hogan, 1992). Surgency is composed of ambition, characterized by individuals who have a “desire for status, power, recognition, and achievement” (as cited in Hogan & Hogan, 1992). Moreover, there has been debate regarding the third dimension, Agreeableness, with Digman and Takemoto Chock arguing for it to be termed Friendly Compliance versus Hostile Noncompliance

instead (1981). Furthermore, a number of researchers have disagreed on interpretations of Consciousness. Hogan, for example, aligns Consciousness with Dependability - traits that include being detail-oriented and responsible (Hogan 1981). On the other hand, some researchers argue that Consciousness also encompasses volitional variables such as being hardworking and persevering. (Costa & McCrae, 1992). Lastly, the fifth dimension, Openness to Experience or Intellect, appears to be the least concrete due to ambiguity and disagreement in its labelling. Costa & McCrae (1992) suggest Openness to experience, while Goldberg (1993) prefers Intellect. Individuals who demonstrate intellect are said to be intelligent, philosophical, and erudite (DeYoung, Shamosh, Green, Braver, & Gray, 2009).

The Big Five is a broad dimension of personality as it suggests that one's personality can be summed up with just five factors. By doing so, it neglects the importance of narrower traits of personality, such as the sub-traits of the Big Five dimension, and their impact on job performance. Some scholars have contests that broad personality dimensions have better predictive validity than narrower traits (Ones & Viswesvaran, 1996). However, research findings have contradicted this claim, instead suggesting that FFM may be too broad for predicting certain dimensions of job performance. For example, a research study found that the narrow traits of conscientiousness, such as cautiousness, dependability, achievement, and order, provided more incremental validity to predicting contextual performance such as job dedication, counterproductive work behavior, and interpersonal facilitation (Dudley, Orvis, Lebiecki, & Cortina, 2006). Moreover, narrow traits also contributed to incremental validity in predicting overall performance in certain occupational groups such as sales personnel and

managers (Dudley et al., 2006). These findings highlight the usefulness of narrow traits in predicting contextual performance and overall performance in certain occupation, as compared to global dimensions of the five factor model.

The Use of the Big Five Model in Person-Organization Fit

The dimension of person-organization fit is another concept that employers consider in making personnel decisions. Employers want to know whether a candidate's characteristics and values are compatible with their organizations, as this can increase job retention and commitment. Thus, assessing a person's personality can help recruiters and employers determine whether there is a match between the individual and the organization (Anderson, Spataro, Flynn, 2008). A study conducted using the FFM and its relation to organization found that certain factors of the model matched certain organizational cultures. The study included four types of organization cultures: clan culture, hierarchy culture, market culture, and adhocracy culture (Gardner, Reithel, Coglisier, Walumbwa, & Foley, 2012) (Appendix M).

It was found that extraversion and agreeableness were positively correlated to clan cultures. This is because extraverted individuals perform well in environments where interaction and communication with others is essential to the occupation. As clan cultures focus on teamwork and loyalty, individuals who score high on agreeableness are more likely to fit well with clan cultures. These are individuals who tend to be co-operative, courteous, and trusting. Agreeableness was negatively related to market cultures, as market cultures tend to have high levels of competition and potential conflict. However,

individuals who are less agreeable are likely to be a good fit for market cultures as they tend to be more competitive in nature (Gardner et al., 2012).

Conscientiousness was found to be positively correlated to hierarchical cultures. Conscientious individuals tend to be attracted to detail oriented and reliable environments, all facets of hierarchical cultures. Neurotic individuals were also found to be a good fit for hierarchical cultures. As neurotic individuals do not respond well to stressful situations, they are likely to find comfort in structured and predictable environment. As a result, neuroticism is also negatively related to adhocratic cultures which are unstructured and ambiguous in nature (Gardner et al., 2012). However, individuals who score high on openness to experiences tend to prefer adhocracy cultures. As these type of individuals are more open-minded and creative, they are likely to fit in an adhocracy culture which emphasizes flexibility, innovation, and entrepreneurship (Gardner et al., 2012).

Thus, determining the personalities of potential candidates can allow companies to understand a candidate's preference for an organizational culture. For example, consulting firms often have a clan culture given that consultants tend to work on teams when staffed on projects. If an individual shows extraverted or agreeable characteristics of personality, they are likely to be a good fit for consulting rather than individuals who score high on neuroticism. Thus, knowledge of an individual's personality can be an important tool in determining cultural fit, which can in turn help organizations make hiring decisions.

Derivative Tests of the Big Five

The following personality tests- Hogan Personality Test, the NEO Personality Inventory, the Caliper Profile, and the Occupational Personality Questionnaire- are all, to some extent, based upon the five-factor model of personality. The Hogan Personality Inventory, based on socio-analytic theory, focuses on measuring how individuals get along with each other. It is used in the business community, particularly for professions that involve leadership. The NEO Personality Inventory is most closely related to the FFM model. It is used both in clinical settings, to assist in identification and diagnosis of adjustment problems, as well as in occupational settings as a predictor of job performance potential. The Caliper Profile, which assesses how individuals perform on four competency categories (Appendix H), and Occupational Personality Questionnaire are both used vocationally to determine if an individual is suited towards a particular job.

The Hogan Personality Test

With respect to criterion-related validity, the Hogan Personality Inventory (HPI) has generally shown strong ratings. Criterion-related validity is seen when scales of the HPI are used to predict specific performance dimensions. For instance, the correlation coefficient between ambitions with specific dimensions of managerial performance was 0.51. However, correlations between specific scales and overall performance were modest. For example, the maximum correlation for overall performance was found between ambition and managerial performance, with an R-coefficient of 0.29. Thus, the HPI appears to be more effective in predicting certain dimensions of work performance rather than overall performance (Prewett et al., 2013). However, researchers have also

highlighted that studies assessing the HPI's criterion validity have limitations as sample sizes tend to be small.

The HPI also lacks sufficient normative information. Normative information refers to population specific properties of a score distribution such as the means and standard deviations of test scores. The provision of this information allows individuals to determine how they fall on a trait, or how their scores compare relative to a relevant population. However, the HPI's manual does not provide mean scores and standard deviations for male and female participants on the HIC scales. There is also no interpretation or significance for the differences between men and women's responses on the inventory (Hogan, n.d.).

Moreover, the occupational scales in the test have been shown to have overlap with the primary scales as well as overlap within the occupational scales themselves. As a result, researchers have advised users to be cautious when interpreting occupational scales. Another limitation was a lack of internal consistency in the HIC scales of the HPI. Finally, a crucial drawback of the HPI is seen in its consideration of response validity, or its consideration of social desirability and faking on the part of test-takers. While the HPI contains a scale that detects responses that are overly-desirable, the HPI's manual does not offer advice on how to use and interpret these scales, and how to adjust scores in light of faking. As a result, the HPI is regarded to have moderately weak response validity (Prewett et al., 2013). With respect to reliability, researchers generally agree that the HPI's seven components, or broad factor measures have strong reliability. The test-retest reliability is also deemed acceptable by researchers (Prewett et al, 2013).

NEO Personality Inventory (NEO-PI-R) and NEO-PI-3

The NEO-PI-R has good internal reliability on both Form S (self-report) and Form R (observer report). Thus, items on each of the five scale are related to the central construct that they intended to measure. Moreover, the test also has high test-retest reliability as the traits measured have good long-term stability (Costa, 1996). With regard to validity, the NEO-PI-R has shown good construct and consensual validity (Costa, 1996). The NEO-PI-R facet scales correlate with other measures of similar constructs. For instance, the Anxiety scale on the NEO-PI-R is correlated to Spielberger State-Trait Personality Inventory and the tensions scale on the Profile of Mood states, thus showing construct validity. Furthermore, consensual validation, defined as the agreement across observers on questionnaire measures (McCrae & Costa, 1987), is strong on the NEO-PI-R as it has both a self-report and observer-report form. For instance, there is evidence of strong correlations between self-reports and spousal ratings on the test on domain scores (correlations range from 0.5 to 0.6) (Costa, 1996).

The latest version of the NEO personality test, the NEO-PI-3, is regarded to have good construct validity and criterion related validity (Prewett et al., 2013). A large number of studies have used the NEO-PI-R for research on personality and job performance. For example, researchers have found that scales on the NEO-PI-R were correlated to flight attendant success, the performance of managers, and salespersons (as cited in Prewett et al., 2013).

Similar to the HPI, the NEO-PI-3 also suffers from weak response validity and a lack of normative information. The use of a Likert Scale leaves the test open to response

distortion. Moreover, the NEO-PI-3 does not include a social desirability scale that checks for the possibility of distortion and faking (Prewett et al, 2013).

Caliper Profile

The Caliper Profile has strong criterion-related validity. Multiple studies have highlighted correlations between its scales and job performance criteria. The correlation coefficients range from 0.29 to 0.39. However, researchers note that no study has examined the correlations between each component of the sub-scales and job performance. This would allow companies to determine which subscales were most important in predicting the job performance criteria. Moreover, unlike the HPI and NEO-PI-3, the Caliper profile has good response validity. This is due to its semi-ipsative, or forced choice items, which are helpful in decreasing response distortion. However, the Caliper Profile suffers poor internal reliability and construct validity. Items on the Caliper Profile scales were taken based on the predictive validities of items from other personality tests. Moreover, the use of a semi-ipsative format also lowers internal reliability as it produces heterogeneous responses (the alpha values for reliability range from .38 to .68). Thus, while the Caliper Profile displays good criterion related and response related validity, it has weak internal reliability and construct validity (Prewett et al, 2013).

Occupational Personality Questionnaire and OPQ-32

With respect to reliability, the OPQ is said to be weak in some aspects as the concepts that the test measures are not unidimensional. The test-retest reliabilities are deemed satisfactory, ranging from 0.7 to 0.9. The OPQ, however, has low internal reliability in

some cases with coefficients ranging from 0.5 to 0.8. On the other hand, the latest version of the OPQ, the OPQ-32, has much better reliability with internal consistency exceeding 0.70. Criterion validity on the OPQ-32 is also regarded as moderately strong. One study that used a sample of 270 managers found that the correlations between scales on the OPQ and performance criteria such as leadership, interpersonal skills, and planning (Prewett et al., 2013). However, response validity was a concern on the normative version of the OPQ (OPQ-32n) as forced-choice measures are less susceptible to distortion. Moreover, the manual of the OPQ-32n does not provide any recommendations on the interpretation or use of the social desirability scale in the test.

Conclusion

Despite limitations within the five-factor model, it is currently the best taxonomy available for classifying personality traits. Its robustness has been highlighted by a number of studies (Goldberg, 1981). Moreover, FFM inventories are shown to be more valid than inventories that do not use the FFM. A study assessing the validity of FFM and non-FFM based personality measures found that inventories using the FFM had greater criterion validity for conscientiousness and emotional stability when compared to non-FFM inventories (Salgado, 2003). While research has been directed into investigating new taxonomies for personality, we still lack a better alternative model to the FFM.

Given the research on the FFM, companies need to keep in mind that specific personality constructs within the Big Five model correlate to certain dimensions of job performance in occupational groups. Therefore, a company administering a personality test in the recruitment of salespersons should check for personality traits like

extraversion, given its correlation to job performance in sales. Derivative tests from the FFM have been shown to have good criterion related validity. That said, researchers have raised concerns over the issue of response distortion in the NEO Personality Inventory as well as the HPI. Thus, companies should also be aware of the types of personality tests available to them and issues of reliability and validity, particularly response distortion, within each of them.

Chapter 5

The Future of Personality Testing

Introduction

This chapter briefly summarizes the current view of personality testing with respect to the issue of low validities in the prediction of job performance criteria. I then proceed to examine how technology impacts the field of personality and facilitates the development of new methodologies of personality assessment. Lastly, I will summarize key findings and takeaways while making recommendations for the use of personality tests in personnel selection.

The Current View on Self- Report Personality Tests in Personnel Selection

Meta-analytic investigations conducted in the 1990's (Barrick & Mount, 1991; Salgado, 1997; Tett, Jackson, & Rothstein, 1991) suggested that personality tests were generally useful in predicting job performance. However, lately, a number of researchers and psychologists have questioned this conclusion, cautioning against the use of personality testing in personnel selection. At a 2004 SIOP conference, editors from five different psychological journals raised the issue of low validities of personality tests in predicting job performance. Although the editors addressed faking on personality tests, their primary concern was the issue of low validities. As Neal Schmitt, editor of the *Journal of Applied Psychology*, states "faking only makes a difference if the measure is valid" (Morgeson et al., 2007). A survey of the methodological studies that investigate the correlation between personality and job performance are generally quite low (Appendix N). The uncorrected average correlations ranged from -0.2 to 0.15. (Morgeson

et al., 2007). As a result, the editors came to a general consensus that it is difficult to justify the use of personality tests in making high-stakes decisions in light of poor validity. Hurtz and Donovan's meta-analytic study (2000) supports this claim. Although their study yielded similar results to Barrick and Mount (1991), their interpretations of results differed. Commenting on the magnitude of the validity coefficient, Hurtz and Donovan state that the correlations tend to be low and moderate (2006) (Appendix N). While Barrick and Mount meta-analytic study (1991) was optimistic about the use of personality tests in personnel selection, Hurtz and Donovan (2006) suggest that the correlations, especially that of conscientiousness and job performance, were not sufficient enough to claim that these factors were predictive of job performance to the extent that Barrick and Mount claimed.

However, there are researchers who disagree with Morgeson et al. point of view (2007). Ones, Dilchert, Viswesvaran, and Judge (2007) assert that Big Five personality constructs are predictive of performance facets such as counterproductive work behaviors ($R=0.44$), organizational citizenship behavior ($R=0.31$), and interpersonal behaviors ($R=0.33$), based on results from meta-analytic investigations. Moreover, the Big Five factors are also good predictors of 1) leadership, with R values ranging from 0.30 to 0.49, 2) some aspects of training performance ($R=0.40$), and 3) entrepreneurship ($R=0.31$). Thus, while self-report personality measures may not be the best predictors of job performance dimensions like task and contextual performance, they are better predictors of other job-related domains such as counterproductive work behaviors and leadership. Therefore, Ones et al. caution against the complete dismissal of self-report personality measures in the domain of personnel selection (2007).

Technology and Its Impact upon Personality testing

The introduction of technology into personality testing has altered the format and administration of personality tests. Organizations have slowly transitioned from paper-and-pencil tests to computerized assessments. Social media analytics and advances in artificial intelligence have automated the process by generating an individual's personality profile through the use of algorithms. The following section traces some of the ways in which technology has been integrated into recruitment, its benefits, and its potential drawbacks.

Online Personality assessments

In the early 1900's, employers relied on paper-and-pencil personality tests; however, with the advent of the internet, we have now shifted to utilizing computer-based assessments. Online assessments improve efficiency, increase the applicant pool, reduce costs, provide immediate feedback, and standardize the recruitment system (Ben-Porath & Butcher, 1986). These assessments are cost-effective as they do not require a test administrator; a candidate can take the test using a mobile phone or laptop on their own time. Moreover, they also ensure increased objectivity as computers do not have a bias and use a standardized algorithm to assess responses. They are less time-consuming as results are computed faster than that of a technician or test administrator. Lastly, computer-based assessments are also more reliable. Their test-retest reliability are always equal to 1 as computers will score responses and generate interpretations in the same way, irrespective of the applicant. Paper and pencil tests, on the other hand, have lower test-retest reliability because they are susceptible to human error when grading or

interpreting responses (Ben-Porath & Butcher, 1986). Due to these advantages, many paper-and-pencil tests, for example the MMPI, have been replaced with online versions.

Social Media

Many companies have adopted social media and social media analytics in personnel selection. A survey in 2010 found that 45% of hiring managers in the United States used information from social media sites to guide hiring decisions (Ihsan & Furnham, 2018). Recruiters collect information from a candidate's social media profile, on sites like Facebook and Twitter, to determine his/her personality based upon their posts and interactions online (Ihsan & Furnham, 2018). IBM, for example, created Watson, a social media analytics tool that uses open text to examine personality. The recruiter simply pastes a candidate's post into Watson, which generates the candidate's basic personality based upon the content of a post. For example, if the post mentions items such as "bars" "drinks" or "friends", the software would likely label the person to be an extrovert given the candidate's involvement in social activities (Ihsan & Furnham, 2018). There are, however, drawbacks to the use of algorithms in determining personality. As Ihsan and Furnham note, the software and systems are sometimes unable to decode the syntax and construction of sentences (2008). If one post reads "I go to Miami because I like to drink and party" while another reads "I prefer not to go to Miami, because of the bars and parties," the algorithm may be unable to differentiate between the content of the two posts (Ihsan & Furnham, 2018). It might label both posts as demonstrating extraversion, when in reality, the author of the second post is disinterested in social activities. Thus, the use of algorithms can result in the "degradation of validity"

which can adversely affect a candidate's prospect of employment (Ihsan & Furnham, 2018).

Facebook, a social media platform used by employers to learn about candidates, has many facets and features that can predict an individual's inclination towards extraversion. For example, a study found that extroverted and agreeable individuals tend to smile in pictures and also tend to have pictures with other people. Moreover, it was found that individuals could accurately predict extraversion and agreeableness from a profile picture based on an individual's expressiveness (Celli, Bruni, & Lepri, 2014).

An individual's Facebook likes and number of Facebook contacts have also been shown to be predictive of personality. myPersonality database, developed by researchers at Cambridge, allowed Facebook users to take psychometric tests. However, a large number of respondents also allowed the database to access some of their personal information, such as their Facebook likes (Ihsan & Furnham, 2018). Using this information, researchers found that an individual's Facebook likes were highly predictive of personality traits. For example it was found that individuals who were more open to experiences and who more extroverted liked more items on Facebook and also posted more status updates (as cited in Ihsan & Furnham, 2018)

There are limitations to relying upon social media to predict a candidate's personality. Although some employers believe that information from social media allows for the analysis of one's true personality, there are cases where this reasoning does not hold true. Facebook users, for example, engage in online impression management where they create, maintain, and modify an "image that reflects one's ideal self" (Hall,

Pennington, & Lueders, 2014). Thus, individuals may use social media as a means to project their ideal self, rather than their true self, which can mislead a third party observer's assessment of their personality. Moreover, only a few personality traits, such as extraversion, appear to be relevant in social media assessments, thus limiting applicability to other dimensions of personality.

Simulations and Gamification

Some firms are abandoning personality tests in favor of simulations that allow employers to see how candidates behave in real life job scenarios. Today, companies use multimedia simulations which incorporate audio, video and 3-D animations when measuring a candidate's personality, task performance, and cognitive abilities. However, a number of tech companies have now adopted virtual work simulations and puzzles in their recruitment process. For example, candidates applying to companies like Amazon and Facebook can attempt online programming puzzles and challenges (Anders, 2011). Tech companies generally look for candidates with superior coding abilities, and these puzzles help them ascertain individuals with good programming skills given the difficulty of the challenge. If the individual is able to successfully solve the puzzle, he/she is called in for an interview. According to recruiters at Facebook, the puzzle established more than just a candidate's aptitude and programming abilities. It reveals personal characteristics like dedication as these puzzles tend to be complex and time-consuming. Moreover, it also allows recruiters to see which candidates veer towards teamwork as conscientious problem solvers annotate their code, while self-absorbed programmers do not provide any explanation for their code (Anders, 2011).

Gamification, which refers to the use of game-design elements in non-game contexts, is an emerging tool being utilized in personnel selections. Gamification makes the recruitment process more enjoyable for potential candidates and also enables companies to put forth the image that they are technology-forward. Academic research on this topic is limited, however, a handful of organizations are now opting to gamify personality tests in their pre-hire process (Ihsan & Furnham, 2018). For example, Unilever partnered with Pymetrics, a gamification solution provider, to develop 13 games which tests an applicant's problem solving, communication, and emotional competencies. Pymetrics takes the top performers in an organization and uses their scores on the test as a baseline for comparison. One of the games on Unilever's test assesses a participant's inclination towards risk. The objective of the game is to collect as much money as possible in a time frame of three minutes. The participant is shown a balloon and can click "pump" to inflate the balloon by 5 cents (Feloni, 2017). The user can click the collect button to collect the money; however, if the balloon pops the user will lose all the money. Thus, a more cautious individual is likely to collect small amounts of money in order to avoid the busting of the balloon, while an adventurous individual will likely inflate the balloon to its limit. Results on these tests are then compared to the benchmark results (Feloni, 2017).

Video Resumes

Although the majority of companies still rely on face-to-face interviews, some companies ask candidates to submit video resumes, which are short messages detailing the work experience and skills of the candidate (Ihsan & Furnham, 2018). Video resumes

allow employers to match the candidate's personality to their resume, a capability that was not provided by a written resume. Research on paper resumes and employers' ability to predict personality from them have revealed that, with the exception of extraversion, the validity of predicting the other Big Five dimensions was low (Ihsan & Furnham, 2018). These results can be explained by the lack of information provided by paper resumes; they provide more insight into a candidate's past experiences and projects rather than personality. However, research conducted on employer's ability to predict personality from a video resume revealed the same findings. Recruiters inaccurately predicted the other personality traits apart from extraversion (Ihsan & Furnham, 2018). In light of these results, companies should use video resumes cautiously in their recruitment process and should generally avoid inferring personality traits apart from extraversion.

Artificial Intelligence

Artificial intelligence (AI) is now being developed and deployed as a recruiting tool by a large number of American companies. A report compiled by Deloitte, found that 33% of the respondents used AI in the hiring process as it was less time consuming and less prone to human errors. Moreover, AI also allows for a larger applicant pool to be considered and also helps differentiate top performers from low performing individuals (Riley, 2018). Hirevue, a software that screens job candidates, is also used by Unilever for pre-hire interviews. Hirevue's flagship product is an automated interview, where candidates record their answers to a set of interview questions on a mobile phone or device. The technology then analyzes the candidate's body posture, intonation,

communication skills, keywords, and records the information for a hiring managers (Feloni, 2017) (Appendix O).

While AI streamlines the recruitment process, reducing both cost and time, some job applicants have expressed reservations to its use. The idea of talking to a screen is dehumanizing for some participants due to the lack of conversation and exchange with a robot. Some candidates have described not feeling “worthwhile” because the company “couldn’t even assign a person for a few minutes” (Buranyi, 2018). Thus, while the use of AI has significant benefits for companies in increasing efficiency of the recruitment candidates, job candidates are at the risk of being disillusioned and feeling dehumanized due to the lack of human interaction.

Key Takeaways and Lessons Learned

Prior to undertaking this thesis, I believed that the Big Five Personality model would have provided much stronger criterion related validity in job performance given its popularity and acceptance. However, as I reviewed the meta-analytic studies, I found that the correlations were not as strong as expected. Given the breadth of research dedicated to personality testing and meta-analytic investigations of criterion validity in job performance, it appears that the Big Five personality constructs are valid predictors of certain job performance criterion (leadership, interpersonal behaviors, and counterproductive behaviors). The correlations for the Big Five factors on certain facets of task and contextual performance are on the moderate to weak side, therefore it is advisable to use other tools, such as simulations, which are able to better predict such criterion. Thus, I propose that personality tests should be used in conjunction with other

instruments, such as cognitive tests, which are shown to have better validity and reliability. For example, a number of companies use the Employee Aptitude Survey to assess a candidate's abilities and thinking skills for personnel selection, vocational guidance, and training. It consists of ten tests which assess a candidate's verbal comprehension, numerical ability, space visualization, manual speed, and symbolic reasoning ("Personnel Selection: Methods: Cognitive", n.d.). Cognitive tests are particularly suited to high complexity jobs as these types of jobs require critical analysis and thinking.

Moreover, if companies decide to incorporate personality testing in their pre-hire process, they should match specific personality traits to specific dimensions of job performance, using existing research on the topic. Morgeson et al. suggest that better criterion validity is achieved when personality traits are related to specific behaviors, rather than broad dimensions of job performance (2007). For example, looking at the correlations between agreeableness and helping behavior will likely produce better validity than comparing agreeableness to overall job performance.

I was also surprised to find how automated personality testing has become and how companies, like Unilever, rely solely on technology in the initial stages of their application process. The fact that candidates applying to Unilever have no human interaction in the first three rounds is both fascinating and disconcerting. This is not to say that the technology does not have benefits. With the advent of technology, interpretations of an individual's personality are now automated using computer software and algorithms. This alleviates concerns over the reliance on self-measure reports of

personality. Rather than having an individual assess his/her personality which can prove problematic in light of faking, technology such as AI, are now being leveraged as objective measures of personality. However, this infiltration of technology in the recruitment process requires more research on its impact on candidates, its validity, and recommendations for future use. As mentioned previously, many candidates are disenfranchised with the use of technology because of the dehumanizing aspect of it.

Finally, in many of the articles I read, it was interesting to note that the many organizations are unaware of the literature behind personality testing and other assessments. As a result, the tests that they choose may not be reflective of the traits that they are looking for in candidates. The wide-spread use of the MBTI, a methodologically flawed test with poor validity, is one such example of the ignorance of personnel assessment literature. Therefore companies should educate themselves on the different tests that are available, what they assess, and potential limitations. Tests should be incorporated based upon the nature of the job, the position, and the values and culture of the organization. For example, a tech company hiring for the position of a coder should ideally administer tests that assess the candidate's technical skills such as programming and coding abilities. A company hiring for a sales position can potentially administer a role-playing exercise where the candidate enacts the role of a salesperson in a job-related situation. Moreover, organizational culture is another aspect that companies should take into account when dividing methods for recruitment. For example, a company that has an adhocracy culture might administer the NEO-PI-3 to assess a participant's inclination towards openness. Given concerns over self-report measures, companies should also consider implementing simulations and situational judgement tests. Research studies have

shown strong correlations with job performance as they allow for employers to assess the actual behavior of a participant. Additionally, simulations also provide information about a candidate's personality and how they interact with clients and co-workers. However, if the personality traits relevant to the job are captured by another instrument, such as an interview or simulation, then the use of a personality test is redundant (Hurtz & Donovan, 2000).

References

- Aarde, N. V., Meiring, D., & Wiernik, B. M. (2017). The validity of the Big Five personality traits for job performance: Meta-analyses of South African studies. *International Journal of Selection and Assessment*, 25(3), 223-239. doi:10.1111/ijsa.12175
- Adler, S. (1987). Toward the more efficient use of assessment center technology in personnel selection. *Journal of Business and Psychology*, 2(1), 74-93. doi:10.1007/bf01061526
- Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The Susceptibility of Overt and Covert Integrity Tests to Coaching and Faking. *Psychological Science*, 7(1), 32-39. doi:10.1111/j.1467-9280.1996.tb00663.x
- Alonso, P., Moscoso, S., & Salgado, J. F. (2017). Structured behavioral interview as a legal guarantee for ensuring equal employment opportunities for women: A meta-analysis. *The European Journal of Psychology Applied to Legal Context*, 9(1), 15-23. doi:10.1016/j.ejpal.2016.03.002
- Aluja, A., García, O., Rossier, J., & García, L. F. (2005). Comparison of the NEO-FFI, the NEO-FFI-R and an alternative short version of the NEO-PI-R (NEO-60) in Swiss and Spanish samples. *Personality and Individual Differences*, 38(3), 591-604. doi:10.1016/j.paid.2004.05.014
- Anders, G. (2011, October 28). Work simulations solve the hiring conundrum. Retrieved from <https://www.theguardian.com/money/2011/oct/28/work-simulations-solve-hiring-conundrum>
- Anderson, C., Spataro, S. E., & Flynn, F. J. (2008). Personality and organizational culture

as determinants of influence. *Journal of Applied Psychology*, 93(3), 702-710.

doi:10.1037/0021-9010.93.3.702

APA Dictionary of Psychology. (n.d.). Retrieved from <https://dictionary.apa.org/neurosis>

Assessment & Selection: Other Assessment Methods. (n.d.). Retrieved from

<https://www.opm.gov/policy-data-oversight/assessment-and-selection/other-assessment-methods/work-samples-and-simulations/>

Assessment Centers Help Companies Identify Future Managers. (n.d.). Retrieved from

<https://www.apa.org/research/action/managers.aspx>

Azarpazhooh, A., Ryding, W. H., & Leake, J. L. (2008). Structured or Unstructured Personnel Interviews? *Healthcare Management Forum*, 21(4), 33-43.

doi:10.1016/s0840-4704(10)60054-3

Barrick, M. R., & Mount, M. K. (1991). The Big Five Personality Dimensions And Job Performance: A Meta-Analysis. *Personnel Psychology*, 44(1), 1-26.

doi:10.1111/j.1744-6570.1991.tb00688.x

Becker, B. (n.d.). Top 5 Behavioral Interview Questions to Ask in 2018. Retrieved from

<https://blog.hubspot.com/marketing/behavioral-interview-questions>

Ben-Porath, Y. S., & Butcher, J. N. (1986). Computers in personality assessment: A brief past, an ebullient present, and an expanding future. *Computers in Human Behavior*, 2(3), 167-182. doi:10.1016/0747-5632(86)90001-4

doi:10.1016/0747-5632(86)90001-4

Ben-Porath, Y. S., Hostetler, K., Butcher, J. N., & Graham, J. R. (1989). New subscales for the MMPI-2 Social Introversion (Si) scale. *Psychological Assessment*, 1(3), 169-174. doi:10.1037//1040-3590.1.3.169

doi:10.1037//1040-3590.1.3.169

Berlinger, J. (2012, December 04). QUIZ: Are You Smart Enough To Work For Thomas

Edison? Retrieved from <https://www.businessinsider.com/thomas-edison-interview-questions-2012-12#who-composed-il-trovatore-2>

Borgatta, E. F. (1964). The structure of personality characteristics. *Behavioral Science*, 9(1), 8-17. doi:10.1002/bs.3830090103

Borgatta, E. F. (1964). The structure of personality characteristics. *Behavioral Science*, 9(1), 8-17. doi:10.1002/bs.3830090103

Borman, W. C., & Motowidlo, S. J. (1997). Task Performance and Contextual Performance: The Meaning for Personnel Selection Research. *Human Performance*, 10(2), 99-109. doi:10.1207/s15327043hup1002_3

Borman, W. C., Bryant, R. H., & Dorio, J. (2010). The Measurement of Task Performance as Criteria in Selection Research. *Handbook of Employee Selection*, 429-447. doi:10.4324/9781315690193-20

Boyce, A. S., Corbet, C., & Adler, S. (2013). Simulations in the Selection Context: Considerations, Challenges, and Opportunities. In M. Fetzner & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 17-41). New York: Springer.

Brataas, A. (1989). Tough Break For Minnesota: Personality Test Has Passed The 'Normals'. Retrieved from http://articles.chicagotribune.com/1989-10-29/features/8901260584_1_mmpi-bell-shaped-curve-minnesota-multiphasic-personality

Buranyi, S. (2018, March 04). How to persuade a robot that you should get the job. Retrieved from <https://www.theguardian.com/technology/2018/mar/04/robots-screen-candidates-for-jobs-artificial-intelligence>

Burch, G. S., & Anderson, N. (2008). Personality as a Predictor of Work-Related

Behavior and Performance: Recent Advances and Directions for Future Research. *International Review of Industrial and Organizational Psychology* 2008, 261-305. doi:10.1002/9780470773277.ch8

Cadman, C., & Brewer, J. (2001). Emotional intelligence: A vital prerequisite for recruitment in nursing. *Journal of Nursing Management*, 9(6), 321-324. doi:10.1046/j.0966-0429.2001.00261.x

Caliper Profile Overview. (n.d.). Retrieved from <http://www.calipermedia.calipercorp.com.s3.amazonaws.com/whitepapers/us/Trait-Training.pdf>

Cattell, H. E. (2001). The Sixteen Personality Factor (16PF) Questionnaire. *Understanding Psychological Assessment*, 187-215. doi:10.1007/978-1-4615-1185-4_10

Cattell, H., & A. D., Mead. (2008). The Sixteen Personality Factor Questionnaire (16PF). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Authors), *The SAGE handbook of personality theory and assessment* (pp. 135-159). Los Angeles, CA: SAGE Publications.

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology*, 38(4), 476-506. doi:10.1037/h0054116

Cavico, F. G., Mujtaba, B., Lawrence, E., & Muffler, S. (2015). Personality Tests in Employment: A Continuing Legal, Ethical, and Practical Quandary. *Advances in Social Sciences Research Journal*, 2(4). doi:10.14738/assrj.24.1004

Celli, F., Bruni, E., & Lepri, B. (2014). Automatic Personality and Interaction Style

Recognition from Facebook Profile Pictures. *Proceedings of the ACM International Conference on Multimedia - MM 14*.

doi:10.1145/2647868.2654977

Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994).

Correcting The 16Pf For Faking: Effects On Criterion-Related Validity And Individual Hiring Decisions. *Personnel Psychology*, 47(4), 847-860.

doi:10.1111/j.1744-6570.1994.tb01581.x

Christiansen, N., & Tett, R. (2014). *Handbook of personality at work*. New York, NY: Brunner-Routledge.

Coe, C. K. (1992). The MBTI: Potential Uses and Misuses in Personnel Administration.

Public Personnel Management, 21(4), 511-522.

doi:10.1177/009102609202100407

Conklin, E. S. (1937). *Three diagnostic scorings for the Thurstone personality schedule*.

Bloomington, IN: Indiana University.

Costa, P. T., & McCrae, R. R. (n.d.). The Revised NEO Personality Inventory (NEO-PI-

R). *The SAGE Handbook of Personality Theory and Assessment: Volume 2 —*

Personality Measurement and Testing, 179-198. doi:10.4135/9781849200479.n9

Costa, P. T. (1996). Work and Personality: Use of the NEO-PI-R in

Industrial/Organisational Psychology. *Applied Psychology*, 45(3), 225-241.

doi:10.1111/j.1464-0597.1996.tb00766.x

Cullen, M. J., & Sackett, P. R. (2004). Integrity Testing in the Workplace. In M. J.

Hilsenroth, D. L. Segal, M. Hersen, E. Heiby, S. Haynes, J. C. Thomas, et al.

(Eds.), *Comprehensive handbook of psychological assessment* (pp. 149-165).

Hoboken, NJ: John Wiley & Sons.

- Daniels, S., & King, E. (2002). The predictive validity of MMPI-2 content scales for small-town police officer performance. *Journal of Police and Criminal Psychology, 17*(2), 54-61. doi:10.1007/bf02807115
- Dennis, P. M. (1984). The Edison questionnaire. *Journal of the History of the Behavioral Sciences, 20*(1), 23-37. doi:10.1002/1520-6696(198401)20:13.0.co;2-k
- Deyoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S., & Gray, J. R. (2009). Intellect as distinct from openness: Differences revealed by fMRI of working memory. *Journal of Personality and Social Psychology, 97*(5), 883-892. doi:10.1037/a0016615
- Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors In The Natural Language Of Personality: Re-Analysis, Comparison, And Interpretation Of Six Major Studies. *Multivariate Behavioral Research, 16*(2), 149-170. doi:10.1207/s15327906mbr1602_2
- Digman, J. M. (1990). Historical antecedents of the five-factor model. *Personality Disorders and the Five-factor Model of Personality.*, 13-18. doi:10.1037/10140-001
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*(1), 40-57. doi:10.1037/0021-9010.91.1.40
- Dvorak, B. J. (1956). The General Aptitude Test Battery. *The Personnel and Guidance Journal, 35*(3), 145-152. doi:10.1002/j.2164-4918.1956.tb01726.x

- Elwood, R. H. (1927). The role of personality traits in selecting a career: The nurse and the college girl. *Journal of Applied Psychology, 11*(3), 199-201.
doi:10.1037/h0070914
- Eysenck, H. J. (1977). Personality and factor analysis: A reply to Guilford. *Psychological Bulletin, 84*(3), 405-411. doi:10.1037/0033-2909.84.3.405
- Faust, Q. C. (1996). Integrity Tests: Do They Have Any Integrity. *Cornell Journal of Law and Public Policy, 6*(1), 211-232.
- Fehring, H. (2001). Contributions and Limitations of Cattell's Sixteen Personality Factor Model.
- Feloni, R. (2017, June 28). Consumer-goods giant Unilever has been hiring employees using brain games and artificial intelligence - and it's a huge success. Retrieved from <https://www.businessinsider.com/unilever-artificial-intelligence-hiring-process-2017-6>
- Ferguson, E., Semper, H., Yates, J., Fitzgerald, J. E., Skatova, A., & James, D. (2014). The 'Dark Side' and 'Bright Side' of Personality: When Too Much Conscientiousness and Too Little Anxiety Are Detrimental with Respect to the Acquisition of Medical Knowledge and Skill. *PLoS ONE, 9*(2).
doi:10.1371/journal.pone.0088606
- Fiske, D. W. (1948). *Consistency of the factorial structures in personality ratings from different sources.*
- Fleming, E. G., & Fleming, C. W. (1929). The validity of the Matthews revision of the Woodworth personal data questionnaire. *The Journal of Abnormal and Social Psychology, 23*(4), 500-506. doi:10.1037/h0075316

- Ford, D., Barik, T., & Parnin, C. (2016). Studying Sustained Attention and Cognitive States with Eye Tracking in Remote Technical Interviews. In *Proceedings of the Third International Workshop on Eye Movements in Programming* (pp. 5-7). Joensuu. Retrieved from https://www.researchgate.net/publication/303487393_Studying_Sustained_Attention_and_Cognitive_States_with_Eye_Tracking_in_Remote_Technical_Interviews.
- Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology*, 9(1), 46-55. doi:10.1007/bf02686767
- Gardner, W. L., Reithel, B. J., Coglisier, C. C., Walumbwa, F. O., & Foley, R. T. (2012). Matching Personality and Organizational Culture. *Management Communication Quarterly*, 26(4), 585-622. doi:10.1177/0893318912450663
- Garrett, H. E., & Schneck, M. R. (1928). A Study of the Discriminative Value of the Woodworth Personal Data Sheet. *The Journal of General Psychology*, 1(3-4), 459-471. doi:10.1080/00221309.1928.9918021
- Gibby, R. E., & Zickar, M. J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, 11(3), 164-184. doi:10.1037/a0013041
- Goldberg, L. R. (2014). What the hell took so long? Donald W. Fiske and the Big Five Factor Structure. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory a festschrift honoring Donald W. Fiske* (pp. 29-43). New York: Psychology Press.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American*

Psychologist, 48(1), 26-34. doi:10.1037//0003-066x.48.1.26

Goldstein, A. M., & Epstein, S. D. (2008). Personality Testing in Employment: Useful Business Tool or Civil Rights Violation. *The Labor Lawyer*, 24(2), 243-252. Retrieved from JSTOR.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36(3), 341-355. doi:10.1108/00483480710731310

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36(3), 341-355. doi:10.1108/00483480710731310

Gurven, M., Rueden, C. V., Massenkoff, M., Kaplan, H., & Vie, M. L. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager-farmers in the Bolivian Amazon. *Journal of Personality and Social Psychology*, 104(2), 354-370. doi:10.1037/a0030841

Hall, J. A., Pennington, N., & Lueders, A. (2013). Impression management and formation on Facebook: A lens model approach. *New Media & Society*, 16(6), 958-982. doi:10.1177/1461444813495166

Haney, C. (1982). Employment Tests and Employment Discrimination: A Dissenting Psychological Opinion. *Berkeley Journal of Employment & Labor Law*, 5(1), 1-86.

Harrell, T. W. (1992). Some history of the Army General Classification Test. *Journal of Applied Psychology*, 77(6), 875-878. doi:10.1037/0021-9010.77.6.875

Hartman, N. S., & Grubb, W. L. (2011). Deliberate Faking on Personality and Emotional

Intelligence Measures. *Psychological Reports*, 108(1), 120-138.

doi:10.2466/03.09.28.pr0.108.1.120-138

Harvey, O. L. (1932). Concerning the Thurstone "Personality Schedule". *The Journal of Social Psychology*, 3(2), 240-251. doi:10.1080/00224545.1932.9919148

Hogan, R. (n.d.). Hogan Personality Inventory. Retrieved from

<http://search.ebscohost.com.ccl.idm.oclc.org/login.aspx?direct=true&db=loh&AN=test.10826&site=ehost-live&scope=site>

Hogan, J., & Hogan, R. (1986). Hogan Personnel Selection Series. Retrieved from

<http://search.ebscohost.com.ccl.idm.oclc.org/login.aspx?direct=true&db=loh&AN=test.977&site=ehost-live&scope=site>

Hogan, R., & Hogan, J. (1995). Hogan Personality Inventory [Revised]. Retrieved from

<http://search.ebscohost.com.ccl.idm.oclc.org/login.aspx?direct=true&db=loh&AN=test.1530&site=ehost-live&scope=site>

Hogan, R., & Hogan, J. (1992). Hogan Personality Inventory. *PsycTESTS Dataset*.

doi:10.1037/t02029-000

Hogan, R., & Hogan, J. (2007). *Hogan Personality Inventory manual*. Tulsa, OK: Hogan Assessment Systems.

Hosie, R. (2017, August 11). Why the Myers-NBriiggs Tests should never be used in

recruitment. Retrieved from <https://www.independent.co.uk/life-style/myers-briggs-test-recruitment-why-never-use-occupation-psychologists-a7888076.html>

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & Mcclloy, R. A. (1990).

Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75(5), 581-595.

doi:10.1037/0021-9010.75.5.581

Humm, D. G., & Wadsworth, J. . G. W. (n.d.). Humm-Wadsworth Temperament Scale, 1940 Edition. Retrieved from <http://search.ebscohost.com.ccl.idm.oclc.org/login.aspx?direct=true&db=loh&AN=test.10213&site=ehost-live&scope=site>

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85(6), 869-879. doi:10.1037/0021-9010.85.6.869

Ihsan, Z., & Furnham, A. (2018). The new technologies in personality assessment: A review. *Consulting Psychology Journal: Practice and Research*, 70(2), 147-166. doi:10.1037/cpb0000106

Interview Process & Preparation | Practice Cases | BCG Careers. (n.d.). Retrieved from <https://www.bcg.com/en-us/careers/roles/consulting/interview-resources.aspx>

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, 2(3), 171-203. doi:10.1002/per.2410020302

John, O. P., & Srivastava, S. (2011). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In O. P. John, R. W. Robinson, & L. A. Pervin (Authors), *Handbook of personality: Theory and research* (pp. 102-138). New York: Guilford.

John, O. P., & Srivastava, S. U. (2011). The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In O. P. John & R. W. Robinson (Eds.), *Handbook of personality: Theory and research*. New York: Guilford.

- Johnson, J. W. (2003). Toward a Better Understanding of the Relationship Between Personality and Individual Job Performance. In M. R. Barrick & A. M. Ryan (Eds.), *Personality and work: Reconsidering the role of personality in organizations* (pp. 83-120). San Francisco, CA: Jossey-Bass.
- Kanfer, R., Ackerman, P. L., Murtha, T., & Goff, M. (2011). Personality and Intelligence in Industrial and Organizational Psychology. In D. H. Saklofske & M. Zeidner (Authors), *International handbook of personality and intelligence* (pp. 577-600). New York: Springer.
- Kinnison, J. (2016, November 08). "Death by HR" – HireVue, Video Interviews, and AI Job Searches. Retrieved from <https://jebkinnison.com/2016/10/26/death-by-hr-hirevue-video-interviews-and-ai-job-searches/>
- Laird, D. A. (1925). Detecting abnormal behavior. *The Journal of Abnormal and Social Psychology*, 20(2), 128-141. doi:10.1037/h0072769
- Latham, G. P., & Sue-Chan, C. (1999). A meta-analysis of the situational interview: An enumerative review of reasons for its validity. *Canadian Psychology/Psychologie Canadienne*, 40(1), 56-67. doi:10.1037/h0086826
- Lievens, F., & Chan, D. (2010). Practical Intelligence, Emotional Intelligence, and Social Intelligence. In J. L. Farr & N. T. Tippins (Authors), *Handbook of employee selection* (pp. 342-364). New York: Routledge.
- Lough, J., & Treuer, K. V. (2013). A critical review of psychological instruments used in police officer selection. *Policing: An International Journal of Police Strategies & Management*, 36(4), 737-751. doi:10.1108/pijpsm-11-2012-0104
- Lunenburg, F. C. (2011). The Art and Science of Personnel Selection: The Use of Tests

and Performance Simulations. *International Journal of Management, Business, and Administration*, 15(1), 1-5.

Lussier, K. (2018). Temperamental workers: Psychology, business, and the Humm-Wadsworth Temperament Scale in interwar America. *History of Psychology*, 21(2), 79-99. doi:10.1037/hop0000081

Mccrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81-90. doi:10.1037//0022-3514.52.1.81

Mccrae, R. R., Costa, J. P., & Martin, T. A. (2005). The NEO-PI-3: A More Readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, 84(3), 261-270. doi:10.1207/s15327752jpa8403_05

Mcdaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79(4), 599-616. doi:10.1037//0021-9010.79.4.599

Meyer, G. J., & Kurtz, J. E. (2006). Advancing Personality Assessment Terminology: Time to Retire "Objective" and "Projective" As Personality Test Descriptors. *Journal of Personality Assessment*, 87(3), 223-225. doi:10.1207/s15327752jpa8703_01

MMPI-2 Overview. (2018, January 09). Retrieved from

<https://www.upress.umn.edu/test-division/mmpi-2>

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Are We Getting Fooled Again? Coming To Terms With Limitations In The Use Of Personality Tests For Personnel Selection. *Personnel*

Psychology, 60(4), 1029-1049. doi:10.1111/j.1744-6570.2007.00100.x

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering The Use Of Personality Tests In Personnel Selection Contexts. *Personnel Psychology*, 60(3), 683-729. doi:10.1111/j.1744-6570.2007.00089.x

Myers, I. B., & Myers, P. B. (1995). *Gifts differing: Understanding personality type*. Mountain View, CA: Consulting Psychologists Press.

Myers-Briggs Type Indicator. (n.d.). Retrieved from <https://www.themyersbriggs.com/en-US/Products-and-Services/Myers-Briggs>

NEO Personality Inventory-3 | SIGMA. (n.d.). Retrieved from <https://www.sigmaassessmentsystems.com/assessments/neo-personality-inventory-3/>

Niece, R. (1983). The Interview and Personnel Selection: Is the Process Valid and Reliable? *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 56(5), 232-235. doi:10.1080/00098655.1983.10113782

Oboyle, E. H., Humphrey, R. H., Pollack, J. M., Hawver, T. H., & Story, P. A. (2010). The relation between emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior*, 32(5), 788-818. doi:10.1002/job.714

Occupational Personality Questionnaire (OPQ32). (n.d.). Retrieved from <https://ptc.bps.org.uk/test-review/occupational-personality-questionnaire-opq32>

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied*

Psychology, 81(6), 660-679. doi:10.1037/0021-9010.81.6.660

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 17(6), 609-626. doi:10.1002/(sici)1099-1379(199611)17:63.0.co;2-k

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In Support Of Personality Assessment In Organizational Settings. *Personnel Psychology*, 60(4), 995-1027. doi:10.1111/j.1744-6570.2007.00099.x

Oostrom, J. K., Melchers, K. G., Ingold, P. V., & Kleinmann, M. (2015). Why Do Situational Interviews Predict Performance? Is it Saying How You Would Behave or Knowing How You Should Behave? *Journal of Business and Psychology*, 31(2), 279-291. doi:10.1007/s10869-015-9410-0

Personnel Selection: Methods: Assessment Centers. (n.d.). Retrieved from <https://www.hr-guide.com/data/G318.htm>

Personnel Selection: Methods: Cognitive Ability Measures. (n.d.). Retrieved from <https://www.hr-guide.com/data/G314.htm>

Piotrowski, C., & Armstrong, T. (2006). Current Recruitment and Selection Practices: A National Survey of Fortune 1000 Firms. *North American Journal of Psychology*, 8(3), 489-496.

Pittenger, D. J. (1993). The Utility of the Myers-Briggs Type Indicator. *Review of Educational Research*, 63(4), 467-487. doi:10.2307/1170497

Porath, B. (n.d.). THE MMPI-2 AND MMPI-2-RF.

Prati, L. M., Douglas, C., Ferris, G. R., Ammeter, A. P., & Buckley, M. R. (2003). Emotional Intelligence, Leadership Effectiveness, And Team Outcomes. *The*

International Journal of Organizational Analysis, 11(1), 21-40.

doi:10.1108/eb028961

- Prewett, M. S., Tett, R. P., & Christiansen, N. D. (2014). A Review and Comparison of 12 Personality Inventories on Key Psychometric Characteristics. In N. D. Christiansen & R. P. Tett (Authors), *Handbook of personality at work* (pp. 191-221). New York, NY: Brunner-Routledge.
- Riggio, R. E. (2018). *Introduction to industrial/organizational psychology*. New York: Routledge, Taylor & Francis Group.
- Riley, T. (2018, March 13). Get ready, this year your next job interview may be with an A.I. robot. Retrieved from <https://www.cnbc.com/2018/03/13/ai-job-recruiting-tools-offered-by-hirevue-my-a-other-start-ups.html>
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83(4), 634-644. doi:10.1037//0021-9010.83.4.634
- Rothmann, S., & Coetzer, E. P. (2003). The big five personality dimensions and job performance. *SA Journal of Industrial Psychology*, 29(1). doi:10.4102/sajip.v29i1.88
- Sackett, P. R., Burris, L. R., & Callahan, C. (1989). Integrity Testing For Personnel Selection: An Update. *Personnel Psychology*, 42(3), 491-529. doi:10.1111/j.1744-6570.1989.tb00666.x
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82(1), 30-43. doi:10.1037//0021-9010.82.1.30

- Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76(3), 323-346. doi:10.1348/096317903769647201
- Salgado, J. F., Moscoso, S., & Alonso, P. (2013). Subdimensional Structure of the Hogan Personality Inventory. *International Journal of Selection and Assessment*, 21(3), 277-285. doi:10.1111/ijsa.12037
- Saville, P., Sik, G., Nyfield, G., Hackston, J., & MacIver, R. (1996). A Demonstration of the Validity of the Occupational Personality Questionnaire (OPQ) in the Measurement of Job Competencies Across Time and in Separate Organisations. *Applied Psychology*, 45(3), 243-262. doi:10.1111/j.1464-0597.1996.tb00767.x
- Saxe, L. (1994). Detection of Deception: Polygraph and Integrity Tests. *Current Directions in Psychological Science*, 3(3), 69-73. doi:10.1111/1467-8721.ep10770416
- Schilpzand, M. C., Herold, D. M., & Shalley, C. E. (2010). Members' Openness to Experience and Teams' Creative Performance. *Small Group Research*, 42(1), 55-76. doi:10.1177/1046496410377509
- Scholte, W., Tiemens, B. G., Verheul, R., Meerman, A., Egger, J., & Hutschemaekers, G. (2012). Predictive Validity of the MMPI-2 Clinical, PSY-5, and RC Scales for Therapy Disruptive Behavior. *Journal of Psychiatric Practice*, 18(6), 420-429. doi:10.1097/01.pra.0000422740.87495.91
- Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, 81(5), 525-531. doi:10.1037/0021-9010.81.5.525

- Scroggins, W. A., Thomas, S. L., & Morris, J. A. (2008). Psychological Testing in Personnel Selection, Part I: A Century of Psychological Testing. *Public Personnel Management, 37*(1), 99-109. doi:10.1177/009102600803700107
- Seedhouse, E. (2013). Minnesota Multiphasic Personality Inventory. *SpringerReference, 233-253*. doi:10.1007/springerreference_184624
- Sekiguchi, T. (2004). Person-Organization Fit and Person-Job Fit in Employee Selection: A review of the Literature. *Osaka Keidai Ronshu, 54*(6), 179-196.
- Saucier, G., & Srivastava, S. (2015). Hat makes a good structural model of personality? Evaluating the big five and alternatives. In M. Mikulincer & P. R. Shaver (Authors), *Personality processes and individual differences* (pp. 283-305). Washington, DC: American Psychological Association.
- Smith, G. T., & Zaploski, T. C. (2009). Construct Validation of Personality Measures. In J. N. Butcher (Ed.), *Oxford Handbook of Personality Assessment* (pp. 81-109). Oxford University Press.
- Stabile, S. J. (2002). The use of Personality Tests as a Hiring Tool: Is the Benefit worth the Cost? *University of Pennsylvania Journal of Business Law, 4*(2), 279-313.
- Strauss, M. E., & Smith, G. T. (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review of Clinical Psychology, 5*(1), 1-25.
doi:10.1146/annurev.clinpsy.032408.153639
- Structured Interviews. (n.d.). Retrieved from <https://www.opm.gov/policy-data-oversight/assessment-and-selection/structured-interviews/>
- Super, D. E. (1942). The Bernreuter Personality Inventory: A review of research. *Psychological Bulletin, 39*(2), 94-125. doi:10.1037/h0058418

- Tett, R. P., Jackson, D. N., & Rothstein, M. (2006). Personality Measures As Predictors Of Job Performance: A Meta-Analytic Review. *Personnel Psychology, 44*(4), 703-742. doi:10.1111/j.1744-6570.1991.tb00696.x
- The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. (1999).
- Thurstone, L. L., & Thurstone, T. G. (n.d.). Personality Schedule, 1929 Edition.
Retrieved from
<http://search.ebscohost.com.ccl.idm.oclc.org/login.aspx?direct=true&db=loh&AN=test.9628&site=ehost-live&scope=sit>
- Tonetti, L. (2011). Circadian Preference and Personality traits: A Mini-review. In M. E. Jordan (Ed.), *Personality traits: Theory, testing and influences*. New York: Nova Science.
- Traxler, A. E. (1941). Current Construction and Evaluation of Personality and Character Tests. *Review of Educational Research, 11*(1), 57-79. doi:10.2307/1167524
- Tuzinski, K. (2013). Simulations for Personnel Selection: An Introduction. In M. Fetzer & K. Tuzinski (Authors), *Simulations for personnel selection*. New York: Springer.
- Vinchur, A. J., & Bryan, L. L. (2012). A History of Personnel Selection and Assessment. *Oxford Handbooks Online*. doi:10.1093/oxfordhb/9780199732579.013.0002
- Vinchur, A. J. (2018). *The early years of industrial and organizational psychology*. Cambridge: Cambridge University Press.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-Analyses of Fakability Estimates: Implications for Personality Measurement. *Educational and Psychological*

Measurement, 59(2), 197-210. doi:10.1177/00131649921969802

Watson, G. (1938). Personality and Character Measurement. *Review of Educational Research*, 8(3), 269-291. doi:10.2307/1167426

Wiggins, J. S., Behrends, R. S., Ben-Porath, Y. S., & Blatt, S. J. (2003). *Paradigms of personality assessment*. New York: Guilford Press.

Wingdor, A. K., & Green, B. F. (1991). Psychological Testing and the Challenge of Criterion. In *Performance Assessment for the Workplace* (Vol. 1, pp. 15-32). Washington D.C.: The National Academic Press.

Witt, L. A., Burke, L. A., Barrick, M. A., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. *Journal of Applied Psychology*, 87(1), 164-169. doi:10.1037//0021-9010.87.1.164

Zee, K. I., Bakker, A. B., & Bakker, P. (2002). Why are structured interviews so rarely used in personnel selection? *Journal of Applied Psychology*, 87(1), 176-184. doi:10.1037//0021-9010.87.1.176

Zickar, M. J., & Kostek, J. A. (2014). History of Personality Testing Within Organizations. In N. Christiansen & R. Tett (Authors), *Handbook of personality at work* (pp. 173-190). New York, NY: Brunner-Routledge.

Appendix A

The Nine Aptitudes for the General Aptitude Battery Test

- G—Intelligence
 - Part 3—Three-Dimensional Space
 - Part 4—Vocabulary
 - Part 6—Arithmetic Reason
- V—Verbal Aptitude
 - Part 4—Vocabulary
- N—Numerical Aptitude
 - Part 2—Computation
 - Part 6—Arithmetic Reason
- S—Spatial Aptitude
 - Part 3—Three-Dimensional Space
- P—Form Perception
 - Part 5—Tool Matching
 - Part 7—Form Matching
- Q—Clerical Perception
 - Part 1—Name Perception
- K—Motor Coordination
 - Part 8—Mark Making
- F—Finger Dexterity
 - Part 11—Assemble
 - Part 12—Disassemble
- M—Manual Dexterity
 - Part 9—Place
 - Part 10—Turn

Appendix B

The questions below are exhibits from the Edison questionnaire devised by Thomas

Edison (Berlinger, 2018)

1. Which countries supply the most mahogany?
2. Who was the Roman Emperor when Jesus was born?
3. What is brass made out of?
4. Who was Leonidas?
5. Where do we get shellac?
6. Who assassinated President Lincoln?

Appendix C

The following table list the primary and global scales of Cattell's 16PF.

Table 7.1 16PF Scale Names and Descriptors

<i>Descriptors of Low Range</i>	<i>Primary Scales</i>	<i>Descriptors of High Range</i>
Reserved, Impersonal, Distant	Warmth (A)	Warm-hearted, Caring, Attentive To Others
Concrete, Lower Mental Capacity	Reasoning (B)	Abstract, Bright, Fast-Learner
Reactive, Affected By Feelings	Emotional Stability (C)	Emotionally Stable, Adaptive, Mature
Deferential, Cooperative, Avoids Conflict	Dominance (E)	Dominant, Forceful, Assertive
Serious, Restrained, Careful	Liveliness (F)	Enthusiastic, Animated, Spontaneous
Expedient, Nonconforming	Rule-Consciousness (G)	Rule-Conscious, Dutiful
Shy, Timid, Threat-Sensitive	Social Boldness (H)	Socially Bold, Venturesome, Thick-Skinned
Tough, Objective, Unsentimental	Sensitivity (I)	Sensitive, Aesthetic, Tender-Minded
Trusting, Unsuspecting, Accepting	Vigilance (L)	Vigilant, Suspicious, Skeptical, Wary
Practical, Grounded, Down-To-Earth	Abstractedness (M)	Abstracted, Imaginative, Idea-Oriented
Forthright, Genuine, Artless	Privateness (N)	Private, Discreet, Non-Disclosing
Self-Assured, Unworried, Complacent	Apprehension (O)	Apprehensive, Self-Doubting, Worried
Traditional, Attached To Familiar	Openness to Change (Q1)	Open To Change, Experimenting
Group-Orientated, Affiliative	Self-Reliance (Q2)	Self-Reliant, Solitary, Individualistic
Tolerates Disorder, Unexacting, Flexible	Perfectionism (Q3)	Perfectionistic, Organized, Self-Disciplined
Relaxed, Placid, Patient	Tension (Q4)	Tense, High Energy, Driven
<i>Global Scales</i>		
Introverted, Socially Inhibited	Extraversion	Extraverted, Socially Participating
Low Anxiety, Unperturbable	Anxiety Neuroticism	High Anxiety, Perturbable
Receptive, Open-Minded, Intuitive	Tough-Mindedness	Tough-Minded, Resolute, Unempathic
Accommodating, Agreeable, Selfless	Independence	Independent, Persuasive, Willful
Unrestrained, Follows Urges	Self-Control	Self-Controlled, Inhibits Urges

Adapted with permission from S.R. Conn and M.L. Rieke (1994). 16PF Fifth Edition Technical Manual. Champaign, IL: Institute for Personality and Ability Testing, Inc.

Appendix D

Components of the FFM.

*Table 1.1
Components of the Five Factor Model*

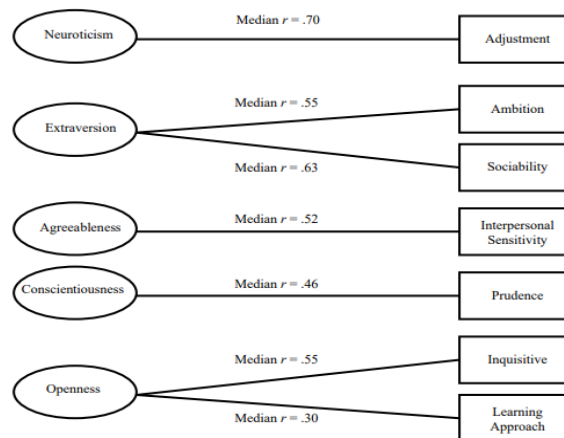
Factor	Definition	ACL Marker Items (a)
I. Surgency	The degree to which a person needs	Quiet, Reserved, Shy vs. Talkative,
II. Agreeableness	The degree to which a person needs pleasant and harmonious relations with	Fault finding, Cold, Unfriendly vs.
III. Conscientiousness	The degree to which a person is willing to comply with conventional rules,	Careless, Disorderly, Frivolous vs. Organized,
IV. Emotional Stability	The degree to which a person experiences the world as threatening	Tense, Anxious, Nervous vs. Stable,
V. Intellect / Openness to Experience	The degree to which a person needs intellectual stimulation,	Commonplace, Narrow interest, Simple vs. Wide interest, Imaginative, Intelligent

(a) The objectives listed here were taken from John's (1990, Table 3.2) listing of factor loadings for selected

Appendix E

The HPI and the FFM Model of Personality

Figure 2.1 Relations between FFM Inventories and the HPI Scales



Note. Median correlation coefficients summarize HPI relations with the NEO PI-R (Goldberg, 2000), Goldberg's (1992) Big-Five Markers (R. Hogan & Hogan, 1995), Personal Characteristics Inventory (Mount & Barrick, 1995), and the Inventario de Personalidad de Cinco Factores (Salgado & Moscoso, 1999). The coefficient ranges are as follows: Adjustment/Neuroticism (.66 to .72); Ambition/Extraversion/Surgency (.39 to .60); Sociability/Extraversion/Surgency (.44 to .64); Interpersonal Sensitivity/Agreeableness (.37 to .61); Prudence/Conscientiousness (.36 to .59); Inquisitive/Openness/Intellect (.33 to .69); Learning Approach/Openness/Intellect (.24 to .35). Reprinted with permissions from the authors. All rights reserved.

Appendix F

HIC subscales of the Hogan Personality Inventory

HPI Subscale Table

	Subscale	Sample Item	Low Scores	High Scores
Adjustment	Empathy	I am often irritated by the faults of others.	Irritated by others' flaws	Seems empathic
	Not Anxious	I am seldom tense or anxious.	Anxious or tense	Seems relaxed
	No Guilt	I rarely feel guilty about things I have done.	Prone to worry about past mistakes	Does not worry about past mistakes
	Calmness	I keep calm in a crisis.	Gets emotional at times	Is calm
	Even Tempered	I rarely lose my temper.	Is temperamental or moody	Is even-tempered
	No Complaints	I rarely complain to others.	Complains about many issues	Does not complain
	Trusting	People really care about one another.	Questions others' intentions	Trusts others
	Good Attachment	There were times I felt like running away	Hostile towards authority	Has positive attitude toward authority
	Ambition	Competitive	I am an ambitious person	Laid back
Self Confident		I am a very self-confident person	Lacks confidence	Confident
Accomplishment		I am known as someone who gets things done	Unhappy with accomplishments	Enjoys self and work
Leadership		In a group I like to take charge of things	Reluctant to assume leadership roles	Willing to assume authority positions
Identity		I know what I want to be	Lacks career direction	Focused career direction
No Social Anxiety		I enjoy talking in front of groups of people	Socially retiring	Confident in social settings
Sociability	Likes Parties	I enjoy going to parties.	Does not enjoy parties	Enjoys social gatherings
	Likes Crowds	Being part of a large crowd is exciting.	Prefers smaller groups	Enjoys large groups
	Experience Seeking	I like a lot of variety in my life.	Unadventurous and prefers little variety	Adventurous, actively seeks out experiences
	Exhibitionistic	I like to be the center of attention.	Avoids the limelight	Enjoys attention and "showing off"
	Entertaining	I am often the life of the party.	Not particularly entertaining	Charming, amusing, good sense of humor
	Easy To Live With	I work well with other people	Not always tolerant and kind-hearted	Perceived as easygoing by others
Interpersonal Sensitivity	Sensitive	I try to see the other persons' point of view	Not very tactful or considerate	Tactful, considerate and
	Caring	I am sensitive to other people's moods	Does not appreciate others' needs	Perceptive and understanding
	Likes People	I enjoy just being with other people	Socially withdrawn	Enjoys others' company
	No Hostility	I don't mind criticizing people when needed	Critical of others	Generally accepting
	Prudence	Moralistic	I always practice what I preach	Prefers to set his/her own rules
Mastery		I do my job as well as I possibly can	Relaxed attitude about his/her work	Concerned with doing a good job
Virtuous		I strive for perfection in everything I do	Willing to admit minor faults	Diligent and precise
Not Autonomous		Other people's opinions of me are important	Independent and feedback resistant	Concerned about how others view him/her
Not Spontaneous		I always know what I will do tomorrow	Spontaneous	Well-planned in his/her approach
Impulse Control		I rarely do things on impulse	Tends to act on impulse	Likes to "play it safe"
Avoids Trouble		When I was in school I rarely caused trouble	Takes unnecessary and negative risks	Considers actions and their consequences
Inquisitive	Science Ability	I am interested in science	Shows little interest in why things happen	Takes an interest in why things happen
	Curiosity	I have taken things apart to see how they work	Low degree of curiosity	High degree of curiosity
	Thrill Seeking	I would like to be a race car driver	Not interested in stimulation/excitement	Wants stimulation, and excitement
	Intellectual Games	I enjoy solving riddles	Not interested in intellectual games	Interested in riddles and puzzles
	Generates Ideas	I am a quick-witted person	Does not see self as an idea generator	Good at generating new ideas
	Culture	I like trying new, exotic types of food	Narrow interests	Wide variety of activities
Learning Approach	Education	As a child, school was easy for me	Does not enjoy traditional education	Positive attitude about traditional education
	Math Ability	I can multiply large numbers quickly	Does not work well with numbers	Works well with numbers
	Good Memory	I can remember details easily	Somewhat forgetful	Can remember things easily
	Reading	I would rather read than watch TV	Does not keep up-to-date	Keeps up-to-date, well-informed

Appendix G

The item below demonstrates the scales and sub-set traits of the NEO-PI-3 as well as the NEO-PI-R (“NEO Personality Inventory-3, n.d.)

Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Anxiety	Warmth	Fantasy	Trust	Competence
Angry Hostility	Gregariousness	Aesthetics	Straightforwardness	Order
Depression	Assertiveness	Feelings	Altruism	Dutifulness
Self-Consciousness	Activity	Actions	Compliance	Achievement Striving
Impulsiveness	Excitement Seeking	Ideas	Modesty	Self-Discipline
Vulnerability	Positive Emotions	Values	Tender-Mindedness	Deliberation

Appendix H

Four competency categories of the Caliper profile and the attributes that comprise each of them.

Leadership/ Persuasiveness	Interpersonal Service Orientation	Problem Solving and Decision Making	Personal Organization and Time Management
Assertiveness	Accommodation	Abstract Reasoning Ability	External Structure
Aggressiveness	Empathy	Flexibility	Cautiousness
Ego-Drive	Gregariousness	Idea Orientation	Risk-Taking
Ego-strength (Resilience)	Skepticism	Openness	Self-Structure
Energy	Sociability		Thoroughness
Level-Headedness			Urgency

Appendix I

Structure of the OPQ

TABLE 1
The Structure of the Occupational Personality Questionnaires (OPQ)

Level I (IMAGES)	Level II (Factor)	Level III (Concept)		
Gregarious (Extrovert, Surgent)	Influential Outspoken Sociable	Persuasive Controlling	Enjoys selling, changes opinions of others, convincing with arguments, negotiates. Takes charge, directs, manages, organises, supervises others.	
		Independent Critical	Has strong views on things, difficult to manage, speaks up, argues. Likes probing the facts, sees the disadvantages, challenges assumptions.	
	Sympathetic (Agreeable)	Empathic	Outgoing Affiliative	Fun loving, humorous, sociable, vibrant, talkative, jovial. Has many friends, enjoys being in groups, likes companionship, shares things with friends.
			Socially Confident Modest (-)	Comfortable with strangers, likes to put others at ease. Reserved about achievements, avoids talking about self.
Imaginative (Intellectual)	Data Rational Conceptual Traditional (-) Innovative	Caring Democratic Behavioural	Considerate to others, helps those in need, sympathetic, tolerant. Encourages others to contribute, consults, listens and refers to others. Analyses thoughts and behaviour, psychologically minded, likes to understand people.	
		Data Rational Artistic	Likes to work with data, operates on facts, enjoys assessing and measuring. Appreciates culture, sensitive to visual arts and music.	
		Conceptual Traditional (-)	Theoretical, intellectually curious, enjoys the complex and abstract. Preserves well-proven methods, prefers the orthodox, disciplined, conventional.	
		Change Orientated	Enjoys doing new things, seeks variety, prefers novelty to routine, accepts changes.	
		Innovative Practical	Generates ideas, shows ingenuity, thinks up solutions. Likes repairing and mending things, enjoys using hands.	

(Continued)

TABLE 1—(Continued)

Level I (IMAGES)	Level II (Factor)	Level III (Concept)	
Methodical (Conscientious)	Methodical	Forward Planning	Prepares well in advance, enjoys target selling, forecasts trends, plans projects.
		Detail Conscious Conscientious	Methodical, keeps things neat and tidy, precise, accurate. Sticks to deadlines, completes jobs, perseveres with routine, likes fixed schedules.
Emotional (Anxious)	Relaxed (-) Emotionally Controlled (-) Optimistic (-)	Relaxed (-) Worrying	Calm, relaxed, cool under pressure, free from anxiety, can switch off. Worries when things go wrong, keyed-up before important events, anxious to do well.
		Tough-Minded (-)	Difficult to hurt or upset, can brush off insults, unaffected by unfair remarks.
		Emotionally Controlled (-) Optimistic (-)	Restrained in showing emotions, keeps feelings back, avoids outbursts.
Achieving (Competitive)	Active Competitive Achieving Decisive	Active Competitive	Cheerful, happy, keeps spirits up despite setbacks.
		Achieving	Has energy, moves quickly, enjoys physical exercise, doesn't sit still. Plays to win, determined to beat others, poor loser.
		Decisive	Ambitious, sets sights high, career-centred, results-orientated.
			Quick at conclusions, weighs things up rapidly, may be hasty, takes risks.

Appendix J

Primary Scales of the Caliper Profile

Primary Scales
<p>Relationships with People:</p> <p>Influence; persuasive, controlling, outspoken, independent minded,</p> <p>Sociability; outgoing, affiliative, socially confident,</p> <p>Empathy; modest, democratic, caring.</p>
<p>Thinking Style:</p> <p>Analysis; data rational, evaluative, behavioural,</p> <p>Creativity and change; conventional, conceptual, innovative, variety seeking, adaptable,</p> <p>Structure; forward thinking, detail conscious, conscientious.</p>
<p>Feelings and Emotions:</p> <p>Emotions: relaxed, worrying, tough minded, optimistic, trusting, emotionally controlled,</p> <p>Dynamism: vigorous, competitive, achieving, decisive.</p>

Appendix K

Results from Barrick and Mount's study in 1991.

TABLE 2
*Meta-Analysis Results for Personality Dimension-Occupation
 Combinations (all Criterion Types Included)*

Occupational group	Total <i>N</i>	Number of <i>r</i> 's	Obs \bar{r}	$\hat{\rho}$	SD_{ρ}	90% C.V.	% Variance accounted
Extraversion							
Professionals	476	4	-.05	-.09	.05	-.03	92
Police	1,496	16	.05	.09	.00	.09	127
Managers	11,335	59	.11	.18	.13	.01	48
Sales	2,316	22	.09	.15	.16	-.05	54
Skilled/Semi-Skilled	3,888	23	.01	.01	.08	-.10	72
Mean (across occupations)			.08	.13	.11	-.01	69 ^a
Emotional stability							
Professionals	518	5	-.07	-.13	.04	-.07	92
Police	1,697	18	.06	.10	.00	.10	138
Managers	10,324	55	.05	.08	.09	-.04	65
Sales	2,486	19	.04	.07	.19	-.18	38
Skilled/ Semi-Skilled	3,694	26	.05	.12	.10	-.06	50
Mean (across occupations)			.05	.08	.10	-.05	63 ^a
Agreeableness							
Professionals	557	7	.01	.02	0	.02	158
Police	1,437	14	.06	.10	0	.10	121
Managers	8,597	47	.05	.10	.03	.06	94
Sales	2,344	16	.00	.00	.24	-.31	25
Skilled/Semi-Skilled	4,585	28	.04	.06	.17	-.16	37
Mean (across occupations)			.04	.07	.09	-.05	54 ^a
Conscientiousness							
Professionals	767	6	.11	.20	.00	.20	106
Police	2,045	19	.13	.22	.20	-.03	40
Managers	10,058	52	.13	.22	.10	.09	64
Sales	2,263	21	.09	.23	.00	.23	150
Skilled/Semi-Skilled	4,588	25	.12	.21	.09	.09	67
Mean (across occupations)			.13	.22	.09	.10	70 ^a
Openness to experience							
Professionals	476	4	-.05	-.08	.04	-.03	94
Police	1,364	13	.00	.00	.00	.00	181
Managers	7,611	37	.05	.08	.16	-.12	37
Sales	1,566	12	-.01	-.02	.16	.18	46
Skilled/Semi-Skilled	3,219	16	.01	.01	.12	-.15	49
Mean (across occupations)			.03	.04	.13	.13	59 ^a

^a An unbiased estimate of mean percentage of variance accounted for across meta-analyses, calculated by taking the reciprocal of the average of reciprocals of individual predicted to observed variance ratios (Hunter & Schmidt, 1990).

Appendix L

Results from Hurtz and Donovan's study

Table 4
Validity Coefficients for Personality Dimensions by Criterion Dimension

Big Five dimension	<i>k</i>	<i>N</i>	\bar{r}	S_r^2	S_e^2	S_{meas}^2	S_{res}^2	% VE	ρ_c	ρ_v	SD_{ρ_v}	90% CV
Task performance												
Conscientiousness	12	2,197	.10	.0138	.0054	.0008	.0076	45	.16	.15	.13	-.02
Emotional Stability	8	1,243	.09	.0015	.0064	.0007	.0000	463	.14	.13	.00	.13
Agreeableness	9	1,754	.05	.0090	.0051	.0002	.0037	59	.08	.07	.09	-.05
Extraversion	9	1,839	.04	.0052	.0049	.0002	.0001	98	.07	.06	.02	.04
Openness to Experience	7	1,176	-.01	.0237	.0060	.0000	.0177	25	-.01	-.01	.20	-.26
Job dedication												
Conscientiousness	17	3,197	.12	.0203	.0052	.0013	.0139	32	.20	.18	.17	-.04
Emotional Stability	15	2,581	.09	.0059	.0058	.0007	.0000	109	.14	.13	.00	.13
Agreeableness	17	3,197	.06	.0096	.0053	.0003	.0040	59	.10	.08	.09	-.03
Extraversion	16	3,130	.03	.0111	.0051	.0001	.0059	47	.05	.05	.11	-.10
Openness to Experience	14	2,514	.01	.0108	.0056	.0000	.0052	52	.01	.01	.11	-.13
Interpersonal facilitation												
Conscientiousness	23	4,301	.11	.0083	.0053	.0010	.0020	76	.18	.16	.07	.07
Emotional Stability	21	3,685	.10	.0046	.0056	.0010	.0000	142	.17	.16	.00	.16
Agreeableness	23	4,301	.11	.0117	.0052	.0012	.0053	55	.20	.17	.11	.03
Extraversion	21	4,155	.06	.0105	.0050	.0004	.0051	52	.11	.10	.11	-.04
Openness to Experience	19	3,539	.03	.0075	.0054	.0001	.0020	73	.05	.05	.07	-.04

Note. *k* = number of validity coefficients; *N* = total sample size; \bar{r} = sample-size weighted mean observed validity; S_r^2 = total observed variance in \bar{r} ; S_e^2 = variance due to sampling error; S_{meas}^2 = variance due to measurement artifacts; S_{res}^2 = residual variance; % VE = percentage of variance accounted for by sampling error and measurement artifacts; ρ_c = true-score correlation; ρ_v = true (operational) validity; SD_{ρ_v} = standard deviation of true validity; CV = credibility value (lower bound of credibility interval for ρ_v).

Appendix M

This table provides characteristics of organizational cultures. The information presented below has been adapted from Gardner et al. (2012)

Organizational culture	Characteristics
Clan culture	Clan cultures emphasize cohesion, morale, participation and loyalty
Market culture	Market cultures emphasize production, competition, and goal achievement
Hierarchy culture	Hierarchy cultures focus on rules, policies, procedures, proficiency, and control
Adhocratic culture	Adhocratic cultures stress innovation, flexibility, creativity, and risk

Appendix N

TABLE 2
Uncorrected Average Correlations Between “Big Five” Personality Measures and Job Performance Criteria

Personality Measure	Hurtz and Donovan (2000)	Salgado (1997)	Barrick and Mount (1991)
Conscientiousness	.15	.10	.13
Extraversion	.06	.06	.06
Agreeableness	.07	-.00	.04
Emotional Stability	.09	.08	.04
Openness to Experience	.03	.00	-.02

Appendix O

An illustration of Hirevue's platform for interviews (Kinnison, 2016)



