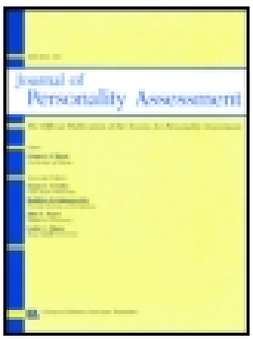

Routledge
Taylor & Francis Group

Journal of Personality Assessment

ISSN: 0022-3891 (Print) 1532-7752 (Online) Journal homepage: http://www.tandfonline.com/loi/hjpa20

Personality Assessment for Employee Development: Ivory Tower or Real World?

Penny Moyle & John Hackston

To cite this article: Penny Moyle & John Hackston (2018): Personality Assessment for Employee Development: Ivory Tower or Real World?, Journal of Personality Assessment, DOI: 10.1080/00223891.2018.1481078

To link to this article: https://doi.org/10.1080/00223891.2018.1481078

Published online: 22 Jun 2018.

JOURNAL OF PERSONALITY ASSESSMENT
https://doi.org/10.1080/00223891.2018.1481078

Routledge
Taylor & Francis Group

SPECIAL SECTION: ORGANIZATIONAL AND CONSULTING PSYCHOLOGY

Personality Assessment for Employee Development: Ivory Tower or Real World?

Penny Moyle[1] and John Hackston[2]

[1]Meyler Campbell, Oxford, UK; [2]OPP Ltd., Oxford, UK

ABSTRACT

The acceptance and popularity of personality assessments in organizational contexts has grown enormously over the last 40 years. Although these are used across many applications, such as executive coaching, team building, and hiring and promotion decisions, the focus of most published research on the use of personality assessments at work is biased toward assessment for employee selection. Reviews have therefore tended to use criteria that are appropriate for selection, neglecting the additional and different criteria that are important in relation to employee development. An illustration of the often-discussed scientist–practitioner divide is that the Myers–Briggs Type Indicator is the most widely known and used personality assessment in organizations, despite harsh criticism by the academic community. This article reviews this debate, and draws implications for the appropriate choice of personality assessments for use in individual and team development, and a new direction for scientific research.

ARTICLE HISTORY
Received 31 August 2017
Revised 1 May 2018

Personality inventories originated in clinical psychology, evolving from highly specific assessments such as the Woodworth Personal Data Sheet (Woodworth, 1917) and the Personality Schedule (Thurstone, 1930) through to multidimensional instruments such as the Bernreuter Personality Inventory (BPI; Bernreuter, 1931) and the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943). In the mid-20th century, personality questionnaires were developed for more general use. Some, such as the Sixteen Personality Factor questionnaire (16PF; Cattell, 1946; Cattell, Cattell, & Cattell, 1993) and the NEO PI (Costa & McCrae, 1985) were constructed on an empirical basis. Others were built on theoretical foundations and clinical experience. The Myers–Briggs Type Indicator (MBTI; Myers, 1962; Myers, McCaulley, Quenk, & Hammer, 1998) was based on the work of Jung (1971).

As defined by Jacobs and Washington (2003), "Employee development refers to an integrated set of planned programs, provided over a period of time, to help assure that all individuals have the competence necessary to perform to their fullest potential in support of the organization's goals" (p. 344). Although personality assessments have featured in employee development for nearly a century, their use over the last 15 years has grown significantly (McDowall & Redman, 2017). Personality-based development is now commonplace at all levels of large organizations, and many smaller ones (Passmore, 2012). Examples include team building, executive coaching, leadership development, communication, and resilience training.

Human resources (HR) practitioners and managers have hundreds of personality tests available to them, many reviewed by respected bodies such as the Buros Center for Testing or the British Psychological Society (BPS). Furnham (2008a) reported that the top ranked assessments used by UK HR practitioners for development were the MBTI, Fundamental Interpersonal Relationship Orientation (FIRO), 16PF, assessments based on the Big Five model of personality, and the Belbin Team Role Self-Perception Inventory, with the most popular being the MBTI, used by over half the group. This overlapped, but contrasted with, the most popular assessments for selection: Big Five assessments, 16PF, Occupational Personality Questionnaire (OPQ), Hogan Personality Inventory (HPI), and Personal Profile Analysis (PPA).

Academic reviews tend to be highly critical of several assessments popular in the development arena (e.g., Chamorro-Premuzic, Winsborough, Sherman, & Hogan, 2016; Furnham, 2008a). Chamorro-Premuzic et al. (2016) are typical, in noting that "there is a substantial gap between what science prescribes and what HR practitioners do, especially around assessment practices" (p. 635). The MBTI attracts particularly severe criticism (e.g., Carter, 2016; Essig, 2014; Grant, 2013; McCrae & Costa, 1989; Michael, 2003; Murphy Paul, 2004; Pittenger, 2005), with views about its continued popularity ranging from concern to consternation and disbelief.

Although less published literature is available on the many other assessments used in employee development, many criticisms leveled at the MBTI also apply to these. For example, the FIRO–B (reviewed by Furnham, 1990, 2008b) and Belbin Team Role Self-Perception Inventory (reviewed by Furnham, Steele, & Pendleton, 1993) are both criticized for poor construct and predictive validity, and poor reliability. Other Jungian-based type indicators, similar to the MBTI, are criticized less often, although they share many of the same features. Examples include the Enneagram (Wagner, 1983), Golden Personality Type Profiler (Golden, 2004), Insights Discovery (Lothian, 1996) and Lumina Spark (Desson, 2017). Strengths inventories,

CONTACT Penny Moyle pennymoyle@hotmail.com Highview, Vernon Ave., Oxford, OX2 9AU, UK.

multirater or 360° feedback instruments, and emotional intelligence (EI) assessments are increasingly popular in organizations, but have also been criticized (e.g., McDowall & Redman, 2017, on strengths inventories; Fletcher, Baldry, & Cunningham-Snell, 1998, on 360° assessments; Conte, 2005, on EI measures).

In this article we examine the popularity of the MBTI, together with the common criticisms of it. This example provides new perspectives on the scientist–practitioner divide in choosing personality assessments for employee development.

Overview of the MBTI assessment

The MBTI (Myers et al., 1998) was developed to make Jung's theory of personality types "understandable and useful in people's lives" (p. 3). In doing so, Myers and Briggs included their own interpretation and extension of Jung's original theory. The standard version (MBTI Step I) sorts individuals according to four dichotomies, as defined in Table 1, which are then combined to create 16 categorical types. There also exists a more elaborate version of the assessment, MBTI Step II, which breaks down each dichotomy into five behavioral facets, each measured on an 11-point scale (Quenk, Hammer, & Majors, 2004). Although MBTI Step II is also widely used and addresses some of the criticisms leveled at the standard assessment, it is less well-known and has attracted little comment in the scientific literature. Furnham (2017) noted that the MBTI is the most widely known and used personality assessment in the world, taken by anywhere from 1.5 million to 5 million people every year. A 2014 *Forbes* article reported that the MBTI is in use by 89 of the *Fortune* Top 100 (Essig, 2014).

Table 1. The four dichotomies of the Myers–Briggs Type Indicator assessment.

Extraversion–Introversion dichotomy (attitudes or orientations of energy)	
Extraversion (E)	Introversion (I)
Directing energy mainly toward the outer world of people and objects	Directing energy mainly toward the inner world of experiences and ideas
Sensing–Intuition dichotomy (functions or processes of perception)	
Sensing (S)	Intuition (N)
Focusing mainly on what can be perceived by the five senses	Focusing mainly on perceiving patterns and relationships
Thinking–Feeling dichotomy (functions or processes of judging)	
Thinking (T)	Feeling (J)
Basing conclusions on logical analysis with a focus on objectivity and detachment	Basing conclusions on personal or social values with a focus on understanding and harmony
Judging–Perceiving dichotomy (attitudes or orientations toward dealing with the outside world)	
Judging (J)	Perceiving (P)
Preferring the decisiveness and closure that result from dealing with the outer world using one of the judging preferences (thinking or feeling)	Preferring the flexibility and spontaneity that results from dealing with the outer world using one of the perceiving processes (sensing or intuition)

Note. Taken with permission from Myers et al. (1998).

The MBTI manuals (Myers et al., 1998; Quenk et al., 2004) document extensive research, including its adaptation and validation in more than 20 languages and cultures, and are regularly supplemented by new research from the test publisher (e.g., Hackston, 2015; Hackston & Dost, 2016; OPP, 2013, 2016). Additional research supporting its use is reported and reviewed elsewhere (e.g., Bayne, 1995; Carlyn, 1977; Hammer & Huszczo 1996; McCaulley, 2000), particularly in the *Journal of Psychological Type* (*JPT*), a peer-reviewed journal, published since 1977, which focuses on research relating to Jungian personality types.

Key criticisms of the MBTI

Despite, or perhaps because of its popularity, the MBTI has been the subject of considerable criticism, both in the academic literature (e.g., Chamorro-Premuzic et al., 2016; Furnham, 1990, 2017; Furnham & Crump, 2014; McCrae & Costa, 1989; Michael, 2003; Pittenger, 2005) and by popular psychology authors (e.g., Carter, 2016; Essig, 2014; Grant, 2013; Murphy Paul, 2004). Whereas some criticisms are backed by scientific analysis and reasoned argument, others are opinion pieces with little substance behind them. For example, Essig (2014) wrote about the "mystery" of the MBTI's popularity, asserting, "[T]he MBTI is pretty much nonsense, sciencey snake oil. As is well-established by research, it has no more reliability and validity than a good Tarot card reading." Chamorro-Premuzic et al. (2016) stated, "In a world driven by accuracy, the Myers–Briggs would not be the most popular assessment tool" (p. 635). Carter (2016) ended with a rallying cry of, "The fight against the MBTI will continue" (p. 30). The most common criticisms are summarized and addressed next.

Trait not type

Critics argue that personality is best described by continuous, normally distributed traits, rather than by discontinuous types (Barbuto, 1997; Furnham, 2017; Pittenger, 2005). The MBTI Step I, in contrast, is designed to sort individuals into one of 16 categories. Several critics of the MBTI state that this categorization does not capture the full range of personality variance and reduces predictive power (e.g., Barbuto, 1997; Grant, 2013), describing type concepts as "out of date" (Furnham, 2017) and as a "misrepresentation of the available evidence" (Pittenger, 2005). Whereas the conceptualization of personality variables as equal-interval continuous or integer-valued quantities (traits) is the mainstream view of academic psychometricians, measurement theorists dispute this stance (Michel, 2000, 2012; Tafreshi, Slaney, & Neufeld, 2016), with Michel characterizing it as "methodologically thought disordered" and "pathological science."

In any case, MBTI and Jungian theory have never suggested that anyone limits behavior to just one side of a dichotomy. On the contrary, the theory posits that we all use both sides, but with a preference for one side over the other, just as we have a preference to write either with our right or left hand, but we can develop skill in using both hands. For example, everyone

needs to act in the external world (extraversion) but also needs time for reflection (introversion). The MBTI Step I questionnaire sets out to capture an individual's underlying preference, but their behavior will also relate to their current situation and past environmental influences. In MBTI theory, we can choose whether to act in an extraverted or an introverted way, although one will be easier, and require less energy (Myers & Myers, 1995).

Dividing personalities into just 16 types is of course, a simplification of human nature. If the goal is to capture maximal variance and to predict behavior from the scores alone, then the MBTI is not the right assessment to use. It does, however, provide simple labels and useful rules of thumb to help people understand individual differences, without overwhelming them with too much information. For those who wish to go further, MBTI Step II captures more of the continuously distributed behavioral differences between people, or a trait tool such as the NEO PI or 16PF can be of benefit. Even critics such as Pittenger (2005) concede that "type-as-a-label" has great utility for this introductory stage of personnel development. Issues of poor practice arise when MBTI Step I scores are erroneously interpreted as if they measure behavior, rather than an indication of categorical preference.

A linked criticism is that if the MBTI dimensions were truly dichotomous, then MBTI continuous scores should have a bimodal distribution, but do not (Arnau, Green, Rosen, Gleaves, & Melancon, 2003; Girelli & Stake, 1993). However, other studies have shown that when item-response theory methods are used to score the MBTI (as with the current Form M version), scores are indeed bimodal (e.g., Harvey & Murray, 1994). In any event, this focus might somewhat miss the point, given the utility of simple categories to facilitate lay understanding.

Test–retest reliability

Critics assert that the MBTI has poor test–retest reliability. For example, Pittenger (2005) noted that a high percentage of people change at least one dichotomy when they take the MBTI questionnaire a second time. However, in looking to replicate the same four-letter type (i.e., all four dimensions simultaneously), such critics are holding the MBTI to a higher level of repeatability than is used for trait measures, which only ever report reliability one scale at a time. Each of the MBTI dimensions shows excellent stability; for example, the U.S. Form M of the MBTI shows test–retest correlations of between .83 and .97 over a 4-week interval, higher than that of many established trait measures, and over intervals greater than 9 months MBTI Form G also showed good stability (.77–.84). Moreover, there is agreement of 84% to 96% for each dichotomy over 4 weeks, with a median of 90% (Myers et al., 1998). The chance of coming out the same type on all four scales would therefore be 0.90^4, or 66%, which is very close to the observed 65% in field research, with 93% of respondents maintaining the same four-letter type, or changing just one dimension (Myers et al., 1998, p. 164).

Salter, Forney, and Evans (2005) noted that MBTI test–retest reliability studies have had mixed results, in part due to "unsophisticated analytical strategies." In their own analysis, they concluded, "if the goal of using the MBTI instrument is to help individuals to become aware of their 'true type' dispositions, which should remain relatively stable over time, then our results seem consistent with that objective" (p. 217). A meta-analytic study of the MBTI by Capraro and Capraro (2002) found strong internal consistency and test–retest reliability. Across all dimensions, median internal consistency reliability was 0.816 (from 50 coefficients) and median test–retest reliability was 0.813 (from 20 coefficients). The lowest reliability was 0.480, on the T-F dimension, from a test–retest study of 17 men; the highest was 0.97, on the S–N, T–F, and J–P dimensions, from a sample of 343 senior managers.

Predictive validity

Pittenger (2005) noted that there is a "conspicuous lack of data demonstrating the incremental validity of the MBTI over other measures of personality" (p. 218). Boyle (1995), Furnham (2017), Grant (2013), and McCrae and Costa (1989) made similar critiques. These criticisms appear to derive from three misconceptions: first, that the purpose of the MBTI is similar to that of personality assessments used for employee selection (predicting job performance); second, that because such information is lacking, the MBTI does not therefore possess any criterion-related validity; and third, that any validity the MBTI does possess does not show any incremental validity over that of other personality instruments. All three assertions are ill-founded.

A central tenet of MBTI theory is that individuals can choose to act against type (or "flex"), if the occasion demands it, and over time they might become very proficient at acting in a nonpreferred way (Myers & Myers, 1995). It is therefore not surprising that an individual's four-letter type preferences might not relate to job performance. The validity of the MBTI has, however, been demonstrated in a range of relevant contexts. Examples include the following.

- Homogeneity within organizations, as predicted by Schneider's (1987) Attraction–Selection–Attrition (ASA) theory (Quintero, Segal, King, & Black, 2009; Thomas, Benne, Marr, Thomas, & Hume, 2000; Wallick, Cambre, & McClugage, 2000).
- Career search (Tinsley, Tinsley, & Rushing, 2002).
- Dealing with conflict (Insko et al., 2001; Kilmann & Thomas, 1975; Mills, Robey, & Smith, 1985).
- Decision making (Gallen, 2006; Haley & Stumpf, 1989; Hough & Ogilvie, 2005).
- Interplay of occupational and organizational membership (Bradley-Geist & Landis, 2012).
- Health, well-being, coping, and stress (Allread & Marras, 2006; Buckworth, Granello, & Belmore, 2002; Du Toit, Coetzee, & Visser, 2005; Horacek & Betts, 1998; Short & Grasha, 1995).
- Relationship with occupational interests (Briggs, Copeland, & Haynes, 2007; Fleenor, 1997; Garden, 1997).
- Ratings of transformational leadership (Brown & Reilly, 2009; Hautala, 2005, 2006; Sundstrom & Busby, 1997).
- Use of technology, e-mail, and social media (Bishop-Clark, Dietz-Uhler, & Fisher, 2006–2007; Bowen,

Ferguson, Lehmann, & Rohde, 2003; Goby, 2006; Hackston & Dost, 2016; Weber, Schaubhut, & Thompson, 2011).

- Working in teams (Amato & Amato, 2005; Choi, Deek, & Im, 2008; Glaman, Jones, & Rozelle, 1996; Hammer & Huszczo, 1996; Schullery & Schullery, 2006).

Whereas the Five-Factor Model (FFM) demonstrates incremental validity over the MBTI in predicting job performance (e.g., Furnham, Jensen, & Crump, 2008), the MBTI has shown incremental validity over trait questionnaires in other situations. For example, Edwards, Lanning, and Hooke (2002) confirmed the incremental validity of the MBTI instrument over the NEO PI–R in predicting attributional adjustment, with no significant effects relating to the NEO. The interaction effect of Judging-Perceiving × Sensing-Intuition × Impressions was significant, $t(265) = 2.45$, $\beta = 0.124$, $p < .026$ (see Table 1 for definitions of Judging-Perceiving and Sensing-Intuition dimensions). Pulver and Kelly (2008) showed that the MBTI assessment added predictive power to the Strong Interest Inventory assessment in students' selection of study majors, improving correct classifications in a discriminant analysis by 3%. A study by Renner, Bendele, Alexandrovicz, and Deakin (2014) used confirmatory factor analysis to demonstrate that the MBTI adds unique explanatory variance over and above the NEO Five-Factor Inventory. In their study, a model that assumed two distinct but correlated factors for each of the NEO–MBTI matched scales (Comparative Fit Indices [CFIs] of 0.720–0.824) described the data better than either a model assuming two orthogonal factors (CFIs of 0.653–0.753) or a single factor (0.652–0.758).

For a psychometric tool used in development, arguably the most important aspect of predictive validity is whether it has demonstrated effective outcomes (Rogers, 2017; Scoular, 2011). The effectiveness of MBTI-based interventions has been shown in many contexts. For example, McPeek et al. (2013) showed positive effects on student grades following MBTI-based training with teachers (Cohen's $d = 0.16$). Katz, Joyner, and Seaman (1999) found that community college students were as likely to change career goals following MBTI feedback as they were following interest inventory feedback, and more likely to change following joint feedback, $\chi^2(3, N = 427) = 10.64$, $p = .01$. Leong, Hardin, and Gaylor (2005) found that medical students reported more certainty in career choice after an MBTI-based workshop than before, $F(1, 107) = 11.71$, $p = .001$, Cohen's $d = 0.29$. Stockill (2014) reported improved ratings of teamwork after an MBTI-based intervention (Cohen's $d = 0.50$). Positive effects of MBTI-based interventions have also been reported in relation to improving communication (Ang, 2002), improving problem-solving style in teams (Sedlock, 2005), and designing residential environments (Schroeder, Warner, & Malone, 1980).

Factor structure and the absence of neuroticism

McCrae and Costa (1989) reported correlations between the MBTI and the NEO PI separately for men and women. Correlations were consistently in the expected direction: E-I and Extraversion ($r = .74$ for men; $r = .69$ for women), S-N and Openness to Experience ($r = .72$ for men; $r = .69$ for women), T-F and Agreeableness ($r = .44$ for men; $r = .46$ for women) and J-P and Conscientiousness ($r = .49$ for men; $r = .46$ for women). Similar results have been found by others with both the NEO and 16PF (Furnham, 1996; Furnham, Moutafi, & Crump, 2003; OPP, 2016; Russell & Karol, 1994).

One interpretation of these findings is that this is a demonstration of construct validity; these four factors emerge from Jung's observations, Myers and Briggs's assessment, and the empirical approach of the NEO and 16PF. However, many critics prefer to highlight that the MBTI is missing an important factor, neuroticism (e.g., McCrae & Costa, 1989; Furnham, 2018). Some go so far as to use this finding as evidence that the MBTI is subsumed by the FFM and therefore redundant (Pittenger, 2005), even though the incremental validity research mentioned earlier contradicts this.

The absence of a measure of neuroticism is a spurious criticism. Although the MBTI framework does include consideration of stress and anxiety (Quenk, 1998, 2002), there is no claim that the questionnaire itself measures this factor of personality, nor that questionnaire results will enable predictions about individuals that relate to state or trait anxiety. Instead, a deliberate decision was made in the assessment's construction not to add in this fifth factor to the assessment, so as to keep focus on the positive and productive differences between people (Myers et al., 1998), an approach that has become a core tenet of the strengths movement (e.g., Peterson & Seligman, 2004).

There is no necessary virtue in an assessment providing full coverage of every aspect of personality. For example, a recent review of predictors of job performance (Schmidt, Oh, & Shaffer, 2016) concluded that only one of the Big Five dimensions (Conscientiousness) consistently provides incremental predictive power. For employee development, many practitioners judge that the positive language associated with the MBTI, and the absence of neuroticism, is of much greater advantage than using a more comprehensive measure; the MBTI is therefore often favored as the first personality assessment to be introduced. Once the ice is broken and where time permits, additional personality measures can be used; these might well include a trait measure of anxiety (Passmore, 2012; Rogers, 2017; Scoular, 2011).

The factor structure and construct validity of the MBTI have also been criticized. For example, Sipps, Alexander, and Freidt (1985) found a six-factor solution, and Saggino and Kline (1996) found that the factor structure of the Italian research version did not fit the MBTI model. However, other studies have supported the four-factor structure. Saggino, Cooper, and Kline (2001) found that the models that best fit the data were the four-factor model (CFI = 0.621) and a five-factor model consisting of the four MBTI dimensions plus an additional factor (CFI = 0.750). Harvey, Murry, and Stamoulis (1995) found a four-factor solution, with goodness-of-fit indexes (GFIs) for oblique models ranging from .744 to .900, as did Bess, Harvey, and Swartz (2003; GFI = 0.854) and Thompson and Borello (1989; GFI = 0.78). Although none of the latter three studies exceeded a GFI of 0.90, all found that a four-factor solution was the best fit to the data.

Fakeability

All self-report personality questionnaires are reliant on some degree of self-awareness and honesty. The MBTI is often

criticized for being highly fakeable (Carter, 2016; Furnham, 1990), and research has demonstrated that in Western cultures and organizations, there is a degree of social pressure to conform to extraverted, sensing, thinking, and judging preferences, as defined in Table 1 (Kendall, 1998). However, while demonstrating that faking can happen, Furnham (1990) also noted that subjects found the MBTI questionnaire difficult to fake, and concluded that if faking occurs, it is not particularly easy to do.

Fakeability is critical when questionnaires are being used for high-stakes assessment, such as to determine future opportunity. Used correctly in a developmental context, there should be no pressure to fake a particular profile, as the primary audience for the results is the individual themselves. Moreover, unlike traditional trait-based questionnaires, the MBTI process does not take the questionnaire results as the final categorization. It is intended to be used as an indicator of an individual's preference. Questionnaire data are one component of feedback with a trained practitioner in exploring what might be an individual's best fit type. During this process, any cultural, social desirability, or other pressures to be of a certain personality type (or to fake the questionnaire results) can be explicitly discussed and resolved. Therefore, fakeability is not a major concern for MBTI use.

Barnum effects

MBTI interpretation and reports have been criticized as using the Barnum effect, where the descriptions of individuals seem insightful but would in fact apply to anyone. For example, Pittenger (1993) stated, "The descriptions of each type are generally flattering and sufficiently vague so that most people will accept the statements as true of themselves" (p. 486). However, Carskadon and Cook (1982) refuted the idea that type descriptions other than one's own might be equally appealing. Individuals were shown four type descriptions and asked to rank order them in terms of their accuracy. Chi-square analysis showed that the distribution of ranks was nonrandom, $\chi^2(3) = 48.98$, $p < .001$, and that a far greater than expected proportion of subjects ranked their assessed description as number one compared to all other descriptions combined, $\chi^2(11) = 59.0$, $p < .001$.

Applying evidence-based practice: The scientist–practitioner divide

Much has been written about the scientist–practitioner divide in occupational and organizational psychology (e.g., Andersen, Herriot, & Hodgkinson, 2001; Gray, Iles, & Watson, 2010). Andersen et al. (2001) are typical in noting

> Practitioners and researchers have often held stereotypical views of each other, with practitioners viewing researchers as interested only in methodological rigor whilst failing to concern themselves with anything in the real world, and researchers damning practitioners for embracing the latest fads, regardless of theory or evidence. (p. 392)

Such debates are generally accompanied by a call for greater evidence-based practice (e.g., Barends, Rousseau, & Briner, 2014; Gifford, 2016), urging more attention to the scientific literature, so as to take advantage of the best available evidence in designing and delivering successful interventions. We agree wholeheartedly with this intent. Barends et al. (2014) recommended that four different kinds of evidence should be considered: scientific, organizational, evidence from practitioners (professional judgment, tacit knowledge), and stakeholder evidence (from people affected by the decision).

Academics and researchers frequently give precedence to scientific evidence, defined by Barends et al. as findings "from empirical studies published in academic journals." Employee selection is an example where science and practice have successfully combined (Barends et al., 2014; Gifford, 2016), with a substantial, well-reviewed, and consolidated body of literature from which practitioners can identify relevant research (e.g., Robertson & Smith, 2001; Schmidt et al., 2016; Schmidt & Hunter, 1998). Unfortunately many writers extrapolate to argue that assessments valid for selection are therefore the most appropriate for all organizational applications, as they mistakenly believe that the value of a personality assessment is always its ability to afford useful predictions of work performance (e.g., Barrick & Mount, 2005; Chamorro-Premuzic et al., 2016; Pittenger, 2005). This is not the case when it comes to employee development.

When practitioners choose a personality assessment by considering organizational factors and subjective experiences, we contend that, rather than them ignoring the scientific evidence, they are in fact taking advantage of the best available evidence. Personal, peer, and colleagues' experiences, rather than being the irrelevant noise of the "latest fads" (Andersen et al., 2001), are actually important and valuable data in choosing an assessment, and might also be the only data available that takes the specific organizational context into account.

It is recognized that practitioners choose different personality assessments for use in development versus selection (Furnham, 2008a; Furnham & Jackson, 2011). However, when researchers ask practitioners, "How valid do you rate this test?" (e.g., Furnham, 2008a), many distinct forms of validity and applications are confounded, with conclusions then drawn that ignore this distinction. Although Furnham and Jackson (2011) lamented the fact that simpler tests have widespread appeal, it could be that test users understand and rightly place a higher weighting on factors other than the psychometric robustness and comprehensiveness of the questionnaire.

Chamorro-Premuzic et al. (2016) principally focused on the need to classify individuals as more or less talented. In doing so, they confused selection and development applications, asserting that the MBTI does not have a place in a "world driven by accuracy." As outlined earlier, the MBTI was never intended to sort the talented from the less talented. Comments that the MBTI is used in selection, and then criticisms on those grounds (e.g., Carter, 2016; Pittenger, 2005), are irrational, and feed confusion about these distinct practices. When the MBTI is used for selection, this is despite repeated explanations and specific training by the test publisher, who will refuse to supply practitioners with product if they are found to be misusing the instrument in this way (OPP, 2017).

Klehe (2004) also noted the difference between academics' recommendations for personnel selection and actual practice, even within their own universities. She understood that it is not a simple matter of education of the organizational client; that

practitioners' choices are a result of weighting multiple, sometimes contradictory institutional pressures. Klehe advocated that researchers develop a practitioner-oriented research agenda with respect for these additional factors. The same is true for research on employee development, where a single focus on assessment accuracy denies the complexity involved in achieving the desired outcomes.

Differences in goals, appropriate evidence, and criteria for assessment choice

In assessing the validity of a personality assessment it is critical to be clear about the application for which it is intended. Who are the results intended for? What outcomes should results lead to? Which aspects of the assessment's validity are most important for that purpose?

Goals of using personality assessment for selection

In selection, the goal of personality assessment is to provide data to contribute to a decision to select out or select in candidates, as part of a multifaceted process, such as an assessment center (Cook, 2004; Smith & Smith, 2005). The tool's ability to accurately measure personality and to predict job performance from those results is therefore critical. The individual who took the assessment might not even see the results.

Psychometric properties, including reliability, construct, content, and predictive validity, are typically quoted as important in choosing a personality assessment (Cicchetti, 1994; Cook, 2004). Other considerations gaining attention include the candidate experience (Ekuma, 2012), acceptance of feedback (Atwater & Brett, 2006; Krings, Jacobshagen, Elfering, & Semmer, 2015), and quality of interpretive reports (De Fruyt & Wille, 2013). Nevertheless, much training in and critique of the use of tests and questionnaires concentrates on construct, content, and predictive validity in a selection context (e.g., British Psychological Society, 2017).

As with any commercial activity, there is a trade-off of cost, time, and quality in reaching the appropriate solution (Klehe, 2004). Organizational clients are often looking for the quickest, cheapest solution or are concerned that potential candidates might be put off by an overly in-depth process. Some psychologists consider the use of brief screening assessments controversial, but this is preferable to organizations basing decisions on an unstructured interview and a resume, which remain the main methods used in many, particularly smaller organizations (Zibarras & Woods, 2010).

Goals of using personality assessment in development

In development, personality results are not used to predict performance, but as a vehicle for increasing self-awareness (Cseh, Davies, & Khilhi, 2013; Rogers, 2017; Scoular, 2011; Tjan, 2012), so that employees can make more conscious choices about their behavior. The personality measure is a starting point for that change, not a predictor of the outcome. The key audience is typically neither the HR practitioner, nor the organization's management, but the individual who took the assessment.

Self-awareness predicts outcomes from well-being (e.g., Harrington & Loffredo 2011) to leadership effectiveness (e.g., Atwater & Yammarino, 1992; Moshavi, Brown, & Dodd, 2003; Van Velsor, Taylor, & Leslie, 1993). In a rare experimental field study, Sutton, Allinson, and Williams (2013) showed that self-awareness improved as a result of training with a personality type instrument (Enneagram); the reflection and insight gained was positively associated with job contentment and enthusiasm, and with improvements in relationships and communication with colleagues. This is consistent with much anecdotal evidence in the HR and business literature, and practice in many organizations (Dierdorff & Rubin, 2015; Drucker, 2005; Grant, Franklin, & Langford, 2002; Tjan, 2012).

Tjan (2012) stated that the best thing leaders can do to improve their effectiveness is to become more aware of what motivates them and their decision making. He noted that "Personality tests like Myers–Briggs, Predictive Index, and StrengthsFinder have gained popularity in recent years, for good reason. It's not that such tests are prefect measures or predictors, but they facilitate self-reflection, which leads to better self-awareness."

In development, the focus is not the scores on the assessment but what is done with those scores. What insights are illuminated? What actions are taken as a result? How are any barriers to change overcome? This wider context might have at least as much to do with the facilitation or coaching skills of the practitioner as with the "scientific rigor" of test results. Simpler measures can have an advantage over more comprehensive models, as they are easier to grasp quickly and can provide more memorable learning for participants. As Rogers (2017) put it, "the unfussy neatness of the MBTI … makes it accessible, memorable and infinitely flexible" (p. 194). For the same reasons, the quality of not just interpretive reports, but also associated materials and resources that explain and reinforce the key learnings, are an essential component of modern employee development interventions. Simple measures lend themselves to high-impact, engaging learning experiences.

That is not to say that accuracy of the assessment is irrelevant; random or meaningless results would be of no value. Reliability and some forms of validity remain important, alongside these other factors. In summary, the criteria for selecting a personality assessment for use in a developmental intervention needs to take into account overlapping, but not the same criteria as those needed for a selection application.

Criteria for choosing a personality assessment to include in a development process

Given the goals and context of using personality assessment in development, the traditional criteria for judging assessments are less relevant than they are for selection, or take on a different emphasis. Assessments still need to be reliable, showing internal consistency and temporal stability, but face validity needs to operate in a different way to engage the individual in developmental actions, and content validity might have a different character. It is often not necessary to cover all aspects of personality, but instead to focus on those relevant to the desired developmental outcome. Construct validity might also have a different emphasis. There should be a clear structure that can

be understood by the end client, but there is no requirement to map onto the FFM. Criterion-related validity should be focused on predicting developmental outcomes rather than predicting job performance. Fakeability is less of a concern, as development is a very different context from high-stakes selection.

Additional criteria are, however, also important when choosing and using personality assessments for development. It is the whole experience that determines whether the intervention is successful, not just the assessment. Assessments should still be critically evaluated against the preceding criteria, but the evidence to be considered goes beyond psychometric properties to include considerations such as the balance of simplicity versus time availability. A very accurate and detailed assessment is of little value if it is too complicated for the test taker to understand, remember, or apply (Rogers, 2017). Interpretive reports and resources that provide accurate, understandable feedback are essential. The concept of user validity (MacIver, Anderson, Costa, & Evers, 2014) has shown that practitioner interpretation of test scores affects the test's validity. This concept can be extended to include the interpretation that a test taker makes of the feedback and reports that he or she receives.

In development, reports that provide nonthreatening, positive feedback tend to be more effective. It is important that employees not only understand, but also are engaged with and accept the results if they are to commit to developmental change. Positive language can be very helpful, as is initially holding back on some uncomfortable truths (e.g., anxiety and other negative dimensions; Atwater & Brett, 2006; Furnham & Varian, 1988; Krings et al., 2015). Having resources such as experiential exercises, videos, interactive Web sites, fun giveaways, and memory aids that create high impact, are also useful. These are not just fashionable gimmicks, but effective ways to reinforce learning. Even something as simple as the systematic use of color can be critical to learning impact (Keller, Gerjets, Scheiter, & Garsoffky, 2006).

Finally, the skill of the practitioner as a coach or as a team facilitator is crucial. As noted by Athanasopoulou and Dopson (2015), "these inventories have no value unless a coach has solid understanding of how to effectively use them" (p. 84). High-quality training and service should be available to support the practitioner to get the most out of whichever assessment they use. In most countries outside the United States, instrument-specific practitioner training is accepted and expected and often results in higher quality practice than experienced from psychologists, who are assumed to have sufficient skills from their degree education to apply any psychometric questionnaire, whether or not it was specifically referenced within their degree courses.

Much of the assessment validity evidence available in the scientific literature is not relevant for developmental applications. Moreover, research on the predictive and criterion-related validity of employee development tends to be less strong than in the selection domain. Very few HR departments or practitioners measure outcomes; the Chartered Institute of Personnel and Development (CIPD, 2015) reported that 51% of HR professionals surveyed did no evaluation on learning and development activities beyond simple satisfaction surveys and only one fifth assessed behavioral change. Moreover, developmental outcomes are particularly difficult to measure, might not be immediate, and might arise from multiple causes. As noted by Athanasopoulou and Dopson (2015), in discussing the effectiveness of executive coaching, any assessment is just one factor and it is difficult to isolate its impact in any scientifically rigorous way.

It can also be difficult to get relevant research published in the academic literature. For example, new editions and revalidation of tests for a new language or culture might not be considered sufficiently cutting edge for academic journals, but are dismissed as self-serving when included in a test publisher's manual or in a specialist journal, such as the *JPT*.

Going beyond the published scientific evidence, then organizational evidence, evidence from practitioners, and stakeholder evidence as described by Barends et al. (2014) are all highly relevant. Referring to the business and management literature and case studies might be useful; although less rigorous than the scientific literature, it is not necessarily invalid. In development, the impact on the end user is key. Views from those who have experienced an assessment in context can provide important insights, and are almost completely neglected in the literature.

We suggest that more research should be carried out into what could be termed *experiential validity*. Rather than relying solely on the perspective of HR practitioners and psychometricians, experiential validity brings the test taker's perspective to the fore, going beyond mere face validity, to determine whether the person completing the assessment experienced the assessment process (including feedback) as personally valuable. Additional components could include the following: Were the intended outcomes from the development achieved? Can key learnings be recalled months or years later? Is there ongoing impact at work? Defining and systematically measuring the components of experiential validity could provide the basis for a new and insightful avenue of assessment validity research that could shed light on the relative utility of assessments for employee development.

Inevitably, some of this information will come from test publishers. Although it is perfectly reasonable to be wary of what publishers say to promote their own products, it is worthwhile remembering that many are represented by psychologists and psychometricians with a depth of academic and practitioner expertise, who regularly present their research in public forums, and who would certainly not see themselves as selling "sciencey snake oil" (Essig, 2014). Additionally, it is worth noting that neither academics nor journalists are necessarily disinterested parties in this debate. Many academic writers have their own assessments, commercial associations with rival test publishers, or sell their own consulting services, and like journalists, want to capture attention with a memorable headline.

Conclusions

Although some personality tools are used for both selection and development, others are used principally in selection or largely (even solely) in development. Many academic reviewers have been highly critical of those assessments most popular in the development arena. Using the MBTI as an example, we have argued that many of these criticisms have been misguided and misleading.

A common thread through much of this critique is a misunderstanding of, or lack of attention to, the important differences

between the requirements of personality assessment in selection and development contexts. This is reflected in a relative lack of academic research on assessment for development. Although there is some commonality between the criteria that are important for choosing a selection assessment, and those relevant for choosing an assessment to be used in development, there are unique features, too. Using the approach recommended by Barends et al. (2014), practitioners should draw on a wide range of evidence to inform their practice. This is in direct contradiction to the advice of many academics. For example, Andersen et al. (2001) are dismissive of much of the evidence that HR practitioners take into account, including "popularist books on emotional intelligence, unvalidated claims in respect of team-building and OD interventions, and self-produced 'validation' studies by less reputable test publishers" (p. 394). We argue that organizational clients are not so much pushing toward so-called popularist science, as using metrics of utility that are highly relevant to their applications. Although academics do conduct pragmatic research, this is most often on topics that are not wholly relevant to the question of employee development. Given that the latter represents a significant investment by organizations, and plays a part in helping clients deal with the present very challenging times, we would encourage such research, enabling academics and practitioners to learn from each other. To this end, we recommend a new branch of research, into what we have termed experiential validity, systematically measuring the perspectives and experiences of test takers in developmental contexts, to identify which assessments have the greatest lasting developmental impact. In this way, academics could not just recognize the importance of stakeholder evidence as recommended by Barends et al. (2014), but could incorporate this perspective into the body of scientific evidence.

Acknowledgments

We are grateful to our colleagues at OPP Ltd. and Meyler Campbell for their support in developing these ideas, and for the encouragement and valuable review from the editor Dr. Hal Shorey, as well as two anonymous reviewers. Their insights have greatly contributed to and improved this article.

References

Allread, W. G., & Marras, W. S. (2006). Does personality affect the risk of developing musculoskeletal discomfort? *Theoretical Issues in Ergonomicfs Science*, *7*(2), 149–167. doi:10.1080/14639220500076504

Amato, C. H., & Amato, L. H. (2005). Enhancing student team effectiveness: Application of Myers-Briggs personality assessment in business courses. *Journal of Marketing Education*, *27*, 41–51. doi:10.1177/0273475304273350

Andersen, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner-researcher divide in Industrial, Work and Organizational (IWO) psychology: Where are we now, and where do we go from here?, *Journal of Occupational and Organizational Psychology*, *74*, 391–411. doi:10.1348/096317901167451

Ang, M. (2002). Advanced communication skills: Conflict management and persuasion. *Academic Medicine*, *77*(11), 1166. doi:10.1097/00001888-200211000-00034

Arnau, R., Green, B., Rosen, D., Gleaves, D., & Melancon, J. (2003). Are Jungian preferences really categorical? An empirical investigation using taxometrical analysis. *Personality and Individual Differences*, *34*, 233–251. doi:10.1016/S0191-8869(02)00040-5

Athanasopoulou, A., & Dopson, S. (2015). *Developing leaders by executive coaching*. Oxford: Oxford University Press.

Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, *45*, 141–164. doi:10.1111/j.1744-6570.1992.tb00848.x

Atwater, L., & Brett, J. (2006). Feedback format: Does it influence manager's reactions to feedback?. *Journal of Occupational and Organizational Psychology*, *79*, 517–532. doi:10.1348/096317905X58656

Barbuto, J. E. (1997). A Critique of the Myers-Briggs Type Indicator and its operationalization of Carl Jung's psychological types. *Psychological Reports, 80*(2), 611–625 doi:10.2466/pr0.1997.80.2.611

Barends, E., Rousseau, D. M., & Briner, R. B. (2014). *Evidence-based management: The basic principles*. Amsterdam: Centre for Evidence-Based Management.

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance, 18*(4), 359–372 doi:10.1207/s15327043hup1804_3

Bayne, R. (1995). *The Myers-Briggs Type Indicator: A critical review and practical guide*. Cheltenham: Stanley Thornes.

Bernreuter, R. G. (1931). *The personality inventory*. Stanford, CA: Stanford University Press.

Bess, T. L., Harvey, R. J., & Swartz, D. (2003). Hierarchical confirmatory factor analysis of the Myers-Briggs Type Indicator. *Proceedings of the Society for Industrial and Organizational Psychology*, 18–23.

Bishop-Clark, C., Dietz-Uhler, B., & Fisher, A. (2006-2007). The effects of personality type on web-based distance learning. *Journal of Educational Technology Systems, 35*(4), 491–506. doi:10.2190/DG67-4287-PR11-37K6

Bowen, P. L., Ferguson, C. B., Lehmann, T. H., & Rohde, F. H. (2003). Cognitive style factors affecting database query performance. *International Journal of Accounting Information Systems*, *4*, 251–273. doi:10.1016/j.accinf.2003.05.002

Boyle, G. J. (1995). Myers-Briggs Type Indicator (MBTI): Some psychometric limitations. *Australian Psychologist, 30*(1), 71–74. doi:10.1111/j.1742-9544.1995.tb01750.x

Bradley-Geist, J. C., & Landis, R. S. (2012). Homogeneity of personality in occupations and organizations: A comparison of alternative statistical tests. *Journal of Business Psychology*, *27*, 149–159. doi:10.1007/s10869-011-9233-6

Briggs, S. P., Copeland, S., & Haynes, D. (2007). Accountants for the 21st century, where are you? A five-year study of accounting students personality preferences. *Critical Perspectives on Accounting*, *18*, 511–537. doi:10.1016/j.cpa.2006.01.013

British Psychological Society. (2017). *The BPS qualifications in test use*. Retrieved from http://ptc.bps.org.uk/sites/ptc.bps.org.uk/files/Documents/Guidelines%20and%20Information/BPS%20qualifications%20in%20Test%20Use%202017_0.pdf

Brown, F. W., & Reilly, M. D. (2009). The Myers-Briggs Type Indicator and transformational leadership. *Journal of Management Development, 28*(10), 916–932. doi:10.1108/02621710911000677

Buckworth, J., Granello, D. H., & Belmore, J. (2002). Incorporating personality assessment into counseling to help students adopt and maintain exercise behaviors. *Journal of College Counseling*, *5*, 15–25. doi:10.1002/j.2161-1882.2002.tb00203.x

Capraro, R. M., & Capraro, M. M. (2002). Myers-Briggs Type Indicator score reliability across studies: A meta-analytic reliability generalization study. *Educational and Psychological Measurement, 62*(4), 590–602. doi:10.1177/0013164402062004004

Carlyn, M. (1977). An assessment of the Myers-Briggs Type Indicator. *Journal of Personality Assessment, 41*(5), 461–473. doi:10.1207/s15327752jpa4105_2

Carskadon, T. G., & Cook, D. D. (1982). Validity of MBTI type descriptions as perceived by recipients unfamiliar with type. *Research in Psychological Type*, *5*, 89–94.

Carter, C. (2016). The Myers-Briggs Type Indicator: Still going strong and still getting it wrong. *OP Matters*, *32*, 29–30.

Cattell, R. B. (1946). *The description and measurement of personality*, New York, NY: Harcourt, Brace & World.

Cattell, R. B., Cattell, A. K., & Cattell, H. E. P. (1993). *16PF fifth edition questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, *9* (3), 621–640. doi:10.1017/iop.2016.6

Chartered Institute of Personnel and Development (2015). Annual Survey Report: Learning and Development 2015. London: CIPD. Retrieved from https://www.cipd.co.uk/knowledge/strategy/development/surveys

Choi, K. S., Deek, F. P., & Im, I. (2008). Exploring the underlying aspects of pair programming: The impact of personality. *Information and Software Technology*, *50*, 1114–1126. doi:10.1016/j.infsof.2007.11.002

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instrument in psychology. *Psychological Assessment, 6*(4), 284–290. doi:10.1037/1040-3590.6.4.284

Conte, J. M. (2005). A review and critique of emotional intelligence measures. *Journal of Organizational Behavior*, *26*, 433–440 doi:10.1002/job.319

Cook, M. (2004). *Personnel selection* (4th ed.). Chichester: John Wiley & Sons.

Costa, , & McCrae, (1985). *The NEO personality inventory manual*. Odessa, FL: Psychological Assessment Resources.

Cseh, M., Davies, E. B., & Khilhi, S. E. (2013). Developing a global mindset: Learning of global leaders. *European Journal of Training and Development*, *37*, 489–499. doi:10.1108/03090591311327303

De Fruyt, F., & Wille, B. (2013). "Hey, This is not like me!" Convergent validity and personal validation of computerized personality reports. *Revue Europeenee de Psychologie Appliquee*, *63*, 287–294 doi:10.1016/j.erap.2013.07.001

Desson, S. (2017). *The Lumina spark technical manual*. Lumina Learning LLP. Retrieved from https://onedrive.live.com/?cid=f5a385ac9de87983&id=F5A385AC9DE87983%21107&authkey=!AIk0-F3Ick0veK0

Dierdorff, E. C., & Rubin, R. S. (2015). We're not very self-aware, especially at work. *Harvard Business Review*, March 2015. Retrieved from https://hbr.org/2015/03/research-were-not-very-self-aware-especially-at-work

Drucker, P. (2005). Managing Oneself. *Harvard Business Review*, January 2005. Retrieved from https://hbr.org/2005/01/managing-oneself

Du Toit, F., Coetzee, S. C., & Visser, D. (2005). The relation between personality type and sense of coherence among technical workers. *South African Business Review, 9*(1), 51–65.

Edwards, J. A., Lanning, K., & Hooke, K. (2002). The MBTI and social information processing: An incremental validity study. *Journal of Personality Assessment, 78*(3), 432–450. doi:10.1207/S15327752JPA7803_04

Ekuma, K. J. (2012). The importance of predictive and face validity in employee selection and ways of maximizing them: An assessment of three selection methods. *International Journal of Business and Management, 7*(22), 115–122 doi:10.5539/ijbm.v7n22p115

Essig, T. (2014). The mysterious popularity of the meaningless Myers-Briggs (MBTI). *Forbes*, September 29, 2014. Retrieved from https://www.forbes.com/sites/toddessig/2014/09/29/the-mysterious-popularity-of-the-meaningless-myers-briggs-mbti/#309d14171c79.

Fleenor, J. W. (1997). The relationship between the MBTI® and measures of personality and performance in management groups. In C. Fitzgerald & L. K. Kirby (Eds.), *Developing leaders: Research and applications in psychological type and leadership development* (pp. 115–138). Boston: Davies-Black Publishing.

Fletcher, C., Baldry, C., & Cunningham-Snell, N. (1998). The psychometric properties of 360 degree feedback: An empirical study and a cautionary tale. *International Journal of Selection and Assessment, 6*(1), 19–34. doi:10.1111/1468-2389.00069

Furnham, A. (1990). The fakeability of the 16PF, Myers-Briggs and FIRO-B personality measures. *Personality and Individual Differences*, *11* (7), 711–716. doi:10.1016/0191-8869(90)90256-Q.

Furnham, A. (1996). The Big Five versus the big four: The relationship between the Myers-Briggs Type Indicator (MBTI) and the NEO-PI five factor model of personality. *Personality and Individual Differences*, *21*, 303–307. doi:10.1016/0191-8869(96)00033-5

Furnham, A. (2008a). HR Professionals' believes about, and knowledge of, assessment techniques and psychometric tests. *International Journal of Assessment and Selection*, *16*, 300–305. doi:10.1111/j.1468-2389.2008.00436.x

Furnham, A. (2008b). Psychometric correlates of FIRO-B scores: Locating three FIRO-B scores in personality factor space. *International Journal of Selection and Assessment*, *16*, 30–45. doi:10.1111/j.1468-2389.2008.00407.x

Furnham, A. (2017). Myers-Briggs Indicator (MBTI). In Virgil Zeigler-Hill & Todd K. Shackelford (Eds). *The SAGE Handbook of personality and individual differences*. New York: Sage.

Furnham, A. (2018). The great divide: Academic vs practitioner criteria for psychometric test choice. *Journal of Personality Assessment*, *100*(5), TBA.

Furnham, A., & Crump, J. (2014). The dark side of the MBTI, *Psychology*, *5*, 166–171. doi:10.4236/psych.2014.52026

Furnham, A., & Jackson, C. (2011). Practitioner reaction to work related psychological tests. *Journal of Managerial Psychology*, *26*, 549–565. doi:10.1108/02683941111164472.

Furnham, A., Jensen, T., & Crump, J. (2008). Personality, intelligence and assessment centre expert ratings. *International Journal of Selection and Assessment*, *16*(4), 356–365. doi:10.1111/j.1468-2389.2008.00441.x

Furnham, A., Moutafi, J., & Crump, J. (2003). The relationship between the revised NEO-Personality Inventory and the Myers-Briggs Type Indicator. *Social Behavior and Personality: An International Journal, 31*(6), 577–548 doi:10.2224/sbp.2003.31.6.577

Furnham, A., Steele, H., & Pendleton, D. (1993)., A psychometric assessment of the Belbin team role self-perception inventory. *Journal of Occupational and Organizational Psychology*, *66*, 245–255. doi:10.1111/j.2044-8325.1993.tb00535.x

Furnham, A., & Varian, C. (1988). Predicting and accepting personality test scores. *Personality and Individual Differences*, *9*, 735–748 doi:10.1016/0191-8869(88)90063-3

Gallen, T. (2006). Managers and strategic decisions: Does the cognitive style matter? *Journal of Management Development*, *25*(2), 118–133. doi:10.1108/02621710610645117

Garden, A. M. (1997). Relationships between MBTI® profiles, motivation profiles and career paths. *Journal of Psychological Type*, *41*, 3–16.

Gifford, J. (2016). *In search of the best available evidence*. London: CIPD.

Girelli, S. A., & Stake, J. E. (1993). Bipolarity in Jungian type theory and the Myers-Briggs Type Indicator. *Journal of Personality Assessment, 60*(2), 290–301. doi:10.1207/s15327752jpa6002_7

Glaman, J. M., Jones, A. P., & Rozelle, R. M. (1996). The effects of co-worker similarity on the emergence of affect in work teams. *Group & Organizational Management, 21*(2), 192–215. doi:10.1177/1059601196212005

Goby, V. P. (2006, February). Personality and online/offline choices: MBTI profiles and favored communication modes in a Singapore study. *CyberPsychology and Behavior, 9*(1), 5–13. doi:10.1089/cpb.2006.9.5

Golden, J. P. (2004). *Golden personality type profiler*. Austin, TX: Psychological Corporation.

Grant, A. (2013). *Say goodbye to MBTI, the fad that won't die*. Retrieved from https://www.linkedin.com/pulse/20130917155206-69244073-say-goodbye-to-mbti-the-fad-that-won-t-die.

Grant, A. M., Franklin, J., & Langford, P. (2002). The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality, 30*(8), 821–835. doi:10.2224/sbp.2002.30.8.821

Gray, D., Iles, P., & Watson, S. (2010). Spanning the HRD academic-practitioner divide – bridging the gap through Mode 2 research. *Journal of European Industrial Training 35*(3), 247–263. doi:10.1108/03090591111120403

Hackston, J. (2015). *Type and work environment: A research study from OPP*. Oxford: OPP. Retrieved from http://www.opp.com/download/item/c881c66e653d40c691e09a5f277c25b7

Hackston, J., & Dost, N. (2016). *Type and Email Communication: A research study from OPP*. Oxford: OPP. Retrieved from http://www.opp.com/-/media/Files/PDFs/White_papers/Type-and-email-survey-report-2016-(1).pdf?la=en

Haley, U. C. V., & Stumpf, S. A. (1989). Cognitive trails in strategic decision-making: Linking theories of personalities and cognitions. *Journal of Management Studies, 26*(5), 477–497. doi:10.1111/j.1467-6486.1989.tb00740.x

Hammer, A. L., & Huszczo, G. E. (1996). Teams. In A. L. Hammer (Ed.), *MBTI® applications: A decade of research on the Myers-Briggs Type Indicator®* (pp. 81–104). Mountain View, CA: CPP, Inc.

Harrington, R., & Loffredo, D. A. (2011). Insight, rumination, and self-reflection as predictors of well-being. *The Journal of Psychology*, *145* (1), 39–57. doi:10.1080/00223980.2010.528072

Harvey, R. J., & Murry, W. D. (1994). Scoring the Myers-Briggs Type Indicator: Empirical comparison of preference score versus latent-trait methods. *Journal of Personality Assessment, 62*(1), 116–129. doi:10.1207/s15327752jpa6201_11

Harvey, R. J., Murry, W. D., & Stamoulis, D. (1995). Unresolved issues in the dimensionality of the Myers-Briggs Type Indicator. *Educational and Psychological Measurement*, *55*, 535–544. doi:10.1177/0013164495055004002

Hathaway, S. R., & McKinley, J. C. (1943). *Manual for the minnesota multiphasic personality inventory*. New York: Psychological Corporation.

Hautala, T. M. (2005). The effects of subordinates' personality on appraisals of transformational leadership. *Journal of Leadership and Organizational Studies, 11*(4), 84–92. doi:10.1177/107179190501100407

Hautala, T. M. (2006). The relationship between personality and transformational leadership. *Journal of Management Development, 25*(8), 777–794. doi:10.1108/02621710610684259

Horacek, T. M., & Betts, N. M. (1998). College students dietary intake and quality according to their Myers Briggs Type Indicator personality preferences. *Journal of Nutrition Education, 30*(6), 387–395. doi:10.1016/S0022-3182(98)70361-9

Hough, J. R., & Ogilvie, D. T. (2005, March). An empirical test of cognitive style and strategic decision outcomes. *Journal of Management Studies, 42*(2), 417–448. doi:10.1111/j.1467-6486.2005.00502.x

Insko, C. A., Schopler, J., Gaertner, L., Wildschut, T., Kozar, R., Pinter, B., Finkel, E. J., Brazil, D. M., Cecil, C. L., & Montoya, M. R. (2001). Inter-individual-intergroup discontinuity reduction through the anticipation of future interaction. *Journal of Personality and Social Psychology*, *80* (1), 95–111. doi:10.1037/0022-3514.80.1.95

Jacobs, R. L., & Washington, C. (2003). Employee development and organizational performance: A review of literature and directions for future research. *Human Resource Development International*, *6*(3), 343–354. doi:10.1080/13678860110096211

Jung, C. G. (1971). *Psychological types* (H.G. Baynes, Trans, revised by R. F. Hull). Collected (Vol. *6*). Princeton, NJ: Princeton University Press. (Original work published in 1921).

Katz, L., Joyner, J. W., & Seaman, N. (1999). Effects of joint interpretation of the Strong Interest Inventory and the Myers-Briggs Type Indicator in career choice. *Journal of Career Assessment*, *7*(3), 281–297. doi:10.1177/106907279900700306

Keller, T., Gerjets, P., Scheiter, K., & Garsoffky, B. (2006). Information visualizations for knowledge acquisition: The impact of dimensionality and color coding. *Computers in Human Behavior, 22*(1), 43–65. doi:10.1016/j.chb.2005.01.006

Kendall, E. (1998). *MBTI (European English Edition) step i manual supplement*. Mountain View, CA; CPP Inc.

Kilmann, R. H., & Thomas, K. W. (1975). Interpersonal conflict-handling behavior as reflections of Jungian personality dimensions. *Psychological Reports*, *37*, 971–980. doi:10.2466/pr0.1975.37.3.971

Klehe, U-C. (2004). Choosing how to choose: Institutional pressures affecting the adoption of personnel selection procedures. *Journal of Selection and Assessment*, *12*, 327–342. doi:10.1111/j.0965-075X.2004.00288.x

Krings, R., Jacobshagen, N., Elfering, A., & Semmer, N. K. (2015). Subtly offending feedback. *Journal of Applied Social Psychology*, *45*, 191–202. doi:10.1111/jasp.12287

Leong, F. T. L., Hardin, E. E., & Gaylor, M. (2005). Career specialty choice: A combined research-intervention project. *Journal of Vocational Behavior*, *67*, 69–86. doi:10.1016/j.jvb.2004.07.004

Lothian, A. M. (1996). *Insights discovery preference evaluator*. Dundee: Insights Learning and Development.

McCaulley, M. H. (2000). Myers-Briggs type Indicator: A bridge between counselling and consulting. *Consulting Psychology Journal: Practice and Research*, *52*, 117–132. doi:10.1037/1061-4087.52.2.117

McCrae, R. R., & Costa, P. T., Jr (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the Five-Factor Model of personality. *Journal of Personality*, *57*(1), 17–40. doi:10.1111/j.1467-6494.1989.tb00759.x

McDowall, , & Redman, (2017). *Psychological assessment – an overview of theoretical, practical and industry trends*. Paper presented at the British Psychological Society Division of Occupational Psychology Annual Conference, Liverpool, January 2017. Retrieved from https://www.youtube.com/watch?v=Sa-kU5qwilE

MacIver, R., Anderson, N., Costa, A., & Evers, A. (2014). Validity of Interpretation: A user validity perspective beyond the test score. *International Journal of Selection and Assessment, 22*(2), 149–164. doi:10.1111/ijsa.12065

McPeek, R. W., Breiner, J., Murphy, E., Brock, C., Grossman, L., Loeb, M., & Tallevi, L. (2013). Student type, teacher type, and type training: CAPT Type and education research 2008–2011 project summary. *Journal of Psychological Type*, *73*(3), 21–54.

Michael, J. (2003). Using the Myers-Briggs Type Indicator as a tool for leadership development? apply with caution. *Journal of Leadership & Organizational Studies*, *10*(1), 68–81. doi:10.1177/107179190301000106

Michell, J. (2000). Normal science, pathological science, and psychometrics. *Theory and Psychology*, *10*(5), 639–667. doi:10.1177/0959354300105004

Michell, J. (2012). "The constantly recurring argument": Inferring quantity from order. *Theory and Psychology*, *22*(3), 255–271. doi:10.1177/0959354311434656

Mills, J., Robey, D., & Smith, L. (1985). Conflict-handling and personality dimensions of project-management personnel. *Psychological Reports*, *57*, 1135–1143. doi:10.2466/pr0.1985.57.3f.1135

Moshavi, D., Brown, F., & Dodd, N. (2003). Leader Self-Awareness and its relationship to subordinate attitudes and performance. *Leadership & Organization Development Journal*, *24*, 407–418. doi:10.1108/01437730310498622

Murphy Paul, A. (2004). *The cult of personality: How personality tests are leading us to miseducate our children, mismanage our companies, and misunderstand ourselves*. New York: Free Press.

Myers, I. B. (1962). *Manual: The Myers-Briggs Type Indicator*, Princeton, NJ: Educational Testing Service.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI Manual: A guide to the development and user of the Myers-Briggs Type Indicator Instrument*. Mountain View, CA: CPP.

Myers, I. B., & Myers, P. B. (1995). *Gifts differing: Understanding personality type*. Mountain View, CA: Davies-Black Publishing.

OPP (2013). *MBTI® Step II™ Instrument European data supplement, January 2013*. Oxford: OPP. Retrieved from http://www.opp.com/download/item/3e9f331e0f0d420e976d0260b38d7242

OPP (2016). *MBTI® Step I™ Instrument European data supplement, December 2016*. Oxford: OPP. Retrieved from http://www.opp.com/download/item/082ac33389fb48d1a659066f56d972df

OPP (2017). *Terms of Business for the purchase of product; Guidelines for the ethical use of tests and questionnaires*. Retrieved from http://www.opp.com/en/About-OPP/Terms-of-Business

Passmore, J. (2012) (ed.). *Psychometrics in coaching: Using psychological and psychometric tools for development* (2nd ed). London: Kogan Page.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York, NY: Oxford University Press.

Pittenger, D. J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research*, *63*, 467–488.

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research*, *57*(3), 210–221. doi:10.1037/1065-9293.57.3.210

Pulver, C. A., & Kelly, K. R. (2008). Incremental validity of the Myers-Briggs Type Indicator in predicting academic major selection of undecided university students. *Journal of Career Assessment*, *16*(4), 441–455. doi:10.1177/1069072708318902

Quenk, N. L. (1998). *In the Grip*. Oxford, UK: OPP.
Quenk, N. L. (2002). *What that really me?*. Palo Alto, CA: Davies Black: CPP.
Quenk, N. L., Hammer, A. L., & Majors, M. S. (2004). *MBTI Step II Manual: Exploring the next level of type (European edition)*. Mountain View, CA: CPP.
Quintero, A. J., Segal, L. S., King, T. S., & Black, K. P. (2009). The personal interview: Assessing the potential for personality similarity to bias the selection of orthopaedic residents. *Academic Medicine*, *84*(10), 1364–1372. doi:10.1097/ACM.0b013e3181b6a9af
Renner, W., Bendele, J. M., Alexandrovicz, R., & Deakin, P. (2014). Does the Myers-Briggs Type Indicator measure anything beyond the NEO Five Factor Inventory? *Journal of Psychological Type*, *74*(1), 1–10.
Robertson, I. T., & Smith, J. M. (2001). Personnel selection. *Journal of Occupational and Organisational Psychology*, *74*(4), 441–472. doi:10.1348/096317901167479
Rogers, J. (2017). *Coaching with personality type: What works*. London, UK: Open University Press.
Russell, M., & Karol, D. (1994). *16PF fifth edition administrator's guide*. Champaign, IL: Institute for Personality and Ability Testing.
Saggino, A., Cooper, C., & Kline, P. (2001). A confirmatory factor analysis of the Myers-Briggs Type Indicator. *Personality & Individual Differences*, *30*, 3–9. doi:10.1016/S0191-8869(00)00004-0
Saggino, A., & Kline, P. (1996). The location of the Myers-Briggs Type Indicator in personality factor space. *Personality & Individual Differences*, *21*, 591–597. doi:10.1016/0191-8869(96)00009-8
Salter, D. W., Forney, D. S., & Evans, N. J. (2005). Two approaches to examining the stability of Myers-Briggs Type Indicator scores. *Measurement and Evaluation in Counseling and Development*, *37*, 208–219. doi:10.1080/07481756.2005.11909761
Schmidt, F. L., Oh, I-S., & Shaffer, J. A. (2016). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings. Manuscript in progress. doi:10.13140/RG.2.2.18843.26400
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124* (2), 262–274. doi:10.1037/0033-2909.124.2.262
Schneider, B. (1987). The people make the place. *Personnel Psychology*, *40* (3), 437–453. doi:10.1111/j.1744-6570.1987.tb00609.x
Schroeder, C., Warner, R., & Malone, D. (1980). Effects of assignment in living units by personality types on environmental perceptions and student development. *Journal of College Student Personnel*, *21*(15), 443–449.
Schullery, N. M., & Schullery, S. E. (2006). Are heterogeneous or homogeneous groups more beneficial to students? *Journal of Management Education*, *30*(4), 542–556. doi:10.1177/1052562905277305
Scoular, A. (2011). *The financial times guide to business coaching*. London: Financial Times Prentice Hall.
Sedlock, J. R. (2005). An exploratory study of the validity of the MBTI team report. *Journal of Psychological Type*, *65*(1), 1–8.
Short, G. J., & Grasha, A. F. (1995). The relationship of MBTI® dimensions to perceptions of stress and coping strategies in managers. *Journal of Psychological Type*, *32*, 13–22.
Sipps, G. J., Alexander, R. A., & Freidt, L. (1985). Item analysis of the myers-briggs type indicator. *Educational & Psychological Measurement, 45*(4), 789–796. doi:10.1177/0013164485454009
Smith, J. M., & Smith, P. (2005). *Testing people at work: Competencies in psychometric testing*. Oxford: BPS Blackwell.
Stockill, R. (2014). *Measuring the impact of training and development workshops: An action orientated approach*. Paper presented at the British Psychological Society Division of Occupational Psychology Annual Conference, Brighton. Retrieved from https://www1.bps.org.uk/system/files/user-files/Division%20of%20Occupational%20Psychology%20Annual%20Conference%202014/dop2014_book_of_abstracts.pdf
Sundstrom, E., & Busby, P. L. (1997). Co-workers' perceptions of eight MBTI® leader types: Comparative analysis of managers' SYMLOG profiles. In C. Fitzgerald & L. K. Kirby (Eds.*), Developing leaders: Research and applications in psychological type and leadership development* (pp. 225–265). Boston: Davies-Black Publishing.
Sutton, A., Allinson, C., & Williams, H. (2013). Personality type and work-related outcomes: An exploratory application of the Enneagram model. *European Management Journal*, *31*, 234–249. doi:10.1016/j.emj.2012.12.004.
Tafreshi, D., Slaney, K. L., & Neufeld, S. D. (2016). Quantification in psychology: Critical analysis of an unreflective practice. *Journal of Theoretical and Philosophical Psychology*, *36* (4), 233–249. doi:10.1037/teo0000048
Tinsley, H. E. A., Tinsley, D. J., & Rushing, J. (2002). Psychological type, decision-making style, and reactions to structured career interventions. *Journal of Career Assessment*, *10*, 258–280. doi:10.1177/1069072702010002008
Thomas, A., Benne, M. R., Marr, M. J., Thomas, E. W., & Hume, R. M. (2000). The evidence remains stable: The MBTI predicts attraction and attrition in an engineering program. *Journal of Psychological Type*, *55*, 35–42.
Thompson, B., & Borello, G. M. (1989). A confirmatory factor analysis of data from the Myers-Briggs Type Indicator. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX, January 1989.
Thurstone, L. L. (1930). A neurotic inventory. *Journal of Social Psychology*, *1*, 3–20. doi:10.1080/00224545.1930.9714128
Tjan, A. K. (2012). How leaders become self-aware. *Harvard Business Review*, July 2012. Retrieved from https://hbr.org/2012/07/how-leaders-become-self-aware
Van Velsor, E. V., Taylor, S., & Leslie, J. B. (1993). An examination of the relationships among self-perception accuracy, self-awareness, gender and leader effectiveness. *Human Resources Management*, *32*(2-3), 249–263. doi:10.1002/hrm.3930320205
Wagner, J. (1983). Reliability and validity study of a Sufi personality typology: The Enneagram. *Journal of Clinical Psychology*, *39*(5), 712–717. doi:10.1002/1097-4679(198309)39:5%3c712::AID-JCLP2270390511%3e3.0.CO;2-3
Wallick, M. M., Cambre, K. M., & McClugage, S. G. (2000, August). Does the admissions committee select medical students in its own image? *Journal of the Louisiana State Medical Society, 152*(8), 393–397.
Weber, A. J., Schaubhut, N. A., & Thompson, R. (2011). *The influence of personality on social media usage*. CPP research paper, Mountain View: CPP.
Woodworth, R. S. (1917). *Personal data sheet*. Chicago: C.H. Stoetling Company.
Zibarras, L. D., & Woods, S. A. (2010). A survey of UK selection practices across different organization sizes and industry sectors. *Journal of Occupational and Organizational Psychology, 83*(2), 499–511. doi:10.1348/096317909X425203