Issues of Assessment, Measurement, and Evaluation

by

Thomas G. Reio, Jr., PhD

Abstract

Assessment, measurement, and evaluation are important if individuals, teams, and organizations ever hope to make correct, data-based decisions that can assist in efforts to keep the organization competitive and solvent. Personality trait, interest, and cognitive ability assessments are extensively used for development purposes and screening and selection decision making, despite concerns about who is trained and credentialed to properly administer and interpret the assessments. There are ethical, moral and legal implications related to assessments that cannot be overlooked. Environmental context, faking, face-to-face versus computer-based testing, cultural diversity, disabilities, second language, stereotype threat, social networking websites, and formative and summative evaluations are explored as issues that impinge upon the proper use of personality trait, interest, and cognitive ability assessments. Research is advocated to develop cross-culturally valid measures and training and credentialing are presented as means to deal effectively with these issues.

Keywords:

Assessment, Measurement, Evaluation, Personality, Faking, Disability, Diversity, Stereotype Threat, Social Networking

Assessment, measurement, and evaluation lies at the heart of best organizational research and practice (Spitzer, 2005; Wang & Wang, 2005). It is hard to imagine how developmental or screening and selection activities would be able to proceed without the benefit of the information gleaned from assessing a characteristic of an individual or team and evaluating its relative importance to the individual, team or organization (McDonald & Hite, 2016). If we were interested in a learner's creativity, for instance, creativity would be assessed first through a series of instruments designed to measure its multidimensional nature (Bang & Reio, 2017). Poor performance on the measures would become a significant concern (evaluation) if being adept at working creatively was a prerequisite to advancing to career-related specializations that require quite a bit of creativity.

Although the terms are often used interchangeably, for the purpose of this article assessment refers to the process of gathering, synthesizing, and deciphering information for the purpose of decision-making and problem solving (Hattie & Leeson, 2013). Measurement, on the other hand, refers to the process of ascertaining a quantitative or qualitative characteristic of an individual, team or group that is of theoretical, research or practical relevance (Michell, 1999; Stevens, 1951). Tests are used as a means to discern the characteristic, such as a spoken question, an observation or a paper-and-pencil or online test (Stemler & Sternberg, 2013). Evaluation, in turn, is the process of combining what we have measured with other relevant information to distinguish the measured characteristic's relevance and importance. With evaluation, therefore, one goes beyond associating numbers or the qualitative characteristics of what we see (measurement) to judging the value of the characteristics relevant to the aims of the evaluation. For example, the work portfolio of a landscape architect student, replete with examples of creative work-related products and accomplishments, is in effect a broad assessment of the individual's competencies and therefore qualifications. An evaluation of the portfolio by a competent teacher or prospective employer would entail contrasting the portfolio's content with meeting some specified aim (e.g., evi-

dence of successfully designing the landscape of an office building in a semi-arid area using xeriscape principles) and determining the extent of its desirability or value.

Assessment, measurement,and evaluation efforts are global and international (Rothstein & Goffin, 2006) and are generally led by human resource (HR) professionals and to a lesser extent managers in workplace contexts. Yet, there are significant issues that call into question the appropriate interpretation of the data generated from these activities and its subsequent use for decision-making and problem solving. The aim of this article will be to highlight nine of the most important of these issues, as suggested by Lundgren, Poell, and Kroon (2019), Heikkila and Reio (2016), Rothstein and Goffin (2006), and Hough and Oswald (2008), including environmental context, faking, face-to-face versus computer-based testing, cultural diversity, and disabilities, second language, stereotype threat, social networking websites, and the purpose of formative and summative evaluations.

The Use of Assessments in Organizations

The use of personality trait, interest, and general cognitive ability assessments, either face-to-face or online, is a major issue because they are used extensively in workplaces for development (Lundgren, Poell, & Kroon, 2019) and screening and selection (Rothstein & Goffin, 2006) purposes. HR professionals sometimes administer tests for selection and potential for advancement purposes at assessment centers, but that is usually not the case for non-managerial applicants (Thornton & Krause, 2009). For developmental settings like executive coaching and management training, personality trait assessments are used as "ice breakers" (e.g., Myers-Briggs Type Indicator; Occupational Personality Questionnaire; Neuroticism, Extraversion, Openness Personality Inventory; Team Management Profile) by trainers in team-building exercises as well as tools to facilitate team-and organizational-learning, management coaching and mentoring, leadership development, and management learning (Lundgren et al., 2019). Although challenges have been made as to the validity and usefulness of personality test assessments for HR and managerial practice in developmental settings (e.g., Ford & Harding, 2007), on the whole, they have been shown to be serviceable when handled appropriately, as in being administered and interpreted by a trained and credentialed professional (Furnham & Jackson, 2011).

With regard to screening and selection, on the other hand, personality trait assessments are used for personnel selection where possible issues related to their validity and usefulness can be magnified because the decisions related to performance on the measure can impact the livelihood of another human being (Hough & Oswald, 2008). Ethical, moral, and legal questions arise when the instruments employed in an assessment have low discriminant and predictive validity and have been administered and interpreted by individuals who are untrained and uncredentialed, especially with psychological tests (Lundgren et al., 2019). It is hard to fathom how an evaluation related to a screening or selection decision can be appropriate when such conditions exist. Still, by a significant degree, personality trait assessments are the most prevalent screening and selection tool used in organizations, followed by interest and cognitive ability assessments (Rossier, 2015). Although each of the nine issues presented in this research has application to interest and cognitive ability assessments, we will focus then primarily on personality trait assessment because of its extensive use by HR professionals and managers (Lundgren et al., 2019).

Personality trait assessment can be especially informative to organizations because traits are indicators of employees' behavioral tendencies in organizational contexts (Ones, Dilchert, Viswesvaran, & Judge, 2007). Personality assessment is a $500 million industry growing currently at 10 percent annually in the US alone (Meinert, 2015). Meinert noted that nearly 60 percent of workers take workplace assessments and 22 percent of organizations use such assessments for pre-employment screening. In the case of employment screening, applicants need to "pass" the personality, interest or cognitive ability assessment to qualify for an interview. In the case of personality, a pass would be earned if the applicant's personality test scores matched the personality profile of what was deemed the pattern of characteristics of a successful employee for the position in question at the organization. For example, the personality profile of a successful customer service manager might be one where the employee is high in conscientiousness, agreeableness and emotional stability and low in aggressiveness and impulsiveness. In a school setting, in contrast, spatial cognition tests (e.g., Cube Perspectives Test, Visual Memory Test, Hidden

Figures Test) might be used to screen prospective medical school students; high spatial ability is best for surgeons in particular (Reio, Czarnolewski, & Eliot, 2004)

Millions of applicants worldwide complete personality trait assessments annually for positions ranging from entry-level workers to mid- and upper-level managers and CEOs (Rothstein & Goffin, 2006). The assessments are useful for gaining evaluative insights into an applicant's conscientiousness, reliability, motivation, risk aversion, values, teamwork and relationship-building propensity, and leadership, to name a few (Heikkila & Reio, 2016). Conscientiousness and positive self-concept, for example, are the traits that best predict job performance (Thorenson, C., Bradley, Bliese, & Thorenson, J., 2004). Conscientious employees tend to be motivated and reliable, and positive employees upbeat and persistent. HR professionals and managers increasingly rely on personality trait assessments because they have been shown also to reliably forecast not only job performance, but good employee and job fit. The thinking is if there is a solid alignment between the employee and job, the employee's attendant job satisfaction, organizational commitment and technical and interpersonal job performance would be enhanced, reducing the likelihood of poor productivity and costly turnover (Hattie & Leeson, 2013).

Personality trait assessments are used extensively in the airline, health, hospitality, food preparation, manufacturing, education, and service industries, as well as in the military (Lundgren et al., 2019). Assessment use is particularly prevalent for entry-level jobs in the food preparation and retail industries (Heikkila & Reio, 2016). Although these positions are desirable for individuals from all age groups, including retirees, younger workers tend to be represented most proportionally for food and retail jobs than the other age groups (Van Iddekinge, Raymark, & Roth, 2005).

Issues and Challenges Associated with Personality Trait Assessments

Although their use is prevalent and increasing (Hattie & Leeson, 2013; Lundgren et al., 2019), there are a number of concerns worldwide about personality trait assessment. Nine of the most salient issues identified in the research literature are presented below.

The environmental context. In Bronfenbrenner's (1979, 1999) bioecological theory, we develop through reciprocal, increasingly complex interactions between the curious, proactive individual and the individuals, objects and symbols in that individual's immediate environment. Thus, to understand the development of individuals and the behaviors that entails, HR professionals and managers need to acknowledge the environment when interpreting the behavior of oneself and others. Bronfenbrenner (1999) decried the lack of suitable research designs and methods of measurement for what he called the "discovery research" (p. 24) necessary for theory generation related to the bioecological theoretical model. Indeed, to date, many social science research designs, which includes those within the field of organizational psychology, still honor hypothesis and theory testing through quantitative methods, rather than theory generation through qualitative methods (Reio, 2016).

Hattie and Leeson (2013) also noted the negligible attention to assessing the interactions with learning contexts, such as what we see in the workplace, and how they interact with individual skills and personality functioning. For instance, it seems unlikely to find a job situation that did not have some sort of environmental or situational constraint impinging upon optimal performance. Situational constraints would include unforeseen circumstances, resource inaccessibility, impossible deadlines, lack of coworker cooperation, and/or uncivil treatment by a supervisor, coworker or vendor. Thus, when making a performance evaluation, these types of situational constraints require acknowledgement through precise measurement if we are to fairly and more accurately evaluate the individual's performance on an internship- or job-related task. Lack of contextual information is a therefore significant barrier to valid data interpretation and evaluation. The issue, then, is two-fold: a) we need to assess the environment to be more predictive of learning and performance or some other desired outcome and b) through rigorous research, design better measures that will more accurately reflect the setting in question.

Faking responses to look better. Faking responses or "gaming a test" can make it relatively easy to claim to be dependable and hard working in job application settings. Faking is prevalent and should be a concern because it may offer unfair advantage to those who do not fake (Lee, Smith, & Geisinger,

2017). In a Croatian study, Galic, Jerniec, and Kovacic (2012) found that university students were able to fake desirable profiles, albeit to a limited degree. A solution to this dilemma might be posting warning messages; in a Chinese setting, posting messages did reduce faking partially, but the effect was minimal and impractical to implement (Fan et al., 2012). Further, as a possible means to reduce the likelihood of faking, Connelly and Ones (2010) reported through their meta-analytic work that observer-rated personality ratings were far better forecasters of job performance than self-report measures of personality. Moreover, Lee et al. (2017) found that faking was linked to job performance as well, but not selection. Taking observer-ratings of personality would be largely impractical in that the observer needs to be trained to perform such assessments, including finding ways to interact with the applicant, and the time required to do them could be costly. Future research is required to refine our knowledge about how best to manage faking far more effectively.

Testing face-to-face versus computer-based. Face-to-face versus computer-based testing remains a relatively unexplored area of assessment. Unquestionably, computer-based testing (which includes web-based, online testing) is growing appreciably because of its use to not only administer tests, but also to collect and present pertinent data, and interpret the results (Hattie & Leeson, 2013). Online versions of tests are increasingly available, particularly personality tests and interest inventories. As with any data collection method, there are limitations to consider. Online security and therefore cheating is clearly an issue because it is difficult to guarantee. This situation becomes problematic in the sense that systematic measurement error may result for two reasons. First, applicants may be selective about what information they choose to disclose online because of security and anonymity concerns. Second, the test taker may not be who they claim to be. Further, there are substantial issues related to web-based versions of tests as they remain relatively untested psychometrically, especially for equivalency and reliability as comparable instruments to paper-and-pencil versions of tests that have undergone rigorous validation. The lack of validity evidence for the web-based version of a test leaves the HR professional or manager in a difficult position because it is not appropriate to apply test norms aligned with paper-and-pencil-based data to those data acquired from online results (Hattie & Leeson, 2013). Future research is required to validate both paper-and-pencil and online versions of personality trait assessments. HR professionals and managers are cautioned to only use a measure that has been validated for both delivery formats.

Cultural diversity. Another issue related to making valid assessments is cultural diversity (Hough & Oswald, 2008). First, we must remember that most of the psychological, sociological, and anthropological theories that undergird our assessment tools are Western-based and therefore may not necessarily be valid tools for use with those from non-Western cultures (Sue, 1999). There is little doubt of the increasing diversity of distinct populations within respective countries and this reality must be acknowledged in our measurement and assessment efforts. Thus, the psychometric properties of assessment tools need to be calibrated with a more global or international perspective. We have discussed earlier how face-to-face tests are not necessarily equivalent to web-based versions of a test; the same idea extends to working with diverse populations. Unless a test has been validated for use with a certain population, there is little reason to support its use for that population, particularly as a tool to make selection and career-based decisions (Hattie & Leeson, 2013; McDonald & Hite, 2016). There are, however, some useful means to address this issue psychometrically (Nimon & Reio, 2011).

Nimon and Reio (2011) introduced the notion of testing for measurement invariance or equivalence, defined as the consistency of measurement across groups (e.g., age, sex, ethnicity, culture, nationality, online vs. face-to-face), to address the issue of construct validity by group. We must be aware that perceived differences or lack of group differences is problematic without establishing the across-group construct validity of the data; thus, the items in an assessment should mean the same thing to each group. Nimon and Reio recommended investigating measurement invariance through a three-step process (the second and third steps would require large sample sizes to run factor analyses). First, compute and compare the reliability estimates of each group for evidence of similarity. Second, conduct exploratory factor analysis by group and determine whether the same number of well-defined factors, accompanied by sufficiently high factor coefficients (> .40), appear per group; subsequently, compute factorial invariance indices (i.e., Salient Variable Similarity Index,

Coefficient of Congruence, and correlations between factor pattern coefficients) to find preliminary measurement invariance evidence. Third, conduct confirmatory factor analysis by group and examine whether the fit indices are comparable per group, followed by performing four measurement invariance tests (configural, weak, strong, and strict). If sufficient evidence of measurement invariance has been found, then the HR professional has reasonable assurance that whatever is being assessed is being perceived similarly by the group or groups in question (construct validity), which renders valid interpretation of the data generated by the measure.

Disabilities. One especially challenging assessment issue can be working with those who have disabilities (Heikkila & Reio, 2016). People with disabilities constitute roughly fifty-four million individuals in the US, making them the largest minority group (Rocco, Bowman, & Bryant, 2014). Rocco et al. also make us aware that roughly a third of individuals without a disability become disabled during their peak working years. Thus, it is curious seeing so few diversity researchers in the organizational psychology field dealing directly with those with disabilities, despite their increasing presence in the workplace.

With passage of the American with Disabilities Act (ADA) in 1990 and its reinforcing amendment in 2016, US workplaces have better representation by those with disabilities. One important practice arising from this law, and one that organizational professionals must be acutely aware of, is that organizations must make reasonable accommodations to make the workplace accessible to employees with physical or mental disabilities. The US Equal Opportunities Commission classifies an individual disabled when they have any physical or mental impairments that substantially limits one or more major life activities, such as being able to perform on the job. Autism, hearing and sight impairment, orthopedic impairment, emotional disturbance, intellectual disability, cerebral palsy, alcoholism, sciatic nerve pain, diabetes, and attentional deficit hyperactivity disorder are examples of disabilities. The question is how to accommodate jobs and testing conditions reasonably across a wide range of disabilities. Obviously, reasonable accommodation would not be possible for all situations; for example, a blind person logically could not be an express delivery driver and a person with a seizure disorder could not be a pilot of a plane.

Noting the increasing prevalence of pre-employment testing for screening purposes, especially in the retail and food industries, HR professionals and managers may need to review the practice of using pre-employment assessment tests for selection and career planning purposes for those with disabilities (Heikkila & Reio, 2016; Rocco et al., 2014). Clearly, individuals with intellectual and emotional disabilities may not succeed in their employment quests if they cannot successfully complete the pre-employment personality assessments. For such individuals, participating at adult education centers might afford learning about and practicing dealing with pre-employment testing conditions and taking personality assessments. For example, students in school-to-work programs, typically 16-21 years-of-age, could work with the school, parents and target hiring organizations to help them develop the test taking skills needed to be assessed correctly and evaluated fairly (Heikkila & Reio, 2016).

Second language. Another assessment challenge has to do with individuals whose primary language is not the same language as the majority culture. In US school settings, for example, children of immigrants constitute one-fifth of all school-aged children, with most of their households not having anyone aged 14 or older who speaks English well (Kopriva & Albers, 2013). Second language students too often perform poorly on achievement tests and are twice as likely as native speakers to drop out of school. Dropping out leaves them ill prepared to compete favorably for well-paying jobs because their language skills tend to be poorly developed as compared to native speakers. English language learners' parents in the US also suffer under testing situations because language-heavy tests like personality assessments are administered in English, leaving them with little hope of demonstrating their true personality assessment scores, thereby limiting possibilities for attaining occupational success. This state of affairs is troubling considering the large majority of immigrants are hardworking, honest, and dependable workers (Garcia, Jensen, & Scribner, 2009). HR professionals and managers must find constructive, productive ways to meet this challenge by helping to create and implement versions of tests used to make selection and promotion decisions that are linguistically valid. Thus, if the individual's primary language is French, and the job does not require strong English speaking skills, then the personality assessment should be in

French to ascertain a truer sense of their personality scores. Valid test scores are vital if suitability and fit evaluations are to be unbiased.

Stereotype threat. A stereotype is an overgeneralization about members of a social group. Stereotype threat, defined as the inhibition of performance due to concern about confirming a stereotype, has been linked with group differences in performance on tests. For stereotype threat to exist, knowledge of the stereotype must exist for the affected party. For example, in the case of older workers, the unfounded stereotype that they are not proficient with computers and new technology is widespread in many workplaces. Based on the older worker's knowledge of the ageist stereotype, an instance of stereotype threat occurs when the older test taker fails to demonstrate their true understanding of how to perform a computer task due to the threat. Likewise, stereotypes that males are not good at writing and verbal tasks lead to males underperforming on quantitative writing and verbal ability assessments when gender-based stereotypes are triggered (Walton & Spencer, 2009). Similar stereotype effects have been found with females with regards to mathematical and spatial ability and non-Asian ethnic minorities' performance on graduate school entrance exams like the Graduate Record Examination (Halpern & Butler, 2013). Therefore, stereotype threat should be an issue to HR professionals, considering that performance on assessments can serve as a means to judge an individual's hiring or promotional fitness or whether he or she should advance from training (Ryan & Sackett, 2013). The point is that test scores can systematically underestimate the test scores (e.g., intellectual ability) of individuals saddled with negative stereotypes (Walton & Spencer, 2009). To lessen the possibility of stereotype threat, HR professionals need to be aware first that it exists and that it can systematically dampen test performance. As a possible solution, Steele (2010) suggested that reducing the likelihood of stereotype threat could be accomplished through attending to the assessment context by using validated instruments, offering a comfortable assessment setting, and describing the assessment's purpose, explaining its use, and familiarizing test takers with the assessment to acclimate them to the testing situation.

Social networking websites. Due to the high stakes associated with recruiting and selecting "the right employee" to fill a job, HR professionals therefore need to be increasingly creative. One relatively new and untested approach to evaluating the personality profiles of current and potential employees is through social networking websites (SNW) like Facebook and LinkedIn. As the fourth most popular online activity (Statista, 2015), organizations are embracing SNWs because they afford the collection of organizationally relevant information, like personality traits, useful for determining job applicant suitability and fit. For example, in one SNW study, conscientiousness was associated with the use of less problematic online posts related to illicit materials and content of a sexual nature (Karl, Peluchette, & Schlaegel, 2010). Moreover, in two studies of university students (justified by the authors because of students' heavy use of Facebook), Kluemper, Rosen, and Mossholder (2012) found that SNW profiles generated from Facebook pages yielded valid other-rated measures of the Big Five personality traits (i.e., extraversion, neuroticism, openness to experience, agreeableness, conscientiousness) (Costa & McCrae, 1992). Thus, they found preliminary support for using SNW profiles to determine the personality traits of college students. Interestingly, after a few hours of training to ensure the fidelity of the study and high inter-rater reliability, the study's evaluators took but 5-10 minutes to complete the Facebook profile evaluations. On the surface, these findings seem promising for busy, cost-conscious HR researchers and managers, but the researchers strongly advised caution to allow for more systematic inquiry into SNW use. Kluemper et al. (2012) also reminded us that there may be unknown legal implications to using SNWs and prospective applicants may simply avoid dealings with the organization because of perceptions that the practice amounts to unethical business practice. Another possibility is that as prospective applicants become more aware of the extent and implication of this organizational practice, just as applicants have been shown to fake face-to-face and online personality and interest tests (Fan et al., 2012; Galic et al., 2012), individuals may also creatively game their Facebook and LinkedIn profiles to appear to be something they are not.

A compelling case could be made that this practice may amount to a situation of social injustice, a label no organization should reasonably risk (Byrd, 2014). Serious, logical thinking must be allowed for considering how this practice is truly equitable and fair and does not violate social norms for reasonable access.

How is it fair that this practice is being used for hiring decisions when many choose not to participate in SNWs? Thus, these individuals may be considered as a less viable candidate for a position simply because they do not possess an SNW for an unknown HR professional to evaluate in the first place. How can we determine equal access when many who are poor or disadvantaged cannot afford a computer or Internet service? How is it fair for those who live in rural or urban areas who have to go the local library to setup a SNW account when there is no "local" library? Even if it exists, without one's own transportation, or being able to be afford taking the bus, equitable access is again denied. The answers to these societal questions are profound because the lack of access, equity and inclusion falls disproportionately on underrepresented minorities (Kormanik & Nwaoma, 2015), the culturally and linguistically diverse (Fenwick, 2015), and those with disabilities (Rocco et al., 2014). HR professionals are in a unique position to need to lead well-designed cross-sectional, mixed-method and longitudinal research efforts to better understand and increase awareness of social justice issues surrounding the use of SNWs to make important decisions about an individual's suitability and fit for a job.

Formative and summative evaluation. As one of the pillars of solid organizational practice, evaluation serves as a meaningful tool for providing ongoing feedback relevant to attaining organizational goals. HR researchers and practitioners are well qualified to plan, implement, analyze and interpret the results of an evaluation. In the field of instructional design, the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model is used to develop instructional courses and training programs. The evaluation piece of this generic systematic process model entails determining whether the problem has been solved, if objectives have been met, the relative impact of the course or product, and what changes may be necessary to do a better job of offering the course or product in the future (Peterson, 2003). On the other hand, there are numerous additional evaluation models (see Reio, Rocco, Smith, & Chang, 2017), but most build extensively upon Kirkpatrick's 4-level evaluation model (reaction, learning, job performance, and organizational impact). Reio et al. (2017) performed an extensive critique of the Kirkpatrick model finding that although the model has its shortcomings, namely it does not address formative and process evaluations, it remains a viable framework for evaluating training programs.

There are two main types of evaluations: formative and summative. Formative evaluation constitutes collecting information and interpreting it during learning, while summative evaluation entails making interpretations about the learning at the end of the learning endeavor. Each provides useful feedback about how one is performing. In the training classroom, formative evaluation is helpful for providing the learner specific feedback about their performance for the purpose of making "on-the-fly" adjustments. In contrast, summative evaluation provides feedback about the degree to which the learner attained the goals of the training class. At a broader organizational level, periodic formative evaluations offer employees the kind of specific feedback needed to advance their skills during a performance appraisal period, while summative evaluations certify whether performance goals have been met at the appraisal period's closing. Consequently, each type of evaluation could be used skillfully by HR professionals in development efforts and by managers to help employees attain performance-related goals and aspirations.

Conclusions

We must be mindful that it is not the test or assessment that is formative or summative, but it is the interpretations that are (Scriven, 1990). Personality, interest, and cognitive skill assessments of course are not suitable for formative evaluations, but they have been shown to be fitting, with caveats, for evaluations (i.e., summative) of prospective candidates for a job or promotion. Thus, HR professionals and managers must be certain that for whatever is being evaluated (e.g., employee-job fit), the data needed to make a proper summative evaluation has been generated from the use of valid measurement tools. The use of widely available survey instruments from online and other non-juried sources is strongly cautioned against unless there is substantial validity evidence for their use with your particular population. Vital decisions about the livelihood of human beings are based on the results of taking tests; as HR professionals, managers, and citizens, we owe it to those taking the tests that they are fair and valid for all or risk contributing to the social injustice issues plagu-

ing our societies. We must acknowledge that issues of interpretation concern the environmental context, faking, mode of test delivery, cultural diversity, presence of a disability, English as a second language, stereotype threat, and if the data were collected from social networking websites. Without acknowledging these issues when interpreting one's data and making the final, summative evaluative decision about a job applicant, the validity of the HR professional and manager's interpretation can be seriously and unnecessarily comprised.

The Author

Thomas G. Reio, Jr., PhD
Florida International University
11200 SW 8th Street
ZEB 358A
Miami, Florida 33199
reiot@fiu.edu
502-226-0088

References

Bang, J., & Reio, T. G., Jr. (2017). Personal accomplishment, mentoring and creative self-efficacy as predictors of creative work involvement: The moderating roles of positive and negative affect. *Journal of Psychology*: *Interdisciplinary and Applied, 151*, 148-170.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (1999). Environments in developmental perspective: Theoretical and operational models. In S. L. Friedman & T. D. Wachs (Eds.), *Measuring environment across the lifespan: Emerging methods and concepts* (pp. 3-28). Washington, DC: American Psychological Association.

Byrd, M. (2014). A social justice paradigm for HRD: Philosophical and theoretical foundations. In N. E. Chalofsky, T. S. Rocco, & M. L. Morris (Eds.), *Handbook of human resource development* (pp. 281-298). New York, NY: Wiley.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*, 1092-1122.

Costa, P., T., Jr., & McCrae, H. H. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13*, 653-665.

Fan, J., Gao, D., Carroll, S. A., Lopen, S. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology, 97*, 866-880.

Fenwick, T. (2015). Conceptualizing critical HRD (CHRD): Tensions, dilemmas and possibilities. In R. F. Poell, T. S. Rocco, & G. L. Roth (Eds.), *The Routledge companion to human resource development* (pp. 113-123). New York, NY: Routledge.

Ford, J., & Harding, N. (2007). Move over management: We are all leaders now. *Management Learning, 38*, 475-493.

Furnham, A., & Jackson, C. J. (2011). Practitioner reactions to work-related psychological tests. *Journal of Managerial Psychology, 26*, 549-565.

Galic, Z., Jerniec, Z., & Kovicic, M. P. (2012). Do applicants fake their personality questionnaire responses and how successful are their attempts? *International Journal of Selection and Assessment, 20*, 229-241.

Garcia, E. E., Jensen, B. T., & Scribner, K. P. (2009). The demographic imperative. *Educational Leadership, 66*(7), 8-13.

Halpern, D. F., & Butler, H. A. (2013). Assessment in higher education admissions and outcomes. In R. F. Geisinger, B. A. Bracken, J. C. Carlson, J. C. Hansen, N. R. Kuncel, S. R. Reise, & M. C. Rodriguez (Eds.), *APA handbook of assessment in psychology:* Vol. 3. *Testing and assessment in school psychology* (pp. 319-336). Washington, DC: American Psychological Association.

Hattie, J., & Leeson, H. (2013). Future directions in assessment and testing in education and psychology. In R. F. Geisinger, B. A. Bracken, J. C. Carlson, J. C. Hansen, N. R. Kuncel, S. R. Reise, & M. C. Rodriguez (Eds.), *APA handbook of assessment in psychology:* Vol. 3. *Testing and assessment in school psychology* (pp. 591-622). Washington, DC: American Psychological Association.

Heikkila, M. R., & Reio Jr., T. G. (2016). The role of computerized personality assessments in students with disabilities' school-to-work transition [reprint]. In I. Management Association (Ed*.), Psychology and mental health: Concepts, methodologies, tools, and applications* (pp. 1613-1628). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0159-6.ch070

Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial-organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology, 1*, 272-290.

Karl, K., Peluchette, J., & Schlaegel, C. (2010). Who's posting Facebook faux pas? A cross-cultural examination of personality differences. *International Journal of Selection and Assessment, 18*, 174–186.

Kluemper, D. H., Rosen, P.A., & Mossholder, K. W. (2012). Social networking websites, personality ratings, and the organizational context: More than meets the eye? *Journal of Applied Social Psychology, 42*, 1143-1172.

Kopriva, R., & Albers, C. A. (2013). Considerations for achievement testing of students with individual needs. In R. F. Geisinger, B. A. Bracken, J. C. Carlson, J. C. Hansen, N. R. Kuncel, S. R. Reise, & M. C. Rodriguez (Eds.), *APA handbook of assessment in psychology*: Vol. 3. *Testing and assessment in school psychology* (pp. 369-390). Washington, DC: American Psychological Association.

Kormanik, M., & Nwaoma, P. C. (2015). Diversity and inclusion initiatives in organizations. In R. F. Poell, T. S. Rocco, & G. L. Roth (Eds.), *The Routledge companion to human resource development* (pp. 307-317). New York, NY: Routledge.

Lee, H., Smith, Z., & Geisinger, K. F. (2017). Faking under a nonlinear relationship between personality assessment scores and job performance. *International Journal of Selection and Assessment*, *25*, 284-298.

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, 29*; 1-22. https://doi.org/10.1002/hrdq.21338

McDonald, K., & Hite, L. (2016). *Career develop ment: A human resource development per spective*. New York, NY: Routledge.

Meinert, D. (2015, June 1). What do personality tests really reveal? *HR Magazine, 34*, 12-16.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge, UK: Cambridge University Press.

Nimon, K., & Reio, T. G., Jr. (2011). Measurement invariance: A foundational principle for quantitative theory building. *Human Resource Development Review, 10*, 198-214.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*, 995–1027.

Peterson, C. (2003). Bringing ADDIE to life: Instructional design at its best. *Journal of Educational Multimedia and Hypermedia, 12*, 227-241.

Reio, T. G., Jr. (2016). Nonexperimental research: Strengths, weaknesses and issues of precision. *European Journal of Training and Development*, *40*, 676-690.

Reio, T. G., Jr., Czarnolewski, M., & Eliot, J. (2004). Handedness and spatial ability: Differential patterns of relationships. *Laterality: Asymmetries of Body, Brain, and Cognition, 9,* 339-358.

Reio, T. G., Jr., Rocco, T. S., Smith, D. S., & Chang, E. (2017). A critique of Kirkpatrick's four-level model after 50 Years. *New Horizons in Adult Education and Human Resource Development*, *29*(2), 35-53.

Rocco, T. S., Bowman, L., & Bryant, L. O. (2014). Disability, health, and wellness programs, and the role of HRD. In N. E. Chalofsky, T. S. Rocco, & M. L. Morris (Eds.), *Handbook of human resource development* (pp. 299-313). New York, NY: Wiley.

Rossier, J. (2015). Personality assessment and career interventions. In P. J. Hartung, M. L. Savickas, & W. B. Walsh (Eds.), *APA handbook of career interventions*: Vol. 1. *Foundations* (pp. 327-350). Washington, DC: American Psychological Association.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16*, 155-180.

Ryan, A. M., & Sackett, P. R. (2013). Stereotype threat in workplaces assessment. In R. F. Geisinger, B. A. Bracken, J. C. Carlson, J. C. Hansen, N. R. Kuncel, S. R. Reise, & M. C. Rodriguez (Eds.), *APA handbook of assessment in psychology:* Vol. 3. *Testing and assessment in school psychology* (pp. 661-674). Washington, DC: American Psychological Association.

Scriven, M. (1990). Beyond formative and summative. In M. McLaughlin & D. Phillips (Eds.), *Evaluation and education at quarter century* (pp. 19-64). Chicago, IL: University of Chicago/National Society for the Study of Education.

Spitzer, D. R. (2005). Learning effectiveness measurement: A new approach for measuring and managing learning to achieve business results. *Advances in Developing Human Resources, 7*, 55-70.

Statista. (2015). *Most popular online activities of adult internet users in the United States as of July 2015*. Retrieved on June 28, 2018, from https://www.statista.com/statistics/183910/internet-activities-of-us-users/

Steele, C. M. (2010). *Whistling Vivaldi and other clues to how stereotypes affect us*. New York, NY: Norton.

Stemler, S. E., & Sternberg, R. J. (2013). The assessment of aptitude. In R. F. Geisinger, B. A. Bracken, J. C. Carlson, J. C. Hansen, N. R. Kuncel, S. R. Reise, & M. C. Rodriguez (Eds.), *APA handbook of assessment in psychology:* Vol. 3. *Testing and assessment in school psychology* (pp. 281-296). Washington, DC: American Psychological Association.

Stevens, S. S. (1951). Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.) *Handbook of experimental psychology* (pp. 1-49). New York, NY: Wiley.

Sue, S. (1999). Science, ethnicity and bias. Where have we gone wrong? *American Psychologist, 54*, 1070-1077.

Thorensen, C. J., Bradley, J., C., Bliese, P. D., & Thorensen, J. D. (2004). The Big Five personality traits and individual job performance and growth trajectories in maintenance and transitional job stages. *Journal of Applied Psychology, 89*, 835-853.

Thornton, G. C., III, & Krause, D. E. (2009). Selection versus development assessment centers. An international survey of design, execution, and evaluation. *International Journal of Human Resource Management, 20*, 478-498.

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*, 536-552.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*, 1132-1139.

Wang, G. G., & Wang, J. (2005). Human resource development evaluation: Emerging market, barriers, and theory building. *Advances in Developing Human Resources, 7*, 22-36.