



The use of personality measures in personnel selection: What does current research support? ☆

Mitchell G. Rothstein ^{a,*}, Richard D. Goffin ^b

^a *Richard Ivey School of Business, University of Western Ontario, London, Ontario, Canada, N6A 3K7*

^b *Department of Psychology, University of Western Ontario, Canada*

Abstract

With an eye toward research and practice, this article reviews and evaluates main trends that have contributed to the increasing use of personality assessment in personnel selection. Research on the ability of personality to predict job performance is covered, including the Five Factor Model of personality versus narrow personality measures, meta-analyses of personality–criterion relationships, moderator effects, mediator effects, and incremental validity of personality over other selection testing methods. Personality and team performance is also covered. Main trends in contemporary research on the extent to which applicant “faking” of personality tests poses a serious threat are explicated, as are promising approaches for contending with applicant faking such as the “faking warning” and the forced-choice method of personality assessment. Finally, internet-based assessment of personality and computer adaptive personality testing are synopsized.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Personality assessment; Personnel selection; Five factor model; Personality and job performance prediction

Personality measures are increasingly being used by managers and human resource professionals to evaluate the suitability of job applicants for positions across many levels in an organization. The growth of this personnel selection practice undoubtedly stems from a series of meta-analytic research studies in the early 1990s in which personality measures were demonstrated to have a level of validity and predictability for personnel selection that historically had not been evident. In this article we briefly review available survey data on the current use of personality measures in personnel selection and discuss the historical context for the growth of this human resource practice. We then review the important trends in research examining the use of personality measures to predict job performance since the publication of the meta-analytic evidence that spurred the resurgence of interest in this topic. Of particular interest throughout this review are the implications for human resource practice in the use of personality measures for personnel selection.

☆ Preparation of this article was supported by grants from The Social Sciences and Humanities Research Council of Canada to Mitchell G. Rothstein and Richard D. Goffin.

* Corresponding author.

E-mail address: mrothstein@ivey.uwo.ca (M.G. Rothstein).

1. Current use of personality measures in personnel selection

Although we can find no reports of research using systematic sampling procedures to determine with any measure of certainty the extent that personality measures are currently being used by organizations as part of their personnel selection practices, a number of surveys of human resource professionals, organizational usage, and industry reports may be combined to provide a reasonably good picture of the degree that such measures are being used. A survey conducted of recruiters in 2003 indicated that 30% of American companies used personality tests to screen job applicants (Heller, 2005). Integrity tests, a particular type of personality assessment, are given to as many as five million job applicants a year (a number that has been growing by 20% a year), and are reported used by 20% of the members of the Society of Human Resource Management (Heller, 2005). Another survey of the Society for Human Resource Management indicated that more than 40% of Fortune 100 companies reported using personality tests for assessing some level of job applicant from front line workers to the CEO (Erickson, 2004). These results seem to indicate a change in attitude among human resource professionals since a survey conducted by Rynes, Colbert, and Brown (2002) in which participants reported more pessimism about the use of personality testing for predicting employee performance. Still another survey indicated that every one of the top 100 companies in Great Britain reported using personality tests as part of their hiring procedure (Faulder, 2005). Beagrie (2005) has estimated that two thirds of medium to large organizations use some type of psychological testing, including aptitude as well as personality, in job applicant screening.

Industry reports are consistent with these surveys indicating increased usage of personality testing. It has been estimated that personality testing is a \$400 million industry in the United States and it is growing at an average of 10% a year (Hsu, 2004). In addition to questions concerning usage of personality testing, numerous surveys have been conducted attempting to determine the reasons for the positive attitude toward personality testing for employment purposes. The most prevalent reason given for using personality testing was their contribution to improving employee fit and reducing turnover by rates as much as 20% (Geller, 2004), 30% (Berta, 2005), 40% (Daniel, 2005), and even 70% (Wagner, 2000). It is of considerable interest that evidence for the validity of personality tests for predicting job performance is rarely cited (see Hoel (2004) for a notable exception) by human resource professionals or recruiters. On the other hand, criticisms of personality testing are often cited in many of the same survey reports, most often with little analysis or understanding of the technical issues or research evidence (e.g., Handler, 2005). For example, the use of the MMPI is often cited for its inability to predict job performance and potential for litigation if used for such purposes (e.g., Heller, 2005; Paul, 2004), despite the fact that this is well known among personality researchers who provide clear guidelines for the proper choice and use of personality tests for employee selection (Daniel, 2005). Thus, it appears that personality testing is clearly increasing in frequency as a component of the personnel selection process, although human resource professionals and recruiters may not entirely appreciate the benefit accrued by this practice nor the complexities of choosing the right test and using it appropriately.

2. Are personality measures valid predictors of job performance? A brief summary of the meta-analytic evidence

The impetus for the numerous meta-analytic studies of personality–job performance relations has most often been based on an influential review of the available research at the time by Guion and Gottier (1965). On the basis of their narrative review, Guion and Gottier concluded that there was little evidence for the validity of personality measures in personnel selection. In the decades following the publication of this paper hundreds of research articles challenged this conclusion and attempted to demonstrate the validity of predicting job performance using a seemingly endless number of personality constructs, a variety of performance criteria, and many diverse jobs and occupations. The first attempt to summarize this literature using meta-analysis was undertaken by Schmitt, Gooding, Noe, and Kirsch (1984). They obtained a mean uncorrected correlation of .15 across all personality traits, performance criteria, and occupations, a finding that led these authors to conclude that personality measures were less valid than other predictors of job performance. By the 1990s however, methodological innovations in meta-analysis and the emergence of a widely accepted taxonomy of personality characteristics, the “five factor model” or FFM (i.e., Extraversion, Agreeableness, Emotional Stability, Conscientiousness, and Openness to Experience), spurred a series of meta-analytic studies that have provided a much more optimistic view of the ability of personality measures to predict job performance.

Two meta-analytic studies of personality-job performance relations have been especially influential (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991). Barrick and Mount (1991) categorized personality measures according to the FFM before examining their validity for predicting job performance in relation to a number of occupational groups and performance criteria. Barrick and Mount found that the estimated true correlation between FFM dimensions of personality and performance across both occupational groups and criterion types ranged from .04 for Openness to Experience to .22 for Conscientiousness. Although correlations in this range may seem relatively modest, nevertheless these results provided a more optimistic view of the potential of personality for predicting job performance and this study had an enormous impact on researchers and practitioners (Mount & Barrick, 1998; Murphy, 1997, 2000). Moreover, correlations of this magnitude can still provide considerable utility to personnel selection decisions (e.g., Cascio, 1991), particularly because the prediction of job performance afforded by personality appears to be incremental to that of other major selection methods (e.g., Goffin, Rothstein, & Johnson, 1996; Schmidt & Hunter, 1998; incremental validity is discussed more fully in a later section). Tett et al.'s meta-analysis of personality–job performance relations had a somewhat different purpose (see Barrick & Mount, 2003) and their main contribution was to highlight the critical importance to validity research of a confirmatory research strategy, in which personality measures were hypothesized a priori to be linked logically or theoretically to specific job performance criteria. Tett et al. determined that validation studies employing a confirmatory research strategy produced validity coefficients that were more than twice as high as studies in which an exploratory strategy was used.

The impact of these meta-analytic studies was partially due to the development of meta-analysis techniques that were better able to cumulate results across studies examining the same relations to estimate the general effect size, while correcting for artifacts such as sampling and measurement errors that typically attenuate results from individual studies. Secondly, these studies provided a clearer understanding of the role of personality in job performance than did previous meta-analyses by examining the effects of personality on different criterion types and in different occupations. Thirdly, the studies benefited from the development of the FFM of personality in which the multitude of personality trait names and scales could be classified effectively into five cogent dimensions that could be more easily understood by researchers and practitioners alike. Thus, results from Barrick and Mount (1991) and Tett et al. (1991) became the foundation for a renewal of interest in both research and practice with respect to the use of personality to predict work-related behavior.

Despite the significant contribution of these groundbreaking studies to understanding personality–job performance relations, it must be acknowledged that they also generated considerable controversy. It is not possible to review here the numerous criticisms and debates that have ensued over the past decade, nor is it necessary given that significant progress has been made toward resolving many of these controversies (e.g., Barrick & Mount, 2003; Rothstein & Jelley, 2003). However, it is necessary to summarize briefly a few of the key issues that have been the focus of much of the controversy in that these issues may inform future use of personality measures in personnel selection by both researchers and practitioners.

At the most fundamental methodological level, the procedure of meta-analysis has itself been much criticized (e.g., Bobko & Stone-Romero, 1998; Hartigan & Wigdor, 1989; Murphy, 1997). In some cases, criticisms of meta-analytic research have been directed at specific applications such as the use of meta-analytic results in police selection (Barrett, Miguel, Hurd, Lueke, & Tan, 2003). For example, Barrett et al. (2003) have argued that selection of personality measures based on meta-analytic findings must ensure that results are based on relevant samples and appropriate tests and performance criteria, especially in the context of police selection. In general, however, the methodological concerns with meta-analysis can be mitigated by a thorough understanding of the technique and its appropriate use. Murphy (2000) has provided an analysis of the key issues to consider to justify making inferences from meta-analyses for research or personnel selection. These issues are (a) the quality of the data base and the quality of the primary studies it contains; (b) whether the studies included in the meta-analysis are representative of the population of potential applications of the predictor; (c) whether a particular test being considered for use is a member of the population of instruments examined in the meta-analysis; and (d) whether the situation intended for use is similar to the situations sampled in the meta-analysis. Although Murphy (2000) points out that many meta-analyses omit such essential information, nevertheless researchers and practitioners have clear guidelines for evaluating meta-analytic results. Assuming that appropriate methodological procedures have been followed, meta-analytic findings are increasingly accepted, especially in the area of personnel selection (Murphy, 2000; Schmidt & Hunter, 2003). A careful consideration of these factors has also been linked to the appropriate use of validity generalization principles for determining the potential value of personality measures as predictors of job performance (Rothstein & Jelley, 2003).

Another controversy important to acknowledge in considering the use of personality measures in personnel selection concerns the role of the FFM of personality. Sixteen meta-analytic studies of personality–job performance relations have been published since 1990 and all have used the FFM of personality in some way or another in their analyses (Barrick & Mount, 2003). Clearly the FFM has facilitated this line of research by providing a taxonomy of personality capable of classifying a huge number, and in many cases a confusing array, of personality trait names into a coherent system of more general but easily understood constructs. However, many researchers have challenged the validity of the FFM as a comprehensive taxonomy of the structure of personality. The most comprehensive critique of the FFM has been provided by Block (1995), but many other critiques have been published in which alternative structures of personality have been proposed based on two factors (Wiggins, 1968), three factors (Eysenck, 1991), six factors (Hogan, 1986), seven factors (Jackson, 1984), eight factors (Hough, 1998a,b), or nine factors (Hough, 1992). In addition, there is a continuing debate on whether or not such “broad” personality dimensions are more or less effective than narrow (i.e., specific traits) personality measures for predicting job performance (see below for a review of this ongoing debate). Once again, it is not possible to review in this context all the controversies and debate surrounding how well the FFM represents the structure of personality. However, for researchers and practitioners interested in the use of personality measures in personnel selection, it is important to recognize that there is more to personality than the FFM. The choice of personality measure to use in a selection context should consider a number of factors, not the least of which is the development of a predictive hypothesis on the relations expected between the personality measure and the performance criterion of interest (Rothstein & Jelley, 2003).

Two other issues made salient by the contribution of meta-analytic studies to understanding personality–job performance research concern the importance of acknowledging the bidirectional nature of many potential personality–job performance relations, and appreciating the potential role of moderators between personality and performance criteria. Regarding the former, Tett et al. (1991) and Tett, Jackson, Rothstein, and Reddon (1994) demonstrated that the nature of many personality constructs is such that negative correlations with performance criteria may be understandable and valid, and that failure to acknowledge this may attenuate results of meta-analyses and limit their use in personnel selection. With respect to the role of moderators, both Barrick and Mount (1991) and Tett et al. (1991) demonstrated that the nature and/or extent of relations between personality and job performance varied significantly depending on a variety of factors. Although Barrick and Mount (1991) are most often cited as demonstrating that Conscientiousness was the best overall predictor of performance across occupations and performance criteria, in fact the contribution of this study is far broader demonstrating that relations between all the FFM dimensions of personality and performance varied according to occupational group and the nature of the performance criterion. Similarly, Tett et al. (1991) demonstrated the critical role of confirmatory versus strictly empirical strategies, and the use of job analysis versus no job analysis, in determining the choice of personality measure to use in selection validation research. The importance of moderators to reveal the full potential of using personality in personnel selection research and practice continues to be an important focus of research since the publication of these meta-analytic studies and will be reviewed further below.

In summary, despite the controversies surrounding meta-analysis and the FFM, the weight of the meta-analytic evidence clearly leads to the conclusion that personality measures may be an important contributor to the prediction of job performance. The impact of these meta-analytic studies has countered the earlier conclusions of Guion and Gottier (1965) and put personality back into research and practice. In the decade or more since these meta-analyses began to be published, research in personality and job performance has continued, creating a wealth of further understanding and implications for the use of personality measures in personnel selection. We next review the important trends in this research, with particular emphasis on the implications for research and practice in human resource management.

3. Current research trends

3.1. *The impact of the FFM on personality–job performance research*

The FFM of personality structure has had a deep impact on personality–job performance research since the series of meta-analytic studies of the 1990s provided support for the use of personality measures in personnel selection. Mount and Barrick (1995) observed that it was the widespread acceptance of the FFM that created much of the optimism for the renewed interest in relations between personality and job performance. “The importance of this taxonomy cannot be overstated as the availability of such a classification scheme provides the long missing conceptual foundation necessary

for scientific advancement in the field” (Mount & Barrick, 1995, p. 190). Goodstein and Lanyon (1999) also credit the FFM for providing a universally accepted set of dimensions for describing human behavior at work and promote their use in organizational settings. Although criticism of the FFM continues, many researchers have accepted it as a reasonable taxonomy of personality characteristics and moved beyond the basic question of whether personality predicts job performance to examine more specific applications (Rothstein & Jelley, 2003). It appears that a considerable amount of new research in this area employs the FFM. Research for the current article involved a computer search of PSYCHINFO and PROQUEST data bases from 1994 to the present and found that of 181 relevant empirical research studies published in this period, 103 (57%) used direct or constructed FFM measures of personality. It is not possible to provide a detailed review of this research here, but following is an analysis of the main trends evident in the continued use of the FFM with respect to the use of these measures in personnel selection as well as some other applications.

3.1.1. Investigations of moderator effects

Although Barrick and Mount (1991) demonstrated that overall the best predictor of job performance across various performance criteria and occupational groups was Conscientiousness, examination of their full meta-analytic results illustrates that the other FFM dimensions varied in their predictive effects depending on the nature of the performance criterion and occupational group. Similarly, Tett et al. (1991) demonstrated that personality–job performance relations were significantly strengthened by the use of a confirmatory research strategy and job analysis. These meta-analytic studies illustrate the importance of moderator effects that underscore the greater potential of personality measures in personnel selection. Investigations of additional moderator variables have continued and provide further insights on how to maximize the predictability of personality measures. For example, Thoresen, Bliese, Bradley, and Thorenson (2004) found that different FFM dimensions predicted pharmaceutical sales depending on the specific nature of the criterion (overall sales versus performance growth) and job stage (maintenance versus transitional). Simmering, Colquitt, Noe, and Porter (2003) determined that Conscientiousness was positively related to employee development, but only when employees felt that the degree of autonomy in their jobs did not fit their needs. The importance of a confirmatory research strategy was reinforced by Nikolaou (2003) who reported that although FFM dimensions were not generally related to overall job performance, Agreeableness and Openness to Experience were related to performance involving interpersonal skills. Hochwarter, Witt, and Kacmar (2000) determined that Conscientiousness was related to performance when employees perceived high levels of organizational politics, but no relations were found among employees perceiving low levels of organizational politics. Barrick and Mount (1993) found that Conscientiousness and Extraversion predicted managerial performance significantly better in jobs categorized as high in autonomy. Finally, it seems that one personality measure may moderate the effects of another. In a study reported by Gellatly and Irving (2001), autonomy moderated the relationships between other personality traits with the contextual performance of managers. In another study of this type (Witt, 2002), Extraversion was related to job performance when employees were also high in Conscientiousness, but with employees low in Conscientiousness, Extraversion was negatively related to performance.

In our view, research investigating moderator effects in personality–job performance relations continue to support one of the main conclusions from Tett et al. (1991), that relations between personality measures and job performance criteria are substantially more likely to be found when a confirmatory research strategy is used. As Rothstein and Jelly (2003) have argued, personality measures are relatively more situationally specific, compared with a measure of general mental ability. This makes the use of validity generalization principles to justify the use of a personality measure in selection more challenging because there may be numerous situational moderators as the above research illustrates. For human resource researchers and practitioners in personnel selection, the key is careful alignment of personality and performance criteria as well as consideration of other potential contextual factors related to the job or organization.

3.1.2. Investigations of mediator effects

Another potential interpretation of the relatively low correlations typically found between personality measures and job performance criteria, in addition to unknown or unmeasured moderator effects, is that personality may only have indirect effects on performance and that there may be stronger relations with mediator variables that in turn are more strongly related to job performance (Rothstein & Jelley, 2003). The logic of this proposition is based on the generally accepted definition of personality as a predisposition to certain types of behavior. Accordingly, if this behavior could be

measured directly, such measures may mediate relations between personality and job performance. Only a small number of research studies have been conducted over the past decade, but results support the existence of mediator effects. For example, [Barrick, Mount, and Strauss \(1993\)](#) found that goal setting behaviors mediated relations between Conscientiousness and job performance in a sample of sales representatives. [Gellatly \(1996\)](#) also examined goal setting behavior in the context of a laboratory task and reported that cognitive processes underlying performance expectancy and goal choice mediated relations between Conscientiousness and task performance. Finally, in another sample of sales representatives, [Barrick, Steward, and Piotrowski \(2002\)](#) determined that measures of cognitive-motivational work orientation, namely striving for success and accomplishment, mediated relations between both Extraversion and Conscientiousness on sales performance.

Collectively these studies illustrate once again that a confirmatory research strategy provides valuable insights to the nature of personality–job performance relations. Such strategies contribute to more comprehensive predictive models and better understanding of how personality affects job performance directly and indirectly. Although relatively few studies of mediator effects have been reported in the literature thus far, existing research indicates that both research and practice in personnel selection would benefit from such studies. Discovering indirect effects of personality on job performance through mediator variables may also help to understand why so many personality–job performance relations are situationally specific which in turn would lead to more effective personnel selection practices.

3.1.3. Investigations of incremental validity

Very few studies on the incremental validity of personality measures over other predictors of job performance have been reported in the research literature in the past decade. In our view, this is unfortunate in that an early study of this phenomenon ([Day & Silverman, 1989](#)) has often been cited as representative of the potential unique contribution of personality measures to the prediction of job performance over other predictors ([Tett et al., 1991, 1994](#)). Although repeated meta-analyses have supported the conclusion that personality predicts job performance ([Barrick & Mount, 2003](#)), from the perspective of human resource researchers and practitioners an important question remaining is to what degree is this prediction incremental in validity and value over other personnel selection techniques. However, in our computer search for relevant research to review, we could find only two empirical studies in the past decade that explicitly examined the incremental validity question. In one study, [Mc Manus and Kelly \(1999\)](#) found that FFM measures of personality provided incremental validity over biodata measures in predicting job performance. A second study demonstrated that personality data provided incremental validity over evaluations of managerial potential provided by an assessment center ([Goffin et al., 1996](#)). Clearly more research is called for in this vital area in order to determine the real potential value of personality measures used in personnel selection. [Schmidt and Hunter \(1998\)](#) provide some optimism in this regard. In a study combining meta-analysis with structural equation modeling, it was estimated that Conscientiousness added significant incremental validity over general mental ability for most jobs. Additional specific studies of the incremental validity of personality are needed to demonstrate that personnel selection practices would benefit from including relevant measures of personality to the assessment of job applicants.

3.1.4. More focused and specific meta-analytic studies

In a recent comprehensive review of meta-analytic studies of personality–job performance relations, [Barrick and Mount \(2003\)](#) observed that the 16 meta-analyses of relations between job performance and FFM personality dimensions conducted in the decade after their 1991 publication, with the exception of some differences in purpose and methodology, produced quite similar conclusions regarding generalizable relations between FFM dimensions and job performance. Furthermore, they conclude that “the point now has been reached where there is no need for future meta-analyses of this type, as they are likely to result in quite similar findings and conclusions” ([Barrick & Mount, 2003](#), p. 208). Apparently, not all researchers in this field are ready to heed Barrick and Mount’s advice. Meta-analyses involving FFM dimensions of personality have continued, albeit they have tended to be focused on more specific issues and unique criterion relations. For example, [Clarke and Robertson \(2005\)](#) conducted a meta-analytic study of the FFM personality constructs and accident involvement in occupational and non-occupational settings. They found that Conscientiousness and Agreeableness were negatively correlated with accident involvement. [Judge, Heller, and Mount \(2002\)](#) examined relations between FFM constructs and job satisfaction. Although they found mean correlations with four of the five FFM factors in the same range as previous meta-analyses with performance criteria, only relations with Neuroticism and Extraversion generalized across studies. [Mol, Born, Willemsen, and Van Der Molen \(2005\)](#) investigated relations between expatriate job performance and FFM personality dimensions and found that

Extraversion, Emotional Stability, Agreeableness, and Conscientiousness were all related with validities in the same range as have been reported with domestic job performance criteria. Judge, Bono, Ilies, and Gerhardt (2002) determined that Neuroticism, Extraversion, Openness to Experience, and Conscientiousness were all related to leadership criteria (leader emergence and leader effectiveness) with Extraversion being the most consistent predictor across studies. Judge and Ilies (2002) examined relations between FFM personality constructs and measures of performance motivation derived from three theories (goal setting, expectancy, and self-efficacy motivation). Results indicated that Neuroticism, Extraversion, and Conscientiousness correlated with performance motivation across studies. In a study that strongly supports conclusions drawn by Tett et al. (1991), Hogan and Holland (2003) aligned FFM personality predictors with specific related performance criteria and found that personality predicted relevant performance criterion variables substantially better than was the case when more general criterion variables were used. Finally, it appears that Barrick and Mount also have an interest in continuing to use meta-analysis to examine more specific criterion relations with the FFM personality dimensions. These authors investigated relations between FFM dimensions and Holland's occupational types and determined that although there were some modest correlations among the two sets of constructs (the strongest relations observed were between Holland's types of enterprising and artistic and the FFM factors of Extraversion and Openness to Experience), by and large it was concluded that the two theoretical approaches to classifying individual differences were relatively distinct (Barrick, Mount, & Gupta, 2003).

It is clear from these continuing meta-analytic studies involving the FFM personality dimensions that the FFM has provided an organizing framework for examining relations between personality and a growing number of work-related variables of interest in addition to job performance. The pattern emerging from these studies is that personality, organized around the FFM, has far ranging effects on an organization beyond relations with job performance. The implication for researchers and practitioners in human resource management is that the assessment of applicant personality in a personnel selection context may provide organizations with predictive information on the likelihood that applicants may be involved in an accident, are likely to be satisfied with their job, will be motivated to perform, and will develop into leaders. Thus, continuing meta-analytic studies of personality organized around the FFM are providing a growing number of valuable implications for personnel selection and more generally, human resource management.

3.1.5. Future trends: unique applications and criterion relations

In addition to the above categories of research involving the FFM of personality and job performance, a wide variety of individual studies have been published in the past decade that do not fall neatly into these categories. There are patterns beginning to appear with some of these studies and undoubtedly these patterns will be the subject of future meta-analyses. For now, however, they may be seen as signs of future trends in personality–job performance research using the FFM of personality. Some of these studies that have appeared over the past decade have already been evaluated in meta-analyses as noted above (e.g., job satisfaction, accident proneness, expatriate performance, leadership), but others may signal emerging trends that if upheld with additional research will have useful implications for human resource researchers and practitioners.

By far the biggest trend in continuing research with the FFM is the search for relations with unique types of criterion measures that although are certainly work-related, are not standard performance criteria. For example, Burke and Witt (2004) found that high Conscientiousness and low Agreeableness were related to high-maintenance employee behavior, defined as chronic and annoying behaviors in the workplace. Cable and Judge (2003) investigated relations between the FFM and upward influence tactic strategies. They reported that Extraversion was related to the use of inspirational appeal and ingratiation, Openness to Experience was related to low use of coalitions, Emotional Stability was related to the use of rational persuasion and low use of inspirational appeal, Agreeableness was related to low use of legitimization or pressure, and Conscientiousness was related to the use of rational appeal. Williams (2004) examined the relation between Openness to Experience and individual creativity in organizations and found that this FFM factor was significantly related to creative performance. Ployhart, Lim, and Chan (2001) distinguished between typical and maximum performance based on ratings from multiple sources and determined that Extraversion was related to both types of performance, but Openness to Experience was the best predictor of maximum performance whereas Neuroticism was the best predictor of typical performance. O'Connell, Doverspike, Norris-Watts, and Hatstrup (2001) reported a significant correlation between Conscientiousness and organizational citizenship behaviors. Lin, Chiu, and Hsieh (2001) investigated relations between the FFM and customer ratings of service quality. They reported significant relations between Openness to Experience and assurance behaviors, Conscientiousness and reliability,

Extraversion and responsiveness, and Agreeableness with both empathy and assurance behaviors. Finally, [LePine and Van Dyne \(2001\)](#) found that Conscientiousness, Extraversion, and Agreeableness were related more strongly to change-oriented communications and cooperative behavior than to task performance.

A second clear trend in FFM research involves exploring linkages with career related issues. For example, [Boudreau, Boswell, Judge, and Bretz \(2001\)](#) found that Agreeableness, Neuroticism, and Openness to Experience were all related positively to job search behaviors over and above situational factors previously shown to affect such behavior. [Judge, Higgins, Thoresen, and Barrick \(1999\)](#) examined relations between the FFM dimensions and career success. In this study Conscientiousness was positively related to both intrinsic (i.e., job satisfaction) and extrinsic (i.e., income and occupational status) career success and Neuroticism was negatively related to extrinsic career success. In another study of job search behavior, [Judge and Cable \(1997\)](#) found a pattern of hypothesized relations between FFM personality constructs and job seekers' organizational culture preferences.

A final potential trend in recent FFM research involves investigations of relations with training effectiveness criteria. [Bartram \(1995\)](#) reported a study in which Emotional Stability and Extraversion were associated with success at military flying training. [Lievens, Harris, Van Keer, and Bisqueret \(2003\)](#) found that Openness to Experience was significantly related to cross cultural training performance in a sample of European expatriate managers.

These emerging trends in FFM research in conjunction with important work-related outcomes suggest important new implications for human resource research and practice. Although it is too early to implement some of these innovative uses of FFM personality measures without additional research, some interesting opportunities are suggested by these recent studies. Most obviously, the discovery of relations between FFM constructs and specific or unique performance criteria open up new opportunities in hiring practices. There are also implications for improving training success and career decisions. As stated at the outset of this section, the FFM of personality structure has been shown to be having a major impact on personality–job performance research and personnel selection practices.

3.2. Are broad or narrow personality measures better for personnel selection?

Although it is clear from the above review that the weight of the meta-analytic evidence over more than a decade of such studies supports the use of personality measures for predicting job performance, and that this research has spawned a growing interest in the FFM of personality as a basis for continuing research, an additional outgrowth of all of this research activity has been a spirited debate on the relative usefulness of broad (e.g., the FFM) versus narrow (i.e., more specific) measures of personality in predicting job performance. Beyond the theoretical and methodological issues raised in this debate, there are important implications for human resource researchers and practitioners in terms of determining the best personality measures to use in a particular selection context. [Rothstein and Jelly \(2003\)](#) have argued that unlike measures of general mental ability, principles of validity generalization are much more complicated to apply to personality measures. What then, are the main issues of relevance regarding the choice of broad versus narrow personality measures for use in personnel selection contexts?

3.2.1. A brief review of the debate

The genesis of the debate on the relative merits of broad versus narrow measures of personality for predicting job performance stemmed from two of the meta-analyses of personality–job performance relations in which somewhat different results were obtained with regard to effects reported for the FFM ([Barrick & Mount, 1991](#); [Tett et al., 1991](#)). Although recently it has been acknowledged by participants on both sides of the debate that the primary purposes of these two meta-analyses were fundamentally different, and that the FFM analysis reported by [Tett et al. \(1991\)](#) was tertiary to their main focus ([Barrick & Mount, 2003](#); [Rothstein & Jelly, 2003](#)), nevertheless these meta-analytic findings initially created a good deal of controversy. The theoretical and methodological issues underlying this debate have been well documented elsewhere. Interested readers may wish to consult the original debate (i.e., [Ones, Mount, Barrick, & Hunter, 1994](#); [Tett et al., 1994](#)) or a subsequent debate initiated by [Ones and Viswesvaran \(1996\)](#) which provoked a number of responses ([Ashton, 1998](#); [Hogan & Roberts, 1996](#); [Paunonen, Rothstein, & Jackson, 1999](#); [Schneider, Hough, & Dunnette, 1996](#)). It is noteworthy however, that in two recent evaluations of these debates, very similar conclusions were reached. [Barrick and Mount \(2003\)](#) characterized the controversy as a debate over the appropriate level of analysis in determining personality–job performance relations, and that the appropriate level will depend on the purpose of the particular prediction context. They concluded that "...a broader, more comprehensive measure is appropriate for predicting an equally broad measure of overall success at work...In contrast, if the purpose is

to enhance understanding, linking specific, lower level facets of FFM constructs to specific, lower level criteria may result in stronger correlations” (Barrick & Mount, 2003, p. 213). Similarly, Rothstein and Jelly (2003) concluded that “there is no compelling evidence that either broad or narrow personality measures are preferable for predicting job performance. Indeed, the evidence reviewed suggests both may be useful under certain circumstances” (p. 246). For Rothstein and Jelly (2003) however, these “circumstances” go beyond matching the appropriate level of analysis between predictor and criterion measures, contending that “...personality measures in selection research should be chosen on the basis of a priori hypotheses regarding their potential relations with appropriate constructs in the job performance domain” (p. 248).

Thus, although there has been vigorous debate on the relative merits of using broad versus narrow personality measures in personnel selection, over the past decade a consensus is growing among researchers in this field that both broad and narrow personality measures may be effective predictors of job performance under the appropriate conditions. This growing consensus has not however, deterred researchers from continuing to compare the effectiveness of broad versus narrow personality predictors, or from investigating unique applications and criterion relations with narrow traits.

3.2.2. *Recent trends in research examining broad and narrow personality predictors*

In a recent discussion of the relations between broad dimensions of personality and job performance, Murphy and Dzieweczynski (2005) point out that the extensive literature on the FFM and job performance has generally produced correlations of very low magnitude. They concluded that “...correlations between measures of the Big Five personality dimensions and measures of job performance are generally quite close to zero” and that “...the Big Five almost always turns out to be fairly poor predictors of performance” (Murphy & Dzieweczynski, 2005, p. 345). These authors further propose that there are three main reasons for why broad measures are such poor predictors of job performance: the absence of theory linking personality to job performance, the difficulty in matching personality to relevant job performance criteria, and the poor quality of so many personality measures. The latter reason is a perpetual problem in personality assessment (Goffin, Rothstein, & Johnston, 2000; Jackson & Rothstein, 1993), and the two other reasons echo Tett et al.’s (1991) meta-analytic findings in which specific (narrow) personality traits were found to predict job performance substantially better when a priori hypotheses, particularly when aided by job analyses, guided the choice of personality predictor. Thus, despite the growing use of the FFM of personality in research predicting job performance reviewed earlier in this paper, it is clear that not all researchers have accepted the FFM as the best measures to use in this research. Specifically, we stated earlier that in our computer search of relevant empirical research on personality–job performance relations published since 1994, we found that 57% used direct or constructed FFM measures of personality. Left unsaid earlier was that the other 43% of new empirical research over this time period has continued to investigate the use of narrow or non-FFM personality traits to predict job performance, with many of these studies designed to demonstrate the incremental validity of narrow traits relative to broad dimensions of personality. It is instructive to briefly review this research.

We can identify four main trends in personality–job performance research in recent years in which narrow measures of personality were of primary interest. First, there have been several studies of the factor structure of broad personality dimensions attempting to identify the narrow facets that comprise these dimensions and compare their validities. For example, Roberts, Chernyshenko, Stark, and Goldberg (2005) factor analyzed 36 scales related to Conscientiousness and determined that six factors underlie this broad dimension. Further, they found that these six facets of Conscientiousness had differential predictive validity with various criteria, and they demonstrated incremental validity over the broad general dimension. Griffin and Hesketh (2004) used factor analysis to determine that two main facets underlay the FFM dimension of Openness to Experience and these two facets were differentially related to job performance. Similarly, Van Iddekinge, Taylor, and Eidson (2005) found eight facets underlying the broad dimension of Integrity and correlations between these facets and job performance ranged from $-.16$ to $.18$. Two of these facets had stronger relations with performance than did the broad dimension of Integrity. Studies of this type continue to challenge the effectiveness of broad personality dimensions for personnel selection, at least with the specific predictors and criteria compared in these studies.

A second trend quite evident in the research on narrow traits is the explicit evaluation of the relative effectiveness of broad versus narrow traits in predicting job performance. Judging from the empirical studies published since the debate began, narrow traits are clearly out performing broad dimensions of personality. Of the eleven studies published in the last decade on this topic and identified in our computer search, all have demonstrated that narrow traits are either better

predictors of job performance than broad dimensions of personality and/or add significant incremental validity over broad dimensions. These studies include comparisons between broad dimensions and the facets that comprise them (e.g., Jenkins & Griffith, 2004; Stewart, 1999; Tett, Steele, & Beauregard, 2003; Vinchur, Schippmann, Switzer, & Roth, 1998) as well as comparisons between specific traits hypothesized to be more closely linked conceptually to a particular performance criterion than the FFM dimensions (e.g., Ashton, 1998; Conte & Gintoft, 2005; Crant, 1995; Lounsbury, Gibson, & Hamrick, 2004).

A third trend that became obvious in the current review is that research investigating relations between many different narrow traits and a wide variety of job performance criteria has continued and not been curtailed by the many meta-analytic reviews that have attempted to summarize previous years of research exclusively in terms of the FFM. There are too many such studies to review here and the range of predictor and criterion variables is too broad to distinguish patterns at this time. Undoubtedly these studies and more will be the subject of future meta-analyses at which time these patterns will be made more salient. At this point however, it may be concluded that a strong interest remains in examining relations between more specific, narrow personality traits and job performance.

One final very recent study is worth mentioning. Both Barrick and Mount (2003) and Rothstein and Jelly (2003) concluded their commentary on the broad versus narrow debate by indicating that both types of personality measures may be effective predictors of job performance under the appropriate conditions. A recent empirical study supports this conclusion. Warr, Bartram, and Martin (2005) found that both the narrow traits of Achievement and Potency, and the broad dimension of (low) Agreeableness were related to different dimensions of sales performance as hypothesized.

It seems, therefore, that for human resource researchers and practitioners the implications of this discussion are straightforward. If both narrow and broad personality measures have the potential to predict job performance, how is this potential realized? The weight of the meta-analytic and more recent empirical evidence is that theoretical or conceptual relations between the personality predictor (regardless of broad or narrow) and criterion of interest should be well understood and articulated. Generally, broader criterion measures may likely fit broader personality measures, although the magnitude of the correlation will likely be low. More specific criteria may be a better fit with narrow personality traits and the magnitude of the correlation can be expected to be larger. However, if there is a sound theoretical or conceptual case for expecting any particular personality construct to be related to a particular performance criterion measure, this would be more important than how broad or narrow the personality measure or criterion is.

3.3. *Personality and team performance*

The study of the impact of personality on team behavior and performance is another area of research that has seen renewed activity in recent years and it is clear that this activity is also a direct result of the meta-analyses conducted during this time period, particularly those focused on the FFM of personality. The study of individual differences in group behavior has a long history, although the individual difference variables in this research have certainly not been confined to personality (Guzzo & Shea, 1992). Moreover, research examining personality linkages to group effectiveness has not produced conclusive results (Driskell, Hogan, & Salas, 1988). One major reason for this has been that the personality variables of interest in these studies may be characterized in much the same way that personality measures had been characterized in personality–job performance research prior to the development of the FFM and the contribution of meta-analysis, that is, a large number of poorly defined traits that could not easily be accumulated into a coherent body of knowledge (Barrick & Mount, 2003; Driskell et al., 1988; Neuman, Wagner, & Christiansen, 1999). However, the FFM has had a similar strong impact on the study of group behavior as it had on personality–job performance research. Of the 16 empirical studies on the role of personality in group behavior/performance conducted over the past decade and obtained in our computer search, 15 involved FFM constructs, and the sixteenth involved Integrity, another very broad personality based measure. In addition, the context of the study of group behavior has shifted a great deal into “team” behavior and performance in the workplace. Organizations have embraced work teams as a critical tool for achieving performance goals in the context of the necessity to respond to increased global competition and rapid technological change (Hackman, 1986; Kichuk & Weisner, 1998; Neuman et al., 1999; Peters, 1988). In many cases teams have changed the fundamental way that work is structured (Kichuk & Weisner, 1998; Tornatsky, 1986). Thus, given the value of teams to organization performance, it is not surprising that research on team effectiveness has received renewed interest, and the importance of selecting effective team members is a major component of this research effort (Baker & Salas, 1994). What then, has been the

contribution of investigations examining the effect of team members' personality on team effectiveness and performance?

As previously mentioned, to review progress made by research over the past decade on relations between personality and team performance is essentially to review the contribution of the FFM in this area, since 15 of 16 studies found for this review involved FFM constructs. The one anomaly among these studies involved a measure of Integrity, although strictly speaking this was not a study of team performance in that although participants in the study were team members, their Integrity scores were correlated with their personal job performance ratings made by their team leaders (Luther, 2000). The other 15 studies could be characterized in many of the same ways as the previous discussion of the general contributions of the FFM to personality–job performance research. There have been investigations of direct prediction of team performance by FFM constructs, mediation effects, studies of incremental validity of the FFM, and studies involving unique criterion measures other than team performance. The small number of studies and the diversity of the criteria that were used precluded defensible meta-analysis of these findings so we offer a brief narrative summary of the findings by FFM dimension.

Overall, Extraversion appears to be the best predictor of team-related behavior and performance. Eleven of the 15 published studies reported significant correlations between Extraversion and various measures including team performance (Barrick, Stewart, Neubert, & Mount, 1998; Barry & Stewart, 1997; Kichuk & Weisner, 1997; Morgeson, Reider, & Campion, 2005; Neuman et al., 1999); group interaction styles (Balthazard, Potter, & Warren, 2004), oral communication (Mohammed & Angell, 2003), emergent leadership (Kickul & Neuman, 2000; Taggar, Hackett, & Saha, 1999), task role behavior (Stewart, Fulmer, & Barrick, 2005), and leadership task performance (Mohammed, Mathieu, & Bartlett, 2002).

Conscientiousness and Emotional Stability are the two other FFM constructs found to be generally good predictors of team-related behavior and performance. Conscientiousness was correlated with team-based performance criteria in eight of the 15 published studies, whereas Emotional Stability was correlated with nine such criteria. Conscientiousness was significantly related to team performance (Barrick et al., 1998; Halfhill, Nielsen, Sundstrom, & Weilbaeher, 2005; Kickul & Neuman, 2000; Morgeson et al., 2005; Neuman et al., 1999; Neuman & Wright, 1999), leadership emergence (Taggar et al., 1999), and task role behavior (Stewart et al., 2005). Emotional Stability was significantly related to team performance (Barrick et al., 1998; Kichuk & Wiesner, 1997; Neuman et al., 1999), ratings of transformational leadership (Lim & Ployhart, 2004), oral communications (Mohammed & Angell, 2003), leadership emergence (Taggar et al., 1999), task role behavior (Stewart et al., 2005), task focus (Bond & Ng, 2004), and leadership task performance (Mohammed et al., 2002).

The remaining two FFM constructs showed poor and/or mixed results with respect to predicting team-related behavior or performance. Openness to Experience was only correlated with team-based performance criteria in three of the 15 published studies, but two of these correlations were positive and one was negative. Agreeableness was correlated with nine team-based performance criteria, but six of these correlations were positive and three negative. Thus, results based on these two dimensions of the FFM appear more unreliable at this time and until more research is conducted to the point where a meta-analysis may be performed to determine if a more consistent picture emerges, no conclusions can be formulated regarding the effectiveness of these two FFM constructs in predicting team behavior or performance. On the other hand, the three FFM constructs of Extraversion, Emotional Stability, and Conscientiousness all show patterns of significant relations with relevant team-based performance criteria that suggest that these personality dimensions have potential for contributing to our understanding of team behavior and performance. However, for researchers and practitioners in human resource management, recommendations for the use of these FFM constructs for selecting team members must be cautious. As discussed previously with respect to the use of FFM measures in predicting individual job performance, the presence of numerous criteria defining team process behavior and performance once again indicates the importance of aligning personality constructs with specific performance measures if the FFM is to contribute to a better understanding of team performance and the selection of more effective team members.

4. Research on faking and personality assessment: cause for optimism

As discussed, in the early 1990s it was established that personality tests are valid predictors of job performance (Barrick & Mount, 1991; Tett et al., 1991). Since that time, it is arguable that the most pervasive concern HR practitioners have had regarding the use of personality testing in personnel selection is that applicants may strategically

“fake” their responses and thereby gravely reduce the usefulness of personality scores (e.g., Christiansen, Burns, & Montgomery, 2005; Goffin & Christiansen, 2003; Holden & Hibbs, 1995; Luther & Thornton, 1999; Ones & Viswesvaran, 1998; Rothstein & Goffin, 2000). Accordingly, research that effectively addressed the issue of faking or “motivated distortion” was called for and the scientific community responded with a gigantic body of studies. We submit that the resulting increase in knowledge of the effects of faking and possible “cures” has been instrumental in the continuing growth of personality assessment in personnel selection. Two main trends can be identified in the faking research and both, ultimately, have provided researchers and practitioners in human resource management with grounds for optimism. In the following section we summarize research on the effects of faking on personality testing within personnel selection contexts. We then review research on suggested approaches for contending with faking.

4.1. *Effects of faking*

Numerous primary studies conducted within simulated or actual personnel selection scenarios (e.g., Furnham, 1990; Goffin & Woods, 1995; Hough, 1998a,b; Jackson, Wroblewski, & Ashton, 2000; Mueller-Hanson, Hegstad, & Thornton, 2003; Rosse, Stecher, Miller, & Levin, 1998; Zalinski & Abrahams, 1979), and a meta-analysis on faking in a variety of contexts (Viswesvaran & Ones, 1999), have converged on the conclusion that test-takers in laboratory situations as well as applicants in applied selection situations can, and do, deliberately increase their scores on desirable personality traits, and decrease their scores on undesirable traits when motivated to present themselves in a positive light. Similarly, a unique survey of recent job applicants that used the randomized response technique (Fox & Tracy, 1986) in order to provide assurances of anonymity found that the base rate of faked responses to the types of items typically comprised by personality tests ranged from 15% to 62% of the sample depending on the nature of the item (Donovan, Dwight, & Hurtz, 2003). Interestingly, the highest rate of faking was for negatively-keyed items that engendered the downplaying of undesirable characteristics.

If faking were uniform among applicants it would have the effect of merely adding (or subtracting) a constant to (or from) everyone’s score, which would mean that candidate rank-ordering, criterion-related validity (i.e., the extent to which personality test scores are related to job performance), and hiring decisions based on personality scores would be unaffected. Unfortunately, this seems not to be the case. The results of several studies suggest that individuals differ in the extent to which they dissimulate (e.g., Donovan et al., 2003; Mueller-Hanson et al., 2003; Pannone, 1984; Rosse et al., 1998). Relatedly, a number of studies have shown that induced faking is associated with a reduction in criterion-related validity (Holden & Jackson, 1981; Jackson et al., 2000; Mueller-Hanson et al., 2003; Topping & O’Gorman, 1997; Worthington & Schlottmann, 1986). Also, Hough’s (1997) comprehensive analysis of criterion-related validities from applied studies found that validities from incumbent samples (wherein the motivation to fake is *not* maximized), were, on average, .07 higher than the respective values from applicant samples (wherein the motivation to distort is likely to be higher). Even in the absence of large effects on criterion-related validity, there is reason to believe that persons who have dissimulated the most may have an increased probability of being hired, resulting in less accurate and less equitable hiring decisions (Christiansen, Goffin, Johnston, & Rothstein, 1994; Mueller-Hanson et al., 2003; Rosse et al., 1998).

Notwithstanding the research just reviewed, there is also abundant grounds for optimism that the usefulness of personality testing in personnel selection is not *neutralized* by faking (e.g., Hogan, 2005; Hough, 1998a,b; Hough & Furnham, 2003; Hough & Ones, 2002; Marcus, 2003). Numerous meta-analyses and large-scale primary studies of personality testing in personnel selection have consistently shown that personality tests have utile levels of criterion-validity even when used in true personnel selection contexts where motivated distortion is very likely to have occurred (e.g., Barrick & Mount, 1991; Goffin et al., 2000; Hough, 1997, 1998a,b; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Tett et al., 1991). Nonetheless, the research reviewed earlier suggests that the usefulness of personality testing in selection may fall short of its full potential as a result of faking. Accordingly, in the next sections we consider recent research on possible strategies for contending with faking, followed by a section that considers faking remedies in light of the underlying psychological processes that may be responsible for their effects.

4.2. *Strategies for contending with faking*

4.2.1. *“Correcting” for faking*

The robust finding that items associated with socially desirable responding are sensitive to “fake good” instructions (e.g., Cattell, Eber, & Tatsuoka, 1970; Goffin & Woods, 1995; Paulhus, 1991; Viswesvaran & Ones,

1999) has led many test publishers to include scales composed of these items in their personality inventories along with the advice that elevated scores on these scales may be indicative of dissimulation (see Goffin & Christiansen, 2003, for a review). Based on the assumption that social desirability may suppress valid trait variance, some test publishers go further and recommend a “correction” for faking that statistically removes the effects of social desirability from candidates’ personality test scores. Until relatively recently (Christiansen et al., 1994), the underlying assumption that such faking “corrections” would improve the criterion-related validity of personality assessment in personnel selection was not tested. However, there is now considerable evidence that faking “corrections” generally do *not* improve validity and that elevated scores on typical social desirability scales may be more a function of valid personality differences than the motivation to fake (e.g., Barrick & Mount, 1996; Christiansen et al., 1994; Ellingson, Sackett, & Hough, 1999; Hough, 1998a,b; McCrae & Costa, 1983; Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Reiss, 1996). Thus, in spite of the fact that 69% of experienced personality test users favored the use of faking corrections in a recent survey (Goffin & Christiansen, 2003), this “remedy” has been contraindicated by considerable empirical research.

One new, possible exception to the accumulation of negative findings regarding “corrections” for faking is Hakstian and Ng’s (2005) development and application of the employment-related motivation distortion (EMD) index. Unlike many social desirability scales, the EMD was designed to capture motivated distortion that is specific to personnel selection contexts, and there is some evidence that personality score corrections based on this index may have higher criterion-related validity (Hakstian & Ng, 2005). We suggest that the EMD itself and the methodology utilized by Hakstian and Ng in its development warrant serious consideration by researchers and practitioners. However, at this early stage it would be premature to suggest that the EMD has solved the problem of “correcting” for faking. A further new development that is worthy of consideration is the operationalization of socially desirable responding as a four-dimensional construct (Paulhus, 2002). Whereas earlier unidimensional and bidimensional operationalizations of social desirability (see Helmes, 2000; Paulhus, 1991 for reviews) have been shown not to improve validity when used in faking corrections (see the research reviewed earlier), the usefulness of the four-dimensional approach has, to our knowledge, not yet been assessed in this regard.

Ultimately, we feel that the difficulties encountered thus far in trying to adequately correct for faking reflect the fact that faking is an intricate process with multiple determinants (e.g., McFarland & Ryan, 2000; Snell, Sydell, & Lueke, 1999). Faking may be manifested in substantially different response patterns depending on the individual differences of the test-takers and their perceptions of the nature of the job they are applying for (e.g., McFarland & Ryan, 2000; Norman, 1963). Therefore, it may not be feasible to develop a single universal faking scale for on which to base score corrections, but the development of multidimensional indices (e.g., Paulhus, 2002), or indices tailored to more specific types of faking (e.g., Hakstian & Ng, 2005) may have value.

4.2.2. *The faking warning*

The faking warning typically comprises a warning to test-takers that advanced, proprietary approaches exist for detecting faking on the personality test that is being used. It may also include the information that as a consequence of faked responses, one’s chances of being hired may be lowered (Dwight & Donovan, 2003; Goffin & Woods, 1995; Rothstein & Goffin, 2000). Rothstein and Goffin reviewed the results of five studies on the faking warning and were led to the conclusion that it had considerable promise for the reduction of faking. Dwight and Donovan meta-analyzed the results of 15 studies, not including three of the studies reviewed by Rothstein and Goffin, and were similarly sanguine as to the benefits of the faking warning, showing that it may reduce faking by 30% on average with larger reductions accompanying warnings that include mention of the consequences of faking detection. Additionally, in their own primary study, Dwight and Donovan provided evidence that the faking warning might improve the accuracy of hiring decisions. Overall, the extant research clearly supports the faking warning as a viable approach to reducing, although not completely eliminating, faking (Dwight & Donovan, 2003; Goffin & Woods, 1995). Also in its favor, the faking warning is inexpensive to add to a selection testing program and can readily be combined with other approaches to faking reduction. Additional research is required to determine whether different strengths of the faking warning are differentially effective. That is, the alleged likelihood of faking being detected and sanctioned could be varied in the warning and studied in relation to the effects on faking suppression.

We also urge researchers to further consider incorporating the “threat of verification” in the faking warning. There is an accumulation of evidence from different sources suggesting that applicants may respond more honestly when they believe their responses will be subject to verification (e.g., Becker & Colquitt, 1992; Donovan et al., 2003; Schmitt &

Kunce, 2002). Thus, in addition to the typical faking warning, and similar to the approach used in Vasilopoulos, Cucina, and McElreath (2005), applicants could be told that *one* of the means of assessing whether faking may have occurred will be to compare the pattern of preferences, work styles, et cetera, indicated in their responses to the personality scale, to the information they have already provided in their resume, and to the impressions conveyed by their references and others. The fact that carefully developed letters of reference may provide valid assessment of personality (e.g., McCarthy & Goffin, 2001) removes this part of the warning from the realm of deception and therefore has the potential to reduce ethical concerns with the faking warning (discussed later). Of course, we would expect that this modified warning would be most effective (a) if personality assessment takes place *after* resumes have been submitted and references have been sought out; and (b), if a substantial percentage of the items on the chosen personality test refer to potentially observable manifestations of the respective traits (e.g., extraversion items often inquire as to one's tendency to assume leadership roles). Vasilopoulos et al. (2005) presented some evidence that the threat of verification in the context of a faking warning may reduce faking. Interestingly, these researchers also showed that the threat of verification tends to increase the "cognitive loading" of personality trait scores. In this context, "cognitive loading" refers to the extent to which cognitive ability (general intelligence) is assessed by the personality test in addition to the personality traits of interest.

In addition to the content of the faking warning itself, logically, the nature of the test-taking conditions may influence the credibility of the warning. In particular, it seems likely that the greater technological sophistication of internet administration, as opposed to paper-and-pencil administration, would strengthen respondents' belief of the faking warning, thereby increasing its potency.

By way of caveats, the potential for the faking warning to reduce the validity of personality scores as a result of test-takers trying too hard to appear as though they are not faking should be investigated (Dwight & Donovan, 2003), as should the earlier-discussed concern that the faking warning might increase the cognitive loading of trait scores. Cognitive loading may have implications with respect to validity because a given personality test score might be, to some extent, indicative of the test-taker's level of cognitive ability as well as his/her personality. This would tend to decrease the ability of the personality test to predict job performance above and beyond cognitive ability. A further consequence of cognitive loading is that personality test scores might have an increased potential to contribute to adverse impact against minority groups. Also, as explained by Goffin and Woods (1995), ethical issues surrounding the use of the faking warning deserve further consideration. Even if faking were completely eliminated by the warning and validity were unequivocally proven to increase, is it a breach of professional ethics for a testing professional to tell job applicants that faking can be detected if, in fact, it cannot? Perhaps the appropriate pairing of the faking warning with approaches that show promise for faking detection provides an answer to this dilemma.

4.2.3. Faking detection

We are heartened that the science of faking detection has made progress on three fronts. First, as already discussed in the "Correcting" for faking section, Hakstian and Ng (2005) as well as Paulhus (2002) have derived improved scales that may be useful in faking detection. Second, collectively, a number of studies have shown that sophisticated measurement and application of the test-taker's latency in responding to personality items might correctly classify a substantial percentage of individuals as either "fakers" or honest responders (Holden, 1995, 1998; Holden & Hibbs, 1995; Robie et al., 2000). Also, compared to the use of social desirability or "faking" scales, response latency measurement is considerably more unobtrusive. The possibility of combining response latency measurement with more typical measures of distortion in order to increase correct classification rates above the levels achievable by either approach has been supported by Dwight and Alliger (1997, as cited in Hough & Furnham, 2003). Nonetheless, the potentially biasing effect of job familiarity on response latencies warrants further research (Vasilopoulos, Reilly, & Leaman, 2000). A further concern is susceptibility to coaching. Interestingly, Robie et al. (2000) reported that test-takers who were coached to avoid the appearance of faking detection in their response latencies were, indeed, generally successful in avoiding detection but also produced personality scores that would not be advantageous to them in a selection situation. This result leaves open the possibility that coaching test-takers on how to "finesse" response latency detection of faking may actually tend to attenuate or avert the effect of faking on trait scores (Robie et al., 2000). Although not compatible with paper-and-pencil test administration, response latency measurement is feasible via internet test administration, which is rapidly expanding in popularity (see section on Internet-based assessment of personality).

Third, in the not too distant future, the application of Item Response Theory (IRT) may prove helpful in identifying persons who are most likely to have faked. Basically, IRT uses a mathematical model to describe the relationship between test-takers' levels on the personality trait being measured and their probability of choosing the various response options of a given personality test item (Crocker & Algina, 1986). Aspects of this mathematical model have the potential to be useful in detecting faking. IRT research on faking is still in its infancy but progress is being made (e.g., Zickar, Gibby, & Robie, 2004; Zickar & Robie, 1999). Perhaps a combination of faking scales, response latency measurement and IRT will one day prove effective in faking detection.

4.2.4. *The forced-choice method of personality assessment*

Typical personality items ask the test-taker to indicate whether or not a particular stimulus (e.g., a statement, adjective, or idea) describes him/her, to what extent it describes him/her, or whether (or to what extent) he/she agrees with the sentiment contained in the stimulus. Dating back to the groundbreaking work of Jackson and colleagues (e.g., Jackson, 1960; Jackson & Messick, 1961, 1962), it is now well-known that test-takers may be influenced by the impression that they feel will be conveyed to others as a result of the responses they provide to personality items (Helmes, 2000; Paulhus, 1991). This is particularly true in personnel selection situations where personality scores are consequential, resulting in an increased tendency to choose item responses that present the self in a positive light (e.g., Jackson et al., 2000; Mueller-Hanson et al., 2003; Rosse et al., 1998).

The forced-choice (FC) approach to personality assessment was proposed as a means of obtaining more honest, self-descriptive responses to personality items by reducing the effect of perceived desirability on response choices. This is achieved by presenting statements in pairs, triplets or quartets that assess different traits, but have been equated with respect to perceived desirability level. The test-taker is instructed to choose the statement that best describes him/her, and, in the case of item quartets, to also indicate the statement that is least self-descriptive. Because the perceived desirability levels of all choices are equal there is no clear benefit to motivated distortion. Test-takers are therefore presumed to respond in a more honest, self-descriptive manner. The Edwards Personal Preference Schedule (Edwards, 1959) and the Gordon Personal Inventory (Gordon, 1956) are two well-known early examples of the FC approach.

Despite the initial enthusiasm expressed for the FC approach, a series of studies conducted throughout the 1950s and 1960s cast doubt on the ability of this approach to validly measure personality traits and to withstand motivated distortion (e.g., Borislow, 1958; Dicken, 1959; Dunnette, McCartney, Carlson, & Kirchner, 1962; Graham, 1958; Norman, 1963). In particular, it was found that participants instructed to fake their responses with respect to a specific target job produced shifts in scores that differed from "respond honestly" conditions. Consequently, the FC approach fell from grace as a personality assessment tool (e.g., Anastasi, 1982) and relatively little new research surfaced for several decades.

Recently, Jackson et al. (2000) breathed new life into the FC approach. Jackson et al. proposed that many of the earlier problems associated with the FC approach might be attributable to poor item development (e.g., inaccurate desirability matching; use of pairs or triplets of items rather than quartets) and dependencies in trait scores (also known as ipsative scoring) that resulted from the manner in which the items were derived. Jackson et al. developed a new FC personality scale that overcame these problems and were able to show in a personnel selection simulation study that it resulted in significantly higher criterion-related validity than a traditional personality measure that assessed the same traits. Moreover, as was the trend in the Stanush (1997) meta-analysis, the FC scale evidenced a significantly smaller shift in scores as a result of "fake good" instructions than did a traditional personality scale. Martin, Bowen, and Hunt (2002) were also able to show that a new forced-choice personality scale was resistant to the shift in scores that usually accompanies fake-good instructions.

Historically, the desirability matching of items in FC scales has been based on ratings of desirability in general, with no particular reference to how desirable the items are with regards to the target job (Rothstein & Goffin, 2000). Understandably, matching statements based only on their general desirability is not the same as matching them in terms of how desirable they are in regards to a given job. As a means of further enhancing the faking resistance of the FC approach in personnel selection contexts, Rothstein and Goffin (2000, p. 235) proposed that "...one could tailor the desirability matching of the statements to the specific job or occupation that the scale would eventually be used to make selection decisions for." Christiansen et al. (2005) adopted such an approach in deriving a FC personality inventory relevant to sales positions. As in Jackson et al. (2000), Christiansen et al. also circumvented the problem of ipsative scoring. In an engaging series of studies, Christiansen et al. were able to provide evidence for the construct validity of their FC measure. Moreover, their results also showed that in a simulated personnel selection context where applicants

were instructed to respond as though in competition for a job, the FC measure correlated more strongly with job performance than a traditional personality measure of the same traits did. Surprisingly, the FC measure evidenced significantly higher criterion-related validity under personnel selection conditions than it did under “respond honestly” conditions.

To summarize, our search revealed only three published studies in which FC personality scales, derived using modern item analytic techniques, were evaluated in terms of faking resistance in a personnel selection context (Christiansen et al., 2005; Jackson et al., 2000; Martin et al., 2002). Despite the paucity of studies, the uniformly positive nature of their findings suggests that the FC approach is worthy of much greater consideration in personnel selection. At present, we are aware of only two commercially available personality scales that are comprised exclusively of forced-choice items and developed for use in personnel selection; the Occupational Personality Questionnaire 3.2i (SHL, 1999), one version of which was used by Martin et al.; and the Employee Screening Questionnaire (Jackson, 2002), an earlier version of which was employed in Jackson et al. (2000). Despite the increased costs of developing FC measures, it is hoped that the positive results evidenced so far will contribute to their proliferation.

Two cautions are pertinent with respect to the FC method. First, Harland (2003) showed that test-taker reactions to the FC approach may be less positive than to traditional personality tests. This finding is reminiscent of the negative reactions of performance raters to the use of FC performance appraisal scales (e.g., Smith & Kendall, 1963). Nonetheless, appropriate communication with test-takers may provide a solution (Harland, 2003). Second, Christiansen et al. (2005) determined that the FC approach, when used in a personnel selection situation, may increase the cognitive loading of trait scores (“cognitive loading” was defined in the section on the faking warning. As was explained in the section on the faking warning where cognitive loading was also a concern, the degree to which cognitive loading may decrease incremental validity and influence adverse impact is deserving of further research.

4.3. Faking remedies, underlying processes, and integrative possibilities

Snell et al. (1999) presented a simple model of applicant faking behavior that provides a useful perspective from which to consider faking and its possible remedies. According to Snell et al.’s model, faking has two main determinants. “Ability to fake” refers to the capacity to fake, whereas “motivation to fake” refers to the willingness to fake (Snell et al., 1999). All faking “remedies” can be seen as primarily targeting one or the other of these determinants. By making successful faking more difficult, the primary effect of the forced-choice approach and the faking correction (if successful) would be to reduce the test-taker’s ability to fake.¹ Nonetheless, sufficient motivation to fake may cause the test-taker to persist in dissimulation attempts despite the challenge presented by forced-choice items, as in the case of the test-taker who desperately needs employment. The faking warning, on the other hand, would tend to reduce the motivation to fake. When confronted by a credible warning that faking attempts might actually reduce hiring prospects, in combination with the increased challenge to faking ability that the FC method provides, the “desperate employment-seeker” might desist from all attempts at faking.

More generally, it is conceivable that even a small degree of ability to fake, when coupled with sufficient motivation could lead to motivated distortion. Similarly, even a very limited amount of motivation to fake may inspire dissimulation if ability to fake is adequate. Logically, then, a faking reduction intervention that effectively targets both the ability and motivation to fake is likely to be more successful than one that targets a single determinant. Surprisingly, our review of the literature failed to find any studies assessing the separate and combined effects of faking interventions targeting both the ability and willingness to fake. Although one study (Hough, 1998a,b) clearly used an approach that targeted both the willingness to fake (e.g., a faking warning) as well as the ability to fake (e.g., a faking “correction”) the necessary controls for studying the separate *and* combined effects of the two interventions were not present. Integrative research of this nature would be most informative as it stands to contribute to knowledge on the practical control of faking as well as further development of conceptual models of faking (e.g., Snell et al., 1999).

¹ Although the primary impact of the forced-choice approach may be on the ability to fake, we acknowledge that it may also impact the motivation to fake to some degree. Clearly, to the extent the test-taker surmises that his/her faking-related behaviors will likely be ineffective, the motivation to fake is likely to wane somewhat.

To summarize, although research suggests that the effects of faking are unlikely to be severe enough to *neutralize* the usefulness of personality tests, faking may possibly lower criterion-related validity and reduce the accuracy of hiring decisions. The human resources researcher or practitioner who is concerned about faking might benefit from employing the faking warning and/or the forced-choice method to attenuate the effects of faking as these two approaches have received the most support from research. Nonetheless, the caveats pertaining to both approaches should be kept in mind as well as the need for additional research.

As discussed earlier, logic dictates that increased knowledge of the effects and control of faking has contributed to increased personality test usage because concern about faking has been an exceedingly persistent impediment. Given that concerns about faking are being addressed through the increased research on this topic, a key practical innovation that is further encouraging the use of personality testing in personnel selection is internet-based assessment.

5. Internet-based assessment of personality

Some have heralded this “the decade of the internet in personnel selection” (Salgado & Moscoso, 2003, p. 194). Accordingly, it would be an understatement to say that the internet has contributed to, and holds great promise for, the continuing growth of personality assessment in personnel selection applications (Stanton, 1999). On a very practical level, internet administration may reduce missing data, allow 24/7 administration of personality tests worldwide, and facilitate instantaneous access to test results for both the test-taker and the manager (Jones & Dages, 2003; Lievens & Harris, 2003; Ployhart, Weekley, Holtz, & Kemp, 2003). As a case in point, internet-based personality assessment allowed a large corporate client of one of the current authors to assess the personality of promising applicants with equal convenience regardless of whether they were located next door or in Saudi Arabia. Moreover, compared to conventional paper-and-pencil personality testing, internet testing largely eliminates printing costs (Lievens & Harris, 2003; Naglieri et al., 2004) and may eliminate the need for human proctors (Bartram & Brown, 2004). Further benefits of the internet include the updating of administration instructions, scoring algorithms, normative data, actual test items, and interpretive score reports with an unprecedented level of speed and efficiency (Jones & Dages, 2003; Naglieri et al., 2004), positive reactions from test-takers (Anderson, 2003; Salgado & Moscoso, 2003), and the potential to increase the representation of minority groups as a result of overall increases in access to applicants (Chapman & Webster, 2003).

5.1. Equivalence of internet-based and paper-and-pencil personality tests

Despite the advantages just discussed, a key issue of ethical and practical importance is whether or not internet-based personality testing will produce assessments that are comparable to paper-and-pencil administration in all important respects (Naglieri et al., 2004). A number of recent empirical investigations of this general issue have been published (e.g., Bartram & Brown, 2004; Buchanan, Johnson, & Goldberg, 2005; Buchanan & Smith, 1999; Cronk & West, 2002; Davis, 1999; McManus & Ferguson, 2003; Pasveer & Ellard, 1998; Ployhart et al., 2003; Salgado & Moscoso, 2003). Overall, the findings of this group of studies point to the general conclusion that internet and paper-and-pencil administration of personality tests will lead to comparable results. However, we thought it prudent to focus our attention on those investigations involving personnel selection scenarios because several researchers have shown that the “high-stakes” nature of such situations impacts responses to personality tests in important ways (Goffin & Woods, 1995; Rosse et al., 1998; Naglieri et al., 2004). Also, we were less interested in those studies in which the samples of participants responding to the internet-based personality test were, by design, not comparable to those who responded to the paper-and-pencil instrument (e.g., McManus & Ferguson, 2003). Consequently, three studies were of particular relevance to our review (Bartram & Brown, 2004; Ployhart et al., 2003; Salgado & Moscoso, 2003).

Bartram and Brown (2004) investigated the comparability of a non-proctored, internet-based administration of the Occupational Personality Questionnaire (OPQ) to a traditional, proctored, paper-and-pencil administration. All participants responded under non-laboratory (i.e., “real”) testing conditions which included personnel selection and development. The target jobs consisted of managerial and professional financial sector and career management positions, and entry-level positions in marketing, sales, and client service management. Five samples (total $n=1127$) responded to the paper-and-pencil version whereas five additional matched samples (total $n=768$) responded to the internet version. Mean internet versus paper-and-pencil differences between the matched samples were examined at the level of Big Five traits (e.g., see Jackson, Paunonen, Fraboni, & Goffin, 1996) and the 32 facet traits comprised by the

Big Five. Mean differences were relatively small ($d = .27$ or less) at the level of the facet scales and smaller still at the level of the Big Five ($d = .20$ or less). Intercorrelations between personality scales were stable across internet and paper-and-pencil versions, suggesting that the internal structure of the test was stable across the different modes of administration. Finally, reliability (internal consistency) and standard error of measurement estimates of the internet-based personality scales were comparable to those of the paper-and-pencil version.² Overall, Bartram, and Brown's results suggested that internet administration did not change the personality test results for better or for worse.

Ployhart et al. (2003) assessed the internet-based versus paper-and-pencil comparability of Conscientiousness, Agreeableness, and Emotional Stability scales, as well as other selection measures. Both test modalities were proctored. The test-takers were all applicants for call-center positions at the same organization wherein paper-and-pencil scales were initially used for selection then later replaced with internet-based scales. A sample of 2544 applicants completed the paper-and-pencil scales whereas 2356 later completed the internet-administered scales. Systematic differences between the two samples could not be ruled out because demographic data were not available. However, Ployhart et al. argued that nontrivial systematic differences in samples were unlikely because the organization's shift to internet-based assessment was rapid, economic conditions were not in flux, and recruiting methods remained constant, as did the test proctors. Standardized mean differences on the three traits ranged from .356 to .447, with the internet-based applicant scores being consistently lower than paper-and-pencil-based, and considerably closer to the means from an incumbent sample who responded to paper-and-pencil scales ($n = 425$). The applicant internet-based personality score distributions were also notably less skewed and kurtotic than the paper-and-pencil-based applicant scores and variances were slightly higher. Further, internal consistency reliability of the trait scores ranged from .64 to .72 for the applicant paper-and-pencil administration, which was very similar to the reliabilities reported for the incumbent paper-and-pencil sample (.63–.73), but considerably lower than the respective internet-based values (.75–.80). Finally, the intercorrelations between the Conscientiousness, Agreeableness, and Emotional Stability scales were somewhat higher in the internet-based applicant administration than in either the applicant paper-and-pencil or incumbent paper-and-pencil administrations, however, the higher reliabilities of the internet-based scales would tend to increase the interscale correlations. All in all, Ployhart et al.'s results suggest that, compared to conventional paper-and-pencil administration, internet administration of a personality measure may produce some non-trivial differences in scores. Moreover, the reliability and normality of scores may be improved by internet administration. Nonetheless, caution is required in interpreting Ployhart et al.'s results because of the aforementioned lack of empirical evidence that the applicant samples responding to the internet-based and paper-and-pencil measures were sufficiently equivalent (e.g., it is not known if the proportion of males and females was comparable).

Salgado and Moscoso (2003) conducted the only within-subjects comparison of internet-based and paper-and-pencil personality tests, within a selection context, of which we are aware. Participants were 162 Spanish undergraduates rather than applicants in the truest sense, but they were informed that their test scores would be used as a basis for selecting individuals for a desirable training program, and participation was a mandatory course component. They completed both the paper-and-pencil and internet-based versions of a Five Factor personality inventory with the administrations of the two versions separated by two or three weeks to prevent carry-over effects, and the order of presentation counterbalanced. The within-subjects design allowed the computation of the correlations between trait scores obtained via the internet versus the paper-and-pencil version. In the psychometric literature, such correlations are referred to as "coefficients of stability and equivalence" and provide strong evidence of the comparability of parallel forms of a test (Aiken, 2000; Anastasi, 1982). The coefficients of stability and equivalence for the five trait scores were high, ranging from .81 to .92. These values are all the more impressive because the time lag between the administration of the two testing modalities was relatively long (two to three weeks), and suggest that the rank-ordering of candidates would not change substantially regardless of which form was administered. Similarly, the standardized mean differences in traits scores across the two forms were very small (.03 to .14) and internal consistency reliability coefficients were very similar, although slightly

² Bartram and Brown (2004) did not have access to the item responses from the samples responding to the paper-and-pencil test. Therefore, matched samples could not be used to compare the internal consistency reliabilities of the internet-based versus paper-and-pencil scales. Internal consistency reliabilities of the internet-based scales were computed using the samples described earlier, whereas the internal consistency reliabilities of the paper-and-pencil scales were computed using the original OPQ standardization sample.

higher on average in the internet modality. Standard deviations of trait scores were also slightly higher in the internet version. Intertrait correlations were remarkably consistent across the two modalities, suggesting that the internal structure of the test was unaffected by the testing mode.

Salgado and Moscoso's (2003) results were generally consistent with those of Bartram and Brown (2004), and Ployhart et al. (2003) in finding that differences between internet and paper-and-pencil administrations of personality tests tend to be small. Moreover, to the extent that such differences occur, they lean towards suggesting that internet administration *improves* the test's properties. Nonetheless, because we could find only three studies that dealt with internet versus paper-and-pencil equivalence of personality testing within a personnel selection context our conclusions must remain tentative at this time. Furthermore, even relatively small internet versus paper-and-pencil differences in mean scores could have important hiring implications if the two modalities are used interchangeably when vetting applicants, and could give rise to appeals. When considering the switch to an internet-based personality test, selection practitioners should carefully evaluate comparability in light of any existing cutoffs. Where feasible, managers would be well-advised to switch entirely from one modality to the other rather than continuing to compare applicant scores based on both modalities. Moreover, any inherent differences that may be introduced by internet administration should be carefully considered. In particular, it appears that some internet platforms may prevent the test-taker from choosing not to respond to an item (e.g., Salgado & Moscoso, 2003). We speculate that this may cause some respondents to dissimulate a response to an item that they otherwise would have left blank, which will result in a net gain in distortion. This speculation is consistent with Richman, Kiesler, Weisband, and Drasgow's (1999) finding that computer administered tests without a backtracking/skipping option engendered more socially desirable responding than computerized tests that did not. If the greater sense of privacy and lack of social pressure that internet-administration affords were to result in less response distortion as some have hypothesized (Richman et al., 1999), this advantage might be neutralized by not allowing the option of skipping items, which is typically possible with paper-and-pencil inventories.

The existence of only three directly pertinent published primary studies makes clear the need for further research on the equivalence of internet and paper-and-pencil administrations of personality tests within personnel selection contexts. In addition to the relatively straightforward methodologies used in the existing studies (described above), Ferrando and Lorenzo-Seva (2005) presented highly sophisticated means of assessing measurement equivalence that might prove useful. Test security is another serious issue for the organization that chooses internet over paper-and-pencil administration of personality tests. Attention must be paid to procedures for reducing unauthorized replication and distribution of test items and scoring procedures. Similarly, procedures for confirming the identity of the test-taker and preventing unauthorized help in responding to the test may be important. Thankfully, the technology of test security appears to be keeping pace with internet testing. We encourage the interested reader to consult Naglieri et al. (2004) for an insightful summary of promising approaches for securing test content and confirming test-taker identity when using internet-based assessment.

The following section highlights an important but seldom exploited potential advantage of internet personality test administration.

5.2. Computer adaptive testing

An exciting possibility that internet administration makes much more feasible is computer adaptive tests (CATs) of personality (Jones & Dages, 2003; Meijer & Nering, 1999; Naglieri et al., 2004). CATs actively monitor the test-taker's responses to each item then selectively administer subsequent items that are most likely to provide responses that are maximally informative in terms of pinpointing the test-taker's level on the respective personality trait. This approach relies on advanced Item Response Theory (IRT) in order to organize and take advantage of detailed information about the individual items on a personality test. There are two main advantages offered by CATs. First, because the computer selectively chooses test items rather than presenting all items, testing time is reduced — often by 50% — compared to a paper-and-pencil test (Meijer & Nering, 1999). This reduction in testing time comes with no loss, and probable gains, in reliability. Consequently, managers could choose to measure twice as many personality traits in the same amount of time, which might allow more comprehensive and precise assessment of personality rather than strict reliance on FFM measures (see discussion above for potential problems with the FFM). Second, as discussed in the Faking detection section, IRT opens up new possibilities in the detection of faking (e.g., Zickar et al., 2004; Zickar & Robie, 1999).

Although the use of CAT is increasing, and the consensus is that CAT's advantages go far beyond offsetting its disadvantages (Meijer & Nering, 1999), there is one clear limitation to implementing CAT personality testing at this point in time. Specifically, although the use of personality CATs has received attention in the research literature (MacDonald & Paunonen, 2002; Meijer & Nering, 1999) our computer search revealed only one published example of a CAT of personality (MacDonald & Paunonen, 2002). Developing a CAT of personality in one's own organization would require considerable upfront investment of resources (Jones & Dages, 2003; Meijer & Nering, 1999), however, we are sanguine that the growth and increasing popularity of CATs for cognitive and achievement testing (e.g., see www.ets.org and www.shl.com) will accelerate the development of personality CATs by major publishers of these tests, making it an attractive option in the very near future.

6. Summary and conclusions

On the basis of our review of recent research on the use of personality measures in personnel selection, we believe the following conclusions are warranted.

1. Numerous meta-analytic studies on personality-job performance relations conducted in the 1990s repeatedly demonstrated that personality measures contribute to the prediction of job performance criteria and if used appropriately, may add value to personnel selection practices.
2. Organizations are increasingly using personality measures as a component of their personnel selection decisions.
3. The Five Factor Model (FFM) of personality has become increasingly popular among researchers and practitioners, contributing to the renewal of interest in personality-job performance relations. However, more specific, narrow personality measures continue to demonstrate equal or greater utility for personnel selection.
4. Choice of an appropriate personality measure for use in predicting job performance should be based on careful consideration of the expected theoretical or conceptual relations between the personality predictor and performance criterion of interest, as well as the appropriate level of analysis between predictor and criterion measures.
5. Realizing the full potential of using personality measures to predict job performance requires consideration of potential moderator and mediator effects due to the situationally specific nature of personality predictors.
6. Although overall validity of personality measures for personnel selection is not seriously affected by applicant attempts to fake their responses, faking may increase the probability of less accurate hiring decisions at the individual level. At this time, research indicates that the most effective ways to limit the effects of faking is to employ a faking warning and/or a forced-choice personality test.
7. Internet administration of personality tests affords many potential advantages in terms of convenience and cost savings. There are only marginal differences between internet and paper-and-pencil administrations of personality tests, although this conclusion must remain tentative due to the limited research available at this time.

References

- Aiken, L. (2000). *Psychological testing and assessment* (10th ed.). Needham Heights, MA: Allyn and Bacon.
- Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: MacMillan.
- Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment*, *11*, 121–136.
- Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior*, *19*(3), 289–303.
- Baker, D. P., & Salas, E. (1994). *The importance of teamwork: In the eye of the beholder?* Paper presented at the Ninth Annual Conference of the Society for Industrial and Organizational Psychology, Nashville, TN.
- Balthazard, P., Potter, R. E., & Warren, J. (2004). Expertise, extraversion and group interaction styles as performance indicators in virtual teams. *Database for Advances in Information Systems*, *35*(1), 41–64.
- Barrett, G. V., Miguel, R. F., Hurd, J. M., Lueke, S. B., & Tan, J. A. (2003). Practical issues in the use of personality tests in police selection. *Public Personnel Management*, *32*(4), 497–517.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26.
- Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationship between the Big Five personality dimensions and job performance. *Journal of Applied Psychology*, *78*(1), 111–118.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personal constructs. *Journal of Applied Psychology*, *81*, 261–272.

- Barrick, M. R., & Mount, M. K. (2003). Impact of meta-analysis methods on understanding personality–performance relations. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 197–222). Mahwah, NJ: Lawrence Erlbaum.
- Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychology*, *56*(1), 45–74.
- Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of Applied Psychology*, *78*(5), 715–722.
- Barrick, M. R., Stewart, G. J., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, *83*(3), 377–391.
- Barrick, M. R., Stewart, G. L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, *87*(1), 43–51.
- Barry, B., & Stewart, G. L. (1997). Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology*, *82*(1), 62–78.
- Bartram, D. (1995). The predictive validity of the EPI and 16PF for military flying training. *Journal of Occupational and Organizational Psychology*, *68*(3), 219–236.
- Bartram, D., & Brown, A. (2004). Online testing: Mode of administration and the stability of OPQ 32i scores. *International Journal of Selection and Assessment*, *12*, 278–284.
- Beagrie, S. (2005). How to... excel at psychometric assessments. *Personnel Today*, *25*.
- Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: Analysis along several dimensions of item type. *Personnel Psychology*, *45*, 389–406.
- Berta, D. (2005). Operators using prescreen tests to overturn turnover. *Nation's Restaurant News*, *39*(24), 22.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *177*(2), 187–215.
- Bobko, P., & Stone-Romero, E. F. (1998). Meta-analysis may be another useful research tool, but it is not a panacea. In G. R. Ferris (Ed.), *Research in personnel and human resources management*, vol. 16 (pp. 359–397). Stamford, CT: JAI.
- Bond, M. H., & Ng, I. W. -C. (2004). The depth of a group's personality resources: Impacts on group process and group performance. *Asian Journal of Social Psychology*, *7*(3), 285–300.
- Borislav, B. (1958). The Edwards Personal Preference Schedule and fakability. *Journal of Applied Psychology*, *42*, 22–27.
- Boudreau, J. W., Boswell, W. R., Judge, T. A., & Bretz Jr., R. D. (2001). Personality and cognitive ability as predictors of job search among employed managers. *Personnel Psychology*, *54*(1), 25–50.
- Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the internet. *European Journal of Psychological Assessment*, *21*, 115–127.
- Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, *90*, 125–144.
- Burke, L. A., & Witt, L. A. (2004). Personality and high-maintenance employee behavior. *Journal of Business and Psychology*, *18*(3), 349–363.
- Cable, D. M., & Judge, T. A. (2003). Managers' upward influence tactic strategies: The role of manager personality and supervisor leadership style. *Journal of Organizational Behavior*, *24*(2), 197–214.
- Cascio, W. F. (1991). *Costing human resources: The financial impact of behavior in organizations*. Boston, MA: PWS-Kent.
- Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing.
- Chapman, D. S., & Webster, J. (2003). The use of technologies in the recruiting, screening, and selection processes for job candidates. *International Journal of Selection and Assessment*, *11*, 113–120.
- Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item format for applicant personality assessment. *Human Performance*, *18*, 267–307.
- Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the Sixteen Personality Factors test for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, *47*, 847–860.
- Clarke, S., & Robertson, I. T. (2005). A meta-analytic review of the Big Five personality factors and accident involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology*, *78*(3), 355–376.
- Conte, J. M., & Gintoft, J. N. (2005). Polychronicity, Big Five personality dimensions, and sales performance. *Human Performance*, *18*(4), 427–444.
- Crant, J. M. (1995). The proactive personality scale and objective job performance among real estate agents. *Journal of Applied Psychology*, *80*(4), 532–537.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt.
- Cronk, B. C., & West, J. L. (2002). Personality research on the Internet: A comparison of web-based traditional instruments in take-home and in-class settings. *Behavior Research Methods, Instruments, and Computers*, *34*, 177–180.
- Daniel, L. (2005, April–June). Use personality tests legally and effectively. *Staffing Management*, *1*(1) Retrieved October 20, 2005, from <http://www.shrm.org/ema/sm/articles/2005/apriljune05cover.asp>.
- Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments, and Computers*, *31*, 177–180.
- Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology*, *42*(1), 25–36.
- Dicken, C. F. (1959). Simulated patterns of the Edwards Personal Preference Schedule. *Journal of Applied Psychology*, *43*, 372–378.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, *16*, 81–106.
- Driskell, J. E., Hogan, R., & Salas, E. (1988). Personality and group performance. *Review of Personality and Social Psychology*, *14*, 91–112.

- Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology, 15*, 13–24.
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*, 1–23.
- Edwards, A. L. (1959). *Edwards Personal Preference Schedule manual*. New York: Psychological Corporation.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155–166.
- Erickson, P. B. (2004, May 16). Employer hiring tests grow sophisticated in quest for insight about applicants. *Knight Ridder Tribune Business News, 1*.
- Eysenck, H. J. (1991). Dimensions of personality: 16, 5, or 3? Criteria for a taxonomy paradigm. *Personality and Individual Differences, 12*, 773–790.
- Faulder, L. (2005, Jan 9). The growing cult of personality tests. *Edmonton Journal, D.6*.
- Ferrando, P. J., & Lorenzo-Seva, U. (2005). IRT-related factor analytic procedures for testing the equivalence of paper-and-pencil and Internet-administered questionnaires. *Psychological Methods, 10*, 193–205.
- Fox, J. A., & Tracy, P. E. (1986). *Randomized response: A method for sensitive surveys*. Beverly Hills, CA: Sage.
- Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology Research and Reviews, 9*, 46–55.
- Gellatly, I. R. (1996). Conscientiousness and task performance: Test of a cognitive process model. *Journal of Applied Psychology, 81*(5), 474–482.
- Gellatly, I. R., & Irving, P. G. (2001). Personality, autonomy, and contextual performance of managers. *Human Performance, 14*(3), 231–245.
- Geller, A. (2004, August 8). Now, tell the computer why you want this job: PCs take lead role in screening hourly workers. *Calgary Herald, F.3*.
- Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A Review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment, 11*, 340–344.
- Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology, 81*, 746–756.
- Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (2000). Personality and job performance: Are personality tests created equal? In R. D. Goffin, & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 249–264). Norwell, MA: Kluwer Academic Publishers.
- Goffin, R. D., & Woods, D. M. (1995). Using personality testing for personnel selection: Faking and test-taking inductions. *International Journal of Selection and Assessment, 3*, 227–236.
- 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.
- Gordon, L. V. (1956). *Gordon personal inventory*. Harcourt, Brace & World: New York, NY.
- Graham, W. R. (1958). Social desirability and forced-choice methods. *Educational and Psychological Measurement, 18*, 387–401.
- Griffin, B., & Hesketh, B. (2004). Why openness to experience is not a good predictor of job performance. *International Journal of Selection and Assessment, 12*(3), 243–251.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135–164.
- Guzzo, R. A., & Shea, G. P. (1992). Group performance and intergroup relations in organizations. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology, vol. 3* (pp. 269–314). Palo Alto, CA: Consulting Psychologists Press.
- Hackman, J. R. (1986). The psychology of self-management in organizations. In M. S. Pallak, & R. Perloff (Eds.), *Psychology and work* (pp. 89–136). Washington, DC: American Psychological Association.
- Hakstian, A. R., & Ng, E. (2005). Employment related motivational distortion: Its nature, measurement, and reduction. *Educational and Psychological Measurement, 65*, 405–441.
- Halfhill, T., Nielsen, T. M., Sundstrom, E., & Weilbaecher, A. (2005). Group personality composition and performance in military service teams. *Military Psychology, 17*(1), 41–54.
- Handler, R. (2005). The new phrenology: A critical look at the \$400 million a year personality-testing industry. *Psychotherapy Networker, 29*(3), 1–5.
- 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*, 285–301.
- Hartigan, J. A., & Wigdor, A. K. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery*. Washington, DC: National Academy Press.
- Heller, M. (2005). Court ruling that employer's integrity test violated ADA could open door to litigation. *Workforce Management, 84*(9), 74–77.
- Helmes, E. (2000). The role of social desirability in the assessment of personality constructs. In R. D. Goffin, & E. Helmes (Eds.), *Problems and solutions in human assessment*. Norwell, MA: Kluwer Academic Publishers.
- Hochwarter, W. A., Witt, L. A., & Kacmar, K. M. (2000). Perceptions of organizational politics as a moderator of the relationship between conscientiousness and job performance. *Journal of Applied Psychology, 85*(3), 472–478.
- Hoel, B. (2004). Predicting performance. *Credit Union Management, 27*(7), 24–26.
- Hogan, R. (1986). *Hogan Personality Inventory manual*. Minneapolis, MN: National Computer System.
- Hogan, R. (2005). In defence of personality measurement: New wine for old whiners. *Human Performance, 18*, 331–341.
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job performance relations. *Journal of Applied Psychology, 88*, 100–112.
- Hogan, R., & Roberts, B. W. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior, 17*, 627–637.
- Holden, R. R. (1995). Response latency detection of fakers on personnel tests. *Canadian Journal of Behavioural Science, 27*, 343–355.
- Holden, R. R. (1998). Detecting fakers on a personnel test: Response latencies versus a standard validity scale. *Journal of Social Behavior and Personality, 13*, 387–398.

- Holden, R. R., & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality*, 29, 362–372.
- Holden, R. R., & Jackson, D. N. (1981). Subtlety, information, and faking effects in personality assessment. *Journal of Clinical Psychology*, 37, 379–386.
- Hough, L. M. (1992). The “Big Five” personality variables—construct confusion: Description versus prediction. *Human Performance*, 5, 139–155.
- Hough, L. M. (1997). Personality at work: Issue and evidence. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131–166). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11, 209–244.
- Hough, L. M. (1998). Personality at work: Issues and evidence. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131–166). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities [Monograph]. *Journal of Applied Psychology*, 75, 581–595.
- Hough, L. M., & Furnham, A. (2003). Use of personality variables in work settings. In W. Borman, D. Ilgen, & R. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology*, vol. 12 (pp. 131–169). Hoboken, NJ: John Wiley & Sons.
- Hough, L. M., & Ones, D. (2002). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology, volume 1: Personnel psychology* (pp. 233–277). Thousand Oaks, CA: Sage.
- Hsu, C. (2004). The testing of America. *U.S. News and World Report*, 137(9), 68–69.
- Jackson, D. N. (1960). Stylistic response determinants in the California Psychological Inventory. *Educational and Psychological Measurement*, 10, 339–346.
- Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Port Huron, MI: Research Psychologists.
- Jackson, D. N. (2002). *Employee screening questionnaire: Manual*. Port Huron, MI: Sigma Assessment Systems.
- Jackson, D. N., & Messick, S. (1961). Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement*, 21, 771–790.
- Jackson, D. N., & Messick, S. (1962). Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, 65, 285–299.
- Jackson, D. N., Paunonen, S. V., Fraboni, M., & Goffin, R. D. (1996). A five-factor versus six-factor model of personality structure. *Personality and Individual Differences*, 20, 33–45.
- Jackson, D. N., & Rothstein, M. G. (1993). Evaluating personality testing in personnel selection. *The Psychologist: Bulletin of the British Psychological Society*, 6, 8–11.
- Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, 13, 371–388.
- Jenkins, M., & Griffith, R. (2004). Using personality constructs to predict performance: Narrow or broad bandwidth. *Journal of Business and Psychology*, 19(2), 255–269.
- Jones, J. W., & Dages, K. D. (2003). Technology trends in staffing and assessment: A practice note. *International Journal of Selection and Assessment*, 11, 247–252.
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87(4), 765–780.
- Judge, T. A., & Cable, D. M. (1997). Applicant personality, organizational culture, and organization attraction. *Personnel Psychology*, 50(2), 359–394.
- Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87(3), 530–541.
- Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52(3), 621–652.
- Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, 87(4), 797–807.
- Kichuk, S. L., & Wiesner, W. H. (1997). The Big Five personality factors and team performance: Implications for selecting successful product design teams. *Journal of Engineering and Technology Management*, 14(3,4), 195–221.
- Kichuk, S. L., & Wiesner, W. H. (1998). Work teams: Selecting members for optimal performance. *Canadian Psychology*, 39(1/2), 23–32.
- Kickul, J., & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and KSAs. *Journal of Business and Psychology*, 15(1), 27–51.
- LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with big five personality characteristics and cognitive ability. *Journal of Applied Psychology*, 86(2), 326–336.
- Lievens, F., & Harris, M. M. (2003). Research on Internet recruiting and testing: Current status and future directions. In C. L. Cooper, & I. T. Robertson (Eds.), *International review of industrial and organizational psychology*, vol. 16 (pp. 131–165). Chichester: John Wiley & Sons, Ltd.
- Lievens, F., Harris, M. M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology*, 88(3), 476–486.
- Lim, B. -C., & Ployhart, R. E. (2004). Transformational leadership: Relations to the Five-Factor model and team performance in typical and maximum contexts. *Journal of Applied Psychology*, 89(4), 610–621.
- Lin, N. -P., Chiu, H. -C., & Hsieh, Y. -C. (2001). Investigating the relationship between service providers’ personality and customers’ perceptions of service quality across gender. *Total Quality Management*, 12(1), 57–67.

- Lounsbury, J. W., Gibson, L. W., & Hamrick, F. L. (2004). The development and validation of a personological measure of work drive. *Journal of Business and Psychology, 18*(4), 427–451.
- Luther, N. (2000). Integrity testing and job performance within high performance work teams: A short note. *Journal of Business and Psychology, 15* (1), 19–25.
- Luther, N. J., & Thornton III, G. C. (1999). Does faking on employment tests matter? *Employment Testing Law and Policy Reporter, 8*, 129–136.
- MacDonald, P., & Paunonen, S. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement, 62*, 921–943.
- Marcus, B. (2003). Personality testing in personnel selection: Is “socially desirable” responding really undesirable? (Persönlichkeitstests in der Personalauswahl: Sind “sozial erwünschte” Antworten wirklich nicht wünschenswert?). *Zeitschrift für Psychologie, 211*, 138–148.
- Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32*, 247–256.
- McCarthy, J. M., & Goffin, R. D. (2001). Improving the validity of letters of recommendation: An investigation of three standardized reference forms. *Military Psychology, 13*, 199–222.
- McCrae, R. R., & Costa Jr., P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882–888.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812–821.
- McManus, M. A., & Ferguson, M. W. (2003). Biodata, personality, and demographic differences of recruits from three sources. *International Journal of Selection and Assessment, 11*, 175–183.
- McManus, M. A., & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry. *Personnel Psychology, 52*(1), 137–148.
- Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement, 23*, 187–194.
- Mohammed, S., & Angell, L. C. (2003). Personality heterogeneity in teams: Which differences make a difference for team performance? *Small Group Research, 34*(6), 651–677.
- Mohammed, S., Mathieu, J. E., & Bartlett, A. L. (2002). Technical–administrative task performance, leadership task performance, and contextual performance: Considering the influence of team- and task-related composition variables. *Journal of Organizational Behavior, 23*(7), 795–814.
- Mol, S. T., Born, M. P., Willemsen, M. E., & Van Der Molen, H. T. (2005). Predicting expatriate job performance for selection purposes: A quantitative review. *Journal of Cross-Cultural Psychology, 36*(5), 590–620.
- Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and team work knowledge. *Personnel Psychology, 58*(3), 583–611.
- Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resource management. In G. Ferris (Ed.), *Research in personnel and human resource management, vol. 13* (pp. 153–200). Stamford, CT: JAI.
- Mount, M. K., & Barrick, M. R. (1998). Five reasons why the “Big Five” article has been frequently cited. *Personnel Psychology, 51*(4), 849–857.
- Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348–355.
- Murphy, K. R. (1997). Meta-analysis and validity generalization. In N. Anderson, & P. Herrio (Eds.), *International handbook of selection and assessment, vol. 13* (pp. 323–342). Chichester, UK: Wiley.
- Murphy, K. R. (2000). Impact of assessments of validity generalization and situational specificity on the science and practice of personnel selection. *International Journal of Selection and Assessment, 8*, 194–206.
- Murphy, K. R., & Dziewieczynski, J. L. (2005). Why don’t measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance, 18*(4), 343–357.
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., et al. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist, 59*, 150–169.
- Neuman, G. A., Wagner, S. H., & Christiansen, N. D. (1999). The relationship between work-team personality composition and the job performance of teams. *Group and Organization Management, 24*(1), 28–45.
- Neuman, G. A., & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology, 84*(3), 376–389.
- Nikolaou, I. (2003). Fitting the person to the organisation: Examining the personality–job performance relationship from a new perspective. *Journal of Managerial Psychology, 18*(7/8), 639–648.
- Norman, W. T. (1963). Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology, 47*, 225–241.
- O’Connell, M. S., Doverspike, D., Norris-Watts, C., & Hattrup, K. (2001). Predictors of organizational citizenship behavior among Mexican retail salespeople. *International Journal of Organizational Analysis, 9*(3), 272–280.
- Ones, D. S., Mount, M. K., Barrick, M. R., & Hunter, J. E. (1994). Personality and job performance: A critique of the Tett, Jackson, and Rothstein (1991) meta-analysis. *Personnel Psychology, 47*(1), 147–156.
- Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*(6), 609–626.
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245–269.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679.
- Pannone, R. D. (1984). Predicting test performance: A content valid approach to screening applicants. *Personnel Psychology, 37*, 507–514.
- Pasveer, K. A., & Ellard, J. H. (1998). The making of a personality inventory: Help from the www. *Behavior Research Methods, Instruments, and Computers, 30*, 309–313.

- Paul, A. M. (2004). *The cult of personality: How personality tests are leading us to mislabel our children, mismanage our companies, and misunderstand ourselves*. New York: Free Press.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes, vol. 1* (pp. 17–59). San Diego: Academic Press.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah NJ: Erlbaum.
- Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior, 20*(3), 389–405.
- Peters, T. J. (1988). *Thriving on chaos*. New York: Knopf.
- Ployhart, R. E., Lim, B. -C., & Chan, K. -Y. (2001). Exploring relations between typical and maximum performance ratings and the five factor model of personality. *Personnel Psychology, 54*(4), 809–843.
- Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology, 56*, 733–752.
- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*, 754–775.
- Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58*(1), 103–139.
- Robie, C., Curtin, P. J., Foster, T. C., Phillips, H. L., Zbylut, M., & Tetrick, L. E. (2000). The effect of coaching on the utility of response latencies in detecting fakers on a personality measure. *Canadian Journal of Behavioural Science, 32*, 226–233.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634–644.
- Rothstein, M. G., & Goffin, R. D. (2000). The assessment of personality constructs in industrial–organizational psychology. In R. D. Goffin, & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 215–248). Norwell, MA: Kluwer Academic.
- Rothstein, M. G., & Jelly, R. B. (2003). The challenge of aggregating studies of personality. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 223–262). Mahwah, NJ: Lawrence Erlbaum.
- Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41*(2), 149–174.
- Salgado, J. F., & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses=perceptions and reactions. *International Journal of Selection and Assessment, 11*, 194–205.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.
- Schmidt, F. L., & Hunter, J. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods, 1975–2001. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 31–66). Mahwah, NJ: Lawrence Erlbaum.
- Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1942 and the investigation of study characteristics. *Personnel Psychology, 37*, 407–422.
- Schmitt, N., & Kuncze, C. (2002). The effects of required elaboration of answers to biodata questions. *Personnel Psychology, 55*, 569–586.
- Schneider, R. J., Hough, L. M., & Dunnette, M. D. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior, 17*(6), 639–665.
- SHL. (1999). *OPQ32 manual and user's guide*. Thames Ditton, UK: SHL Group.
- Simmering, M. J., Colquitt, J. A., Noe, R. A., & Porter, C. O. L. H. (2003). Conscientiousness, autonomy fit, and development: A longitudinal study. *Journal of Applied Psychology, 88*(5), 954–963.
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*, 149–155.
- Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resources Management Review, 9*, 219–242.
- Stanton, J. M. (1999). Validity and related issues in web-based hiring. *The Industrial–Organizational Psychologist, 36*, 69–71.
- Stanush, P. L. (1997). *Factors that influence the susceptibility of self-report inventories to distortion: A meta-analytic investigation*. Unpublished doctoral dissertation, Texas A&M University, College Station, TX.
- Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology, 84*(6), 959–968.
- Stewart, G. L., Fulmer, I. S., & Barrick, M. R. (2005). An exploration of member roles as a multilevel linking mechanism for individual traits and team outcomes. *Personnel Psychology, 58*(2), 343–365.
- Taggar, S., Hackett, R., & Saha, S. (1999). Leadership emergence in autonomous work teams: Antecedents and outcomes. *Personnel Psychology, 52* (4), 899–926.
- Tett, R. P., Jackson, D. N., & Rothstein, M. G. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703–742.
- Tett, R. P., Jackson, D. N., Rothstein, M. G., & Reddon, J. R. (1994). Meta-analysis of personality–job performance relations: A reply to Ones, Mount, Barrick, and Hunter (1994). *Personnel Psychology, 47*(1), 157–172.
- Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality–job performance relationship. *Journal of Organizational Behavior, 24*(3), 335–356.
- Thoresen, C. J., Bliese, P. D., Bradley, J. C., & Thoresen, J. D. (2004). The big five personality traits and individual job performance growth trajectories in maintenance and transitional job stages. *Journal of Applied Psychology, 89*(5), 835–853.

- Topping, G. D., & O’Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences*, 23, 117–124.
- Tornatsky, L. G. (1986). Technological change and the structure of work. In M. S. Pallak, & R. Perloff (Eds.), *Psychology and work* (pp. 89–136). Washington, DC: American Psychological Association.
- Van Iddekinge, C. H., Taylor, M. A., & Eidson, C. E. J. (2005). Broad versus narrow facets of integrity: Predictive validity and subgroup differences. *Human Performance*, 18(2), 151–177.
- Vasilopoulos, N. L., Cucina, J. M., & McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology*, 90, 306–322.
- Vasilopoulos, N. L., Reilly, R. R., & Leaman, J. A. (2000). The influence of job familiarity and impression management on self-report measure response latencies and scale scores. *Journal of Applied Psychology*, 85, 50–64.
- Vincur, A. J., Schippmann, J. S., Switzer III, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, 83(4), 586–597.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197–210.
- Wagner, W. F. (2000). All skill, no finesse. *Workforce*, 79(6), 108–116.
- Warr, P., Bartram, D., & Martin, T. (2005). Personality and sales performance: Situational variation and interactions between traits. *International Journal of Selection and Assessment*, 13(1), 87–91.
- Wiggins, J. S. (1968). Personality structure. *Annual Review of Psychology*, 19, 293–350.
- Williams, S. D. (2004). Personality, attitude, and leader influences on divergent thinking and creativity in organizations. *European Journal of Innovation Management*, 7(3), 187–204.
- Witt, L. A. (2002). The interactive effects of extraversion and conscientiousness on performance. *Journal of Management*, 28(6), 835–851.
- Worthington, D. L., & Schlottmann, R. S. (1986). The predictive validity of subtle and obvious empirically derived psychology test items under faking conditions. *Journal of Personality Assessment*, 50, 171–181.
- Zalinski, J. S., & Abrahams, N. M. (1979). The effects of item context in faking personnel selection inventories. *Personnel Psychology*, 32, 161–166.
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods*, 7, 168–190.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, 84, 551–563.