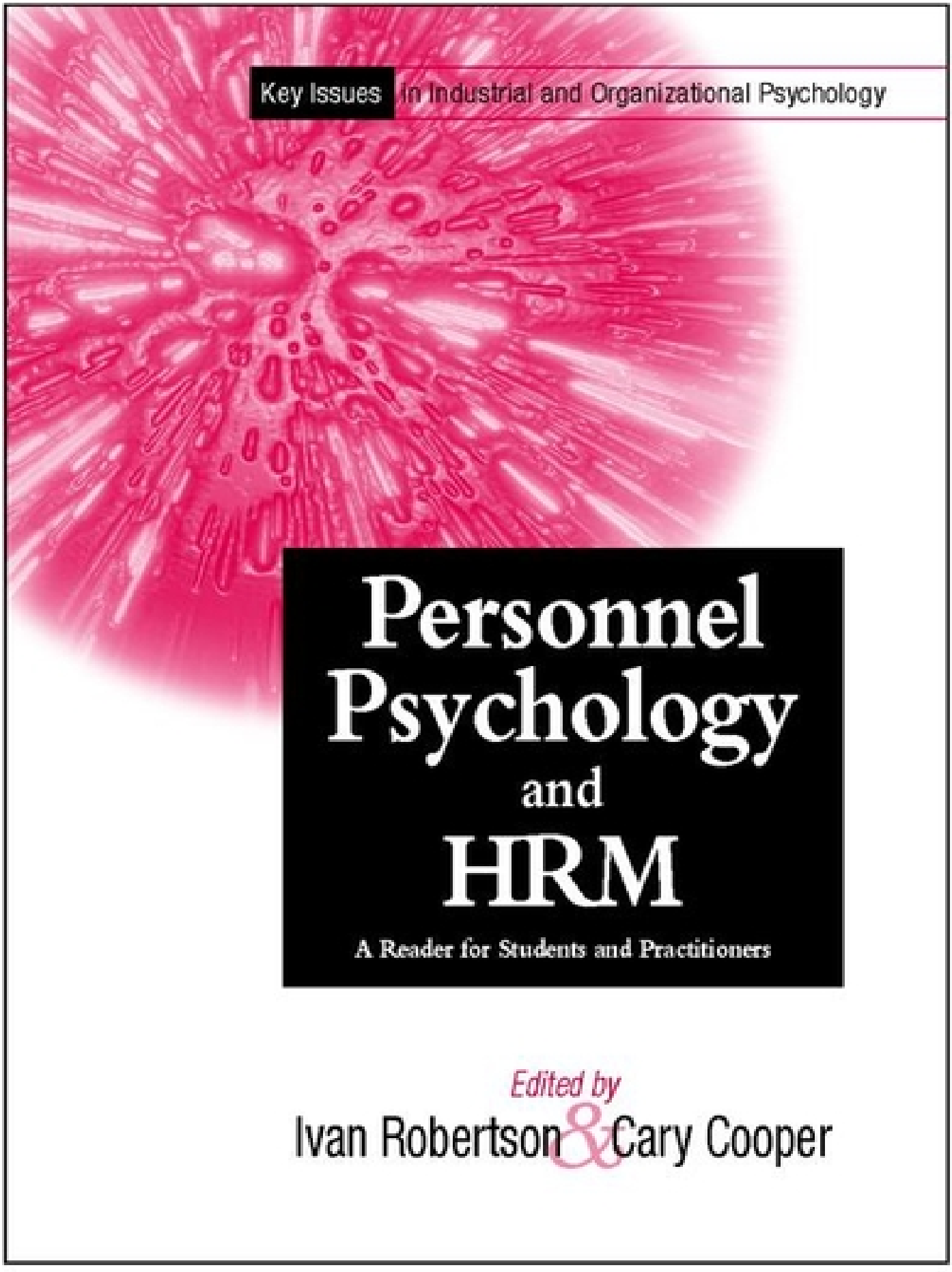
Key Issues in Industrial and Organizational Psychology
Personnel Psychology and HRM
A Reader for Students and Practitioners
Edited by
Ivan Robertson & Cary Cooper

INTRODUCTION

I/O psychology has its origins in attempts to make better decisions about how to deal with people at work. In particular personnel selection, training and other key personnel practices have always been of central interest to I/O psychologists. This remains true in the 21st century, despite the significant growth in the influence and expertise of specialist personnel, or human resource practitioners and researchers. This volume covers a number of key topics at the interface of human resource (HR) management and I/O psychology. Although attempts to identify clear boundaries between HR management and I/O psychology are unproductive and futile, it is worth noting the distinctive expertise and point of view that I/O psychologists bring to HR issues. In general terms the essence of this is an approach to research which is driven by the hypothetico-deductive method. This approach includes the application of strong measurement techniques and the empirical testing of ideas, principles and theories. The contributions of I/O psychologists to personnel and HR issues are reflected in three main areas of organizational life: performance; attachment and well being. Topics concerned with well-being are the subject of another volume in this collection (Well-being in Organizations). Performance and attachment issues specifically related to teams or groups and organization-wide matters are also dealt with in another volume (Organizational Psychology and Development).

This volume concentrates on topics at the individual level. The chapters included explore the research in a range of topics. The chapters are written by leading scholars with international reputations in their fields, from Australia,

Belgium, Canada, Spain, the UK and the USA. The book is divided into two broad sections: Personnel Psychology and Human Resource Management. As noted above the dividing lines between these fields are not firm and many researchers and practitioners from either field work in similar areas. The subdivision adopted below is merely a convenient way of grouping the chapters into related themes and does not reflect a firm distinction between the fields.

The chapters under the Personnel Psychology heading focus on selection and assessment, beginning with two chapters that provide an overview of research on personnel selection methods and their usage. The other three chapters look at more specific aspects of selection and assessment: assessment centres; multiple source (360 degree) feedback systems and cognitive ability testing. All of the chapters in this section provide a good indication of the way in which psychologists in this field have employed the empirical approach to personnel issues mentioned above. Taken together these five chapters provide an up to date and comprehensive view of the status of research in personnel selection and assessment.

The chapters under the Human Resource Management heading look at a wider set of issues related to attachment and performance in organizations. The first chapter in this section provides a thorough review of learning strategies and occupational training and together with the first chapter in the first section provides an indication of the state of the art in the two main approaches to delivering effective individual performance at work: selection and training. Chapters on absence, turnover, commitment and psychological contracts cover the major issues relating to attachment to organizations.

All of the topics are thoroughly reviewed, drawing on material from the leading volumes of the International Review of Industrial and Organizational Psychology. We hope that this compendium of quality reviews will help to improve research and practice in I/O psychology. In the end, we must attempt to move in the direction of understanding the simple truth of John Raskin’s comments about work in 1852: ‘In order that people may be happy in their work, these three things are needed: they must be fit for it; they must not do too much of it; and they must have a sense of success in it.’

Part I
PERSONNEL PSYCHOLOGY

Chapter 1
Personnel Selection Methods

Jesús F. Salgado
Universidad de Santiago de Compostela

INTRODUCTION

This review considers the literature that appeared from 1991 to Autumn 1997. I have also included some materials published before 1991 which may be relevant to this review and were not previously cited in reviews of personnel selection methods in IRIOP, as well as some new materials currently under publication.

In these last six years, a number of topics appear specially relevant in the personnel selection field. First, more attention has been given to the theoretical aspects concerning the validity of predictors of job performance. Second, interviews based on a behavioral and structured format (e.g., situational, behavior description, behavioral structured interviews) have been the subject of a great deal of investigation. Third, personality measures have received a lot of interest from researchers in personnel selection and appear to be predictors of job performance against previous beliefs. Fourth, advances in methodology have also been of interest (e.g., improvements in meta-analysis and utility analysis). In addition to these issues, this review covers traditional topics of personnel selection methods: abilities, biodata, assessment centers, simulations, references and job analysis.

In order to find the literature for this review I used three strategies. First, I conducted several computer-aided searches in the PsycLit data base, using the main concepts

as entry words (e.g., personnel selection, ability, personality, job performance) and cross-intersecting between them. Second, I conducted hand-made searches in the most relevant journals which cover the personnel selection area (e.g., *Journal of Applied Psychology, Personnel Psychology, Journal of Occupational and Organizational Psychology)*. Third, I wrote to over one hundred researchers asking for relevant articles.

Several publications must also be remarked upon in this introduction. Dunnette and Hough (1990–1994) have edited the second edition of the 'Bible' of Industrial and Organizational Psychology, Smith and Robertson (1993) have published the third edition of the personnel selection 'Bible' in Great Britain. Anderson and Herriot (1997) have edited an *International Handbook of Selection and Assessment* with contributors from 20 American, Asian and European countries. Other relevant books were edited or published by Cook (1993), Murphy (1996), Rumsay, Walker and Harris (1994), Schmitt and Borman (1993), and Schuler, Farr and Smith (1993). Also, a new journal has been published since 1993: the *International Journal of Selection and Assessment* (Neil Anderson, Editor). Furthermore, two commemorations must be noted. In 1992 the American Psychological Association celebrated its Centennial, and the *Journal of Applied Psychology) (JAP)* published several articles related to the date (Austin & Villanova, 1992; Harrel, 1992; Katzell & Austin, 1992; Landy, 1992). The Society of Industrial and Organizational Psychology (APA's Division 14) celebrated its golden anniversary in 1996. To commemorate this date *JAP* published five articles on the history of applied psychology (Benjamin, 1997; Farr & Tesluk, 1997; Koppes, 1997; Landy, 1997; Van de Walter, 1997).

Finally, I must acknowledge to being a little biased in my review. The number of articles, book chapters, congress

presentations, and so on, was so large, that I had to be somewhat selective, and to be selective in this field is very difficult and risky. Because of the fact that American literature on personnel selection is more well known than the European contributions, I tended to show the current European research whenever possible. I made this choice taking into account the fact that American literature is very well covered in the *Annual Review* chapters on personnel selection. Thus, I thought that showing the European contributions could open doors for possible cooperation between researchers from both continents.

PERFORMANCE AND JOB ANALYSIS ISSUES

Four advances were particularly relevant on performance topics and job analysis in the years that this review covers. One was the contribution made by J.P. Campbell (1994; Campbell, McCloy, Oppler & Sager, 1993; Campbell, Gasser & Oswald, 1996), who outlined a multidimensional theory of job performance. The second advance was the suggestion made by Borman and Motowidlo (1993) that job performance must be distinguished as task performance and contextual performance, The third advance was the result of work done by C. Viswesvaran (1993), who sustained that a general factor of job performance was responsible for most of the covariation among job performance dimensions. Furthermore, research was focused on investigating psychometric characteristics of job performance ratings (Salgado & Moscoso, 1996; Viswesvaran, Ones & Schmidt, 1996). An extended and comprehensive review of performance evaluation in work setting was written by Arvey and Murphy (1998) as an *Annual Review* chapter. The fourth remarkable advance was related to the development of a prototype to build an

Occupational Information Network (O*NET; Peterson, Mumford, Borman, 1995).

Job Performance

Based on the findings of Project A, Campbell (1994; Campbell et al., 1993) has developed a conception of job performance characterized as a multidimensional construct. The theory is based mainly on four notions: '(1) the assumption of a general job performance factor cannot represent the best fit; (2) the notion of an ultimate criterion has no meaning; (3) the subjective versus objective distinction is false; and (4) there is a critically important distinction to be made between performance and the results of performance' (Campbell et al., 1993, p.38). To Campbell, performance is the same as behavior, and only includes actions or behaviors relevant to the organization's goals. Performance is not the consequence of action, it is the action itself. Therefore, it is necessary to distinguish between performance and consequences of performance (e.g., effectiveness, productivity or utility). Campbell et al. (1993) suggested that job performance is composed of eight factors, common to all jobs, but with different relevance for each one. The factors are: (1) job-specific task proficiency; (2) non-job-specific task proficiency; (3) written and oral communication task proficiency; (4) demonstration of effort; (5) maintenance of personal discipline; (6) facilitation of peer and team performance; (7) supervision/leadership; and (8) management/ administration. Furthermore, according to Campbell et al. (1993) there are three determinants of individual differences in performance: (a) Declarative Knowledge (i.e., knowledge about facts and things); (b) Procedural Knowledge and Skill (i.e. cognitive skill, interpersonal skill); and (c) Motivation (i.e., combined choice effect to expend effort, choice of effort level to be expended, and

choice to persist in the expenditure of that level of effort). A confirmatory factor analysis has given empirical support to his model (McCloy, Campbell & Cudeck, 1994).

A second relevant advance with regards to job performance is the suggestion made by Borman and Motowidlo (1993) to distinguish job performance as two different clusters of behaviors. They suggest that job performance includes both in-role behaviors as well as extra-role behaviors. The first behaviors assess task performance and the second behaviors assess what they call 'contextual' performance. The contextual performance construct is closely related to other constructs known as prosocial organizational behavior (Brief & Motowidlo, 1986), organizational citizenship (Munene, 1995; Organ, 1988), organizational spontaneity (George & Brief, 1992) and personal initiative (Frese, Fay, Hilburger et al., 1997). Borman and Motowidlo suggest that task performance is the incumbent proficiency when performing technical or services activities while contextual performance is the contribution made by the incumbent beyond the explicit job requirements (e.g., extra effort, cooperativeness, etc.). Borman and Motowidlo have also suggested that typical measures of job performance Include not only task proficiency but also contextual proficiency. However, these two dimensions of job performance are not related to the same predictors. Task proficiency should be related to predictors of knowledge and aptitude (e.g., cognitive ability tests, job knowledge tests) while contextual performance should be related to motivational predictors (e.g., personality factors). Borman and Motowidlo (1997) have provided a taxonomy of contextual performance which is composed of five categories: (1) persisting with enthusiasm and extra effort where necessary to complete one's task activities successfully; (2) volunteering to carry out task activities that are not formally part of one's job; (3) helping and

cooperation with others; (4) following organizational rules and procedures; (5) endorsing, supporting, and defending organizational objectives. According to Motowidlo, Borman and Schmit (1997), contextual and task performance are different in three ways. First, task activities vary across jobs whereas contextual activities are more general across jobs. Second, task activities are role-prescribed whereas contextual activities tend to be extra-role behaviors. Third, cognitive ability is a probable antecedent of task performance while personality and motivational variables are more probable antecedents of contextual performance.

Several studies supported the hypothesized differences between task performance and contextual performance. For example, Motowidlo and Van Scotter (1994) found that task and contextual performance contributed independently to overall performance. They also found that task proficiency was better predicted using cognitive ability while personality dimensions predicted contextual performance. Van Scotter and Motowidlo (1996) have made an effort to refine the construct of contextual performance by dividing it into two elements: interpersonal facilitation and job dedication. Interpersonal facilitation is the set of behaviors that contribute to organizational goal accomplishment (e.g., altruism, helping coworkers, acts that improve moral, cooperation, removing barriers). Job dedication consists of self-disciplined behaviors (e.g., following rules, working hard, taking initiative to problem-solving). Van Scotter and Motowidlo hypothesized that interpersonal facilitation was related to agreeableness, extraversion, high social confidence and positive affectivity. They also suggested that job dedication covaried with motivational and volitional traits such as conscientiousness, generalized expectancy of task success and goal orientation. Van Scotter and Motowidlo found that task performance correlated with experience, job knowledge,

conscientiousness and expectancy of task success. Interpersonal facilitation correlated with experience, job knowledge, conscientiousness, expectancy of task success, extraversion, agreeableness and positive affectivity. However, the motivational elements defined as job dedication were associated with task performance and correlated with predictors such as experience, job knowledge and conscientiousness. Based on these findings. Van Scotter and Motowidlo suggested that the definitions of task performance and interpersonal facilitation should be revised in order to include the motivational elements of job dedication.

A different view of job performance was suggested by Viswesvaran (1993). In his research, Viswesvaran (1993) found that 13 job content areas represented comprehensively the job performance domain. The dimensions were: (1) overall job performance; (2) job productivity; (3) quality; (4) problem-solving and leadership; (5) communication skills; (6) administrative skills; (7) effort; (8) interpersonal skills; (9) job knowledge; (10) compliance and acceptance of authority; (11) absenteeism; (12) accidents; and (13) tenure. Psychometric meta-analysis was used to develop the correlation matrix between these content areas, and a confirmatory factor analysis was used to obtain the latent structure of job performance. Confirmatory analysis was conducted on two matrices, one with the intercorrelations between the organizational record-based dimensions (e.g., absenteeism, accidents) and another with the intercorrelations between ratings of the different dimensions of job performance. For each matrix, there was a general factor underlying most performance measures, although this general factor does not fully explain the relationship among performance ratings.

Viswesvaran, Ones and Schmidt (1996) have carried out the most comprehensive meta-analysis on the reliability of job performance ratings up to date. Using the ten dimensions of job performance ratings found by Viswesvaran, they compared the interrater and intrarater reliabilities of the ratings. For overall job performance, the most significant results showed that interrater reliability of supervisory rating was 0.52 (sample size weighted) and 0.68 (frequency weighted), the interrater peer rating reliabilities were 0.42 and 0.44, respectively. The stability coefficients for supervisory ratings of overall job performance were 0.81 (sample size weighted) and 0.84 (frequency weighted). The alpha reliabilities for supervisory ratings were 0.86 and 0.84 (sample size and frequency weighted respectively). For peer ratings the alpha coefficients were 0.85 and 0.81 (sample size and frequency weighted respectively). Viswesvaran, Ones and Schmidt (1996) also found that 20% to 30% of rating variance of job performance dimensions of the average rater was specific to the rater. In connection with the criterion reliability, Salgado and Moscoso (1996) also carried out a study on interrater reliability of job performance ratings in validity studies of personnel selection, using setting (civilian versus military) and criterion composition (single versus composite) as moderators of reliability. They conducted three meta-analyses in which interrater reliability (frequency weighted) was estimated for civilian composite criteria, military composite criteria and a single civilian criterion. The average reliabilities were 0.64, 0.53 and 0.40 respectively. These results therefore suggest that the setting (civilian vs military) and the criterion composition (single vs composite) are relevant moderators of interrater reliability of job performance ratings.

An interesting experimental study was carried out by Sanchez and De La Torre (1996) to check the relation between rating and behavioral accuracy in performance appraisal. They found that the relationship between behavioral and rating accuracy was confined to dimensional judgments of strengths and weaknesses. Sanchez and De La Torre suggested that judgments were related to behavioral memories depending on how relevant and easily accessible these memories were.

Job Analysis

An important advance in job analysis was carried out by Peterson et al. (1995), who developed a prototype to build a database of occupational information (O*NET). This prototype incorporates comprehensive information of both worker and work attributes, while the current Dictionary of Occupational Titles is based only on task descriptions. In the O*NET, jobs will be described showing (a) worker requirements, (b) worker characteristics, (c) experience requirements, (d) occupational requirements, (e) labor market characteristics, and (f) occupation-specific requirements. In order to measure the variables used in the descriptions of jobs, nine questionnaires were developed: (1) skills; (2) knowledge; (3) training, education, licensure and experience; (4) generalized work activities; (5) work context; (6) organizational context; (7) abilities; (8) occupational values; and (9) work styles.

Goldstein, Zedeck and Schneider (1993) explored job analysis as a prerequisite for establishing content validity of personnel selection procedures. Goldstein, Zedeck and Schneider concluded that the job analysis-content validity process could be questionable if attention was not paid to the series of processes and judgments involved in establishing content validity. Connected with this last point, Arthur, Doverspike and Barrett (1996) presented a method

for determining the Relative Content Contribution (RCC) of each test in a battery, using a content-validity approach. The method uses job analysis information (e.g., KSAOs) to estimate the specific weights to be given to the tests. The RCC weight for each single test is a function of (a) importance of the work behavior, (b) time spent performing the work behavior, (c) consequence of error or cost associated with an error in the performance of the work behavior, and (d) time-to-proficiency. Arthur, Doverspike and Barrett presented the formulas to calculate RCC index and an illustrative case. Sanchez and Fraser (1992) examined the issue of which task scales to retain in task analysis. They examined interrater reliability, convergence among task scales and measurement of task importance. Sanchez and Fraser selected four typical task dimensions: (1) relative time spent; (2) difficulty in learning the task; (3) criticallty or consequences of error; and (4)overall importance. The conclusions suggested that the choice of scales for task analysis has to be made considering that (a) the interrater reliability was over 0.60, (b) criticality and importance were highly redundant scales, (c) ratings of time spent, difficulty in learning and importance or criticality provided independent information, and (d) there was no clearly superior algorithm in order to make a composite.

Two other advances which must be mentioned are those concerned with job analysis and the ‘Big Five’ personality dimensions. In France, Rolland and Mogenet (1994) developed the ‘Description in 5 Dimensions (D5D)’ system, a computerized job analysis system based on the five factor model of personality. The D5D system provides a personality-related job profile based on ipsative scores. Also, applicant personality may be assessed with the D5D and a profile computed. With the D5D, different profile comparisons can be made using graphical presentation

(e.g., job profile–applicant profile; group profile-applicant profile). The second advance was carried out by Raymark, Schmit and Guion (1997), who developed the Personality-Related Position Requirements Form (PPRF), a job analysis questionnaire to identify relevant personality traits for job performance. They noted that typical inventories of job analysis did not include items from which to generate hypotheses concerning personality variables that would be relevant for a job. For these reasons, the PPRF is based on the Big Five personality taxonomy and it is composed of 12 subdimensions. The first results indicated that the PPRF was a highly reliable instrument. Internal consistency reliabilities were mostly in the eighties and interrater reliability ranged from 0.85 to 0.97. Raymark, Schmit and Guion suggested that PPRF could be potentially useful to identify personality variables to be measured as predictors but PPRF could also be useful to develop criterion measures related to the personality dimensions associated with job behaviors.

COGNITIVE ABILITY

Abilities have been the most prominent predictors in personnel selection since the first decade of this century. However, Landy, Shankster and Kohler (1994) pointed out that I/O psychologists are currently doing little research in the development and validation of new cognitive ability tests beyond the classic ones. According to Landy, Shankster and Kohler (1994), there is slow progress as a result in the area.

Two principal lines of work may be distinguished in this topic which are related to personnel selection. One is that of psychometric-*g* proponents, who essentially suggest that specific abilities do not add incremental validity beyond *g* (e.g. Olea & Ree, 1994; Ree, Earles & Teachout, 1994;

Schmidt, 1994). The second line is associated with the proponents of multiple abilities (e.g., Ackerman & Kanfer, 1993; Ackerman & Goff, 1995).

In connection with the first line of research, Carretta, Ree and their colleagues are carrying out a program on this topic, with emphasis on construct validity and criterion validity. The results of Carretta and Ree's research program show a strong support for *g* validity (see Ree & Carretta, 1997, 1998). Ree and Carretta (1994a) have shown that the Armed Services Vocational Aptitude Battery (ASVAB) has a hierarchical structure, including a higher order factor for *g* and the three lower order factors of speed (NO, CS), verbal/math (AR, WK, PC, MK), and technical knowledge (GS, AS, MC, EM). All ten tests load on *g*. The percentage of total variance accounted for by these factors was 63.8% for *g*, 6.2% for speed, 2.4% for verbal/math and 7.7% for technical knowledge. Stauffer, Ree and Carretta (1996), using hierarchic confirmatory factor analysis, showed that a battery containing 10 traditional paper-and-pencil aptitude tests (ASVAB) and a second battery of 25 cognitive-component-based tests (based on Kyllonen's 1993, 1994 work) appear to measure the same factor which they identified as *g*. In this study, the correlation between the higher order factors from the two batteries was 0.994, indicating that both measured *g*. Also, Carretta, Perry and Ree (1996) showed the validity of several cognitive-components-based tests for predicting aircraft pilot performance.

The Air Force Officer Qualifying Test (AFOQT) has been validated for the selection of pilots and navigators, as well as many other jobs within the Air Forces (Carretta & Ree, 1995; 1996; Olea & Ree, 1994). For pilot and navigator training, the criteria included passing–failing training, average flying, performance rankings, composites of specific check side performances, and day and night

celestial navigational tests,, Olea and Ree (1994) showed that the major source of validity for the AFOQT came from its measurement of *g*. They also showed that the incremental validity of the AFOQT subtest beyond *g* was approximately 0.02 (0.462 vs 0.482 for the summed composite criterion) for navigators across five individual training criteria and a summed composite of five training criteria,, The incremental validity of non-*g* portions of the AFOQT for predicting pilot training success was about 0.08 (0.314 vs 0.398) for the summed composite criterion.

Ackerman and his colleagues (Ackerman & Goff, 1995; Ackerman & Heggestad, 1997) have developed a theory of intelligence that endeavours to integrate literature regarding the relations among personality, interest and intellectual development. The theory is known as PPIK. It is made up of four components: Intelligence-as-process, Personality, Interest and Intelligence as knowledge, In this theory, the construct of intelligence as typical performance is basic (Ackerman, 1994). According to Ackerman, one reason why intelligence tests do not show a correlation higher than 0.5 or 0.6 for academic performance or job performance is that intelligence is measured with a ‘maximal’ paradigm while academic and occupational performance takes place in a ‘typical’ environment (Ackerman & Goff, 1995). A measure of intelligence as typical performance should be more highly correlated with crystalized abilities (e.g., knowledge) and intelligence as maximal performance with fluid abilities (e.g., abstract reasoning). Some results appear to support this last prediction (Ackerman & Goff, 1995; Ackerman, Kanfer & Goff, 1995; Goff & Ackerman, 1992). Ackerman and Goff (1995) and Ackerman and Heggestad (1997) provided a set of extensive meta-analyses of the correlations between cognitive abilities, personality dimensions and career interests. Ackerman and Kanfer (1993) designed the

Aptitude Assessment Battery (AAB), which is a cognitive test battery. The AAB contains nine ability tests and an 18-item measure of motivational skills. This battery was used for the selection of air traffic controllers and demonstrated its validity.

A relevant study was conducted by Levine, Spector, Menon, Narayanon and Canon-Bowers (1996). In a meta-analysis of craft jobs in the utility industry (e.g., electrical assembly, telephone technicians, mechanical jobs), they found that cognitive ability tests (e.g., tests of verbal ability, numerical ability, reasoning ability and memory tests) showed a corrected average validity of 0.43 ($k = 149$, $n = 12504$) for predicting job performance, with some validity variation across job families. Using training criteria, the corrected average validity was 0.67 ($k = 52$, $n = 5872$), a remarkably higher value than that found for job performance. Taken together these two findings confirmed Hunter and Hunter's (1984) results, which found that the average corrected validity of cognitive ability tests was 0.45 for performance ratings and 0.54 for training success.

An interesting result was found by Pettersen and Tziner (1995). They used two specific intelligence tests (Otis and Beta) to predict job performance in a manufacturing enterprise. They found that the predictive validity of both tests was stable over time. Another finding was that job complexity affected the predictive validity, with a higher validity when job complexity increased. Pettersen and Tziner suggested that the intelligence tests were a strong predictor, particularly with complex jobs. Pettersen and Tziner's findings contradicted the proposition of Hulin, Henry and Noon (1990), who proposed that validity varied as a function of time on the job. Contrary to Hulin, Henry and Noon (1990), Pettersen and Tziner's results suggest that predictive validity of intelligence is stable and does not change significantly as a function of time, as long as the job

content remains the same. Salgado (1998a) found that cognitive ability tests were predictors of training proficiency in four samples of pilot trainees. He also found that all observed variability in 15 validity coefficients was accounted for by sampling error variance. Schuler, Moser, Diemand and Funke (1995) found that grades at the end of an apprenticeship training in a financial organization were predicted by cognitive abilities tests ($r = 0.55$; validity corrected for attenuation). Bartram and Baxter (1996) found that MICROPAT, a computer-based test battery, predicted flying grading and a pass-fail training criterion in a civilian pilot sample.

In summary, research carried out in the 1990s on cognitive abilities shows that they continue to be one of the best (if not the best) predictors of job performance and training performance. A future line of research will be to check Borman and Motowidlo's (1993) suggestion that cognitive abilities predict task performance and not contextual performance. Some recent studies indicate that this last suggestion may be correct but more research seems necessary.

PHYSICAL AND PSYCHOMOTOR ABILITIES

Not much literature on physical and psychomotor abilities has been published in the past few years, although some relevant contributions must be commented upon. J. Hogan (1991a) comprehensively reviewed the literature on physical, aptitudes and integrated knowledge from different disciplines, including personnel selection, differential psychology, industrial engineering, biomechanics, work physiology and law. Hogan's review is probably the best review of physical aptitudes up to date. J. Hogan (1991b) also carried out several studies on the

structure of physical requirements of occupational tasks. Her results showed that three components—muscular strength, cardiovascular endurance and movement quality—described those requirements. This three-component structure was found in both job analysis and test performance data and it appeared to be independent of jobs, raters or incumbents' performance levels. J. Hogan, Arneson and Petersons (1992) developed and validated a physical performance test battery to select high-pressure cleaning workers. The battery included two types of predictor measures: ability tests and work sample simulations. J. Hogan and Quigley (1994) examined the effects of training and nontraining preparation on physical ability test-taking. Their results showed that preparation was capable of maximizing selection opportunities of qualified applicants for physically demanding jobs. Also they found that preparation could reduce adverse impact. Arvey, Landon, Nutting and Maxwell (1992) showed that eight physical ability tests used for police officers' selection were related to two constructs: strength and endurance. Blakley, Quiñones, Crawford and Jago (1994) reported the results of a meta-analysis of six studies in which isometric strength tests were used as selection procedures. They found that isometric strength tests were valid predictors of supervisory ratings of physical performance ($rho = 0.32$) and performance on work simulations ($rho = 0.55$).

With regard to psychomotor abilities, Carretta and Ree (1994) and Ree and Carretta (1994b) showed that the BAT, a psychomotor test battery, added a small incremental validity beyond the validity of cognitive tests. They suggested that this small incremental validity was due to the commonality of measurement. They also found that the psychomotor scores were *g* saturated. In a meta-analysis of validity studies carried out in Spain using psychomotor tests as predictors, Salgado (1994a) found a corrected

average of 0.42. Salgado (1995) found that a psychomotor test (Rotary Pursuit Test) predicted the performance in a practical driving examination. Using cumulative techniques he found a validity of 0.26 (uncorrected for unreliability and range restriction). He also found that there was no variability across four samples of drivers. Levine et al. (1996), in a meta-analysis of craft jobs in the utility industry (e.g., electrical assembly, telephone technician, mechanical jobs), found a corrected average validity of 0.34 ($k = 147$, $n = 12677$) for predicting job performance, with some validity variation in job families. For example, electrical assembly jobs and mechanical jobs were best predicted (0.49 and 0.42 respectively). Using training criteria, the corrected average validity was 0.35 ($k = 41$, $n = 3042$), a similar value to the one found for job performance. In the case of training criteria, newly created electrical assembly and mechanical jobs presented the best validities (0.58 and 0.60 respectively).

In summary, recent evidence indicates that physical and psychomotor ability tests can be useful tools in physically demanding jobs. Also, some findings suggest that psychomotor adds little variance over cognitive ability tests. This may not be the case for physical ability tests. However, the current data are not conclusive.

JOB KNOWLEDGE TESTS

Under this rubric, three types of measures can be included: job knowledge tests (Hunter & Hunter, 1984), tacit knowledge tests (Sternberg, Wagner, Williams & Horvath, 1995), and situational judgment tests (Pulakos & Schmitt, 1996). In connection with job knowledge and tacit knowledge constructs, an interesting debate took place between Schmidt and Hunter (1993) and Sternberg and Wagner (1993). According to Schmidt and Hunter (1993)

tacit knowledge is not a new construct. It is just changing the name of job knowledge which is a broader construct.

Dye, Reck and McDaniel (1993) examined the validity of written job knowledge tests as predictors of job performance and training success. Their database consisted of 502 validity coefficients and a total sample of 363 528 persons. They found a corrected mean validity of 0.45 for job performance and 0.47 for training success. Dye, Reck and McDaniel also found that validity was moderated by job-test similarity and job complexity. For the first moderator, results showed that validity was almost double (*rho* = 0.62 and *rho* = 0.76 for job performance and training success respectively) for high job-test similarity what it was for low job-test similarity (*rho* = 0.35 and *rho* = 0.46 for job performance and training success respectively). For the second moderator results showed that validity was greater for jobs of high complexity than for jobs of low complexity. For example, against a job performance criterion the corrected validities were 0.57 versus 0.39 respectively, for high and low complexity jobs. As regards training success, validities were 0.57 versus 0.46 for high and low job complexity respectively. Dye, Reck and McDaniel suggested to employers that when job knowledge tests were to be used, it was better to make them as job-specific as possible. They also suggested that job knowledge tests were better for high complexity jobs (jobs that demanded greater levels of knowledge and required greater judgment and synthesis of job knowledge). Borman, Hanson, Oppler et al. (1993) showed that job knowledge mediated in relations between cognitive ability and job performance ratings. This mediated effect was demonstrated previously by Hunter (1986). In connection with this last point, Pulakos, Schmitt and Chan (1996) found that job knowledge, assessed with a 117-itern measure, had direct and indirect effects on performance

ratings made by supervisors. The job knowledge test had a direct effect of 0.11 and an indirect effect of 0.04 on job performance.

The second line of research (i.e., practical intelligence) is associated with Sternberg, Wagner and their colleagues (Sternberg, 1997; Sternberg & Wagner, 1993; Sternberg et al., 1995; Wagner, 1994; Wagner & Sternberg, 1990). These authors suggested distinguishing between academic intelligence (as measured with intelligence test scores) and practical intelligence or common sense. In addition to the distinction between academic and practical kinds of intelligence, Sternberg and Wagner made a similar distinction between two types of knowledge: formal academic knowledge and tacit knowledge. The first is knowledge sampled by the ability tests. Tacit knowledge is action-oriented knowledge, acquired without direct help from others, that allows individuals to achieve goals they personally value. Tacit knowledge has three characteristics: (1) it is procedural in nature; (2) it is relevant to the attainment of goals people value; (3) it is acquired with little help from others. Tacit knowledge is typically measured with instruments containing a set of work-related situations, each having between 5 and 20 response items. Each situation poses a problem for the participant to solve by rating the various response items. Sternberg and Wagner also suggest that work problems are often (a) unformulated or in need of reformulation, (b) of personal interest, (c) lacking in information necessary for solution, (d) related to everyday experience, (e) poorly defined, (f) characterized by multiple ‘correct’ solutions, each with liabilities as well as assets, and (g) characterized by multiple methods for picking a problem solution. According to Sternberg (1997; Sternberg et al., 1995), the correlations between tacit knowledge and job criteria typically ranged from 0.2 to 0.4. However, a problem with

these last studies is that the samples were of a small size and restricted range. Furthermore, Sue-Chan, Latham and Evans (1997) found that a tacit knowledge test for nurses had no incremental validity beyond *g* for predicting the first year and second year grade-point average.

The third type of job knowledge measures commented upon here are the situational judgment tests. Motowidlo and Tippins (1993), using a situational inventory which they named as low-fidelity simulation, found an overall validity of 0.28. Sanchez and Fraser (1993) developed and validated the Corporate Social Style Inventory (CSSI), a measure of customer service skills. The CSSI consisted of multiple-choice items that presented individuals with hypothetical situations and asked respondents how they would handle each situation. The results of eight studies indicated that CSSI had an acceptable test-retest reliability and showed criterion validity as well as incremental validity over cognitive ability to predict job-performance ratings. Pulakos and Schmitt (1996) found that a situational judgment test (SJT), a practical intelligence test, was a valid predictor of job performance. They also found that the correlation between cognitive ability measures and the SJT was very small (r=0.13, r = 0.07 and r=0.17). Moreover, the SJT showed incremental validity beyond the cognitive ability tests. Pulakos, Schmitt and Chan (1996) tested a model of supervisory performance ratings that included measures of cognitive ability, practical intelligence, job knowledge, task proficiency, achievement orientation, and performance ratings. They suggested that cognitive ability and practical intelligence were different constructs and that both had influenced performance ratings through job knowledge. In addition, the model proposed that practical intelligence directly influenced performance ratings. The results supported the model and showed that practical intelligence had a direct effect of 0.128 and an indirect

effect of 0.024. Cognitive ability had an indirect effect of 0.124. McDaniel, Finnegan, Morgeson, Campion and Braveman (1997) examined the criterion validity of situational judgment (SJ) tests. They found that the corrected validity was 0.56. Furthermore, McDaniel et al. (1997) reported that SI tests had an average corrected correlation of 0.53 with general mental ability tests. This correlation reached 0.68 when the SI tests were based on a job analysis and fell to 0.29 when the tests were not based on a job analysis. These findings have important consequences for the constructs of situational judgment, tacit knowledge and practical intelligence. An important percentage of their predictive capacity can be due to the correlation with general mental ability, contradicting the opinion of 'street smart' supporters (e.g., Sternberg et al., 1995; Wagner, 1994) and giving support to Schmidt and Hunter's (1993) position.

PERSONALITY

Personality and personality measures have received considerable interest in the past six years, in part due to the consolidation of the Five Factor Model of personality as a paradigm of the area. Barrick and Mount (1991), in a large-sample meta-analysis of the Big Five dimensions, found that the validity of conscientiousness construct was 0.13 (0.22 corrected for criterion reliability and range restriction) and generalized across occupations and criteria. Tett, Jackson and Rothstein (1991), using studies in which the choice of personality measures was based on a job analysis or guided by hypotheses, found that confirmatory strategies moderated the validity of the Big Five. Hough (1992, 1997a) found that seven personality factors were needed to cover the personality domain in a more comprehensive fashion, and she also found that

neuroticism, dependability and achievement (both facets of conscientiousness) generalized validity across occupations and criteria. More recently, Mount and Barrick (1995) suggested that previous meta-analyses may have understated the true validity of conscientiousness and affirmed that values of 0.18 (0.32 corrected for criterion reliability, range restriction and imperfect construct measurement) for overall job proficiency and 0.17 (0.30 corrected for criterion reliability, range restriction and imperfect construct measurement) for training proficiency criterion were probably closer to the true validity. All these meta-analyses were carried out using studies conducted with American samples. In order to know if validity generalization of the Big Five has geographical boundaries, Salgado (1997a, b, 1998b) conducted some meta-analyses using only European studies. He found that two factors generalized validity across occupations and criteria: conscientiousness and emotional stability. As a whole, all these meta-analyses suggest that two personality dimensions, conscientiousness and emotional stability (neuroticism) are valid predictors for all jobs and criteria. Recently, Barrick and Mount (1996) confirmed this suggestion in two samples in which conscientiousness and emotional stability were valid predictors of supervisory ratings and voluntary turnover in the two samples.

A number of articles provided evidence of criterion-related validity for some of the most popular Big Five-based questionnaires. For example, Piedmont and Weistein (1994) found validity evidence for concientiousness as it is assessed with the NEO-PI-R (Costa and McCrae, 1992) and Salgado and Rumbo (1997) found that a composite of the Big Five, assessed with the NEO-FFI (Costa and McCrae, 1992), was a valid predictor for financial service managers. Costa (1996) made a summary of reliability and validity of the NEO-PI-R and discussed its use in industrial and

organizational psychology. Costa, McCrae and Kay (1995) also presented a NEO-PI-R-based profiler to be used in job analysis. Barrick and Mount (1993; Barrick, Mount & Strauss, 1993; Mount & Barrick, 1995) found evidence for the PCI (Barrick and Mount, 1996). Hogan and Hogan (1995) published the second edition of the manual for the Hogan Personality Inventory (HPI) that included a large list of studies in which the HPI was used and the corresponding validity found. In the European context, several researchers have presented empirical evidence supporting work-oriented personality questionnaires based on the five factor model. Bartram (1993) presented validity evidence for the ICES, Rolland and Magenet (1994) developed the 'Description in Five Dimensions' questionnaire and Salgado (1996a) showed construct-validity evidence of the IP/5F (Salgado, 1994b).

Two additional lines of evidence in favor of personality measures as predictors of work behavior were presented by Robertson and his associates and Day and Bedeian (1991, 1995). Robertson and Kinder (1993; see also Salgado, 1996b) found that personality predicted a large number of job competencies and validity was generalized across occupations. Nyfield, Gibbons, Baron and Robertson (1995) found that the Big Five dimensions which were assessed using the scales of the Occupational Personality Questionnaire (OPQ; Saville, Holdsworth, Nyfield et al., 1984) predicted managerial competencies across three different countries: USA, UK and Turkey. Saville, Sik, Nyfield et al. (1996) also presented a demonstration of the OPQ validity to predict job competencies in UK managers. Day and Bedeian (1991, 1995) observed interactions between personality factors (conscientiousness and agreeableness) and psychological climate dimensions to predict job performance, job satisfaction and tenure.

A number of articles checked the validity of complex personality variables such as integrity and customer service orientation. Ones, Viswesvaran and Schmidt (1993), in a large meta-analysis of integrity tests, found a validity value of 0.42 for predicting job performance and a value of 0.38 for predicting training proficiency. In two samples of prisoners previously employed in upper-level positions of authority, Collins and Schmidt (1993) found that the best single predictor that differentiated between offenders and nonoffenders was a personality-based integrity test. Ones (1993), using meta-analytic techniques, demonstrated that integrity measures were composed of conscientiousness, agreeableness and neuroticism, and that the correlation between integrity and cognitive ability was practically null. Murphy and Lee (1994) showed that conscientiousness explained a small percentage of integrity test validity. Sackett and Wanek (1996) presented the fourth report in a series of reviews of the use of integrity, checking the literature on criterion validity, construct validity (including the Big Five, cognitive ability, moral reasoning and social desirability), applicant reactions and legal effects. Hogan and Brinkmeyer (1997) investigated whether overt and personality-based integrity tests had the same psychological meaning. In a large applicant sample (n = 2168), they showed that overt and personality-based integrity tests were made up of four themes: (1) punitive attitudes; (2) admissions of illegal drug use; (3) reliability; and (4) theft admissions. Correlating the HPI, punitive attitudes, admissions of illegal drug use, theft admissions and the Abbreviated Reid Report, Hogan and Brinkmeyer (1997) also found that punitive attitudes and theft admission components correlated with prudence (conscientiousness) adjustment (emotional stability) and likeability (agreeableness). However, there were correlations with no practical significance between illegal drug use component and HPI scales. As a whole, the

Abbreviated Reid Report correlated with prudence (r = 0.50), adjustment (r = 0.44), likeability (r = 0.26), ambition (r = 0.28) and school success (r = 0.26). These findings are a solid indication of the psychological meaning of an overt integrity test.

Frei and McDaniel (1997) carried out a meta-analysis of service-oriented questionnaires. The criterion validity was based on 41 coefficients with a total sample size of 6945. The mean validity corrected for criterion unreliability and range restriction was 0.50 (observed r=0.24). They found strong relationships between the Big Five and customer service orientation measures. Emotional stability correlated 0.63 (n = 15479), agreeableness correlated 0.61 (n = 14786) and conscientiousness correlated 0.42 (n = 95057). They also found that customer service measures correlated with sales orientation (r = 0.31, n = 2101) but did not correlate with cognitive ability (r = 0.06, n = 753). Furnham (1995) found that the Customer Service Questionnaire (CSQ; Saville & Holdsworth, 1992) predicted job performance in an air company. Furthermore, Furnham and Coveney (1996) reported that CSQ correlated primarily with conscientiousness and neuroticism as measured by the NEO-PI-R. Ones and Viswesvaran (1996b) indicated that service-orientation measures were based on emotional stability, agreeableness and conscientiousness, and they also found a large correlation between customer service scales and integrity tests.

Taking together the meta-analyses of the Big Five validity (e.g., Barrick & Mount, 1991; Hough, 1992, 1997b; Salgado, 1997a,b, 1998b; Tett, Jackson & Rothstein, 1991) with the findings of the integrity tests meta-analysis (e.g., Ones, Viswesvaran & Schmidt, 1993) and with the findings of service-orientation measures (e.g., Frei & McDaniel, 1997), as well as the results of research carried out to analyze the constructs underlined on integrity and