CANADIAN PSYCHOLOGICAL ASSOCIATION
SOCIÉTÉ CANADIENNE DE PSYCHOLOGIE

Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement

ISSN: 0008-400X

2021, Vol. 53, No. 2, 175–186
https://doi.org/10.1037/cbs0000230

SPECIAL ISSUE

Using Personality Feedback for Work-Related Development and Performance Improvement: A Rapid Evidence Assessment

R. Blake Jelley
Faculty of Business, University of Prince Edward Island

The personality-related literature in industrial and organizational (I–O) psychology and management has focused on the validity of personality scales for predicting criteria like job performance, and corresponding applications for personnel selection. In parallel, and subject to much less research attention, personality inventories have been used for developmental purposes in group workshops and individual coaching. Personality-assessment feedback represents modern means through which to follow ancient advice to "know thyself" (Emre, *The personality brokers: The strange history of Myers-Briggs and the birth of personality testing.* 2019, p. xviii). However, task-based feedback intervention theory (Kluger & DeNisi, *Psychological Bulletin*, *119*, 1996, 254–284) casts doubt on the effectiveness of self-focused feedback cues for performance improvement. The present article synthesized studies on the impact of work-related personality-feedback interventions (PFIs) on performance- or development-related criteria through a Rapid Evidence Assessment (Barends et al., *CEBMa guideline for rapid evidence assessments in management and organizations, version 1.0*, 2017). Structured queries of five databases plus reference list and forward-citation searches identified a diverse set of 12 empirical studies. Most studies used a type-based rather than a trait-based assessment tool, with the Myers-Briggs Type Indicator (MBTI) being the most popular. No study examined PFI impact on observer ratings of job performance. Though there were some indications of beneficial effects, researchers have not yet established the causal impact of PFIs on performance-related criteria, and the possibility exists of negative side-effects. Practitioners are urged to apply PFIs with caution, temper expectations, evaluate intervention effectiveness, and share results—possibly as part of researcher–practitioner partnerships.

Public Significance Statement
Each year, millions of people around the world complete personality tests as part of training or coaching programs. This review shows that the alleged benefits of personality-feedback interventions have not been well established.

Keywords: personality feedback, test, personality change, performance, self-awareness

Supplemental materials: https://doi.org/10.1037/cbs0000230.supp

Organizations use personality inventories in training programs and coaching engagements to support the development of their employees, leaders, and teams. Individuals enroll in similar self-improvement programs. Personality assessments can provide concepts and language to improve people's descriptions and understanding of human behavior while comprehensive personality-feedback processes can guide the development of action plans for performance improvement (Goodstein & Lanyon, 1999). Such interventions assume that feedback based on a personality inventory facilitates individuals' awareness, development, and management of the self and their interactions with other people (e.g., co-workers). The quality (and extent) of evidence on the performance-related effects of workplace-related personality-feedback interventions (PFIs) was the focus of the present Rapid Evidence Assessment (REA). An REA is a mid-range form of a systematic review that provides a rigorous, transparent, and efficient summary of available evidence (Barends et al., 2017). Questions and comments about the use of personality tests for development, from members of various organizations, prompted this review.

Personality and Personality-Feedback Interventions

Personality refers to "individuals' characteristic patterns of thought, emotion, and behavior together with the psychological

Conditional Acceptance (Minor Revision): November 2020.
The author has no known conflicts of interest to disclose.
Thanks to everyone who provided assistance with this project. A guest editor and three anonymous reviewers provided valuable feedback and suggestions. Christine MacLauchlan and Inter-Library Loan staff at UPEI and our partner institutions located and shared resources. Donald Moses provided a welcomed check on the search process and facilitated access to resources. Jennifer Jelley, Hayden Woodley, and Tina Saksida provided constructive comments on drafts of this article.
Correspondence concerning this article should be addressed to R. Blake Jelley, Faculty of Business, University of Prince Edward Island, 550 University Avenue, Charlottetown, PE, C1A 4P3. Canada. Email: bjelley@upei.ca

mechanisms—hidden or not—behind those patterns" (Funder, 2001, p. 198). Characteristic tendencies are "thought to drive and direct behavior" and influence people's reactions to situations (Christiansen & Tett, 2013, p. 2), including their behavior in organizations. Even though most research attention has been on decision-making applications (e.g., selection), the word personality referred originally to awareness of the self—consistent with a developmental focus (Zickar & Kostek, 2013). Millions of people complete personality inventories for developmental purposes each year (Moyle & Hackston, 2018; Zickar & Kostek, 2013) and may account for approximately 40% of an industry worth (at least) hundreds of millions of dollars annually (Lundgren et al., 2017). Emre (2019) reported a higher estimate of $2 billion annually for the segment of the industry involving the Myers-Briggs Type Indicator (MBTI).

The Personality Applications Matrix Model (PAMM) outlines various uses of personality data for developmental as well as decision-making purposes at the individual, team, and organizational levels of intervention (Church et al., 2015). The following depiction of developmental PFIs at individual and team levels is based on Church et al. (2015) and other PFI literature. A PFI as defined herein employs one or more standardized, self-report measure(s) of personality. Participant responses provide data for the intervention in which personality feedback serves diagnostic and change-energizing functions. Other assessments may be used in conjunction with a PFI (e.g., 360° feedback). A trained feedback provider or coach is often involved as the individual reflects on their tendencies and plans functional adaptations to their behavior. Self-exploration and action planning may occur in one-on-one coaching, as part of a workshop with other people, or otherwise (e.g., self-study or digital intervention). Within teams, PFI sharing and facilitation seek to increase mutual understanding and guide functional interactions among team members. A PFI may inform people's enacted behaviors, increase appreciation of individual differences, and facilitate functional communications and relationships.

Self-awareness is typically invoked as a mediator of (presumed) PFI effects. Church et al. (2015, p. 152) asserted that "an emphasis on enhancing self-awareness and development planning should always remain a critical (if not required) component of any use of personality tools" involving data collected from individuals. Despite proponents' apparent optimism, the probable impact of PFIs is uncertain. According to Kluger and DeNisi's (1996) task-related feedback-intervention theory and meta-analysis, feedback interventions do not always improve performance. Cues that direct attention to self-related processes (beyond a given task) may be less functional or even detrimental compared to cues that direct attention to lower-level task-motivation or task-learning processes (Kluger & DeNisi, 1996). Clarity is needed on the extent of beneficial versus adverse effects arising from PFIs, given their focus on the self.

The need for systematic, developmental-effect research is heightened by the controversy that surrounds some PFI tools. Personality inventories vary in conceptual foundations and psychometric quality. In particular, the MBTI has been the subject of spirited debate. Emre (2019) characterized corporate interests involving that popular, yet frequently criticized indicator as operating at the "shady crossroads of industrial psychology and self-care" (p. xvi). Prewett et al.'s (2013) psychometric analysis of the MBTI and 11 other popular instruments seems to provide a sober, well-reasoned guide. Prewett et al. included the MBTI in their lowest or "overall modest evaluations" (p. 222) category of personality inventories.

Critics have focused on the MBTI even though other instruments share similar features and limitations (Moyle & Hackston, 2018). One controversial issue is a focus on types (e.g., Briner, 2018). Some personality inventories classify respondents into mutually exclusive "type" groups, whereas "trait" inventories focus on respondents' relative standing, on a continuum, on each trait in a given inventory (Lundgren et al., 2017). Those approaches are not entirely incompatible. Sub-groups of people with distinct configurations of trait scores can be identified (e.g., Espinoza et al., 2020) and type classifications are often based on underlying dimensions. For example, the MBTI (Step I) purports 16 types based on combinations of four binary preference domains (Prewett et al., 2013). Nevertheless, some employees have reported that the resulting type groupings can be constraining (Garrety, 2007). Trait-based PFIs are likely to avoid that problem.

The Five-Factor Model (FFM) provides a widely recognized set of trait categories (i.e., conscientiousness, agreeableness, neuroticism/emotional stability, openness to experience, and extraversion), and has developmental applications (Church et al., 2015; Lundgren et al., 2017; McCormick & Burch, 2008). For example, the Hogan Personality Inventory (HPI) adopts a seven-factor variation of the FFM (Prewett et al., 2013). Some PFIs have featured the HPI and the Hogan Development Survey (HDS). The HPI is one of four inventories that Prewett et al. included in their most favorable "overall positive evaluations" category (p. 222). The HDS departs from the FFM—which does not subsume all traits—to assess maladaptive tendencies likely to manifest under certain conditions (e.g., stressful situations). Those dark-side traits may be malleable (Church et al., 2016; Nelson & Hogan, 2009).

PFIs are often depicted as catalysts for functional behavioral changes rather than attempts to change personality per se (McCormick & Burch, 2008). However, the "relatively enduring" nature of personality does not mean that personality traits are necessarily "unchanging or unchangeable" (Nye & Roberts, 2013, p. 796). In addition to mean-level trait changes across the life course, personality can change relatively quickly through intervention (Roberts et al., 2017). The personality-change perspective could be a future direction for organizational behavior (Tasselli et al., 2018) that may or may not (e.g., Stieger et al., 2020) involve personality feedback.

Inaccurate feedback based on a psychometrically inferior tool could be misleading, unhelpful, and even harmful (Briner, 2018). However, on their own, psychometric critiques fail to document the adverse effects that critics fear may arise from (some) PFIs. Psychometric critiques neither answer developmental-effect research questions nor demonstrate that psychometrically revered inventories better facilitate development than do lower-quality tools. Consequently, this REA was not focused on the psychometric critique of personality inventories. Researchers must subject to direct empirical scrutiny the alleged positive, "astonishing" effects that factions of "true believers" have offered in support of a given PFI tool (e.g., Emre, 2019, p. 268).

The outcomes purported to be developed through a PFI are not always made clear. The nature of effects may depend on nuances of the client, intervention, or personality inventory. Generally, a development program reflects a planned, programmatic approach to growth, learning, functional change, and improvement of

competence or performance that, in the context of employee development, is usually linked to organizational goals (Hullinger & DiGirolamo, 2018; Moyle & Hackston, 2018). Consequently, the present REA considered studies that examined a variety of performance or development criteria. The purpose of this REA was to synthesize existing findings on developmental and performance-related effects stemming from participants' receipt of personality feedback (i.e., PFIs) to inform future practice and research.

Research Questions

The central research question was: What is known (in the formal-research literature) about the effects of PFIs on work-related development or performance? Supplementary questions included: Are effects of PFIs contingent upon factors such as the particular instrument used, its theoretical framework, psychometric quality, trait-based versus type-based approach, use of bright- versus dark-side traits, item content (e.g., invasiveness), method of feedback delivery (e.g., oral, written, or electronic), or facilitation (e.g., coaching process, coach qualifications, training features)? How does personality feedback exert its presumed effects?

Method: Search Strategy and Results

I conducted a mid-range systematic literature search (Barends et al., 2017) to identify REA-relevant studies. Preliminary searches helped to refine inclusion criteria and search terms.

Inclusion Criteria

The present REA focused on investigations of workplace-related programs or products that provide feedback to respondents of a standardized, self-report personality test (i.e., a PFI). Included studies had to have investigated a developmental PFI rather than personality measures as predictors or moderators of criteria (e.g., job performance). Studies of bogus (false) personality feedback and studies of real personality feedback in non-developmental (e.g., lab) settings were not eligible. Student samples were eligible if the students received a PFI.

Personality feedback in therapeutic, clinical, or non-work life-coaching settings was outside the current scope (see Hanson & Poston, 2011). Coaching-related PFIs were included as per the distinction between coaching and therapy offered in the International Coach Federation's guidelines for therapeutic referral (Hullinger & DiGirolamo, 2018). Coaching involves future-oriented, thought-provoking partnerships between a coach and a well-functioning client to maximize the latter's potential, learning, development, or performance.

I prioritized studies that investigated the impact of PFIs on measures of participants' subsequent behavior, performance (e.g., in-role, contextual, dysfunctional), or results (e.g., business impact). Studies of PFIs' effects on indicators of learning or development were eligible, including studies that used measures of personality as dependent variables. Studies of participant reactions to PFIs were also tracked, but not included formally in the REA.

Empirical studies in English or French published from 1990 to early 2020 were eligible for this REA, reflecting a period of resurgence for personality in organizations (e.g., Zickar & Kostek, 2013) and the researcher's language limitations. There were no geographic or industry restrictions. Studies at the individual, team, or organizational levels of analysis were eligible.

Search Strategy and Study Selection

This REA started with structured searches of the PsycINFO, Business Source Complete (BSC), Canadian Business and Current Affairs (CBCA), Education Resources Information Center (ERIC), and Academic Search Complete (ASC) databases. I identified potentially relevant empirical studies for full-text consideration and possible use in subsequent reference-list and forward-citation searches, as well as conceptual works closely related to PFIs.

Phase-one queries used only the exact term "personality feedback" within each database. No REA-included studies were identified. Previous researchers have used that term differently than how it is used herein (e.g., studies of manipulated personality feedback on recipients' acceptance of feedback or another phenomenon vs. developmental PFIs).

A more nuanced examination of the databases was developed and executed as a planned, second-phase set of structured searches. Unlike an exhaustive systematic review, phase-two database searches were confined to peer-reviewed sources. Making concessions about the scope of a review is consistent with Barends et al.'s (2017) guidelines because REAs are designed to be rigorous, yet efficient mid-range reviews. As detailed in the online supplement, phase-two queries involved identifying records that, in their titles or abstracts, included "personality" and synonyms of two concepts (i.e., test and work-related development), along with other limiters. "Feedback" was included with an extensive set of development-related terms since preliminary searches showed that using feedback as a required term would eliminate some promising records. Broader development-related terms (e.g., train* OR develop* OR coach*), beyond those specified in the online supplement, proved unwieldy and resulted in many irrelevant records. Searches did not include any limiters related to dependent variables or research designs given the apparent paucity of research in the area and desire to uncover relevant conceptual as well as empirical sources for consideration. Titles and abstracts of identified records were screened for relevance to PFI evaluation, and 81 unique records were selected for full-text review based on database queries.

Sources were mostly peer-reviewed journal articles. Theses, dissertations, books, white papers, and test manuals were beyond the scope of this REA, with two exceptions—chapters that critically reviewed research in related areas (Batey et al., 2012; Walck, 1997). Phase-three searches involved considering in-text citations and reference sections of 24 sources (identified in the online supplement). Phase four consisted of forward-citation searches of those same sources via Google Scholar (i.e., subsequent studies that cited those sources). Six REA-included studies were identified via phase two searches and two additional studies were identified in each of phases three and four—phases that exceeded REA guidelines (Barends et al., 2017). Finally, two included sources were identified by other means (e.g., studies I happened to recall). The online supplement details the search process. Overall, I reviewed the full text of more than 130 sources.

Data Extraction and Critical Appraisal of Studies

REA techniques can be applied to various research questions, including questions about the impact of an intervention on criteria of interest (Barends et al., 2017). Barends et al. adapted from other sources a six-level (E to AA) evidence-classification system for impact-related REAs based on research designs' relative abilities to satisfy three conditions for cause-effect inference (i.e., co-variation, temporal precedence, and isolation from alternative explanations). Case studies (E, lowest), cross-sectional studies (D), meta-analyses of cross-sectional studies (C), non-randomized controlled before-after studies (B), randomized control trials (A), and meta-analyses of randomized control trials (AA, highest) are examples of research designs at various levels of methodological appropriateness for cause-effect inferences. The quality of research-procedure implementation was also considered when estimating evidence levels, with category downgrades available in cases of weaknesses that undermined a study's credibility (Barends et al., 2017). Not all included studies were designed specifically to evaluate PFIs, so a level-of-evidence estimate should not necessarily be viewed as an overall evaluation of a given source. Table 1 provides a summary of REA-included studies.

Given the heterogeneous nature of research designs, dependent variables, and reporting practices, quantitative effect sizes are not included in Table 1. Comments about effect sizes are noted based on information in the included studies and effect size interpretation guidelines (i.e., Cohen's rules of thumb, as summarized in Barends et al., 2017, p. 20). Large effects are apparent to anyone, medium effects are apparent to alert observers, and small effects can only be detected via careful examination (see Barends et al., 2017 for a summary of quantitative indices).

Included-study limitations (and strengths) noted in this review are not exhaustive. Common limitations included: Sample sizes tended to be small and based on non-probability sampling; randomized experiments were rare and subject to other limitations; isolating PFI effects from other intervention components (e.g., training, coaching, or assessment programs) was a challenge. Reporting problems concerning data collection procedures and results (e.g., lack of basic descriptive statistics) were surprisingly common, but not universal.

Findings

Nine of the 12 REA-included studies used type-based PFI tools. Sutton et al. (2015) used an Enneagram model (nine types), but most (six) studies employed the MBTI or conceptually similar instruments (i.e., Rekar Munro & Laiken, 2003; Stefansdottir & Sutherland, 2005). Studies evaluating effects of trait-based PFIs were uncommon and not necessarily designed with PFI evaluation as the primary focus (i.e., Church et al., 2016; Ellingson et al., 2007). Two studies used HPI and HDS trait scales (Church et al., 2016; Mansi, 2007).

The 12 REA-included studies were published between 2003 and 2018. Neither the volume nor the methodological quality of PFI research seemed to improve during that time. No included study used performance ratings by observers as dependent variables. Sections below highlight the main findings from type-based and trait-based PFIs (see also Table 1). Three studies with B-level ratings (Church et al., 2016; Ellingson et al., 2007; McPeek et al., 2013) were relatively strong among included studies but had limitations with respect to cause-effect inferences. Effect sizes in B-level studies tended to be small, with the exception of a large effect on the set of HDS (dark-side-trait) dependent variables in Church et al. (2016).

Four studies were rated at the C-level—two studies with possible medium effects (Rekar Munro & Laiken, 2003; Varvel et al., 2004) and two studies with small effects or a combination of favorable and unfavorable effects (Stefansdottir & Sutherland, 2005; Sutton et al., 2015). Two studies (Sedlock, 2005; Waite & McKinney, 2018) provided D-level evidence and little information about the nature of (possible) changes resulting from PFIs. Three studies (Mansi, 2007; Segovia, 2016; Waite & McKinney, 2015) received E-level ratings. Optimistic estimates of PFI effects (i.e., Mansi, 2007; Segovia, 2016) were found among studies rated E.

Dedicated sections for mediators and moderators, recommended by Barends et al. (2017), are not included herein given the limited available evidence. Barends et al. also recommended including a section for definitions and conceptual clarifications arising from an REA. In the present review, that section is focused on distinguishing different forms of self-awareness.

Clarification of the Self-Awareness Construct Domain

In PFI-related literature, authors often present self-awareness in a generic, positive light that ignores nuances identified by self-awareness scholars (Sutton et al., 2015). Self-awareness is "a higher-level concept which includes the extent to which people are consciously aware of their interactions or relationships with others and of their internal states" (Sutton et al., 2015, p. 611). The authors addressed the potential "sadder-but-wiser" (p. 611) self-awareness paradox by differentiating styles of self-awareness (e.g., reflective with positive outcomes vs. ruminative with adverse outcomes). Sutton et al. prioritized dispositional self-awareness (self-consciousness or self-attentiveness) and alluded to different measurement approaches (i.e., self-reports, used in their study, vs. comparing self-ratings to others' ratings).

Type-Based Personality Feedback Interventions

Sutton et al. (2015) compared Enneagram personality-based self-awareness training (a PFI) and generic self-awareness training on measures of dispositional self-attentiveness (reflection and rumination) and job-related well-being ($N = 79$). Results showed that self-awareness improvement is not necessarily straightforward. For example, "job satisfaction was significantly lower in the longer term [partial $\eta^2 = .42$] and participants in the generic training had significantly lower rumination scores [partial $\eta^2 = .11$] than those in the Enneagram training" (Sutton et al., 2015, p. 617). PFI participants reported less rumination in their qualitative feedback, whereas generic participants' comments noted greater reflection. Feedback about the training programs was variable, but some respondents reported positive effects and described behavioral changes (Sutton et al., 2015). Confusing reporting of some results complicates study interpretation, but the idea of comparing a PFI to generic, conceptually similar self-awareness training is laudable.

Stefansdottir and Sutherland (2005) investigated an Insights Discovery System program (similar to the MBTI) as part of stress management training for college instructors and staff members ($N = 18$). The authors did not report effect sizes, but descriptive

Table 1
Data Extraction From REA-Included Studies

Author(s) & Year	Sector/Sample	Study Summary	Effect Size(s)/ Contributions	Limitations	Level
Church et al. (2016)	$N = 207$ senior leaders in a global consumer products firm; archival study	Design: [quasi-experiment; pre-post personality measures (HPI, HDS, MVPI)]. T1 assessments (personality; 360° feedback) were for development. T2 assessments were for development ($n = 79$) or decision-making ($n = 128$) purposes. A study of assessment stability that included a PFI (90-min feedback session with a trained feedback provider). Some dark-side traits (HDS scales) improved over time. No significant changes in core traits (HPI) scales. Authors reported that MVPI results were difficult to interpret.	Reported η^2 ESs (HPI small-medium, HDS large, MVPI large). Reported means & SDs. Some evidence of expected stability versus change for different traits. PFI may improve maladaptive tendencies.	DVs were self-report personality scales. PFI effects not isolated versus effects from other factors (e.g., 360° feedback). Univariate ESs were not reported (may be smaller?). Authors noted issues that may underestimate PFI effects.	B
Ellingson et al. (2007)	$N = 713$ from multiple organizations and jobs (mostly managerial, professional, or technical); USA archival study	Design: [quasi-experiment; pre-post CPI scales]. Study of intentional response distortion by test-taking purpose. Selection or development purpose at T1 and T2 (4 groups). Development purpose included a PFI (feedback from an assessor) as part of a battery of assessments, including 360° feedback. Studied effects of time, feedback, and response distortion. CPI feedback increased T2 scores (T1 development-T2 development, $n = 77$).	Reported d ESs. Very small feedback effect (isolated from time effect—i.e., test, retake, and maturation effects).	DVs were self-report personality scales. PFI effects not isolated versus effects from other factors (e.g., 360° feedback). Are all CPI scales expected to change (at all or in the same direction)?	B
Mansi (2007)	One valuable, interpersonally problematic senior manager referred by the company CEO	Design: [individual case study]. HPI and HDS were used with person-centered and cognitive behavioral counseling/coaching to mitigate maladaptive tendencies. Feedback was "very positive" (e.g., less irritable & hostile; more thoughtful). Author reported that personality-feedback reports were suited to the client who valued evidence-based information.	No ES but could be medium (change noticed). Author described use of PFI (process details, including benefits and challenges).	Behavioral-effects feedback was indirect. Low clarity of data procedures. Value added by PFI versus counseling alone? Individual case with the coach as researcher.	E
McPeek et al. (2013)	K-12 teachers (& students) in the USA. Type-trained teachers; $N = 35$ from four schools of 10 in a large-scale research program	Design: [multi-site study; experiment/quasi-experiments, pre-post or between-group comparisons]. Large-scale study of MBTI that included an element on type training (e.g., MBTI feedback and interpretation) for volunteer teachers (10-hr PFI) who received $200 or $400 via a Myers and Briggs Foundation grant. Teachers were trained near the beginning of a school year. DVs = students' peer-relative z-scores for grades (small, positive effects) and proficiency tests (no consistent effects).	Various d ESs reported but not overall d. No or small positive effects. A complex, ambitious study. Authors sought meaningful comparisons and noted study limitations	Small sample despite the scale of the project. No overall ES reported. Impact the missing ESs might have on overall ES (lower)? However, are students' results sensitive enough to detect PFI effects, as the authors noted?	B

(*table continues*)

Table 1 (*continued*)

Author(s) & Year	Sector/Sample	Study Summary	Effect Size(s)/ Contributions	Limitations	Level
Rekar Munro and Laiken (2003)	Participants in teams (4–5) from various sectors and professions. $N = 118$ from Ontario, Canada randomly assigned to teams; focus groups ($n = 31$)	Design: [experiment; exp vs. control—performance and satisfaction (post-tests) across four weekly exercises; end-of-study focus groups.] 14 experimental teams were trained in a team development model using personality styles (Keirsey Temperament Sorter ˜ MBTI-related) and other components (e.g., action planning, process checks). Four weekly decision-making exercises were completed individually and then in teams. Satisfaction surveys collected weekly. Favorable satisfaction, performance, and feedback reported for experimental teams.	No ESs reported; positive, maybe medium ES (?). PFI and type theory central to team training. Experimental teams did better (vs. controls) in areas related to PFI (e.g., valuing differences, managing conflict)	Abbreviated report (incomplete reporting, e.g., of quantitative scoring and results). Not able to isolate PFI effects versus other training components (e.g., action planning, action research, process checks).	C
Sedlock (2005)	Members of intact work teams from diverse organizations in the USA ($N = 51$, including nine team leaders)	Design: [descriptive, exploratory survey about the MBTI Team Report]. The test publisher provided information about the study to consultants who, in turn, could decide whether to invite client-team members to participate. The 19-item questionnaire included two items about action-plan implementation and results ($N = 47$), in addition to reactions. Median responses for implementation and results items were 2 on a 0–4 scale.	No ES. Self-reported actions based on the report did not seem high (at scale midpoint vs. mostly 3s for other reactions). Diverse sample. Included action-oriented items.	Convenience sample from publisher-consultant referrals (biased?). Self-reports versus independent assessments of actions. Use of medians (underestimate effects?).	D
Segovia (2016)	Managers in one high-technology company. $N > 300$ trained over 2 years. Interview and survey. Ns not reported.	Design: [company-level case study; interviews, surveys?]. One sentence in the Abstract described the method (few details). The company's Empowerment Camp involved 60-hr of training over 12 weeks plus weekly webinars. It featured the MBTI and associated workbooks and included course-long company-related projects and other components. Variables and data collection methods were not specified. Included an example of how to use type differences to understand others and manage meeting behavior.	No (defensible) ESs reported. 98% agreed training made them more effective in their jobs. Claimed potential savings of more than $20 million (?).	Inadequate reporting of data collection and results. No isolation of PFI effects (vs. other components). Post-intervention account of (alleged) benefits. A wait-listed control quasi-experiment or other research design could have given more clarity.	E
Stefansdottir and Sutherland (2005)	Instructors and staff volunteers from two colleges in Iceland ($N = 18$ of 30 participants)	Design: [pre-post quasi-experiment]. A study of stress management training that included an Icelandic translation of the Insights Discovery System-Evaluator (Jungian types ˜ MBTI). "Preference awareness education" (a PFI) included two, 2-hr workshops 7-weeks apart, and four homework tasks. Participants completed the Insights Evaluator prior to and received a personal report following the first workshop.	ESs were not reported but seemed small and not always favorable. The PFI seemed thorough. PFI might work for certain profile types or pre-existing stress (high) or job satisfaction (low) levels, as per the authors. More work needed.	Small sample size. Exploratory moderator analyses (60 = statistical fishing) provide only possibilities to investigate further. The homework tasks about how participants had or could better use knowledge could have provided qualitative examples of uses of PFI (not reported).	C

Table 1 (*continued*)

Author(s) & Year	Sector/Sample	Study Summary	Effect Size(s)/ Contributions	Limitations	Level
		Pre- (2–4 days before the first workshop) and post- (2–4 days after the second workshop) stress and job satisfaction levels were measured via the Occupational Stress Indicator. There were no total-sample pre-post differences.			
Sutton et al. (2015)	A convenience (snowball) sample of full-time employees, mostly British (N = 79 of 88 self-awareness workshop participants)	Design: [between-groups and repeated measures, mixed-methods quasi-experiment; post-study responses]. Studied self-awareness change through Enneagram (PFI) versus generic self-awareness workshop conditions. One PFI and the generic workshop condition involved one pre-test and two post-tests. Different PFI conditions examined pre-testing effects. DVs = self-attentiveness (reflection & rumination) and job-related enthusiasm, contentment, and satisfaction. Behavioral actions and improvements in self-awareness were noted, but some inconsistent results.	"Small- to medium-sized effects of training on both self-awareness and job-related well-being" (p. 622). (But effects were not always favorable.) Qualitative data and coding. Focused on PFI evaluation.	Self-reported DVs. Small sample size (for the number of conditions). Some reporting problems (e.g., assignment of participants to conditions, no descriptive statistics by condition, quantitative results confusion, missing ESs, need to verify ES interpretation).	C
Varvel et al. (2004)	Senior engineering design students (various fields). USA (N~60 or 188?)	Design: [quasi-experiment (?), post-test; MBTI-training]. A study of team effectiveness and the MBTI included a PFI-related comparison (not the focus). N = 188 students in self-managed project teams (one or two semesters). N for PFI-related analyses might be approximately 60 based on ANOVA results (team or individual level?). Participants completed the MBTI within the first 2 weeks and the Team Effectiveness Questionnaire (TEQ) within the last 3 weeks of the (first) semester. Approximately half of the students received MBTI training (1-hr) from a certified facilitator within the first month. The other half received that training at the same time they completed the TEQ. Early versus late MBTI training (a PFI) affected (positively?) TEQ scores (performance, communication, interdependence, psychological safety, and attitude).	No ESs reported. I estimate, on average, a medium effect on TEQ scales (based on their Table 4 and other estimates, where required). The study focused on other research questions but provided (some) PFI evaluation data by manipulating when MBTI training was given.	Self-reported DVs. Reporting problems (Ns, number of teams, level of analysis, effect sizes, means, SDs). Grade data were not analyzed in terms of PFI effects. Some confusion about TEQ constructs mentioned but not analyzed (i.e., exclusion of common purpose, goal clarity, role clarity?).	C
Waite and McKinney (2015)	N = 14 female nursing students at a private university in Northeastern USA	Design: [pre-post MBTI in a leadership development program]. The program involved six one-credit courses over 5 months. It was based on Kouzes and Posner's leadership framework, and also included chances to learn about and work with personality styles. Authors	One or two students moved from one binary preference to its opposite in each of four MBTI domains. 9/16 MBTI type profiles changed (± 7%–8%).	Data (MBTI stability-change) versus discussion (behavioral & relational implications). Data not reported for seven types (no changes?). Small convenience sample.	E

(*table continues*)

Table 1 (*continued*)

Author(s) & Year	Sector/Sample	Study Summary	Effect Size(s)/ Contributions	Limitations	Level
		alluded to students' reflections about increased self-awareness, recognition of differences, more demanding to enact non-preferred styles.			
Waite and McKinney (2018)	*N* = 18 nursing and other health profession students (USA)	Design: [pre-post MBTI in a leadership development program]. The program consisted of three courses over 9 months (fall, winter, and spring) and featured personality types. Pre-post MBTI data and faculty members' documentation of students' perceptions provided data. Discussion themes: increased self-awareness and need for adaptability, differences as sources of (potential) conflict, proactive steps for team-optimization.	Stability versus (little) change of MBTI profiles. Included qualitative themes and more information about identifying themes than in the authors' 2015 study.	Lack of qualitative (or quantitative) data (e.g., illustrative quotes) RE: themes and details of the approach/method. Seems to provide preliminary ideas RE: development of leadership skills and team building.	D

Note. Not all studies were designed as PFI evaluations and may offer higher-level evidence for other research questions. T1 = time-one; T2 = time-two; 360 = multisource assessment; ANOVA = analysis of variance; CPI = California Psychological Inventory; DV = dependent variable; ES = effect size; HDS = Hogan Development Survey; HPI = Hogan Personality Inventory; MBTI = Myers-Briggs Type Indicator; MVPI = Motives, Values, Preferences Inventory; PFI = personality-feedback intervention; *SD* = standard deviation; TEQ = Team Effectiveness Questionnaire. Levels of evidence range from AA (highest) to E (lowest).

statistics suggest the effects on stress and satisfaction scales were generally small and not always favorable.

Sedlock (2005) surveyed team members ($N = 47$) from diverse organizations about the MBTI Team Report by using test publisher and MBTI-consultant referrals. Among the reaction measures Sedlock assessed were two that seem relevant to this REA. Median responses for self-reported action–plan–implementation and results items were two (Some) on a 0–4 scale.

Rekar Munro and Laiken (2003) randomly assigned participants ($N = 118$) to teams. The experimental group was trained in a team development model that focused on personality style using the Keirsey Temperament Sorter (similar to the MBTI). The authors reported that quantitative and qualitative (focus group, $n = 31$) data supported the value of the training, but details about performance scoring and results reporting were lacking. It appeared that effect sizes were medium, possibly larger. However, isolating a PFI-specific effect from the influence of other components (e.g., action planning and process checks) was not possible.

Varvel et al. (2004) had senior engineering design students ($N = 188$) in self-managed project teams complete brief (1-hr) MBTI training either in the first month or near the end of their respective courses. The authors did not report PFI effects on available grades (performance data), but MBTI training seemed to improve team effectiveness ratings (performance, communication, interdependence, psychological safety, and attitude; no significant effect on conflict). Varvel et al. (2004) did not report effect sizes, the direction of effects (presumably positive), or descriptive statistics. Also, the level of analysis was not always clear. The average effect size appeared to be medium (lowest for conflict; highest for attitude) based on information presented in Table 4.

McPeek et al. (2013) described a ten-site research program on psychological-type that included (MBTI) training (10-hr) for K-12 teachers. The PFI-related sample size was reduced (to $N = 35$ teachers) after researchers identified four protocol-following schools and meaningful comparison data. A site that used random assignment was not interpretable because student caliber differed between MBTI-trained teachers (worse students) versus untrained teachers (better students). Other sites used various quasi-experimental comparisons. McPeek et al. (2013, p. 45) reported "significant, albeit small, positive effects upon student grades" and "no consistent post-training effect upon standardized test scores." Despite its limitations (e.g., not reporting all effect sizes), the McPeek et al. study represents a credible set of attempts to evaluate PFIs.

Segovia's (2016) research paper designed to "gauge the success" (p. 29) of intensive MBTI training in one firm's Empowerment Camp provided optimistic estimates (e.g., $20 million potential savings) but few details of how the post-intervention, self-reported accounts of apparent benefits were derived. Also, reported benefits were likely affected by other training components (e.g., work-related projects). PFI effects were not isolated clearly.

In four REA-included studies, personality inventories were administered before and after participants received a PFI. Waite and McKinney (2015, 2018) reported studies of mostly nursing students ($N = 14$ & 18, respectively) in leadership development programs that employed the MBTI. One or two students in each sample changed dichotomous preferences in most areas with similar, slight changes in some MBTI-type profiles. Unfortunately, the pre–post MBTI data did not align with the authors' main discussion points (e.g., interpersonal communications, behavioral adjustments, conflict management). Waite and McKinney (2018) took a step

toward clearer reporting of qualitative methods by explaining that faculty members had documented, over time, the discussions from which the authors identified themes.

Trait-Based Personality Feedback Interventions

Two pre-post personality studies used trait data from different archives. Ellingson et al. (2007) studied response distortion on the California Psychological Inventory (CPI). Respondents ($N = 713$) completed the CPI at two times for the same or different purposes that differed in motivation to distort responses (selection = high vs. development = low). Development participants completed a battery of assessments (i.e., CPI, cognitive tests, and assessment-center exercises) and received an assessor's feedback. Ellingson et al. statistically controlled differences in test–retake time lapses and isolated various effects, including that of feedback (average $d = .07$). The authors noted that " . . . the experience of receiving feedback on the CPI at an earlier point in time also contributed to a small increase in scores" (p. 392). However, the effect was very small and confounded with other assessments. It is also unclear as to whether increased scores are always better on CPI scales. Examining traits that individuals target for change (vs. all scales on an inventory) might provide more sensitive criteria (Martin et al., 2014).

Church et al. (2016) studied the stability of different personality assessments (e.g., HPI & HDS) over time and as a function of different test–retake purposes (development vs. decision-making). Senior leaders ($N = 207$) had participated in personality and 360° feedback assessments initially for development only. Personality feedback involved one-on-one 90-min sessions with a trained feedback provider. Some participants re-took assessments early for a decision-making purpose ($n = 128$), whereas others' re-takes were after 5 years (development-only, $n = 79$). Statistical checks reduced concerns about the between-group time-lapse difference. Time-related η^2 effect sizes for three sets of dependent variables were small-to-medium (.05) on the HPI, and large on the HDS (.15) and Motives, Values, Preferences Inventory (MVPI, .17). Core personality (HPI) scores did not change significantly. MVPI scales are less central to personality and had a confusing pattern of results (Church et al., 2016). Changes in dark-side traits (HDS scales) were observed as a function of time (but not purpose). Declines in characteristics "referring to cynical, arrogant, micromanaging, and sycophantic behaviors" suggest "that personality self-awareness and subsequent development planning do help leaders modify derailing tendencies" (Church et al., 2016, p. 469). PFI impact is clouded by possible non-feedback time effects (*cf.* Ellingson et al., 2007) and developmental uses of 360° feedback. Behavioral measures such as 360 ratings would also be useful for evaluating self-awareness interventions (Church et al., 2016).

Mansi (2007) provided behavioral evidence via a case study. The client was a valuable but interpersonally problematic senior manager referred by the company's CEO. The HPI and HDS were used with counseling techniques. The evidence-level estimate is low given that this study involved an individual case, combined personality feedback, and other techniques, did not collect behavioral data directly from observers, and combined the roles of coach and researcher. Nevertheless, Mansi's process explanations seem useful (see also Nelson & Hogan, 2009) and feedback from organization members, given to the client or the CEO, was "very positive" (p. 57; e.g., client was less angry, irritable, and hostile; more thoughtful and clear with expectations).

Synthesis and Discussion

Batey et al. (2012, p. 51) considered the developmental use of a broad range of psychometric tools (e.g., cognitive ability, personality) and concluded that "there is little evidence on how psychometrics can support individual development." Neither Batey et al. nor an earlier review of MBTI research (Walck, 1997) identified PFI-evaluation studies. The present REA provides a comprehensive and transparent review of the published evidence on PFI's performance-related effects to date. No REA-included study examined PFI's impact on observers' (e.g., supervisors') performance ratings. Included studies used heterogeneous criteria, and research designs did not clearly establish cause–effect relations. Effect sizes were not always reported and varied in terms of size and direction. The inconclusive evidence to date does not mean that PFIs do not (or cannot) work but is troubling given the decades-long popularity of these interventions and their potential for negative side-effects. Possible adverse effects include "pigeonholing" or confining typologies (see Garrety, 2007; Lundgren et al., 2017; Lundgren, Kroon et al., 2019) and other detrimental effects of self-focused feedback (Kluger & DeNisi, 1996) such as increases in negative rumination (Sutton et al., 2015) which may include a sense "there is something wrong with me" based on a PFI (Lundgren, Kroon et al., 2019, p. 366).

Half of the MBTI-included studies (i.e., McPeek et al., 2013; Sedlock, 2005; and Segovia, 2016) had connections to MBTI-interested institutions. There may be concerns about potential conflicts of interest and, in two studies, dissemination of results via a journal closely linked to that perspective and line of products (i.e., *Journal of Psychological Type*). On the other hand, the small investments made in MBTI development-intervention research seem disproportionate to the size of the MBTI industry. Going forward, further investments in independent research are needed. Psychometric and substantive critique of the MBTI was not the current focus (e.g., Prewett et al., 2013), but the present REA extends concerns about the MBTI to the limited and weak existing evidence that it facilitates development, behavior change, or performance improvement. That criticism applies to other PFI tools, as well.

Directions for Future Research

Psychologists and management scholars have rarely disseminated PFI evaluation research, so there are many options for competent contributions. Studies that help isolate the effects of PFIs (e.g., via experiments or quasi-experiments) are needed. The PAMM urges researchers and practitioners to extend work on personality in organizations beyond the individual level (Church et al., 2015). This REA has revealed the need for developmental research at all levels. The Canadian I–O psychology community is well-positioned to contribute to PFI research and practice and, I hope, will join with scholars and practitioners around the world to invigorate PFI research. Whether one identifies with personality or another area (e.g., assessment, communication, conflict, leadership, organizational change, performance management, teams), there are opportunities for practical and theoretical contributions. For example, research on moderators could

reveal the features that work best as PFIs, for whom they work, and under what conditions. Determining the individual and joint effects of PFIs and other developmental techniques (e.g., coaching, 360° feedback, leadership training) would clarify the value added by individual components (Michael, 2003). That information could have practical implications in terms of intervention design and resource allocation. More nuanced examinations of PFI-effect mediation via different styles of self-awareness (Sutton et al., 2015) might better integrate with and possibly refine feedback intervention theory (Kluger & DeNisi, 1996). Such work should be of interest to reputable outlets in psychology and management that have not yet featured investigations of this popular practice.

PFI research can be challenging to do. Some difficulties stem from the nature of PFIs that, by design, involve customized, diagnostic feedback. Participants are likely to focus on different goals and action plans. As one of this article's reviewers observed, " . . . in a single study, you might have an introverted worker who wishes to be better at social networking, a disagreeable manager who wishes to [be] better at retaining [their] temper, or a radical thinker who is trying to bring [themself] back 'down to earth.'" That diversity requires careful consideration of relevant criteria. Recent studies of targeted personality change (Martin et al., 2014; Stieger et al., 2020) may be instructive, whether researchers are interested in personality criteria or other effects (e.g., behavioral measures). Nevertheless, both basic and applied researchers should consider PFIs' purported effects and choose dependent variables accordingly.

Even finding research sites, ensuring compliance with research protocols, and recruiting participants are difficult tasks for PFI researchers (Lundgren, Kroon, et al., 2019; McPeek et al., 2013; Sutton et al., 2015). Partnerships involving parties from university, organization, publishing firm, or consultancy settings could promote sharing of data access and expertise. Expect some conflicting interests and tensions among such parties (Lundgren et al., 2017), but constructive collaborations are possible. Academics and practitioners could offer insights on measurement, (quasi-) experimental manipulations, and research designs that balance rigor with data collection practicalities. For example, participants in a PFI group could be compared, on relevant dependent variables, to a wait-listed control group in a switching replications design. That is one evaluation option when withholding a treatment is not acceptable, yet logistics prevent all participants from getting a PFI at the same time.

We must remember that the public learns about personality through PFIs and popular treatments of the topic (e.g., Emre, 2019). We need to extend our knowledge base and share the best of what we know.

Limitations

The major limitation of the present REA was having one person search for, identify, extract data from, and critically appraise the REA-included studies. Multiple coders are expected in a systematic review and are recommended for some REA steps (Barends et al., 2017). Judgment calls were required—other coders may have made different inclusion decisions and evidence-level estimates. In particular, I excluded Krieshok et al. (2000) who used a work-related PFI with a clinical (substance abuse) sample and Martin et al. (2014) because their personality-change PFI with a general population did not seem to be work related.

This mid-range review serves as a snapshot of PFI studies that have been refined and adjudicated through a peer-review process. This REA considered published articles and was not an exhaustive systematic review of all available evidence. Future searches could include books, test manuals, and unpublished sources (e.g., theses; gray literature). Hopefully, this article will also encourage practitioners to help refine our understanding of PFIs and their effects by (publicly) sharing previously unpublished investigations in this area (e.g., white papers or internal reports), in addition to conducting or facilitating new primary research. The PFI-practitioner community likely has much to offer.

I may have missed PFI-behavior implications noted in predominantly reaction-focused studies or illustrative (anecdotal) cases. Likewise, synthesizing PFI-reactions research was beyond the present scope. However, participant reactions to instruments, feedback reports, and facilitation (e.g., training programs, coaching engagements) are important considerations (e.g., Garrety, 2007; Harland, 2003; Lundgren, Kroon et al., 2019; Moyle & Hackston, 2018).

Future reviews could expand the set of databases and the number of sources used for reference-list and forward-citation searches. However, conducting a full-fledged systematic review at this point seems unlikely to change the overall conclusion that there is not yet compelling evidence that PFIs improve workplace behavior, performance, or results. Primary studies of PFI effects are needed and practical outcomes represent a fruitful avenue for research (see also Church et al., 2016; Sutton et al., 2015).

Implications for Practice

Lundgren, Poell et al. (2019) used qualitative methods to explore human resource development professionals' ($N = 18$) approaches to personality test use. The authors offered six preliminary categories (i.e., ethical-protective, scientific-selective, cautious-avoiding, cautious-embracing, user friendly pragmatic, and knowledgeable-accommodating). Familiarity with different approaches, as well as associated insights, tips, and cautions (Lundgren, Poell et al., 2019), may be useful to practitioners navigating PFIs, either for light-use or in-depth development (Lundgren et al., 2017). Similarly, Lundgren et al.'s (2017) study of dynamics and tensions in the testing industry have implications for understanding PFIs in practice that likely extend beyond their Western European context. Professional testing and ethical standards are general considerations (Lundgren et al., 2017). Guidelines for developmental applications have been offered (e.g., Batey et al., 2012) and sources cited herein can support thoughtful PFI use.

Practitioners should use PFIs with caution, ask questions (e.g., testing purpose, psychometrics), and temper expectations about performance improvement. Ask about evidence of effectiveness, the value added by a PFI to development efforts (Michael, 2003), and possible side-effects. Remain open to new evidence yet be skeptical of anyone who suggests that "research supports" a given personality instrument for development or performance improvement. Considering the unsettled evidence to date, ask "how will we know if or when a PFI has 'worked?'" Practitioners are urged to gather PFI-effectiveness evidence, disseminate findings, or provide researchers with access to sites or data. Practitioners and researchers should appreciate, critique, and extend the work of the PFI pioneers cited herein.

Conclusion

Clear empirical evidence does not yet exist that PFIs help members of the workforce—individually or as collectives—develop and perform effectively. Given the popularity of personality inventories for developmental purposes, this is an area in need of concerted research attention.

Résumé

La littérature au sujet de la personnalité au sein de la psychologie industrielle et organisationnelle (I-O) s'est attardée sur la validité des échelles de personnalité pour la prévision de critères comme le rendement au travail, et les applications connexes pour les besoins de la sélection du personnel. Parallèlement, mais faisant l'objet de beaucoup moins de recherches, les inventaires de personnalité ont été utilisés à des fins de perfectionnement dans le cadre d'ateliers de groupe et de coaching individuel. La rétroaction sur l'évaluation de la personnalité représente un moyen moderne par lequel on applique l'antique conseil d'apprendre à « se connaître soi-même » (Emre, *The personality brokers: The strange history of Myers-Briggs and the birth of personality testing.* 2019, p. xviii). Toutefois, la théorie de l'intervention avec rétroaction fondée sur des tâches (Kluger & DeNisi, *Psychological Bulletin, 119,* 1996, 254–284) soulève des doutes quant à l'efficacité d'une rétroaction autofocalisée au chapitre de l'amélioration du rendement. Le présent article résume des études menées sur les répercussions d'interventions reliées au travail et basées sur la personnalité pour des critères relatifs au rendement ou au perfectionnement, au moyen d'une évaluation rapide des données probantes (Barends et al., *CEBMa guideline for rapid evidence assessments in management and organizations,* version 1.0, 2017). Au terme de requêtes structurées dans cinq bases de données et dans des listes de références et de compilations de citations, 12 études empiriques diversifiées ont été recensées. La plupart d'entre elles ont utilisé un outil d'évaluation basé sur les types plutôt que sur les traits; le Myers-Briggs Type Indicator (MBTI) a été l'indicateur de personnalité le plus utilisé. Aucune étude n'a examiné l'incidence des interventions basées sur la personnalité dans les évaluations des observateurs sur le rendement au travail. Bien que certains indices fassent état d'effets bénéfiques, les chercheurs n'ont pas encore établi l'effet des interventions basées sur la personnalité sur les critères reliés au rendement, et la possibilité d'effets négatifs demeure. On exhorte les praticiens à utiliser les interventions basées sur la personnalité avec prudence, à atténuer les attentes, à évaluer l'efficacité des interventions et à faire connaître les résultats — peut-être dans le cadre de partenariats entre chercheurs et praticiens.

Mots-clés : rétroaction sur la personnalité, test, changement de personnalité, rendementc, conscience de soi

References

References marked with an asterisk indicate studies included in the REA and references marked with a hashtag indicate studies cited only in the supplemental materials

Barends, E., Rousseau, D. M., & Briner, R. B. (Eds.). (2017). *CEBMa guideline for rapid evidence assessments in management and organizations, version 1.0*. Center for Evidence Based Management. https://cebma.org/wp-content/uploads/CEBMa-REA-Guideline.pdf

Batey, M., Walker, A., & Hughes, D. (2012). Psychometric tools in development—do they work and how? In J. Passmore (Ed.), *Psychometrics in coaching* (2nd ed., pp. 49–58). KoganPage.

Briner, R. B. (2018, November 15). *Do personality-typing tools have a place in HR?* HR Magazine, November. https://www.hrmagazine.co.uk/article-details/do-personality-typing-tools-have-a-place-in-hr

#Cavanagh, M., & Palmer, S. (2014). Editorial: Changes. *International Coaching Psychology Review*, *9*(1), 4–6.

Christiansen, N. D., & Tett, R. P. (2013). The long and winding road: An introduction to the handbook of personality at work. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 1–7). Routledge. https://doi.org/10.4324/9780203526910

*Church, A. H., Rotolo, C. T., Margulies, A., Del Giudice, M. J., Ginther, N. M., Levine, R., Novakoske, J., & Tuller, M. D. (2016). Does purpose matter? The stability of personality assessments in organization development and talent management applications over time. *The Journal of Applied Behavioral Science*, *52*(4), 450–481. https://doi.org/10.1177/0021886316668748

Church, A. H., Rotolo, C. T., Margulies, A., Del Giudice, M. J., Ginther, N. M., Levine, R., Novakoske, J., & Tuller, M. D. (2015). The role of personality in organization development: A multi-level framework for applying personality to individual, team, and organizational change. *Research in Organizational Change and Development*, *23*, 91–166. https://doi.org/10.1108/S0897-301620150000023003

#Coe, C. K. (1992). The MBTI: Potential uses and misuses in personnel administration. *Public Personnel Management*, *21*(4), 511–522. https://doi.org/10.1177/009102609202100407

*Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology*, *92*(2), 386–395. https://doi.org/10.1037/0021-9010.92.2.386

Emre, M. (2019). *The personality brokers: The strange history of Myers-Briggs and the birth of personality testing* [Vintage Canada edition]. Penguin Random House. (Original work published 2018. Doubleday; Random House Canada).

Espinoza, J. A., Daljeet, K. N., & Meyer, J. P. (2020). Establishing the structure and replicability of personality profiles using the HEXACO-PI-R. *Nature Human Behaviour*, *4*, 713–724. https://doi.org/10.1038/s41562-020-0853-6

Funder, D. C. (2001). Personality. *Annual Review of Psychology*, *52*, 197–221. https://doi.org/10.1146/annurev.psych.52.1.197

Garrety, K. (2007). Beyond ISTJ: A discourse-analytic study of the use of the Myers-Briggs Type Indicator as an organisational change device in an Australian industrial firm. *Asia Pacific Journal of Human Resources*, *45*(2), 218–234. https://doi.org/10.1177/1038411107079117.

Goodstein, L. D., & Lanyon, R. I. (1999). Applications of personality assessment to the workplace: A review. *Journal of Business and Psychology*, *13*(3), 291–322. https://doi.org/10.1023/A:1022941331649

#Gordon, S., & Palmer, S. (2014). Editorial. *International Coaching Psychology Review*, *9*(2), 116–117.

Hanson, W. E., & Poston, J. M. (2011). Building confidence in psychological assessment as a therapeutic intervention: An empirically based reply to Lilienfeld, Garb, and Wood (2011). *Psychological Assessment*, *23*(4), 1056–1062. https://doi.org/10.1037/a0025656

Harland, L. K. (2003). Using personality tests in leadership development: Test format effects and the mitigating impact of explanations and feedback. *Human Resource Development Quarterly*, *14*(3), 285–301. https://doi.org/10.1002/hrdq.1067

Hullinger, A. M., & DiGirolamo, J. A. (2018). *Referring a client to therapy: A set of guidelines*. International Coach Federation. https://coachfederation.org/app/uploads/2018/05/Whitepaper-Client-Referral.pdf

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Krieshok, T. S., Ulven, J. C., Hecox, J. L., & Wettersten, K. (2000). Résumé therapy and vocational test feedback: Tailoring interventions to self-efficacy outcomes. *Journal of Career Assessment*, *8*(3), 267–281. https://doi.org/10.1177/106907270000800305

#Lee, F. K., Johnston, J. A., & Dougherty, T. W. (2000). Using the five-factor model of personality to enhance career development and organizational functioning in the workplace. *Journal of Career Assessment*, *8*(4), 419–427. https://doi.org/10.1177/106907270000800411

Lundgren, H., Kroon, B., & Poell, R. F. (2017). Personality testing and workplace training: Exploring stakeholders, products and purpose in Western Europe. *European Journal of Training and Development*, *41*(3), 198–221. https://doi.org/10.1108/EJTD-03-2016-0015

Lundgren, H., Kroon, B., & Poell, R. F. (2019). Pigeonholing or learning instrument? Test takers' reactions to personality testing in management development. *European Journal of Training and Development*, *43*(3/4), 354–374. https://doi.org/10.1108/EJTD-09-2018-0091

Lundgren, H., Poell, R. F., & Kroon, B. (2019). "This is not a test": How do human resource development professionals use personality tests as tools of their professional practice? *Human Resource Development Quarterly*, *30*, 175–196. https://doi.org/10.1002/hrdq.21338

*Mansi, A. (2007). Executive coaching and psychometrics: A case study evaluating the use of the Hogan Personality Inventory (HPI) and the Hogan Development Survey (HDS) in senior management coaching. *The Coaching Psychologist*, *3*(2), 53–58.

Martin, L. S., Oades, L. G., & Caputi, P. (2014). Intentional personality change coaching: A randomised controlled trial of participant selected personality facet change using the five-factor model of personality. *International Coaching Psychology Review*, *9*(2), 196–209.

#McCarthy, A. M., & Garavan, T. N. (1999). Developing self-awareness in the managerial career development process: The value of 360-degree feedback and the MBTI. *Journal of European Industrial Training*, *23*(9), 437–445. https://doi.org/10.1108/03090599910302613

McCormick, I., & Burch, G. S. J. (2008). Personality-Focused coaching for leadership development. *Consulting Psychology Journal: Practice and Research*, *60*(3), 267–278. https://doi.org/10.1037/1065-9293.60.3.267

*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. https://doi.org/10.1177/107179190301000106

Moyle, P., & Hackston, J. (2018). Personality assessment for employee development: Ivory tower or real world? *Journal of Personality Assessment*, *100*(5), 507–517. https://doi.org/10.1080/00223891.2018.1481078

Nelson, E., & Hogan, R. (2009). Coaching on the dark side. *International Coaching Psychology Review*, *4*(1), 9–21.

Nye, C. D., & Roberts, B. W. (2013). A developmental perspective on the importance of personality for understanding workplace behavior. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 796–818). Routledge.

Prewett, M. S., Tett, R. P., & Christiansen, N. D. (2013). A review and comparison of 12 personality inventories on key psychometric characteristics. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 191–225). Routledge.

#Randall, R. M., Kwong, L., Kuivila, T., Levine, B., & Kogan, M. (2017, May/June). Building physicians with self-awareness. *Physician Leadership Journal*, *4*(3), 40–44.

*Rekar Munro, C., & Laiken, M. E. (2003). Developing and sustaining high performance teams: Managing difference through an action research approach. *OD Practitioner*, *35*(4), 62–67.

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, *143*(2), 117–141. https://doi.org/10.1037/bul0000088

*Sedlock, J. R. (2005). An exploratory study of the validity of the MBTI team report. *Journal of Psychological Type*, *65*(1), 1–8.

*Segovia, M. (2016). JDSU uses personality typology during management Empowerment Camps. *Industrial and Commercial Training*, *48*(1), 29–32. https://doi.org/10.1108/ICT-07-2015-0048

#Shorey, H. S. (2018). Introduction to the special issue on the role of personality assessment in consulting to organizations. *Journal of Personality Assessment*, *100*(5), 493–497. https://doi.org/10.1080/00223891.2018.1481859

*Stefansdottir, S. I., & Sutherland, V. J. (2005). Preference awareness education as stress management training for academic staff. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *21*, 311–323. https://doi.org/10.1002/smi.1071

Stieger, M., Wepfer, S., Rüegger, D., Kowatsch, T., Roberts, B. W., & Allemand, M. (2020). Becoming more conscientious or more open to experience? Effects of a two-week smartphone-based intervention for personality change. *European Journal of Personality*, *34*, 345–366. https://doi.org/10.1002/per.2267

*Sutton, A., Williams, H. M., & Allinson, C. W. (2015). A longitudinal, mixed method evaluation of self-awareness training in the workplace. *European Journal of Training and Development*, *39*(7), 610–627. https://doi.org/10.1108/EJTD-04-2015-0031

Tasselli, S., Kilduff, M., & Landis, B. (2018). Personality change: Implications for organizational behavior. *The Academy of Management Annals*, *12*(2), 467–493. https://doi.org/10.5465/annals.2016.0008

*Varvel, T., Adams, S. G., Pridie, S. J., & Ruiz Ulloa, B. C. (2004). Team effectiveness and individual Myers-Briggs personality dimensions. *Journal of Management Engineering*, *20*(4), 141–146. https://doi.org/10.1061/(ASCE)0742-597X(2004)20:4(141)

*Waite, R., & McKinney, N. (2015). Knowing your preference: The nexus of personality and leadership. *Creative Nursing*, *21*(3), 172–178. https://doi.org/10.1891/1078-4535.21.3.172

*Waite, R., & McKinney, N. S. (2018). Personality typology: Understanding your preferences and striving for team effectiveness. *The ABNF Journal*, *29*(1), 8–16.

Walck, C. L. (1997). Using the MBTI in management and leadership: A review of the literature. In C. Fitzgerald & L. K. Kirby (Eds.), *Developing leaders: Research and applications in psychological type and leadership development* (pp. 63–114). Davies-Black Publishing.

#Wood, R., & Butterworth, A. (1997). Fair use of psychometrics in guidance. *Journal of Education and Work*, *10*(3), 281–291. https://doi.org/10.1080/1363908970100305

Zickar, M. J., & Kostek, J. A. (2013). History of personality testing within organizations. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 173–190). Routledge.

Received May 15, 2020
Revision received November 24, 2020
Accepted December 11, 2020 ■