Stockholm
University

Utilizing research in the practice of personnel selection

General mental ability, personality, and job performance

Sofia Sjöberg

© Sofia Sjöberg, Stockholm University 2014

ISBN 978-91-7447-883-9

Printed in Sweden by Printcenter US-AB, Stockholm, 2014
Distributor: Department of Psychology, Stockholm University, Sweden.

To my family — for support,
inspiration and joy!

Abstract

Identifying and hiring the highest performers is essential for organizations to remain competitive. Research has provided effective guidelines for this but important aspects of these evidence-based processes have yet to gain acceptance among practitioners. The general aim of this thesis was to help narrowing the gap between research and practice concerning personnel selection decisions. The first study compared the validity estimates of general mental ability (GMA) and the five factor model of personality traits as predictors of job performance, finding that, when the recently developed indirect correction for range restriction was applied, GMA was an even stronger predictor of job performance than previously found, while the predictive validity of the personality traits remained at similar levels. The approach used for data collection and combination is crucial to forming an overall assessment of applicants for selection decisions and has a great impact on the validity of the decision. The second study compared the financial outcomes of applying a mechanical or clinical approach to combining predictor scores. The results showed that the mechanical approach can result in a substantial increase in overall utility. The third study examined the potential influences that practitioners' cognitive decision-making style, accountability for the assessment process, and responsibility for the selection decision had on their hiring approach preferences. The results showed that practitioners scoring high on intuitive decision-making style preferred a clinical hiring approach, while the contextual aspects did not impact practitioners' preferences. While more research may be needed on practitioner preferences for a particular approach, the overall results of this thesis support and strengthen the predictive validity of GMA and personality traits, and indicate that the mechanical approach to data combination provides increased utility for organizations.

Keywords: Personnel selection, job performance, correction for range restriction, general mental ability, personality, clinical and mechanical data collection, clinical and mechanical data combination, utility, preference for hiring approach.

Acknowledgements

I guess every thesis has its own history of *why* it came about. My inspiration for this thesis grew slowly as I became aware of the extensive gap between what was suggested by research and what went on in selection practice. Inspired by the work of prominent but somewhat disillusioned researchers in this field and Professor Emeritus Hunter Mabon's stubborn arguing for the use and importance of utility analysis both in general and in relation to personnel selection, the consequences of this gap became tangible, and without this awakening this thesis would certainly never have been written.

True inspiration also comes from those rare practitioners who, I know, struggle as consultants, internal human resource professionals, and hiring managers. You argue for the implementation of evidence-based methods and processes both within yourselves, within your organizations, and towards customers. You make this work worthwhile so keep up the good work – you are the true role models bringing professionalism into the personnel selection industry.

A thesis is rarely produced by the author alone and this thesis is certainly no exception. *How* this thesis came about was through a collective effort. First, I would like to thank my supervisor Professor Magnus Sverke for all his support in the writing of this thesis. For his patient guidance and thorough feedback at all levels, I am forever thankful. I am also deeply thankful to my co-supervisor and friend Associate Professor Katharina Näswall who stuck by me although she left to live and work on the other side of the globe. Your positive attitude and encouragement when I most needed it kept me from giving up several times. Thank you!

A special thank goes to the co-authors for one of my papers, Kristina Langhammer and Thomas Lindevall – for their fun, creative, and energetic work. Together we proved that it is possible to accomplish a lot with scarce resources.

I also want to thank David Speeckaert for, with patience and confidence, helping me with the language and structure of the content, and Sofia Halldén for helping me out with my pressparagraph. Communicating research is a challenge but with your help I think we have done a terrific job!

I feel somewhat sorry for all of the other Ph.D students and Ph.D students to come since the best roommate ever, Niklas Hansen, has already been taken and used. Thank you Niklas for many laughs, inspiring discussions, and support in times of doubt. You made my time at the department a sheer joy!

My positive experiences at the department are, of course, also thanks to all my colleagues at the Department of Psychology at Stockholm University who have provided many inspiring lunch discussions and an intellectual and challenging work environment. Being around such intelligent people is truly rewarding.

I want to thank the two reviewers of my thesis, Associate Professor Claudia Bernhard-Oettel and Professor Gerry Larsson, for their valuable comments, and Associate Professor Johnny Hellgren for help with preparing my defense.

I have shared my time working on my thesis with my work developing psychological tests, first at Assessio and currently at Pearson Assessment. Without all my dear and exceptionally competent colleagues I would never have considered starting my thesis work and certainly not finished it. Thank you all, you are the best!

I would like to dedicate a special thanks to Katarina Forssén, my informal mentor, colleague, and supervisor. Katarina hired me, an inexperienced master student, almost fifteen years ago, and then hired me again ten years later when in the middle of my thesis work, despite knowing that my focus would be somewhere else the following years. This has given me invaluable confidence in my competence and value in a non-academic work setting. In addition, she has also been a true resource when writing the kappa — I owe you more than a box of wine!

My final note is directed to my (neglected) friends and big family. I want to thank my parents, Helene and Per, for always believing in me, providing me with stability and encouraging critical thinking without telling me what to do or how to live my life. I also want to thank the best sister in the world, Anja, and my big (well, at least on paper) brothers, Johan and Martin, and their families for just being themselves and always being there for me. A special thanks also goes to my stepchildren, Simon, Sara, and Hugo, who have enriched my life in a very special way and prepared me for the role of being a parent.

Finally, the most sincere and deepest thanks go to my true inspiration, my co-author, colleague, opponent, friend, and dearest husband: Anders Sjöberg — my best selection decision ever! Absolutely nothing would have been possible without you. At last, as everybody knows, the best things come in twos: my magnificent girls, Stella and Blenda, have taught me about priorities and made me aware of the true meaning of life. This is for you.

Sofia Sjöberg
Stockholm, March 2014

List of Studies

I. Sjöberg, S., Sjöberg, A., Näswall, K., & Sverke, M. (2012). Using individual differences to predict job performance: Correcting for direct and indirect restriction of range. *Scandinavian Journal of Psychology, 53*, 368–373.

Reprinted with permission. © The Scandinavian Psychological Associations, Blackwell Publishing Ltd.

II. Sjöberg, S. (submitted). The utility gain of leaving professional judgment outside of prediction: Clinical versus mechanical interpretation of GMA and personality. *Academy of Management Perspectives.*

III. Sjöberg, S., Langhammer, K., Sjöberg, A., & Lindevall, T., (submitted). Preference for hiring approach: Cognitive style or context dependent? *International Journal of Selection and Assessment.*

Contents

Introduction 1
  General aim 5

Job performance 7
  Theory of job performance 7
  Task performance 8
  Contextual performance 8
  Counterproductive work behavior 9

Individual differences 10
  General mental ability and its usefulness in personnel selection 10
  Personality and its usefulness in personnel selection 13
    Measurement of personality 13
    The five factor model of personality 15
    Personality and job performance 17

Threats to validity in the context of personnel selection 19
  Measurement error in personnel selection 19
  Constructs versus methods as predictors 20
  Correction for range restriction 21

A utility framework in personnel selection 23
  Utility analysis 23
  The monetary value of job performance 25
  The use of multiple predictors 28

Decision making in personnel selection 31
  Data collection 32
  Data combination 33
  Data collection and combination for prediction 34

Summary of studies 37
  Study I 37
  Study II 39
  Study III 40

Discussion 43
  Data collection 44

  Data combination..47
  Practitioners' preferences for hiring approach 49
  Methodological considerations and future research 50
  Conclusions 52

References 54

Introduction

The research and practice of personnel selection has involved researchers, psychologists, and practitioners from a wide range of backgrounds for more than a hundred years (Vinchur, 2007). The use of tests by psychologists for employee selection began early in the 20th century in connection with the employment needs of the military during World War I (Thorndike & Hagen, 1961). Since then, the scientific literature related to personnel selection has continued to grow, the number of selection decisions in practice has steadily increased, and the financial impact of selection decisions for organizations, societies, and individuals has become more pronounced (e.g., Cascio, 2000; Cascio & Bodreau, 2008). Over the last few decades, the nature of work has shifted and organizational assets have become mainly dependent upon intellectual or human capital rather than fixed and physical assets. This shift has increased the financial importance of human resource management in general, and today all aspects of staffing, including personnel selection, play a crucial role for organizational survival and competitiveness (e.g., Kim & Ployhart, 2013). Therefore, attracting, identifying, hiring, and retaining the most suitable and productive employees are now among the biggest challenges for modern organizations (e.g., Cascio, 2000; Cascio & Bodreau, 2008).

Fortunately, the research-based knowledge on the proper procedures for assessing applicants and for making informed selection decisions based on these assessments is extensive. This accumulated knowledge constitutes a general framework which not only can be used to identify the individual differences that are characteristic of good performers, but which also describes the procedures for effectively assessing individual differences for personnel selection purposes (Barrick, Mount, & Judge, 2001; Salgado & Anderson, 2003). This knowledge can be characterized by three major components. The first is that the assessment of individuals carried out for selection purposes concerns the measurable differences between individuals that are relevant to job performance. This implies that non-measurable variables, as well as variables where there is no variation between individuals, are excluded and that other factors, such as formally required education, work experience, and so forth, are taken into account prior to the phase of assessing individual differences. Second, the activities carried out and methods used when assessing the individual differences should contribute to maximizing the predictive validity and thus maximize the accurateness of the ranking of the applicants on job performance. Activities serving other purposes than prediction of job performance, for example, extensive feedback processes, are not included since they do not aim to predict job performance. And third, in order to evaluate and compare the efficiency of the predictors, processes, and selection decisions, a financial aspect is required. Without analyzing the overall gain and cost with different selection alternatives, there is no possibility of making informed and sound decisions of how to design selection processes.

Given the above framework, it is clear that personnel selection is dependent upon having an accurate measurement of individual differences among applicants. Research on individual differences, specifically intelligence and personality traits, started with the pioneering work of researchers such as Cattell (1890), Spearman (1904), Binet and Simon (1916), and Allport and Odbert (1936). With Sir Francis Galton (1822-1911) laying the groundwork for both individual differences and the calculation of the correlation coefficient, much of the progress in this area can be attributed to major advancements in quantification, measurement, and methodology over the years. Hunter and Schmidt's (2004) development of a methodology for conducting meta-analyses, along with Hunter, Schmidt, and Le's (2006) methodology for correcting for the restriction of range (Thorndike, 1949), constitute important contemporary contributions to this and other fields of research.

During the first part of the 20th century, research and practice concerning the measurement of individual differences in connection to performance were closely aligned as developments in research were being applied in selection practice at this time. The relevance of using a psychometric approach when developing psychological tests for measuring personality, and especially intelligence for the prediction of job performance, was established and became the standard among practitioners (Vinchur, 2007). By the middle of the century, however, there was considerable disillusionment among researchers over the use of tests designed to measure individual differences for personnel selection. Due to variations in the results of different empirical studies on predictive validity related to job performance, the perception arose that this method of testing was not as effective as had been previously claimed (Hale, 1992). This debate was fueled by Mischel's (1968) argument that the individual's behavior was highly dependent upon situational cues, rather than expressed consistently across diverse situations that differed in meaning. This argument called into question the very foundation of trait theory and thereby its relevance for prediction and measurement, regardless of the method used. In addition, there were critics who considered it to be dehumanizing to workers to mechanically aggregate their predictor scores into a unified assessment by using an algorithm (Viteles, 1932). Eventually, however, the use of professional judgment as the basis for the aggregation of data was endorsed by parts of the research community and became the main approach in practice. This discussion has continued ever since and the disagreement between Viteles (1925) and Freyd (1926), who argued for the mechanical approach, in the early 20th century could just as well be taken out of contemporary literature. In the decades following the period of severe critique and skepticism during the 1960s, the status and relevance of individual traits were reclaimed in this area of research, at least to a certain extent. In recent decades, there have been advancements and refinements in the methods for reaching reliable, valid, and fair selection decisions based on predictions using individual differences (Barrick, Mount, & Judge, 2001; Salgado & Anderson, 2003). However, with a few exceptions, practitioners have been reluctant to utilize these advancements (Highhouse, 2008).

Personnel selection decisions are generally preceded by several considerations and decisions by the practitioner concerning what constructs to measure, what method(s) to use for measurement, how to aggregate or combine the data into an overall assessment, how to rank order applicants, and, finally, how to make the actual selection decision. A lot of effort has been put into refining the first of these, focusing on identifying relevant predictors of behavior and performance at work and establishing their level of importance. Two aspects of individual characteristics have appeared to be more important than others. The first is intelligence or, as it is more commonly referred to in work psychology, general mental ability (Spearman, 1904), which has become a central predictor due to having the strongest relationship with general job performance (Schmidt & Hunter, 2004). In addition to general mental ability, personality has been shown to be relevant for the prediction of job performance, for at least some personality traits, and for some aspects of job performance (Barrick & Mount, 1991).

Research addressing the issue of how to measure these constructs suggests that standardized psychological tests are the most feasible and cost-efficient way of assessing general mental ability and personality for the purposes of personnel selection (Schmidt & Hunter, 1998), and that test scores from this combination of predictors maximizes the predictive validity for job performance (Schmidt, Schaffer & Oh, 2008). This conclusion is based on research showing that both general mental ability and personality predict job performance (albeit to very different extents), they were found to be mutually independent and non-correlated, and the cost of implementing them as predictors was low. Altogether, this combination of predictors maximizes the predictive strength at a low cost, and represents a method with wide applicability compared to other competing methods (e.g., interviews, assessment centers, work samples).

However, the conceptions among practitioners about what should be measured and what the most reliable and efficient methods are for measurement often stand in contrast to what scientific evidence argues.

Personality, for example, is among practitioners often believed to be extremely important for job performance, while general mental ability is considered less important, if relevant at all. In addition, the trust in standardized methods, especially psychological tests, is often low and the perceived need for less standardized methods, such as the unstructured interview and reference checks is common among practitioners (Langhammer, 2013).

Another inevitable activity when making personnel selection decisions is the aggregation, or combination of data into a unified and overall assessment. No matter how data is collected, the data needs to be combined into a unified assessment to enable decision-making. From a scientific point of view, it is clear that in order to reach the maximized level of validity for a set of predictors, each applicant's set of predictor scores needs to be weighted according to the established validity for each predictor. Thus, the predictive validity estimates provide information about how the predictors, general mental ability and personality, relate to job performance, and also correspond to the optimal weighting of applicants' predictor scores in order to reach the joint and maximized validity. This highlights two important aspects of the assessment and selection process: the accuracy of the predictive validity estimates is crucial, and the weights need to be defined in an algorithm and applied mechanically to each applicant's set of predictor scores. However, early research by Thorndike (1949) indicated that predictive validity estimates may lack specificity due to a restriction of range in the study samples that are used for deriving these estimates, employee hirings are almost always based on some third unknown variable, and he suggested a methodology to correct for this.

The suggested methodology however, has been hard to apply due to the requirement of information on this third, and per definition, unknown variable. More recent development of this methodology has provided the opportunity to correct for range restriction without information on the third variable and thereby obtain more accurate predictive validity estimates (Schmidt et al., 2008). However, application of this methodology for general mental ability and a comprehensive personality model for the prediction of job performance is still lacking. The use of a comprehensive personality model, instead of only using some specific traits, corresponds more closely to what is used in practice and would facilitate making this research knowledge more transferrable to practitioners. Estimating the predictive validity of general mental ability and personality traits for job performance would maximize the predictive validity for the predictor measures, illustrate the relative importance of these characteristics, and provide the weights that can be used to combine applicant scores into an overall assessment. In addition, a more correct estimate of the overall or joint predictive validity for general mental ability and personality would increase accuracy when analyzing the financial outcome of using these measures for personnel selection.

With accurate optimal predictive validity estimates available, they can be used in an algorithm which is applied equally to each applicant's set of predictor scores in order to maximize the predictive validity and thus the accuracy of the selection decision (Hattrup, 2013). Such a mechanical approach for combining data is, however, very uncommon in selection practice. The common way of combining data in selection practice is to apply a clinical approach for combining data, which can be characterized as the opposite of the mechanical weighting. When a clinical approach is applied, the professional's judgment is utilized for combining each applicant's data, or predictor scores, into an overall assessment. Research comparing the two approaches for aggregating data for prediction purposes has consistently shown that predictive validity is much lower with the clinical approach (Grove, Zald, Lebow, Snitz, & Nelson, 2000) but this research evidence has had little effect on selection practice (Highhouse, 2008). One possible reason for the mechanical approach not being used in practice is that the research comparing the mechanical and the clinical approaches for data combination is communicated in statistical terms and therefore is difficult to interpret into real world outcomes. One potential way of making such research more accessible, and thus increasing the likelihood of acceptance, awareness and thus impact, is to present the findings in financial terms.

The gaps between research and practice concerning the personnel selection process have existed for many years and have become a research subject of its own. Why practitioners prefer a clinical hiring approach rather than using an evidence-based mechanical approach is a fascinating question that remains to be answered. Can this preference be explained by practitioners' general decision-making style and/or are there contextual factors that affect the way selection processes are designed? Overall, more knowledge and understanding about what underlies the choices of practitioners is needed, especially with regard to what it is that determines their preferences for a certain hiring approach. Such knowledge may be useful for guiding actions encouraging a greater use of evidence-based selection practices thereby narrowing the gap between research and practice in employee selection.

General aim

Over the last century, a lot of effort has been put into investigating how valid, cost-efficient, and fair selection decisions can be made, and there are few questions in the field of human resource management with more evidence pointing in the same direction. The considerable gap between research and practice persists, however, and in some aspects practice not only diverges from the research-based recommendations, but stands in direct conflict with the preponderance of evidence (Kuncel, 2008).

The overall objective of this thesis is to contribute to a narrowing of this research–practice gap. In order to reach this objective, the process of personnel selection needs to be discussed from several perspectives which need to be related to each other in order to contribute to a greater understanding of effective and accurate selection practice. The perspectives concern the choice of predictors, the framing and process of measurement, and how to combine predictor scores into an overall assessment. Altogether, the goal is to provide a roadmap for how to assess and use measures of individual differences in order to execute valid, reliable, cost-efficient, and fair selection decisions. An examination is also made of factors that determine the assessment and selection-making preferences of practitioners providing information about how to facilitate the adoption of a more evidence-based approach to personnel selection in practice. Steering the overall objective are three specific aims which correspond to the three empirical studies conducted.

The first aim of the thesis is to investigate the relationship between individual differences and job performance and to provide accurate estimates for how to weight the predictor scores in order to maximize the prediction of job performance. Being able to collect the correct data and combine it properly is a pre-requisite for making evidence-based and cost efficient assessments and selection decisions. Such estimates are calculated and presented in Study I.

The second aim is to investigate and compare the mechanical approach with the more commonly applied clinical approach for data combination with regard to financial outcomes. The two approaches differ extensively in regard to their predictive validity and the cost of applying them. By applying utility-analysis to estimate the overall financial outcome of selection methods and processes, the real world outcome of these differences become more evident and accessible for practitioners than communication based on predictive validity coefficients. Such a comparison is conducted and presented in Study II.

However, in order to bring about a change in applied approach among practitioners, towards a more evidence-based approach, increased knowledge about what underlies their current hiring selection preferences is called for. The third aim of this thesis is therefore to increase our understanding of what it is that drives practitioners' preferences for a certain hiring approach. This is done by investigating the potential influence of cognitive style in decision-making along with whether there are circumstances in the selection context, such as accountability for the assessment process or responsibility for the selection decision, which might affect the preference for hiring approach. This is the focus of Study III.

Job performance

The main objective of collecting and combining data about applicants in personnel selection is to predict job performance. Rank-ordering applicants as accurately as possible according to predictions of their future job performance and making hiring decisions based on this rank-ordering constitutes the very essence of personnel selection (Schmidt & Hunter, 1998). The underlying presumption that the performance levels of individuals will differ within an organization is central to this type of prediction. If this were not true, there would be no need trying to differentiate among applicants with regard to future performance, simply because there would not be any differences to predict – and personnel selection as a strategic process would not be an issue at all. The existence of individual differences in performance can be seen when considering the importance of, for example, education, training, and experience for many jobs; hiring a surgeon without proper medical education and training would, for example, be unthinkable and clearly inappropriate. That everybody, after taking formal requirements such as education into consideration, would perform equally well is perhaps just as unthinkable and has, more importantly, proven to be an incorrect assumption. Individuals with equivalent formal backgrounds contribute with different levels of job performance. The challenge lies in finding out what more specifically underlies these differences in levels of job performance and transforming this knowledge into applicable strategies that can be utilized for personnel selection purposes. However, before discussing the prediction of differences in job performance any further, the concept of performance needs to be elaborated.

Theory of job performance

The theoretical model used in this thesis, defining job performance as hierarchically organized with the construct general job performance as the highest order and most generalizable factor at the apex of the performance taxonomy and with primary job performance domains located at the level below, has gained strong empirical support (Viswesvaran, Schmidt, & Ones, 2005). In reference to occupational settings, the general factor of job performance is defined as "scalable actions, behavior and outcomes that employees engage in or bring about that are linked with and contribute to organizational goals" (Viswesvaran & Ones, 2000, p. 216). The general factor of job performance is an aggregation of the primary performance domains and there is strong support for a structure of three primary job performance domains – task performance, contextual performance, and avoidance of counterproductive work behaviors (Rotundo & Sackett, 2002). All three domains contribute to the general factor of job performance and represent three distinctly different aspects of human behavior and performance in the workplace.

Task performance

The performance domain of task performance concerns the performance of core job tasks. More specifically, task performance refers to activities and behaviors that, firstly, contribute to the core objectives of the organization in terms of the production of a good or the provision of a service and, secondly, that are formally recognized as part of the job (Borman & Motowidlo, 1993; Conway, 1999). In general, this corresponds to behaviors that contribute to the accomplishment of duties and responsibilities associated with a given job.

Contextual performance

Contextual performance, on the other hand, represents willful behaviors that contribute to the effective functioning of organizations by supporting the overall organizational, social, or psychological environment (Borman & Motowidlo, 1993). As such, contextual performance serves as the catalyst for task activities and processes (Borman & Motowidlo, 1993), through behaviors that demonstrate effort and show perseverance in helping and supporting peers, and that facilitate team performance (Campbell, 1990). It refers to behaviors involving cooperation, communication, the exchange of job-related information, constructive suggestion making, and spreading goodwill (George & Brief, 1992), as well as endorsing, supporting, and defending organizational objectives (Borman & Motowidlo, 1993). This aspect of performance is often conceptualized as organizational citizenship behavior (OCB), which represents "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the efficient and effective functioning of the organization" (Organ, 1988, p. 4).

Counterproductive work behavior

The performance domain of counterproductive work behavior, or personal deviance (Robinson & Bennett, 1995; 2000), concerns "any intentional behavior on the part of an organization member viewed by the organization as contrary to its legitimate interests" (Sackett & DeVore, 2001, p. 145). This domain includes voluntary behaviors that have a negative outcome for the organization. These behaviors can either be directed towards the organization (e.g., destruction or theft of company property or equipment) or they may be negative actions that harm coworkers in the organization (e.g., harassment, verbal or physical abuse). Selecting employees with a view to reducing counterproductive work behavior is vital for organizations due to the monetary and societal consequences that are associated with this type of behavior (Sackett & DeVore, 2001).

Although task and contextual performance are conceptually different, they generally show strong positive correlations with each other, ranging between 0.45 and 0.65 (Hoffman, Blair, Meriac, & Woehr, 2007). Measures of counterproductive work behavior, however, generally have a moderately negative correlation with task performance and a low negative correlation with contextual performance (Berry, Ones, & Sackett, 2007).

Thus, the concept of general job performance reflects the overall contribution of each employee to the organization, as it takes core task effectiveness, positive contribution to the social and psychological climates, and the absence of destructive and counterproductive behaviors into consideration. As such, general job performance concerns the expected combined value of an employee's employment-related productive and unproductive behaviors at an organization over a certain period of time. Job performance can be measured with objective performance based measures (e.g, if objective goals are met) or by measuring work behavior using supervisor ratings (e.g., task performance). In meta-analysis both types of measures are considered.

Individual differences

The fact that individuals differ in their levels of job performance makes it essential for organizations, and applicants, to identify and hire the highest performers. Identifying the factors that might predict job performance level is therefore important, and research indicate that a great deal of the variation in job performance seems to be explained by individual differences in basic characteristics. At a broader level, individual differences discussed in connection with understanding human behavior in organizations typically concern the four domains of cognitive and physical abilities, personality, interests, and core self-concepts. While all four of these domains can be expressed through behaviors and choices that are directly relevant to work in organizations, not all of them are important determinants for the criterion of interest in this thesis, job performance (Murphy, 2012). Rather than being predictive of job performance, physical abilities have a limited span of influence in general, interests influence occupational and organizational choices as well as satisfaction with and commitment to jobs and organizations, and core self-concepts influence persistence, adaptability, creativity, and motivation. On the other hand, certain aspects of personality and, especially, general mental ability, have proven to be important determinants for job performance (Schmidt & Hunter, 1998).

General mental ability and its usefulness in personnel selection

Individual differences in general mental ability contribute significantly to explaining differences between people in many vital areas of life (Gottfredson, 1997a; Hemmingsson, Melin, Allebeck, & Lundberg, 2006; Jensen, 1998; Neisser, 1996). Since the publication of Spearman's paper, "'General Intelligence,' Objectively Determined and Measured" in 1904, more than a century of empirical research has demonstrated the pervasive influence of cognitive ability in such various areas as academic achievement, occupational attainment, delinquency, socioeconomic status, racial prejudice, divorce, and even age of death (Gottfredson, 1997a; Jensen, 1998). Based on the positive correlation between students' rankings in math and language, Spearman (1904) suggested that this shared variance represents a general factor, the g factor, of cognitive ability. Spearman stated in his two factor theory that individual differences in the true score of any measurement of any ability are attributed to only two factors, the g factor, which is common to all mental ability assessments, and a second factor that is specific to each and every measurement of mental ability. In 1939, Holzinger and Swineford proposed the first hierarchical model of intelligence with a general factor at the top and several uncorrelated specific ability factors below, and although Spearman's g factor and the hierarchical model have been criticized (e.g., Thurstone 1931, 1944, 1947), accumulated research has provided solid evidence for the robustness and soundness of the hierarchical model and for the relevance of the g factor (Carroll, 1993). Competing theories (e.g., Guilford, 1988; Sternberg, 1985) are certainly not absent, but suffer, at least for the time being, from a lack of empirical support (Jensen, 1998).

Note that while Spearman (1904) labeled this general ability factor as the g factor, some later scholars have used other terms to refer to the same construct, such as general mental ability, cognitive ability, and intelligence. The preferred label is likely to depend on the context, in work and organizational psychology for example, and particularly within the branch of personnel selection, general mental ability is often considered the most suitable expression. The term cognitive ability is usually associated with clinical psychology and the term intelligence is often, for historical reasons, negatively charged. Regardless of which of these labels is used, the underlying construct being referred to is the same.

Today, it is fair to say that there is broad consensus in the scientific community concerning the hierarchical structure of cognitive ability, the existence of the general mental ability factor, and the definition of the construct. As might be evident from what has been said above, general mental ability does not represent a narrow academic intelligence, but will manifest itself in almost any realm of activity that involves active information processing. A definition proven to be useful in applied psychology is the one presented by Gottfredson (1997b), which was first published in the *Wall Street Journal* in 1994 as part of an editorial written by Gottfredson and signed by a number of colleagues. In their words,

> Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—'catching on', 'making sense' of things, or 'figuring out' what to do. (p.13)

Defining general mental ability is important for understanding its breadth of impact on human behavior, but it does not, in itself, help explain why individuals high in general mental ability exhibit higher job performance. By exploring this question, research has identified learning as the proximal determinant of overall job performance. This implies that the acquisition of job knowledge mediates the relationship between general mental ability and job performance (Borman, Hanson, Oppler, Pulakos, & White, 1993; Schmidt, Hunter, & Outerbridge, 1986). Thus, individuals with a high level of general mental ability are more proficient at acquiring knowledge about the job, learning from experience, and utilizing this knowledge, and therefore perform better than individuals with lower levels of general mental ability. Knowing how general mental ability affects the acquisition of job knowledge is essential to a framework for understanding why general mental ability has such a central role in job performance.

Spearman's (1927) formulation of the original g theory included the assumption that the g factor should influence performance on a wide range of tests and tasks, an assumption that has been supported (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004). Some of these supportive results, based on test batteries, have been found to be clearly related to performance in academic settings where acquisition and mastery of new knowledge and skills are a major focus, and have also been shown to be strong predictors of job training as well as job performance (Schmidt & Hunter, 1998). Furthermore, meta-analytic findings indicate that measures of general mental ability predict job performance in a variety of different jobs, organizations, occupations, and countries, which contributes to its universal importance to job performance (Salgado & Anderson, 2003; Salgado, Anderson, Moscoso, Bertua, de Fruyt, & Rolland, 2004; Schmidt & Hunter, 1998).

The degree to which a person is able to learn, adapt, deal with complexity, and process job-relevant information, therefore appears to determine his or her work behavior in general. In addition, the relationship between general mental ability and job performance has been found to be linear, which implies that higher levels of general mental ability are consistently related to higher levels of job performance, and that there is no point where a higher level of general mental ability is negatively related to job performance (Sackett, Borneman, & Connelly, 2008). In addition, the validity of general mental ability tends to increase with job complexity: general mental ability predicts low-complexity jobs in the 0.20s, medium-complexity jobs in the 0.50s, and high-complexity jobs in the 0.70s. Demographics such as ethnicity and gender as well as organizational, national, and cultural settings have not been found to moderate the relationship between general mental ability and job performance (Hunter & Hunter, 1984).

It should be mentioned that the empirical support for using lower order factors for personnel selection purposes (i.e., for predicting job performance) is not as convincing. Specific abilities are generally less important for explaining behavior than general mental ability and research suggests that the incremental validity of specific abilities (defined as ability factors unrelated to the general factor) in the prediction of performance and training outcomes is minimal when more general factors are taken into account (Ree, Earles, & Teachout, 1994). The reason for this might be that most tasks that involve active information processing rely on a range of abilities rather than one specific ability.

Personality and its usefulness in personnel selection

Although contemporary research points out general mental ability as the most important aspect of individual differences when it comes to predicting job performance, certain aspects of personality have also been identified as relevant and useful predictors (Barrick & Mount, 1991). The fascination with personality and the attempts to describe and understand it have a long history. As early as 500 B.C., the Greek philosopher Empedocles defined the four elements (earth, water, air, and fire) and used them as a basis for categorizing and describing differences in human dispositions. Subsequent and well-known scholars like Hippocrates (460–370 B.C.), Plato (427–347 B.C.) and Aristotle (384–322 B.C.), and later on Freud (1856–1939), Adler (1870–1937), and Jung (1875–1961) have all influenced the development of personality psychology and illustrate the diversity of disciplines (e. g., philosophy, economics, and medicine) in which there has been an interest in personality as a phenomenon. Contributions from different areas of psychological research and other disciplines have resulted in a diversity of theoretical and methodological approaches, and a large number of themes and controversies have emerged over the years. Some theoretical and methodological milestones and advancements, in combination with the impact of contemporary cultural and social and political contexts, have had a substantial impact on the contemporary use of personality in personnel selection.

Measurement of personality

The methodological and measurement-oriented approach to personality started to emerge in the beginning of the 20th century, partly as a result of psychologists' efforts to follow in the footsteps of the more "exact sciences." Based on the work on individual differences by Cattell, Thorndike, and Terman, and inspired by the use of items and scales for measuring intelligence, inventories and questionnaires designed to measure particular personality characteristics started to be developed and used for personnel selection purposes (Freyd, 1926). In the decades that followed, interest in personality research and measurement brought about a significant growth in the development of these inventories, mainly as a result of an actual need for effective tools in personnel selection practice.

In 1921, Allport and Allport introduced the concept of *trait,* holding that individuals generally can be characterized in terms of relatively enduring patterns of thoughts, feelings, and actions. At the same time, they made a distinction between personality and character – two concepts that had been used interchangeably until then. Although Allport (1935) argued for the recognition of personality as a research field of its own, rather than being a subfield of abnormal or social psychology, it took several decades before the first handbook of personality (Borgatta & Lambert, 1968) was published.

One of the first attempts at identifying a basic structure of personality was made by Allport and Odbert (1936), who introduced the lexical approach, which was facilitated by theoretical and statistical advancements of the time. According to the lexical hypothesis, a comprehensive specification of personality traits or dimensions can be obtained through an examination of the interpersonal adjectives found in natural language (Goldberg, 1995). In total, Allport and Odbert (1936) identified about 18,000 adjectives in the English vocabulary describing personality. Using the recently developed method of factor analysis for their data reduction, 30 empirically based traits were identified. A combination of theoretical and statistical advancements thus made possible the first attempt to identify a basic structure of personality traits. Researchers such Cattell (1943, 1945) and Fiske (1949) continued the work started by Allport and Odbert (1936), but it was not until the 1960s, when Tupes and Christal (1961) applied factor analysis on large amounts of data from the U.S. Air Force, that a framework for personality, as we recognize it today, was established; a personality measurement model that consisted of five broad factors.

In the 1960s, however, general criticism against trait theory and measurement grew strong. The criticism was mainly based on arguments relating to the existence and measurement of personality traits for the prediction of work criteria. Guion and Gottier (1965) stated that the predictive validity of measures of personality (i.e., personality tests) for job performance was too low to be of practical use. Mischel (1968) argued that there were no such things as inherent and stable personality traits that could create stable patterns of thoughts, feelings or actions across situations or contexts. From Mischel's perspective, feelings, thoughts, and behavior were born out of and dependent on the context. For those who adopted this perspective, the measurement of personality and the use of personality traits for predictive purposes lost its relevance. These theoretical controversies, together with the contemporary political and cultural movements (not least of all in Scandinavia), made the study, measurement, and use of individual differences in all fields of psychology less acceptable, and difficult to embrace. This also resulted in psychologists turning their backs against the study and measurement of individual differences, at least for some time (Mabon, 2002).

The five factor model of personality

Personality research in general started to recover during the 1980s when Costa and McCrae developed the NEO Personality Inventory (1985) using a questionnaire approach for measurement. The convergence between Costa and McCrae's (1985, 1992) research, the lexical approach, and other self-rating approaches gave new support for the relevance of Tupes and Christal's (1961) five broad factors, emotional stability, extraversion, openness for experience, agreeableness, and conscientiousness. Goldberg (1981) was the first to use the term "Big Five" when describing these factors in order to emphasize the abstraction and broadness of each trait and to avoid giving the impression that personality could be reduced to only five traits. Taken together, this work represents the first steps towards a joint conceptualization of the five traits, the five factor model of personality (Wiggins & Trapnell, 1997).

The five factor model may at a first glance seem simple, basic, and non-dynamic. It should be noted, however, that the multitude of relationships amongst them contribute to an extensive complexity and that there are important differences between the personality traits (McCrae & Costa, 2008). The five-factor model traits differ in level of abstraction; some traits are more manifest and some are more abstract. The traits also differ in character; some have a more descriptive character while some are more explanatory. Some traits are primarily expressed in the interpersonal space, while others define and affect internal processes such as thoughts and feelings. The traits overlap to different extents, they are related to each other in different ways, and they interact in unique and non-linear fashions when they are related to different criteria. In addition, different ways of measuring the underlying constructs (e.g., via self-reports and interviews) affect estimated relationships with other variables. Some aspects of this complexity become apparent when considering the core definition of each construct (John, Naumann & Soto, 2008). A brief definition of the constructs and descriptions that commonly underlie self-report measures of the five factor model traits will follow, and some caveats regarding the operationalization of certain constructs will be highlighted.

The construct of *neuroticism*, which is an intra-individual trait, is defined as an enduring tendency to experience negative emotional states, such as anxiety, anger, guilt, disgust, and depression (John, Naumann & Soto, 2008). Individuals who score high on neuroticism are on average more likely than others to interpret situations as threatening or hopelessly difficult, and to respond to them in accordance with the underlying trait – being anxious, moody, hostile, irritable, personally insecure, sad or even depressed. An individual with a low level of neuroticism is typically emotionally stable, calm, and relaxed, and handles stress without a problem. Measures of this trait are often reversed to represent its opposite, *emotional stability,* in order to ease interpretation and to better fit into non-clinical settings. Emotional stability is often pointed out as being somewhat of a driver for the other traits.

The inter-individual trait *extraversion* is defined as the degree of energy a person directs towards the outer world, including his or her social surroundings. Individuals scoring high on extraversion are characterized as being excitable, sociable, active, and talkative (Watson & Clark, 1997). Overall, they display high amounts of emotional expressiveness and need social interaction and stimulation from the outer world. Individuals scoring low on extraversion are more reserved, independent, and even-tempered. Extraversion is often thought to include aspects such as warmth, empathy, and assertiveness, which would be to presume certain "character" features for their social interaction with others. The term extraversion, however, is used to refer to individuals' degree of social energy, the amount of attention they direct towards social surroundings, and their need and capacity to interact with other people.

The intra-individual trait of *openness to experience* is characterized by openness towards the person's own inner feelings and emotional states (McCrae & Costa, 2008). This includes the ability to differentiate between nuances of emotions, to be open to the flow of emotions, and to have a tendency to give emotional states a prominent role or for it to be a determinant of behavior. The need and search for stimulation of one's inner world of emotional experiences may be accompanied by explicit behaviors, such as engaging in physical activities, preferably new ones, or it may also involve mental activities (e.g., meditation, intellectual reasoning). The focus for the search for stimuli, which can be directed towards the outer world or towards the inner mental world, is likely to be determined by the individual's degree of extraversion. High levels of openness to experience are typical of individuals who have an active imagination, are attentive to inner feelings, and who are intellectually curious. Individuals with low levels do not have the same need or urge for new emotional experiences. They are likely to be more conventional and conservative, and tend to settle for what is familiar and previously experienced.

The construct of *agreeableness*, an inter-individual trait, refers to the character of an individual's relationships (Graziano & Eisenberg, 1997). The construct includes attributes such as trust, altruism, kindness, affection, warmth and other pro-social behaviors. A high level of agreeableness corresponds to being altruistic, helpful, warm, compliant, and modest towards others. A high need for consensus, as well as a need for approval and popularity, are characteristics that might lead to conflict avoidance and difficulties with expressing and standing by personal opinions. A low level of agreeableness, on the other hand, characterizes someone who is more reserved, independent, skeptical, and critical towards others' intentions. Having such a low need for others' approval, in combination with a typical distinct and straightforward communication style, is likely to result in these individuals being perceived as less agreeable and somewhat antagonistic.

The construct of *conscientiousness* concerns how individuals approach to their tasks, commitments, and undertakings (Hogan & Ones, 1997). The level of achievement- and performance-striving and the preferred way of accomplishing goals are included in the definition. Effectiveness, persistence, and goal-directed behaviors are typical for those high in this trait. They also tend to be organized, mindful of details, reliable, self-disciplined, and self-motivated. Others typically describe them as competent, dutiful, self-disciplined, and well organized. Individuals with low levels are likely to have a more liberal attitude towards rules, obligations, commitments, and tasks and tend to be less self-motivated. They are often described by others as flexible and free-spirited, but also as careless, irresponsible, lazy, impulsive, low in achievement striving and lacking in ambition.

Personality and job performance

Since its resurgence in the 1980s, most research on personality and job performance has been conducted using trait theory as the theoretical framework and the five factor model as the measurement model. The development and use of meta-analysis (Hunter & Schmidt, 1990) has provided stable and generalizable support for the basic tenets of trait theory, postulating that traits exist, can be quantitatively assessed, show some degree of cross-situational consistency, are biologically based, and are meaningful for explaining important aspects of human behavior (John, Naumann & Soto, 2008). Altogether, this has contributed to the broad acceptance of the five factor model as a useful and relevant framework for empirical research on the relationship between personality and job performance (Murray, 2005).

Prior to the emergence of meta-analysis, general and valid conclusions concerning the relationships between personality traits and job performance were difficult to draw (Mischel, 1968). But with the meta-analytic technique, single studies – based on small samples and with uncertain and non-generalizable results – became possible to assemble, re-analyze, and summarize. In contrast to previous research, stable and generalizable estimates of the relationships between personality and job performance became possible to establish, resulting in the conclusion that personality may indeed matter (Barrick & Mount, 1991). Continued use of meta-analysis has further established the five factor model as the measurement model with the strongest empirical support with regard to its ability to predict work-related behavior (Barrick, Mount, & Judge, 2001), including job performance (Barrick & Mount, 1991). In addition, it has been shown that the measurement method of self-reported test data on personality contribute with incremental validity in job performance over and above general mental ability in a more efficient way compared to competing alternative selection methods (Schmidt & Hunter, 1998).

The relationships between different personality traits and job performance however, have been found to vary in magnitude. Several meta-analyses have shown that conscientiousness and emotional stability predict job performance consistently for most or all jobs (Barrick & Mount, 1991; Barrick, Mount & Judge, 2001; Hurtz & Donovan, 2000; Mount & Barrick, 1995; Salgado, 1997, 2003). The willingness to follow rules and exert effort (high conscientiousness) and the capacity to allocate resources to accomplish tasks (high emotional stability) have been suggested as a “will do” component, that is, a motivational component generalizable across jobs and tasks (Barrick & Mount, 2005). The remaining personality traits seem to predict success in specific occupations and for specific tasks. For example, it has been shown that extraversion predicts high performance in jobs involving intensive teamwork but poor performance in jobs where negotiations are an important part. Extraversion also predicts job performance in jobs which require considerable interaction with others (such as sales and management jobs), but is less relevant in other jobs. Agreeableness predicts high performance in jobs involving teamwork and low performance where negotiations play an important part, while openness to experience has not been found to have predictive impact across the criteria. These three factors are therefore not universal predictors for all jobs and criteria (Barrick & Mount, 1991).

Threats to validity in the context of personnel selection

The concept of validity in psychology is broad and spans across multiple definitions and aspects of psychological measurement (Markus & Borsboom, 2013). The most essential idea regarding validity is that it refers to the degree to which evidence supports inferences one proposes to draw about the target of assessment (Putka, & Sackett, 2010) implying that depending on the purpose of assessment, one out of several aspects of validity may be more relevant than others. In personnel selection, validity is almost synonymous with the examination of the relationship between assessment results and job performance scores (Schmitt, Arnold & Nieminen, 2010). The relationship between predictor scores and the criteria is central because the purpose of assessment in relation to personnel selection is to predict future performance. There are however several sources of threats to valid, thus accurate, and efficient personnel selection decisions, some of these are especially important to acknowledge and should to be taken into account when comparing the validity and efficiency of different selection methods.

Measurement error in personnel selection

Research investigating the validity of predictors is traditionally done at the construct level set out to explain the theoretical relationships between predictors and outcome variables (e.g., Judge, Bono, Ilies, & Gerhard, 2002). Correlational estimates of this kind provide valuable knowledge about how constructs such as general mental ability and the personality traits relate to the construct of job performance, and are based on the assumption that both predictors and outcome variables are free from measurement error.

The goal of validation research in personnel selection, however, is to establish the actual relationship between the predictor scores (e.g., GMA and personality) and the construct domain (e.g., job performance) in applied practice and the assumption of no measurement error overestimates the relationship between predictors and outcome since all measures in applied practice suffer from measurement error to some extent. In order to avoid this overestimation, each predictor's *operational* validity needs to be calculated (Binning & Barrett, 1989) by correcting for measurement error in the criteria but not in the predictor score. The criterion score represents the construct domain, that is, the conceptual specification of work behaviors, and should be corrected for measurement error (Binning & Barrett, 1989), while the predictor score, a score that could be used in the selection decision, should not be corrected. Operational validity more closely corresponds to the validity of the predictors in practice and should constitute the estimate upon which comparisons among competing predictors, thus selection methods, are done.

Constructs versus methods as predictors

Another important aspect to consider concerning the validity of predictors for selection purposes is what the predictors represent. In the previous discussion on general mental ability and personality there was no need to distinguish between the constructs and their predictors. This is quite logical due to the fact that the constructs of general mental ability and the five factor model personality traits correspond in a one-to-one relationship to the scales in the self-report tests and, thus, to the actual measurement or selection methods. Predictors of this kind, which represent a single construct *and* a measurement method, make comparisons straightforward, relevant, and possible to evaluate from a utility perspective. However, this clear operationalization and straightforward connection between constructs, operationalization, and measurement, or selection methods, is not characteristic of all predictors.

To start with, some predictors represent constructs (e. g., conscientiousness) while other predictors represent methods (e. g., an interview). Comparisons among both of these types of predictors are equally important but serve different purposes. Comparing the predictive validity of constructs, such as general mental ability and the five factor model personality traits, is important for increasing our understanding of the underlying nature of these constructs and their relationships with each other and related phenomenon, and for more general theory building. In applied selection practice, this type of knowledge, concerning why the constructs relate to each other and to job performance, is essential for choosing and combining the most appropriate and effective methods for measuring the relevant constructs. Sometimes the choice of measurement method is made without taking the underlying constructs into consideration. The assumption that different methods automatically measure different aspects of human nature, that is, different constructs, is still in circulation despite evidence showing that different selection methods, such as interviews and tests, often measure the same underlying constructs (Roth & Huffcut, 2013). Working under this assumption could likely lead to the use of needlessly expensive data collection methods, methods which lack incremental validity, or multiple methods which overlap on the construct level – approaches which would have a severe negative impact on overall utility. In selection practice, the validity of selection methods as predictors, rather than constructs, constitutes crucial information. It is the predictive validity of the selection method, not the underlying constructs, that drives the actual gain and cost of implementing the selection method into the selection process.

Predictors not only differ in what they represent, they also differ in composition. Some predictors represent methods that measure more than one construct, such as the interview, which traditionally measures constructs such as agreeableness, extraversion, and general mental ability. Other predictors represent methods that actually include several measurement methods, such as assessment centers, which traditionally include both exercises (which in turn measure multiple constructs) and general mental ability tests. Overall, this makes it more complicated to compare them from an overall utility perspective.

Nonetheless, when establishing the magnitude of the relationships between predictors and general job performance, constructs and measurement methods have often been confounded (Hough, 2001), as comparisons between predictors representing single constructs, such as conscientiousness and general mental ability, and predictors representing methods, such as interviews, references, and work samples, are often made (e.g., Hunter & Hunter, 1984; Schmidt & Hunter; 1998). The bottom line is that comparisons between predictors representing constructs and predictors representing methods (e.g., Hunter & Hunter, 1984; Schmitt, Gooding, Noe, and & Kirsch, 1984) are based upon un-equal pre-requisites and thus encourages invalid conclusions about their efficiency. Valid conclusions based on comparison among predictors must be made between predictors representing *either* constructs or methods, thus, within in the domain of predictors. In order to evaluate the overall utility of a selection method, the predictors need to represent the selection methods rather than the constructs, although the predictive validity of constructs provides the foundation and theoretical rational for combining selection methods.

Correction for range restriction

Research on individual differences as predictors of job performance has been, and still is, very much reliant on methodological improvements and statistical advancements concerning measurement. The existence of effective measurement tools that provide accurate predictor scores and accurate estimates of predictor validity is a prerequisite for studying individual differences and making use of them in practice.

One such recent advancement is the refined methodology correcting for range restriction presented by Hunter and Schmidt (2004). The problem with range restriction in samples used for estimating the relationship between individual differences as predictors of job performance has been known for a long time (e.g., Binning & Barrett, 1989; Hunter & Schmidt, 2004; Rydberg, 1963; Thorndike, 1949). This problem arises when the samples used for validity estimation are not gathered through random sampling of the general population. Almost all of the study samples in this area of research consist of employed individuals, who, as such, with little exception, have undergone some kind of selection personnel selection process before being hired (Sackett & Yang, 2000). Although a selection process do not explicitly set out to make hiring decisions based on applicants levels' of general mental ability and five factor model personality trait, it is likely that the hiring decision, to some extent and indirectly, has been affected by these predictors causing range restriction in the predictors (e.g., Binning & Barrett, 1989; Hunter & Schmidt, 2004; Rydberg, 1963) and decreasing the correlations between predictors and criteria (Sackett & Yang, 2000).

The most known and frequently applied formula to correct for range restriction in the research area of personnel selection (Hunter et al., 2006) is based on the assumption that those who were selected for hiring from among the applicants were directly chosen through top-down selection based on their test scores, thus, that no other, unaccounted for, predictors have influenced the selection decision (Thorndike's, 1949). This, however, is very rare in practice. Job applicants are typically selected for the job based on some other *indirect* unknown variable which is correlated with the test score, such as a composite of unmeasured variables indicative of applicant performance. To overcome this, Hunter and Schmidt (2004) suggested a procedure where measurement error in both the test score and the criterion is corrected for range restriction, and later in the process measurement error in the predictor score is reintroduced, thus transforming back to operational validity. This process makes it possible to correct for the indirect effect of a third variable without having data for the variable. When indirect range restriction is corrected for, estimates of predictive validity are more accurate (Schmidt et al., 2008) and therefore provide more specific knowledge about how strongly general mental ability and personality traits predict job performance. The increased accuracy allows more reliable generalizations to be made regarding the relationships between general mental ability and self-reported five factor model personality traits and job performance, and it also increases precision when estimating the financial utility of using single and groups of predictors in personnel selection processes.

A utility framework in personnel selection

The degree of difficulty in convincing practitioners to utilize evidence-based decision-making aids is likely to depend on several factors (Highhouse, 2008). One such factor could be that research findings are mainly presented in statistical terms, as in, for example, coefficients expressing relationships between predictors and criteria. Research results that are communicated in statistical terms are often difficult for most non-researchers, such as human resource practitioners, to grasp and translate into potential organizational outcomes.

Statistical estimates alone are also difficult for practitioners to use when justifying and communicating the efficiency of different personnel selection activities internally within their own organizations as well as with customers. Under the intense pressure of global competition, human resource executives are often pressed to justify personnel selection activities in more tangible ways. This can be done, for instance, by considering the relative payoffs – the utilities – of different alternatives in relation to organizational objectives in terms of monetary value. In addition, comparing predictors, representing single or multiple selection methods, solely based on their level of validity leaves out important factors that are necessary for evaluating the *overall utility* of alternatives. Clearly, methods and activities connected to the process of personnel selection are not ends in themselves; rather, they are means for reaching the overarching aim of the organization, namely to increase *utility* within the organization. This objective is reached by, as cost efficiently as possible, rank-ordering applicants according to predicted future job performance, and then hiring the top ranked applicants.

Utility analysis

Utility analysis provides a framework to guide decisions about investment in human capital (Bodreau & Ramstad, 2003). The concept of utility in the context of personnel selection refers to the estimation of realized gains or losses resulting from implementing selection methods and procedures and at its simplest level, utility can be defined as gains minus costs (Cascio & Bodreau, 2008). Utility analysis takes into account three important parameters of the selection process: quantity, quality, and cost (Cascio & Bodreau, 2008), which are crucial for the overall payoff, utility, to the organization. Quantity refers to the number of applicants and the number of selected employees, quality refers to the validity of the selection decision, and cost refers to the average cost of the selection process for each applicant. Utilizing and measuring these parameters is thus necessary for evaluating and estimating the actual contribution of a selection process to the organization. Typically, however, selection processes are not evaluated by taking these key parameters into consideration, resulting in a failure to assess the actual overall utility of the selection process. Rather, the focus is often directed towards the tangible cost of implementing the selection method(s) without taking into account the costs associated with errors in selection decisions.

Utility analysis allows practitioners to consider the potential gains and costs of using different and competing selection methods and processes. More importantly, the utility framework forces the practitioner to define the goals of the selection process clearly, to enumerate the expected consequences or outcomes of the selection decisions, and to attach differing utilities or values to each part of the selection process. Such an approach is likely to contribute to making decisions that rest upon a greater foundation of sound and logical reasoning and thoughtfulness. Without utility as an overarching framework, the outcome of comparisons between selection methods and processes becomes arbitrary with the risk of choosing a selection process which in the end causes financial loss for the organization. Thus, it is reasonable to suggest that a utility framework is a prerequisite for making relevant comparisons between selection strategies and for making logical choices among them.

The strategy that maximizes the expected overall utility for an organization should be chosen (Guion, 1991; Brealy, Myers, & Allen, 2006) and there are multiple approaches to calculating utility (Taylor & Russell, 1939; Naylor & Shine, 1965; Raju, Burke, & Normand, 1990). The most versatile and commonly used model was developed by Brogden (1946; 1949), further refined by Cronbach and Gleser (1965), and is now referred to as the Brogden-Cronbach-Gleser (BCG) model (Cascio & Bodreau, 2008). The BCG model integrates the concept of selection system cost as well as the gain or loss of the selection process in monetary terms – aspects that were not included in previous utility models. The BCG model comprises several factors, and different forms or variations of the original BCG equation can be applied depending on the actual contextual purpose (e.g., Le, Oh, Schaffer, & Schmidt, 2007). The underlying assumptions and contributing factors, however, remain the same and can be seen in the following classic equation (Brogden, 1949):

$$\overline{U} = r_{xy} SD_y \overline{Z}_x \tag{1}$$

The outcome of this equation, denoted $\overline{U}$, represents the margin utility, that is, the financial difference of implementing competing selection strategies. What the financial difference represents depends upon the first factor in the equation, r_{xy}, which represents the validity expressed as a correlation coefficient. This coefficient could represent the relationship between a predictor and the criterion, in which case the comparison will be made against a selection strategy with the validity of zero, thus a random selection strategy. The validity coefficient could also represent the increase in validity, thus the incremental validity, of applying a new selection strategy compared to an old strategy.

Validity is an important factor in all instances of utility analysis because it constitutes the driver behind the overall utility gain. According to the BCG model, there is a positive and linear relationship between validity and utility – higher validity corresponds to increased utility. In addition, the BCG model states that utility is proportional to validity, implying that if the cost of implementing the selection process is zero, the gain in utility will correspond to the increase in validity (Cronbach & Gleser, 1965). In addition, the BCG model states that the linearity holds at all selection ratios (Brogden, 1946).

The second factor in the equation, $\overline{Z}_x$, represents the average level of performance for the group of selected applicants using the new improved selection strategy and expresses this in standardized scores. $\overline{Z}_x$ is dependent on the selection ratio, which is the ratio of the number of selected applicants to the total number of applicants. As the selection ratio decreases, $\overline{Z}_x$ increases due to the increased impact of validity. Increased selection ratio, on the other hand, brings about less of an increase in $\overline{Z}_x$ since the level of validity will have a more limited impact. Naylor and Shine (1965) computed the values of $\overline{Z}_x$ for different levels of selection ratios, so that $\overline{Z}_x$ can be read and added to the equation using the selection ratio.

The monetary value of job performance

The third factor in Equation 1, SD_y, represents the standard deviation of job performance and expresses the variability in job performance in financial terms. The extent to which the level of job performance can vary depends on two main factors: the nature of the job and the extent to which the job permits individual autonomy and discretion (Cabrera & Raju, 2001). For jobs with rigidly specified requirements (for example, jobs in fast food restaurants) individual differences in general mental ability and personality will have less noticeable effects on performance compared to, for example, a sales job with a high degree of autonomy and with considerable discretion due to more flexible job requirements (Cascio & Bodreau, 2008).

The variability of job performance also depends on the complexity of the job, and increases as a function of complexity (Hunter, Schmidt, & Judiesch, 1990). The more complex a job is, the more difficult it becomes to explicitly specify the procedures that comprise the job's performance, and, as a result, individual differences become more important determinants of the variability of job performance. The variability of job performance also depends on the relative importance of the job position for the functioning of an organization. With some jobs, performance differences are vital to the successful achievement of the strategic goals of an organization; for example, performance differences in a product development manager role at a telecom company will potentially have more of an impact than performance differences in a cashier role for a supermarket chain. Thus, the monetary value of variation in employee performance determines the likely payoff of investments in the personnel selection process and translates performance improvements into money. Without SD_y, the effect of a change in criterion performance could be expressed only in terms of standard Z-score units, and would thus not allow measuring if a selection process, or improvement of a selection process, justifies the cost of the process.

As of yet, there is no perfect measure of the value of performance variation. Early on, the BCG model had been underutilized before better methods for estimating SD_y were developed. In 1979, Schmidt, Hunter, McKenzie, and Muldrow introduced a model for estimating SD_y as a percentage of annual salary (Schmidt & Hunter, 1983). In 1986, Schmidt, Hunter, Outerbridge, and Trattner followed up on this work and empirically estimated SD_y as a percentage of mean output. They argued for a conservative SD_y based on 40% of the average annual salary. For example, in dollars, an annual salary of \$50,000 would correspond to an SD_y of job performance of \$20,000. Given this, a high performer, defined as performing one standard deviation above average, would contribute to the organization with a job performance worth \$70,000 (\$50,000 plus \$20,000). A poor performer, on the other hand, performing one standard deviation below average, would contribute \$30,000 (\$50,000 minus \$20,000) worth to the organization. Although there are empirical estimations of the magnitude of SD_y for specific jobs (Schmidt & Hunter, 1983), 40% is recommended as a rule of thumb when time and/or resources do not permit estimating the standard deviation for a specific context (Cascio & Bodreau, 2008). The application of SD_y as 40% of the annual salary generally generates conservative estimates of job performance in terms of monetary value (Judiesch, Schmidt, & Mount, 1996). This might not be an issue when comparing selection methods, since the evaluation of margin utility is done through a relative comparison of the competing alternatives. Thus, the decision of which selection method to choose would not be affected although the absolute level of utility would be. Also, a study by Hazer and Highhouse (1997) showed that managers were most favorable toward utility analysis when the 40% salary method was used compared to less conservatively calculated SD_y.

With Equation 1 providing the expected utility gain of each selected applicant per year, the next step is to figure in the number of applicants being hired and the average number of years they would be expected to stay with the organization. For this purpose, Equation 1 can be expanded into Equation 2, which includes $\overline{T}$, representing average tenure, and N_s, representing the number of hired employees.

$$\overline{U} = \overline{T}N_s\, r_{xy}SD_y\overline{Z}_x \tag{2}$$

Continuing with the above example, if both $\overline{T}$ and N_s were 3, the high performers would contribute with a job performance worth \$630,000 (3 x 3 x \$70,000), while the low performers would contribute with \$270,000 (3 x 3 x 30,000). The difference in utility gain is thus \$360,000.

Implementing a selection method or process will almost inevitably incur a cost, which needs to also be specified in the utility equation. The cost which can depend, for example, on the business model of an assessment provider or the approach of an internal assessor, can often be divided into fixed and relative costs. For example, developing a unique test or a structured interview is initially costly, but the cost per applicant will decrease over time as it is used with more applicants. Among assessment service providers, some sell their products at a fixed price per applicant, while others offer a fixed price subscription which allows an infinite number of administrations. Selection activities requiring that practitioners to spend time on each applicant are generally more costly compared to activities not demanding the involvement of a practitioner for each individual assessment and the cost rarely decrease with the number of applicants. For example, the cost per applicant when using an interview as a data collection method or a clinical approach for data combination will increase in a linear fashion with the number of applicants rather than the opposite. Regardless of what the costs are based on, they are all figured into the total cost of the selection process for all applicants, which is denoted as C in Equation 3.

$$\overline{U} = \overline{T}N_s\, r_{xy}SD_y\overline{Z}_x - C \tag{3}$$

In summary, utility analysis using the BCG model provides an estimation of the increased monetary value of job performance when using certain selection strategies. The margin utility depends on how selective organizations can be (the selection ratio), how accurate predictions of future job performance can be (the validity of comparing selections strategies), and the extent to which differences in levels of job performance can translate into differences in monetary value to the organization (the variation, SD_y, of job performance). With utility as a framework, individual predictors, selection methods, and data combination approaches can be compared, and the actual quality of selection decisions, manifested in financial outcomes for the organization, thereby becomes tangible.

The use of multiple predictors

The BCG model can be used to calculate utility at several levels; it may be used for estimating the utility of the overall selection process or for evaluating the utility of specific parts in a selection process. Of course, the utility of the specific parts and processes comprises the overall utility, but there are some important aspects to pay attention to concerning the gain–cost relationship in these distinct and specific aspects (Schmidt & Hunter, 1998). Aspects of data collection are central to utility, the choice of predictors, including the degree of overlap between them at the construct level, and the method(s) used for measurement. An individual predictor, representing a selection method, needs to have a relationship with the criterion; higher predictor validity generates higher utility. The increase in utility generated by higher validity should, however, not exceed the cost of implementing the method as this would result in a financial loss. If the predictor (i.e., the method) has no relationship with the criterion, it should be excluded, because no validity will inevitably correspond to the lack of utility. Implementation of a method with no validity will cause a financial loss to the organization corresponding to the cost of implementing the method (Cascio & Bodreau, 2008).

If multiple predictors are used they should, in addition to the above, be as mutually independent as possible. In terms of utility, utilizing overlapping predictors (i.e., predictors sharing explained variance in the criterion) is to retrieve (and thus pay for) the same information and data about applicants more than once, at least to a certain extent. It goes without saying that the implementation of a method with predictors that do not explain any additional variance in the criterion also causes the organization to incur financial loss. If the identified predictors overlap (and the methods thus compete in explaining variance), the most cost-effective way of collecting the information should be used. A lot of methods generate the same data about applicants, but at very different costs. One example is the overlap between scores from a test of general mental ability and employment interviews (Roth & Huffcut, 2013). To a large extent, these two methods measure the same underlying construct. A psychological test, however, traditionally provides the data at a much lower cost than an interview could, making the test the preferred alternative from a utility perspective. Variance in an interview that is not shared with the general mental ability test score can of course provide incremental validity, but this contribution needs to be established and set into relation to the implementation cost. Whether the same level of incremental validity could be reached using a different method with greater cost efficiency should be taken into consideration.

Although using a certain predictor or predictors over others can in some cases both provide improved incremental validity and be more cost-efficient, in other cases small incremental validities can generate high utility but still not be feasible because the cost of adding the predictor would exceed the expected gains in job performance of the hired workers. The goal is for the incremental validity units to be associated with higher levels of job performance to the extent that it exceeds the cost of implementing the method in the selection process.

The extensive research establishing univariate relationships between predictors and job performance and lack of overlap among measures, in combination with cost-efficiency, has demonstrated the superior effectiveness of combining scores from tests of general mental ability with scores from tests of the five factor personality dimensions for predicting job performance and, thus, make personnel selection decisions, compared to other methods (Schmidt & Hunter, 1998). In order for this approach to be effective, not only should data on general mental ability and personality traits be *collected* by the use of psychological tests; data must also be *combined* accordingly. The predictive, and per definition maximized, validity of the aggregated assessment can thus only be reached if each set of predictor scores are calculated using an empirical and optimal regression-based model, utilizing an algorithm for combining the scores. Deviations from this mechanical approach for combining data will inevitably lower the predictive validity of the overall assessment, and recent research has estimated this decrease in validity to be extensive for work-related criteria (Kuncel, Klieger, Connelly, & Ones, 2013). However, the financial impact of this decrease in validity has not been estimated. Considering the general lack of acceptance for the mechanical approach when combining applicants set predictor scores into overall assessment for prediction, this constitutes an opportunity for selection practice to improve the utility of selection processes.

Another important aspect to notice is that the research showing that relations between selection methods and performance criteria are almost always linear (Coward & Sackett, 1990), imply that compensatory selection models yield the highest selection utility. In a compensatory model, all applicants are given all predictors and an applicant can make up for being low on one predictor by being high on another (or others). This means the desired validity estimate for each predictor is that in the initial (complete) applicant pool and the use of indirect range restriction corrections estimates such validities. When multiple predictors are used, the applicant pool validity of each predictor should be estimated and used along with the applicant pool predictor intercorrelations (computed without range restriction corrections) in a regression equation predicting job performance (Hunter, Schmidt & Le, 2006). Thereby the regression equation is the statistical implementation of the compensatory selection model.

In practice multiple hurdle models are common, usually because of administrative convenience, rejecting applicants who falls below the cut score on any predictor at that point, no matter how high his or her scores are on predictors previously administered (or would have been on those yet to come in the sequence). The non-compensatory feature means that use of multiple hurdle models is equivalent to making the false assumption of nonlinear relationships between predictor and criteria, which imply that traditional selection utility models do not apply.

Decision making in personnel selection

Making the most accurate decision about whether to hire, or not hire, a person is at the very heart of personnel selection. Such decisions are based on comparisons between available applicants, done through rank-ordering applicants based on their probable level of future job performance. The accuracy of any hiring decision is determined by the level of predictive validity of the selection process, which in turn relies on the accuracy of two of its constituting parts – data collection and data combination (Sawyer, 1966).

Identifying relevant constructs and deciding what method to use for measurement relates to data *collection*. Data collection is an essential, but not sufficient, part of the process towards a selection decision. Personnel selection decisions are (explicitly or implicitly) based on predictions of future job performance and all prediction-based decision-making, whether done by a psychologist, a human resource consultant, a physician, a psychiatrist, or any other professional, requires that the collected data is *combined* into a unified assessment (e.g., Meehl, 1954; Sawyer, 1966; Grove & Lloyd, 2006). The collected data must be interpreted, integrated, or aggregated into an overall assessment in order to serve its intended purpose – to establish a prediction of the future outcome. Thus, the process of data combination constitutes a second, necessary part of all decision-making processes, including personnel selection decisions. The distinction between data collection and data combination provides a conceptual framework for understanding how different approaches to data collection and data combination can affect the outcome of the selection process, defined as the level of accuracy in the final selection decision.

There are two main approaches to data collection and data combination, the clinical and the mechanical approaches (e.g., Dawes, Faust, & Meehl, 1989; Freyd, 1926; Grove et al., 2000; Meehl, 1954; Sawyer, 1966; Viteles, 1925). The choice of approach to use with each of these data-related steps affects the outcome of the selection process in terms of the level of accuracy in the final selection process. These two approaches represent mutually exclusive ways of collecting and combining data for decision-making purposes in the context of personnel selection. In essence, the clinical approach rests upon professional judgment, while the mechanical approach rests upon explicit and standardized rules and procedures (Sawyer, 1966). A practitioner can act in accordance with one of the categories, combining the different approaches to data collection and data combination. The four different combinations of clinical and mechanical approaches to data collection and data combination are,

1. Mechanical collection and mechanical combination
2. Clinical collection and clinical combination
3. Clinical collection and mechanical combination
4. Mechanical collection and clinical combination

Data collection

The process of data collection refers to the process of collecting information, or data, about each applicant and results in a set of predictor scores for each applicant. The process of data collection is preceded by identifying the constructs and method(s) that are to be used to measure them.

When it comes to utilizing individual differences as the central factors for the prediction of job performance, which is the focus of the present thesis, the research is quite unambiguous. General mental ability and the personality traits operationalized in the five factor model have been shown to be the most relevant predictors, and the most effective and reliable way of gathering information about these predictors is through the use of standardized measurement tools – psychological tests (Schmidt & Hunter, 1998). Using standardized psychological tests as a data collection method represents a *mechanical approach*. In psychological tests, as well as in other standardized methods, items, response alternatives, and scoring rules are standardized and explicit in order for all of the applicants to have an equal opportunity to provide the same information about themselves in the same manner. The questions and how they are presented as well as the response alternatives are the same for all applicants and it is thus not possible for an applicant to provide more or less data about him or herself in any other way. In addition, the responses are analyzed in a consistent manner by applying the same scoring rules to all applicants. The collected data is thereby not affected by human, or professional, judgment.

A *clinical approach* implies that the data collection process, overall or at some point, is not standardized and that the data has been generated by, or affected by, the practitioner's judgment. An example would be the unstandardized interview, where different applicants are given different questions, or applicants' responses are interpreted by the interviewer and thus not scored systematically according to predefined scoring rules. Data collection methods classified as clinical do not apply explicit and pre-defined rationales in a standardized manner, and the data that is collected is dependent on the practitioner's judgment. As a consequence, the process leading to the outcome is implicit and cannot be fully replicated.

Data combination

Data collection alone would be sufficient for the purpose of describing the applicants, but decision-making based on predictions requires an additional process to take place, namely the combination of collected data into a unified and aggregated assessment. (The exception would be a situation where only one predictor is used for rank-ordering and decision-making, a rare situation in practice even if there are situations where this approach might be appropriate.) Research indicates that the type of data combination method used has a greater impact on predictive validity than the type of data collection method used (Sawyer, 1966; Grove et al., 2000). Despite this, and a bit surprisingly, the process of data combination is rarely under scrutiny.

Simply put, the process of data combination corresponds directly to the prediction itself. It defines how the individual scores on each predictor are to be weighted in order to maximize the predictive validity for the criterion at hand. Each predictive validity estimate constitutes the weight the predictor score needs to be given in order for the overall assessment to reach the highest possible level of predictive validity when combining the predictors at hand. Deviations from such weighting will inevitably decrease the overall predictive validity when using the predictors at hand (Kuncel et al., 2013). In addition, this implies that weights incorporated into an algorithm constitute the explanation and rationale for the decision itself.

As with data collection, data combination can be performed by applying either a clinical or mechanical approach (e.g., Dawes et al., 1989; Freyd, 1926; Grove et al., 2000; Meehl, 1954; Viteles, 1925). The primary defining characteristic of the clinical data combination approach is that the data combination rests on professional judgment (Dawes et al., 1989). This means that the process of combining data into an overall assessment is done internally within the practitioner's mind or by consensus through group discussions. The outcome, the overall assessment, is thus dependent on practitioner judgment. The bases for judgment in this approach are the practitioner's prior professional experience, interpretation of cues, insight and intuition, and prior knowledge, including whatever relevant research findings known to, and interpreted by, the practitioner. The actual combination of data, that is, the "weighting" of each predictor score or information unit, is done implicitly and through the professional's mental processing. The rationale behind utilizing clinical data combination as an approach thereby rests on the faith that professionals have in their own competencies for making holistic assessments for prediction purposes. Given the characteristics of the clinical approach to data combination, it follows that the process can be neither transparent nor fully reproducible.

Applying the mechanical approach to data combination eliminates human judgment from the process of combining each applicant's set of predictor scores into an overall assessment for prediction (Dawes et al., 1989). The mechanical combination of data is based solely on explicit and pre-specified algorithms that specify the rules for combining the data. The rules govern the weight of each predictor, and they are uniformly applied to each applicant's set of predictor scores. The primary feature of the mechanical approach is that the algorithm states the logic behind the combination of data and the rationale, or rules behind the decision-making, making it transparent as well as consistent across all applicants. The results of mechanical data combination are therefore reproducible, as the same algorithm will produce the same predictions for a given set of data (Vrieze & Grove, 2009).

The features of clinical and mechanical data combining are mutually exclusive and opposed to each other; hence, a decision-maker cannot act in accordance with both of these incompatible prediction methods simultaneously (Meehl, 1986). In a hypothetical situation in which the mechanical and clinical predictions are in disagreement, there is rarely an "in between" option that can be chosen (Meehl, 1998). Selection decisions are thus dichotomous – an individual is either chosen for hire or not, and, in a multiple hurdle selection process, the applicant passes to the next phase or not. Nevertheless, practitioners often state that they apply both approaches to data combination. Most often, this means applying a mechanical approach for some part of the process of combining data, and then adding a clinical approach right before the actual decision-making. From a strictly empirical view, the rules of the mechanical data combination process must be adhered to in all stages of the decision-making process – otherwise, the process will be classified as clinical (Sawyer, 1966).

Data collection and combination for prediction

Data collection, as a function, affects predictive validity although not to same extent as the function of data combination. Although the differences in levels of validity between applying a mechanical approach and a clinical approach when collecting data are not extensive, the mechanical approach tends to lead to a better financial outcome due to the lower cost of implementation. In addition, practitioners tend to be more accepting of using mechanical data collection methods compared to a mechanical approach when combining data.

Empirical research comparing clinical and mechanical data combination for prediction purposes has been ongoing across a range of fields and criteria for almost a century (e.g., Dawes et al., 1989; Freyd, 1926; Grove & Meehl, 1996; Grove et al., 2000; Meehl, 1954; Reichenbach, 1938; Sarbin, 1941; Viteles, 1925). Research has consistently demonstrated improvements in the accuracy of the mechanical approach over the clinical approach (Dawes et al., 1989; Goldberg, 1968; Grove & Meehl, 1996; Grove et al., 2000; Gough, 1962; Meehl, 1965, 1967; Sarbin, 1943; 1944; Sawyer, 1966; Sines, 1970).

The first quantitative review comparing clinical and mechanical data combination methods for the purpose of predicting human performance in work and academic settings was published recently by Kuncel et al. (2013). This meta-analysis set out to estimate the magnitude of loss in predictive strength when applying a clinical instead of a mechanical approach to combining data. In agreement with results from the latest meta-analysis carried out on a broader range of criteria (Grove et al., 2000), the results from the Kuncel et al. (2013) study demonstrated a sizable and consistent loss in predictive validity for work (job performance, advancement, and training) and academic (grade point average and diverse non-grade measures) criteria when combining the same set of data clinically as compared to mechanically.

The difference in validity was especially pronounced for job performance criteria, where a clinical combination of data decreased the predictive validity by more than 50%, compared to a mechanical data combination approach. Despite the consistent and massive support for the superiority of the mechanical data combination approach for prediction-based decision-making, clinical data combination is still the predominant approach in practice (Vrieze & Grove, 2009). In 1998, Paul Meehl stated:

> The clinical-statistical problem ("what is the optimal method of combining data for predictive purposes?") has also been solved, although most clinicians haven't caught on yet. A meta-analysis on 136 research comparisons of informal with algorithmic data combination, conducted by Will Grove, has settled this question. I do not know of any controversy in the social sciences in which the evidence is so massive, diverse, and consistent. It is a sad commentary on the scholarly habits of our profession that many textbooks (and even encyclopaedias) persist in saying that the question is still open. I suppose they would be saying that if we had 1360 studies. (Meehl, 1998)

The acceptance of the mechanical approach is very low, not only among practitioners in the personnel selection field but in other fields of psychological assessment as well (Vrieze & Grove, 2009). Not only does there seem to be a resistance to applying the mechanical approach, there is also a strong belief in the actual *need* for applying a clinical approach. One recent example from the academic literature is the article by Pretz and Totz (2007) who starts their article by asking the reader to consider a context which requires the use of intuition defined as human judgment, and then suggests the field of personnel selection. Understanding the mechanisms behind professionals' preferences for the choice of predictors, concerning what to measure and how to measure it, as well as how to combine predictor data and make selection decisions, is an important step towards influencing practitioners of personnel selection to become more empirically based. Highhouse (2008) describes the inability of researchers to convince professionals to use appropriate decision-making aids, in terms of combining predictors in accordance with empirical evidence, as the greatest failure in the history of industrial and organizational psychology.

Summary of studies

The process of personnel selection comprises several steps before a valid and cost-efficient selection decision can be made. The three empirical studies in this thesis address several of these steps, including identifying predictors, determining what types of measures to use, and how to combine the predictor scores in order to arrive at the most accurate and cost-effective prediction of job performance. In addition to this, possible determinants of practitioners' hiring approach preferences are investigated.

Study I

Using individual differences to predict job performance: Correcting for direct and indirect restriction of range.

The research investigating the relationships between individual differences and job performance is extensive and supports the use of general mental ability and personality factors as predictors of job performance for personnel selection purposes (Schmidt & Hunter, 1998; Barrick, Mount, & Judge, 2001; Barrick & Mount, 2003; Schmidt et al., 2008; Hogan & Holland, 2003; Ones & Viswesvaran, 2001). However, it was further found that the validity estimates presented in previous studies lack accuracy, but that a recently developed methodology that corrects for range restriction may amend this problem (Le et al., 2007).

The aim of Study I was to investigate the relationships between general mental ability (Spearman, 1904) and the personality traits of the five factor model (Costa & McCrae, 1985) and job performance (Viswesvaran & Ones, 2000) when applying and comparing the old and new methods for range restriction correction. Traditionally, study samples used in this research consist of job incumbents, who are rarely randomly selected from the general population (Sackett & Yang, 2000). Typically, they have undergone some kind of selection process before being hired, which causes a restriction of range in predictors such as general mental ability and personality (e.g., Binning & Barrett, 1989; Hunter & Schmidt, 2004; Rydberg, 1963; Thorndike, 1949), and in turn leads to decreased correlations between predictors and criteria. Also, since different predictors can suffer from range restriction to different extents, the relative importance of the predictors can be distorted if the restriction in range is not corrected.

In previous research, this has been taken care of by correcting for *direct* range restriction and measurement error in the criterion. The formula used when correcting for direct range restriction, however, is based on the assumption that those chosen for hiring from among the job applicants are selected directly based on their top-down scorings, which is extremely rare in practice (Sackett & Yang, 2000). Rather, applicants are typically selected for the job based on some other, indirect and unknown, variable which is correlated with the test score, usually a composite of unmeasured variables of applicant performance (Le et al., 2007). Correction for *indirect* range restriction, however, enables correcting for this indirect and unknown variable and thus results in more accurate estimates of the relationships between general mental ability, personality traits, and job performance (Hunter & Schmidt, 2004).

After re-analyzing previously published meta-analytic data (presented in studies by Judge, Jackson, Shaw, Scott & Rich, 2007; Mount, Barrick, Scullen & Rounds, 2005; and Schmidt et al., 2008) and correcting for both direct and indirect restriction of range, job performance was regressed on general mental ability and the five personality traits; by varying the order of inclusion, the incremental validity of each block over the other was estimated.

The results showed that the contribution of general mental ability was substantially larger than that of personality, regardless of correction method and that the absolute level of validity increased substantially when correcting for indirect range restriction as compared to the traditional method for correcting for direct range restriction. Mainly general mental ability contributed to the increase in predictive validity while none of the personality predictors contributed to the same extent; the incremental validity of all five personality predictors when controlling for general mental ability was modest.

The results imply that correcting for direct range restriction produces an underestimation of the operational validity of individual differences in predicting job performance. More specifically, the old method underestimates the validity of general mental ability in predicting job performance. The results thereby not only confirm the role of general mental ability as the most important predictor of job performance but also show that its incremental validity compared to personality is greater than previously identified, which points to measures of general mental ability having greater utility in the selection process. The relative order of importance of the five personality traits was found to be in line with previous research; conscientiousness stands out as the most important predictor of performance. However, compared to general mental ability’s contribution to the increase in validity, conscientiousness did not contribute half as much.

The regression weights presented in Study I constitute optimal weights which specify how scores for each predictor should be weighted when combined in order to reach the maximum level of predictive validity. Thus, these estimates can be used for the mechanical weighting of scores derived from measures of general mental ability and the five factor model personality traits, which will assist in reaching the most accurate prediction of job performance. In addition, accurate estimates of operational validity for individual predictors and of the incremental validity for personality, when all five personality traits are taken into consideration, are necessary for accurately determining the financial costs of implementing these methods in a selection process, and for making relevant comparisons with other competing selection methods.

Study II

The utility gain of leaving professional judgment outside of prediction: Clinical versus mechanical interpretation of GMA and personality.

Despite the large amount of research supporting the validity and effectiveness of using a mechanical, algorithm-based approach for combining predictor scores (Dawes et al., 1989; Gough, 1962; Goldberg, 1968; Grove & Meehl, 1996; Grove et al., 2000; Kuncel et al., 2013; Meehl, 1965, 1967; Sarbin, 1943; Sawyer, 1966; Sines, 1970), the procedures used by practitioners often reflect a preference for a clinical approach to data combination with a focus on using professional judgment (Vrieze & Grove, 2009).

One possible way of increasing the use of established personnel selection research methods in applied practice is to encourage the acceptance of research results by presenting them in financial terms through the use of utility analysis (Cascio, 2000). An increased awareness of the factors that may impact the differences in utility between the personnel selection options is also a prerequisite for evaluating, comparing, and making rational choices about which data collecting and combining approach to use for prediction purposes.

The aim of Study II was to estimate and illustrate in financial terms the differences between clinical and mechanical approaches to data combination, using general mental ability and the personality traits of the five factor model as the predictors of job performance. Using Brogden's (1949) and Cronbach and Gleser's (1965) utility equation, Study II utilized the operational validity (R = .71) from Study I to represent the validity for the mechanical data combination of general mental ability and the five factor model personality traits. The meta-analytically estimated decrease of R = .16

for combining predictors clinically instead of mechanically (Kuncel et al., 2013) corresponds to a predictive validity of $R = .55$ for combining the same set of predictors clinically. By applying Schmidt, Hunter, Outerbridge, and Trattner's (1986) standard deviation of job performance as 40% of the annual salary of an employee, the margin utility of the data combination approach was analyzed for two levels of selection ratios: 30% and 70% respectively (Schmidt & Hunter, 1998). The cost of applying a clinical and a mechanical approach for data combination was taken into account in the utility analyses as well as tenure, the number of applicants, and selection ratio.

The results showed that the differences in validity between the data combination approaches can contribute to an extensive financial gain of applying a mechanical data combination approach compared to a clinical approach. The illustrated margin utilities between the two data combination methods are sizable and likely to have a severe impact on the success or failure of organizations. Emphasizing the importance of the data combination process and the superiority of the mechanical approach in utility terms can help professionals become aware of how to increase organizational performance by improving their selection practice. This study also illustrates how highly generalizable meta-analytic estimates can be utilized (cf. Study I, Kuncel et al., 2013; Schmidt et al., 2008) in order to estimate utility. Communicating research evidence in terms of utility serves the purpose of increasing its acceptance and awareness among practitioners, who can in turn refer to this knowledge when discussing human resource issues, including personnel selection, with business people.

Study III

Preference for hiring approach: Cognitive style or context dependent?

The aim of Study III was to investigate if and to what degree decision-making style (Scott & Bruce, 1995), accountability for the selection process, and responsibility for the selection decision (Tetlock, 1985) predict preference for hiring approach categorized into four alternatives representing combinations of clinical or mechanical data collection and clinical or mechanical data combination respectively. Previous research has indicated that individual differences in cognitive style are related to practitioners' preferences for hiring approach, and, specifically, that an intuitive decision-making style is related to a preference for a clinical hiring approach (Lodato, Highhouse, & Brooks, 2011). Decision-making researchers in other areas have argued that the contextual factors of accountability (Arkes, 1991) and responsibility (Tetlock, 1985) may lead to more accurate judgment.

In personnel selection practice, accountability is likely to be especially relevant for human resource professionals who may be held accountable for their hiring procedures and expected to be able to describe how they arrived at the overall suitability and comparative merits of each applicant. Responsibility is a contextual factor which puts more pressure on hiring managers, who are usually responsible for the actual outcome of the hiring decision.

The first overall proposition in Study III stated that decision-making style would predict preferred hiring approach and, specifically, that practitioners scoring high on intuitive decision-making style would tend to prefer a clinical hiring approach, while those scoring high on rational decision-making style would tend to prefer a mechanical hiring approach (Lodato et al., 2011). The second overall proposition stated that the selection context, in terms of procedural accountability in combination with responsibility for the selection decision, would predict preferred hiring approach. More specifically, it was hypothesized that practitioners who are held accountable for the selection process and responsible for the selection decision would be more likely to prefer a mechanical hiring approach compared to professionals who lack accountability and responsibility.

Postings on four Swedish human resource groups on the social networks LinkedIn and Facebook provided 168 respondents to the web-based questionnaire. The majority of the respondents reported that personnel selection was their main job task. After responding to background questions and two scales measuring intuitive and rational decision-making styles, the respondents were asked to imagine themselves in a selection context with 20 applicants remaining to be assessed for individual suitability. Next, each respondent was randomly assigned one of the four possible contextual scenarios. They were to imagine themselves as an employer who is accountable for the procedure and responsible for the decision, as an external human resource professional who is accountable for the process but not responsible for the decision, as an employer responsible for the decision but who hires external help to carry out and assume accountable for the process, or as an administrative assistant with no accountability for the process and no responsibility for the decision. The respondents were then asked to choose which hiring approach they preferred from among the four alternatives that were comprised of the combinations of clinical or mechanical data collection and clinical or mechanical data combination. Due to the small sample size, however, hiring approach was dichotomized prior to analysis into representing a "pure" mechanical approach (both data collection and combination are mechanical) and a clinical approach (both or either data collection or combination is clinical).

A logistic regression analysis was computed to predict the binary hiring approach variable (Menard, 2001) using two models: the first model used intuitive and rational decision-making styles as predictors, while in the second model the four context-dependent categories were added as predictors. The results showed partial support for the first overall proposition, in that individuals scoring high on intuitive decision-making style preferred a clinical hiring approach. However, individuals scoring high on rational decision-making style did not prefer a mechanical hiring approach. The results did not support the second overall proposition, as selection context in terms of procedural accountability in combination with responsibility for the selection decision was not found to predict preferred hiring approach.

Discussion

Making accurate hiring decisions is beneficial for all levels of society and the way individual assessments are executed and selection decisions are made represents important aspects of a legally secure and effective modern working life. At its best, personnel selection contributes extensively to the financial well-being of organizations and thus economies worldwide, and selection decisions based on accuracy and fair treatment provide individuals with appropriate work opportunities, increase diversity and help prevent discrimination in working life. At its worst, when inaccurate and irrelevant assessments and predictions are used, personnel selection can prevent individuals from being offered suitable work opportunities, allowing discrimination and unfair treatment and the financial loss for organizations can be extensive due to suboptimal job performance and cost-inefficient selection processes.

The overall objective of this thesis was to contribute to a narrowing of the research–practice gap in the area of personnel selection. In order to do this, the thesis has focused on tying together the important and closely aligned aspects of the selection process that are critical to successful hiring. By focusing on the financial outcomes of some of the components of the selection process which research has identified as crucial, but that are not always applied in practice, the goal has been to create awareness among practitioners and facilitate the improvement of selection work.

The specific aims of this thesis were investigated and discussed in the three empirical studies. The aim of the first study was to establish and provide more accurate estimates of the relationships between both general mental ability and the five factor model personality traits and job performance by applying a recently developed methodology for range restriction correction that utilizes previously published meta-analytic data. The aim of the second study was to determine and compare the marginal utilities of applying a mechanical and a clinical approach using the predictor weights from the first study for combining predictor scores. The aim of the third study was to increase our understanding of what underlies practitioners' preferences for mechanical or clinical hiring approaches to personnel selection by examining the influence of general cognitive style in decision-making and of the contextual aspects of both accountability for the selection process and responsibility for the selection decision.

In the following, results from the three studies will be discussed and somewhat integrated. Aspects of data collection and the choosing of predictors will be taken up first before looking at the combination of predictor scores, as it is these functions that more closely reflect how selection work is organized in practice. Each general discussion is followed by suggested practical implications within each function.

Data collection

The overarching aim of personnel selection is to contribute to the utility of organizations by hiring the most productive employees as cost-efficiently as possible. This requires collecting relevant information about applicants and accurately combining the data into an overall assessment for prediction-based decision-making. Accurate estimation of the predictive validity of general mental ability and personality traits, separately and jointly (i.e., studied as separate predictors or a group of predictors), in relation to job performance constitutes the groundwork for personnel selection and has several practical implications.

In the first study included in this thesis, the predictive validities of general mental ability and the five factor model personality traits for job performance were investigated. The re-analysis of previously published meta-analytic data indicated that the predictive validity of general mental ability for job performance had been underestimated in previous research due to a lack of specificity in the range restriction correction method used (Hunter & Schmidt, 2004). The predictive validity for general mental ability increased by 26% when the more accurate correction method for indirect range restriction was applied as compared to estimates based on the previous methodology for direct range restriction correction. This finding strengthens the position of general mental ability as the most important individual trait for predicting performance, enhancing the importance of general mental ability in solving problems at work, suggesting that this is crucial to job performance. The increased validity for general mental ability found in the study also confirms previous research which suggests that employee selection is traditionally based on an indirect and unknown third variable that is correlated with general mental ability, and that correcting for this indirect restriction of range is important for arriving at accurate validity estimates.

Personality, on the other hand, was found to have a modest role in predicting performance, with regard to both individual trait predictors and the predictor for the combined construct, which corresponds well with previous research (Schmidt, Schaffer & Oh, 2008). The results in this study thereby add support for personality being of less importance than general mental ability, as it was found to add a lesser amount of additional incremental validity to the prediction of performance than did general mental ability. Although the joint personality predictors, as compared to general mental ability, was found to be modestly associated with job performance, the associations found with the specific personality predictors differed. In accordance with previous research, the results showed that conscientiousness was the strongest personality contributor of incremental validity and thus the most important trait beyond that of general mental ability. For predicting job performance, conscientiousness was almost twice as important as the second most influential predictor among the personality variables, represented by openness (and it should be noted that this relationship was negative), but it was still not even half as important as general mental ability.

That none of the personality predictors showed increased validity after indirectly correcting for range restriction implies that when it comes to personality, applicants are in general not chosen systematically. If they had been, correction for indirect range restriction would have increased the predictive validity for the personality predictors. This implies that the generic relationship between conscientiousness and job performance, that is, that conscientiousness predicts performance across jobs, which has extensive empirical support (Barrick & Mount, 1991; Hurtz & Donovan, 2000; Mount & Barrick, 1995; Salgado, 1997, 2003), does not seem to have influenced selection practice.

Although some of the five factor model personality traits were only modest predictors of job performance, an examination of the lower-order personality traits, the subfacets of the five factor model traits, might have revealed greater predictive validity. There is convincing research that specific subfacets can be utilized either as individual predictors or combined and thereby more effectively predict work related criteria (Judge, Rodell, Klinger, Simon & Crawford, 2013). Notice, however, that unless these subfacets are separated from the general five factor model structure when implementing them in selection practice, the cost of implementing the method will remain the same.

Increasing the precision of estimations of the relationships between general mental ability and performance and between personality and performance allows for a more accurate evaluation and comparison of the utility of using them as predictors in selection practice. The substantive increase in validity found for general mental ability implies that tests measuring this trait are even more accurate for job performance prediction and more cost-efficient than what previous research has indicated. The difference in terms of efficiency between general mental ability tests and other competing selection methods has thereby increased. This strengthens the position of general mental ability tests as the most efficient predictor and selection method, confirming the role of other selection methods as complementary methods.

The sustained level of validity found for personality implies that there is no increase in utility for the use of personality tests in personnel selection compared to what previous research has already indicated (Schmidt & Hunter, 1998). Nevertheless, even with its modest contributions to incremental validity, personality measures still contribute to the utility of selection processes. The relationship found between personality and job performance was primarily due to the contribution of conscientiousness, which supports previous findings that a combination of measures of general mental ability and conscientiousness should be used. In order to evaluate the contribution of the other four personality traits to overall utility, the incremental validity of these traits would need to be set in relation to the cost of implementation. Even a small amount of incremental validity can contribute to a positive margin utility given that the cost of implementation is low.

In addition to being necessary for accurately evaluating the utility of both single (general mental ability) predictors and groups (personality traits) of predictors, the validity estimates for individual traits are useful in selection practice and can serve as a solution to a practical problem. The estimates correspond to the weights by which applicants' predictor scores could be weighted in order to reach the maximized level of validity of predicting job performance using these six predictors.

Based on the fact that conscientiousness showed the highest predictive validity and greatest contribution to incremental validity among the personality traits, it could be argued that it would be relevant to measure and include only conscientiousness in the prediction, and it is possible that from a strict utility perspective this could be true. However, this may be difficult to implement in practice for two reasons. First, most personality tests that are available to organizations are fixed and include a set of scales with associated subscales corresponding to a theoretical structure. Organizations rarely have the opportunity to choose, and thus only pay for, the specific scales, or predictors, that would provide them with the highest predictive validity as well as cost-efficiency. Second, an approach utilizing only conscientiousness would be lacking in face validity and thereby not requested by organizations or human resource practitioners. In the present study, all five personality traits, and their weights, were included in order to be more in line with contemporary practice and the type of implementation that is perhaps more likely to occur in selection processes. The main focus has been on the logic behind the way of choosing among predictors and selection methods and combining them for prediction purposes in relation utility. The predictors as well as the criteria could, at a theoretical level, change in number or be substituted with other predictors (e.g., subfacet scores) or criteria (e.g., counter productive work behavior), depending on the purpose of the prediction.

Data combination

Collecting relevant data is a prerequisite for making valid selection decisions, but it is clearly not enough. Data, in this case applicants' scores on general mental ability and the five personality traits, inevitably needs to be interpreted or combined into an overall assessment upon which the decision to hire or not is based. Although previous research comparing the two approaches to combining data, clinical and mechanical, has shown that the mechanical approach is superior to the clinical approach in terms of its validity for predicting performance, the aspect of financial impact had not been previously examined, at least not to my knowledge. Study II in this thesis illustrated that the difference in margin utility between the two data combination approaches could be extensive when examining fairly common selection scenarios involving the standard deviation of job performance and selection ratios.

In the second study, the validity estimates from Kuncel et al. (2013) were applied to different selection scenarios, illustrating the differences financial implications associated with applying clinical and mechanical approaches to data combination in personnel selection. This study's findings on margin utilities draw attention to the financial impact of data collection in general, an aspect that is often overlooked in selection practice. Expressing this research outcome in monetary value highlights the importance of combining the collected data accurately, and that the effect of data combination method could be sizable. Personnel selection practitioners are provided with a more accessible rationale for changing this aspect of selection practice.

In addition, presenting research results in financial terms encourages the discussion of selection decisions to focus more on financial decision making than on specific data collection and data combination activities. This transition is necessary if rational and beneficial choices are to be made from among the selection options, and if the objective of personnel selection, in terms of increasing organizational utility, is to be reached. The utility framework, and outcomes like the one illustrated in the second study, points out that activities connected to data collection and data combination are not ends in themselves but, rather, represent the means of rank-ordering job applicants on job performance as accurately and effectively as possible. Not following recommendations for best practice can thus lead to incorrect hiring decisions that may result in very serious financial consequences for an organization.

Increased profit for the organization is not the only reason for encouraging practitioners to increase their use of the mechanical data combination approach, although this has been the main focus of the empirical research in this thesis. Even for organizations with objectives other than profit gain, there is still compelling evidence for using professional and effective selection processes. In the public sector, for example, taxpayer money is often spent on less efficient selection processes which often contribute to poor selection decisions. In turn, this leads to a low or even negative margin utility, incurring larger costs than necessary for taxpayers. In addition, there are other reasons for using a mechanical data combination approach that are more ethical in character.

Making the most accurate selection decisions carries with it an ethical aspect in that by offering the job to the most suitable applicant or applicants instead of to relatively less suitable applicants, individuals are receiving fair access to work opportunities and careers that are suitable for them. The ethical perspective, however, is more often used when advocating the necessity of using a clinical approach for combining the predictor scores. Incorporating professional judgment seems to be viewed as a way to compensate for what they may consider to be dehumanizing and therefore unethical about the mechanical approach, and is thus more of a question of the ethical views of the practitioner. However, when also considering the ethics of lost job opportunities for individuals, the question of what represents good ethics becomes less clear cut. A reevaluation of the ethics of not using the most relevant and accurate predictors and combining them as efficiently as possible may be in order.

The function of data combination constitutes the actual prediction; it specifies the rules for how the predictor scores are to be combined in order to reach maximum predictive validity. These optimal predictor weights thus represent the logic and rationale behind the actual selection decision, and the mechanical combination itself guarantees that this logic is equally applied across all applicants. By utilizing these two features, the two-fold problem that is often associated with the clinical data combination approach is avoided, namely the tendency of practitioners to assign non-optimal weights to predictors, and the tendency to apply those non-optimal weights inconsistently across applicants (Meehl, 1998). Together with the rationale behind the choice of predictors, or rather the constructs represented by the predictors, mechanical data combination is required in order to answer the question of "why" an applicant did or did not get the job, and ensures that the selection decisions are based on the same accurate information across applicants. This makes it possible for feedback to be given to applicants that explains the reasons behind their outcomes, which is reasonable to expect from the applicant's perspective and also in line with good ethics and professionalism within the personnel selection community. Being able to provide an explicit explanation of the rationale behind a hiring decision can also be used to demonstrate that an applicant was not rejected based on unlawful grounds such as gender or age discrimination. From a practitioner's perspective the mechanical approach with empirically derived weights can be viewed as a useful tool for fair treatment and as an "insurance policy" in cases where a hiring-decision disagreement could take on legal proportions. In other circumstances, the algorithm with its predictor weights could be viewed as an empirically based job profile outlining the sought-after traits, which could be used to explicitly explain the grounds of the selection process to applicants prior to their participation. As such, the mechanical approach for data combination can ensure that the job profile does not change during the course of the selection process or across candidates.

Note also that although Study II uses a proper linear model, a model in which predictor variables are given weights in such a way that the resulting linear composite optimally predicts some criterion of interest, even improper linear models, models in which the weights of the predictor variables are obtained by some nonoptimal method; for example obtained on the basis of intuition or set to be equal, has proven to be superior to clinical data combination for prediction (Dawes, 1979). In fact, unit (i.e., equal) weighting seem quite robust for making such predictions (Hattrup, 2013) implying that calculating simple average of standardized predictor scores rank-order job applicants effectively, that it is the actual mechanical combination, rather than the derivation of weights, which is of importance, and that this is an improvement that comes at a very low cost.

In addition, having transparent and explicit rules for how selection decisions are made also makes it possible to conduct a thorough evaluation of the process. Without this, relevant improvements and corrections in the data combination process would be difficult to make.

Practitioners’ preferences for hiring approach

The extensive research in the field of personnel selection has contributed with some important and general recommendations for best practice. Several of these guidelines, however, have only gained modest acceptance from practitioners and are not yet fully incorporated in practice. Examples of this concern the selection of accurate and relevant constructs to measure, the use of standardized and cost-efficient methods as predictors of the constructs, and utilizing mechanical methods for combining predictor scores (see Study I and II). Given the fact that it has been well documented that selection decisions based on relevant predictors and transparent and standardized rules for combining predictor scores are beneficial to all stakeholders (applicants, organizations, and practitioners) – and that the process can be argued to be in line with good ethics and professionalism – why is there still such an extensive gap between research and practice (cf. Highhouse, 2008)?

This is not an easy question to answer and it likely depends on several factors (Born & Scholaris, 2005). The prevalent use of less efficient predictors has been suggested to be a result of both preference and lack of knowledge (Kuncel, 2008). Study III in this thesis focused specifically on investigating factors that might influence practitioners’ hiring approach preferences, defined as four combinations of clinical and mechanical approaches to data collection and data combination. The results were in line with previous research (Lodato et al., 2011), showing that practitioners with an intuitive decision-making style preferred a clinical hiring approach (i.e., preferred a clinical approach to data collection and/or data combination). Overall, very few of the participating practitioners preferred a mechanical hiring approach (i.e., preferred a mechanical approach to both data collection and data combination), a situation very common among practitioners (see, e.g., Vrieze & Grove, 2009). These relationships held regardless of the degree of accountability for the selection process and of the level of responsibility for the selection decision, which excludes these factors from being potential aspects to address in order to try to change practitioners' preferences for using the mechanical approach.

The consistent preference that practitioners show for utilizing human judgment rather than an algorithm-based approach for data combination may be due to individuals using faulty heuristics for reasoning and judging evidence that can result in all sorts of inaccuracies (Tversky & Kahneman, 1974). In the context of data combination for personnel selection purposes, one such faulty heuristic could be that the mechanical approach, in contrast to the hypothesis in Study III, actually invokes a perceived *lack* of control and flexibility rather than the other way around, and that the clinical approach invokes a perceived increase in control. If so, practitioners who feel greater pressure in their work may be more drawn to a clinical hiring approach, as it could give them the impression of having greater control over the selection process.

Methodological considerations and future research

The methods used to arrive at the results presented in this thesis warrant some attention along with a discussion of how the results and implications may be used in future studies. In the following, both the general limiting aspects and the specific potential limitations of each study will be discussed.

The results in this thesis are based on cross-sectional data; the data on predictors and criteria was collected at the same point in time, which often raises the question of causality between predictors and criteria. In personnel selection, however, the objective of the predictors is not to contribute with causal inferences, it is to predict. Predictive research emphasizes practical applications, such as selection decisions, whereas explanatory research, which may be both relevant and interesting from a broader perspective, focuses on achieving a theoretical understanding of the phenomenon of interest (Venter & Maxwell, 2000).

With regard to the generalizability of the results, the situation is quite different for the three empirical studies. Since Study I was based on a reanalysis of previously published meta-analytic data, it suffers from the same strengths and limitations as the original meta-analyses regarding inclusion of studies, sample size, sample composition, and so forth (Schmidt et al., 2008). Meta-analytic estimates, which are based on extensive data sets, are more likely to be more stable, accurate, and generalizable than estimates based on small samples in single studies. It is still possible, however, that the established weights may be too general for the actual measures of personality and general mental ability used in the selection practice, and that they may lack ecological validity in a specific selection setting using corresponding predictors. Future studies of specific measures of personality and general mental ability would be called for to shed more light on the impact of each of these factors on job performance.

The generalizability of the results in Study III suffers because of the small and non-randomized sample used in this study. Since the sampling was done by convenience, there is no way to rule out the possibility that this may have affected the results. However, skewness in representativity concerning hiring approach is more likely to have favored a mechanical hiring approach (attracting more evidence-based practitioners) but this does not seem to have been the case since the overall preferences for a mechanical hiring approach were low. Furthermore, the reliance on artificial scenarios may limit the generalizability of the results in Study II. Future studies should focus on investigating the preferences of various groups of service providers and clients in the personnel selection industry by asking them to relate to their own experiences. For example, Born and Scholarios (2005) has suggested to use bounded rationality theory (Simon, 1957), the idea that in decision-making, rationality of individuals is limited by the information they have, the cognitive limitations of their minds, and the finite amount of time they have to make a decision, to understand the complexity of personnel selection decisions. Born and Scholarios (2005) conceptualize the process of selection decision making as a series of overlapping layers: the individual decision maker, the organizational context, and the wider environment, and suggests that interactions between these layers shape selection decisions. Such a model could be applied in order to investigate the possible complexity of selection decisions that goes beyond the individual decision maker.

The results in Study II are dependent on certain assumptions which may have implications for the generalizability of the results. These assumptions are embedded in the BCG model (Cascio & Bodreau, 2008) for estimating the margin utility, which, for example, assumes that there is a linear relationship between validity and utility, and that there are certain estimated costs for implementing data collection methods. Traditionally, these assumptions are not tainted by excesses (the estimated cost of applying the clinical approach, for example, is based on a single professional making the prediction, while in practice, the more expensive method of consensus meetings is common) but future studies should examine whether these assumptions can be confirmed in selection practice.

Another general potential limiting aspect concerns the operationalization of predictors, general mental ability and personality, and the criterion of job performance. General mental ability, in terms of its theory, operationalization, measurement, and relevance, lacks any strong competitors, and it is unlikely that even if it were operationalized in some other way, the results would have been affected to the extent that it would change recommendations for selection practice. Concerning the use of the five factor model of personality, however, it should be noted that despite it being widely acknowledged as a relevant and useful conceptual framework for the study of personality traits, there is some disagreement concerning it. There is, for example, strong support that higher-order factors above the five factor level (De Young, 2006; Digman, 1997; Markon, Krueger, & Watson, 2005; Mount et al., 2005), as well as lower order subfacets (Judge, Rodell, Klinger, Simon & Crawford, 2013; Ones & Viswesvaran, 2001), are useful for the prediction of job performance and other work-related criteria. This includes empirical support for the incremental validity of combining specific lower order facets beyond what the broader five factor model traits can accomplish. However, this seems to be true for more narrow and specific criteria, criteria at the same level of abstraction as the subfacet predictors (Ones & Viswesvaran, 2001).

General job performance is the traditional criterion to use in research on individual differences, performance at the workplace, and personnel selection, and there is consistent support for the relevance of using a general, overall construct of job performance. This seems especially true if the predictors are at a higher level of abstraction, such as general mental ability and the five factor model traits. However, most studies, including the primary studies included in the meta-analysis underlying Study I in this thesis, are based on supervisory ratings. With this type of rating, contextual performance tends to be more heavily weighted than task performance, which may affect the results (Rotundo & Sackett, 2002).

Future studies are also needed to further explore if there are other individual differences, such as values and interests, which can provide overall utility in a selection process. A recent meta-analysis of the relation between interests and job performance showed substantial validity estimates (Van Iddekinge, Roth, Putka, & Lanivich, 2011), suggesting that vocational interests may hold more promise for predicting job performance than what research has suggested up until this point in time.

Conclusions

Successful hiring is crucial from many aspects and for many stakeholders; it is of financial importance for organizations, societies, and taxpayers, and of personal importance for applicants and personnel selection practitioners.

Unfortunately the gap between research and practice is wide, especially with regard to how collected data is combined, suggesting that there is a need for improvement of current selection practice. With the aim of contributing to a narrowing of this research–practice gap, this thesis has focused on some important and closely aligned aspects of the selection process which are critical to successful hiring.

The results in Study I indicated that when the more accurate correction for range restriction introduced by Hunter et al. (2006) was applied, the validity increased for general mental ability as a predictor of job performance while the personality traits showed no or modest increase. The relative importance of general mental ability as compared to personality thereby increased, which should be taken into account in selection practice.

In Study II, where mechanical and clinical approaches to data combination were compared in financial terms using general mental ability and five factor model personality traits as predictors of job performance, it was illustrated that the financial gain of applying a mechanical data combination approach (using the empirically established estimates from Study I) can be extensive and have serious impact in many organizations.

Despite the previously established superiority of the mechanical approach concerning its validity for prediction, the clinical approach to combining data is most common in practice. The results in this thesis show that practitioners' preference for a clinical approach was determined by their general cognitive decision-making style rather than by the context-related aspects of being accountable for the assessment process or responsible for the selection decision. Thus, it is not likely that holding practitioners accountable for the assessment process or responsible for the selection decision would help shift their preferences more towards using an evidence-based mechanical hiring approach.

References

Allport, F. H., & Allport, G. W. (1921). Personality traits: Their classification and measurement. *Journal of Abnormal and Social Psychology, 16*, 1–40.

Allport, G. W. (1935). Attitudes. In C. Murchison (ed.), *Handbook of social psychology*, (pp. 798–844). Worcester, MA: Clark University Press.

Allport, G. W. & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs, 47*, i–171.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110*, 486–498.

Barrick, M. R., Mount, M. K. (1991). The Big Five personality dimensions and job performance. *Personnel Psychology, 44*, 1–26.

Barrick, M. R., & Mount, M. K. (2003). Impact of meta-analysis on understanding personality and performance relations. In K. Murphy (Ed.), *The impact of validity generalization methods on personnel selection* (pp. 197–221). Mahwah, NJ: Erlbaum.

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

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and job performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9,* 9–30.

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, Organizational deviance, and their common correlates: A Review and meta-analysis. *Journal of Applied Psychology, 92*, 410–424.

Binet, A., & Simon, T. (1916). *The Development of Intelligence in Children.* New York, NY: Arno Press.

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74*, 478–494.

Bodreau, J. W., & Ramstad, P. M. (2003). Strategic industrial and organizational psychology and the role of utility analysis models. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Volume Eds.), *Handbook of Psychology*, Volume 12, *Industrial and Organizational Psychology* (pp. 193-221). Hoboken, NJ: Wiley.

Borgatta, E. F., & Lambert, W. W. (1968). *Handbook of personality theory and research.* Chicago, IL: Rand McNally.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco, CA: Jossey-Bass.

Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L. A. (1993). Role of early supervisor experience in supervisor performance. *Journal of Applied psychology, 78,* 443–449.

Born, M. Ph., & Scholarios, D. (2005). Decision making in selection. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection* (pp. 267–290). Malden, MA: Blackwell Publishing.

Brealy, R. A., & Myers, S. C., & Allen, F. (2006). *Principles of corporate finance* (8th ed.). Burr Ridge, IL: Irwin/McGraw-Hill.

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology, 37*, 64–76.

Brogden, H. E. (1949). When testing pays off. *Personnel Psychology, 2*, 171–185.

Cabrera, E. F., & Raju, N. S. (2001). Utility analysis: Current trends and future directions. *International Journal of Selection and Assessment, 9*, 92–102.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (pp. 687-732). Palo Alto, CA: Consulting Psychologists Press, Inc.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factoranalytic studies.* Cambridge, UK: Cambridge University Press.

Cascio, W. F. (2000). *Costing human resources. The financial impact of behavior in organizations (4th ed.).* Cincinnati, OH: South-Western College Publishing.

Cascio, W. F., & Bodreau, J. (2008). *Investing in people. Financial impact of human resource initiatives.* New Jersey, NJ: Pearson Education.

Cattell, J. M. (1890). Mental tests and measurements. *Mind, 15*, 373–381.

Cattell, R. B. (1943). The description of personality. Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology, 38*, 476–507.

Cattell, R. B. (1945). The description of personality: Principles and findings in a factor analysis. *American Journal of Psychology, 58*, 69–90.

Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology, 84,* 3–13.

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

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Coward, M. W., & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology, 75,* 297–300.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and PersonnelDecisions* (2nd ed.). Urbana, IL: University of Illinois Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist 34,* 571–582.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgement. *Science, 243*, 1668–1674.

De Young, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*, 1138–1151.

Digman, M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*, 1246–1256.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology, 44*, 329–344.

Freyd, M. (1926). The statistical viewpoint in vocational selection. *Journal of Applied Psychology, 4*, 349–356.

George, J. M., & Brief, A. P. (1992). Feeling good-doing good: A conceptual analysis of the mood at work-organizational spontaneity relationship. *Psychological Bulletin, 112,* 310–329.

Goldberg, E (1968). Are women prejudiced against women? *Transaction, 5*, 28–30.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (pp. 141–165). Beverly Hills, CA: Sage.

Goldberg, A. E. (1995) *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago, IL: Chicago University Press.

Gottfredson, L. (1997a). Why g matters: The complexity of everyday life. *Intelligence, 24,* 79–132.

Gottfredson, L. S. (1997b). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*, 13–23.

Gough, H. G. (1962). Clinical versus statistical prediction in psychology. In L. Postman (Ed.), *Psychology in the making* (pp. 526–584). New York, NY: Knopf.

Graziano, W. G., & Eisenberg, N. H. (1997). Agreeableness: A dimension of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality. Theory and research* (pp. 795–825). New York, NY: Guilford press.

Grove, W. M., & Lloyd, M. L. (2006). Meehl's contribution to clinical versus statistical prediction. *Journal of Abnormal Psychology, 115*, 192–194.

Grove, W. H., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical statistical controversy. *Psychology, Public Policy, and law, 2*, 293–323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 1*, 19–30.

Guilford, J. P. (1988). Some changes in the structure of intellect model. *Educational and Psychological Measurement, 48,* 1–4.

Guion, R. M. (1991). Personnel assessment, selection, and placement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* 2nd ed, (Vol. 2, pp. 327–397). Palo Alto, CA: Consulting Psychologists Press.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18,* 135–164.

Hale, M. (1992). *Human science and the social order. Hugo Munstberg and the origins of applied psychology.* Philadelphia, NJ: Temple University Press.

Hattrup, K. (2013). Using composite score in personnel selection. In N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection* (pp. 297–319). Oxford, UK: Oxford University Press.

Hazer, J T., & Highhouse, S. (1997). Factors influencing managers' reactions to utility analysis: effects of SD_y method, information frame, and focal intervention. *Journal of Applied Psychology, 82*, 104–112.

Hemmingsson, T., Melin, B., Allebeck, P., & Lundberg, I. (2006) The association between cognitive ability measured at ages 18–20 and mortality during 30 years of follow-up—a prospective observational study among Swedish males born 1949–51. *International Journal of Epideomiology, 35,* 665–670.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology, 1,* 333–342.

Hoffman, B. J., Blair, C., Meriac, J., & Woehr, D. J. (2007). Expanding the criterion domain? A meta-analysis of the OCB literature. *Journal of Applied Psychology, 92,* 555–566.

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*, 100–112.

Hogan, J., & Ones, D. S. (1997). Conscientiousness and integrity at work. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of* *personality. Theory and research* (pp. 849–872). New York, NY: Guilford press.

Holzinger, K., & Swineford, F. (1939). A study in factor analysis: The stability of a bifactor solution. *Supplementary Educational Monograph, 48*, xi–91.

Hough, L. M. (2001). I/Owes its advances to personality. In B. W. Roberts & R. T. Hogan (Eds.), *Personality psychology in the workplace* (pp. 19–44). Washington: American Psychological Association.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis.* Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology, 75*, 28–42.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings, *Journal of Applied Psychology, 91,* 594–612.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879.

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology, 96*, 1167–1194.

Jensen, A. R. (1998). *The g Factor. The Science of Mental Ability*. Westport, CT: Praeger Publisher.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality. Theory and research* (pp. 114–156). New York, NY: Guilford press.

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence, 34*, 95–107.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765–780.

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology, 92*, 107–127.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology, 98,* 875–925.

Judiesch, M. K., Schmidt, F. L., & Mount, K. K. (1996). An improved method for estimating utility. *Journal of Human Resource Costing and Accounting*, 1.2, 31–42.

Kim, Y., & Ployhart, R. E. (2013, December 30). The Effects of staffing and training on firm productivity and profit growth before, during, and after the great recession. *Journal of Applied Psychology,* doi: 10.1037/a0035408.

Kuncel, N. R. (2008). Some new (and old) suggestion for improving personnel selection. *Industrial and Organizational Psychology, 1,* 343–346.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology, 98*, 1060–1072.

Langhammer, K. (2013). *Employee selection: Mechanisms behind practitioners' preference for hiring practices.* Doctoral thesis, Stockholm university, Stockholm, Sweden.

Le, H., Oh, I., Shaffer, J. A., & Schmidt, F. L. (2007). Implications of methodological advances for the practice of personnel selection: How practitioners benefit from recent developments in meta-analysis. *Academy of Management Perspectives, 21*, 6–15.

Lodato, M. A., Highhouse, S., & Brooks, M. E. (2011). Predicting practitioner preferences for intuition–based hiring. *Journal of Managerial Psychology, 26*, 352–365.

Mabon, H. (2002). *Arbetspsykologisk testing. Om urvalsmetoder i arbetslivet*. Stockholm, Sweden: Psykologiförlaget AB.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*, 139–157.

Markus, K., & Borsboom, D. (2013). *Frontiers of validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality. Theory and research* (pp. 159–181). New York, NY: Guilford press.

Meehl, P. E. (1954). *Clinical versus statistical prediction.* Minneapolis, MN: University of Minnesota.

Meehl, P. E. (1965). Seer over sign: The first good example. *Journal of Experimental Research in Personality, 1,* 27–32.

Meehl, P. E. (1967). What can the clinician do well? In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 594–599). New York, NY: McGraw-Hill.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50*, 370–375.

Meehl, P. E. (1998, May 23rd). The power of quantitative thinking. Speech delivered upon receipt of the James McKeen Cattell Fellow award at the meeting of the American Psychological Society, Washington, D.C.

Menard, S. (2001). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage Publicatons, Inc.

Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.

Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resources management. *Research in Personnel and Human Resources Management, 13*, 153–200.

Mount, M. K., Barrick, M. R., Scullen, S. M., & Rounds, J. (2005). Higher order dimensions of the Big Five personality traits and the Big Six vocational interest types. *Personnel Psychology, 58*, 447–478.

Murphy, K. R. (2012). Individual differences. In N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection* (pp. 48–67). Oxford, UK: Oxford University Press.

Murray, R. B. (2005). Yes, personality matters: Moving on to more important matters. *Human performance, 18*, 359–372.

Naylor, J. C., & Shine, L. C. (1965). A table for determining the increase in in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology, 3*, 33–42.

Neisser, U. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77–101.

Ones, D., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment, 9,* 31–39.

Organ, D. W. 1988. *Organizational citizenship behavior: The good soldier syndrome.* Lexington, MA: Lexington Books.

Pretz, J. E., & Totz, K. S. (2007). Measuring individual differences in affective, heuristic, and holistic intuition. *Personality and Individual Differences, 43*, 1247–1257.

Putka, D. J., & Sackett, P. T. (2010). Reliability and validity. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of Employee Selection* (pp. 9–49). New York, NY: Routledge.

Raju, N. S., Burke, M. J., & Normand, J. (1990). A new approach for utility analysis. *Journal of Applied Psychology, 75*, 3-12.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology, 79,* 518–524.

Reichenbach, H. (1938). *Experience and prediction*. Chicago, IL: University of Chicago Press.

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviours: A multidimensional scaling study. *Academy of Management Journal, 38*, 555–572.

Robinson, S. L., & Bennett, R. J. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology, 85*, 555–572.

Roth, P. L., & Huffcut, A. I. (2013). A meta-analysis of interviews and cognitive ability. *Journal of Personnel Psychology, 12,* 157–169.

Rotundo, M. & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology, 87*, 66–80.

Rydberg, S. (1963). *Bias in prediction. On correction methods.* Stockholm, Sweden: Almqvist Wiksell Boktryckeri AB.

Sackett, P. R., & DeVore, C. J. (2001). Counterproductive behaviours at work. In N. Anderson, D. S. Ones, H. K. Sinangil, & V. Viswesvaran (Eds.), *International Handbook of Work Psychology* (Vol. 1, pp. 145–164. London, UK: Sage Publications.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112–118.

Sackett, P. R., Borneman, M. J., & Connelly, S. B. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63, 215*–227.

Salgado, J. F., & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European Community. *Journal of Work and Organizational Psychology, 12*, 1–17.

Salgado, J. F. (1997). The five-factor model of personality and job performance in the European Community. *Journal of Applied Psychology, 82*, 30–43.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychological, 76*, 323–346.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F., & Rolland, J. P. (2004). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology, 88,* 1068–1081.

Sarbin, T. R. (1941). Clinical psychology – Art or science? *Psychometrica*, *6*, 391–400.

Sarbin, T. R. (1943). A contribution to the study of actuarial and individual methods of prediction. *American Journal of Sociology, 48*, 593–602.

Sarbin, T. R. (1944). The logic of prediction in psychology. *Psychological Review, 51*, 210–228.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178–200.

Schmidt, F. L., & Hunter, J. E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology, 68*, 407–414.

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. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Applied Psychology, 86*, 162–173.

Schmidt, F. L., Schaffer, J. A., & Oh, I. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology, 61*, 827–868.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432-439

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Trattner, M. H. (1986). The economic impact of job selection methods on size, productivity, and payroll costs of the federal work force: An empirically based demonstration. *Personnel Psychology, 39*, 1–29.

Schmitt, N. W., Arnold, J. D., & Nieminen, L. R. G. (2010). Validation strategies for primary studies. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of Employee Selection* (pp. 51–71). New York, NY: Routledge.

Schmitt, N., Gooding, R. Z., Noe, R. D., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*, 407-422.

Scott, S. G., & Bruce, R. A. (1995). Decision making style: The development of a new measure. *Educational and Psychological Measurement, 55*, 818–831.

Simon, H. A. (1957). *Administrative Behavior* (2nd ed.). New York, NY: Macmillan.

Sines, J. O. (1970). Actuarial versus clinical prediction in psychopathology. *British Journal of Psychiatry, 116,* 129–144.

Sjöberg, S., Sjöberg, A., Näswall, K., & Sverke, M. (2012). Using individual differences to predict job performance: Correcting for direct and indirect range restriction. *Journal of Scandinavian Psychology, 53*, 368–373.

Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology, 15*, 201–292.

Spearman, C. (1927). *The abilities of man*. New York, NY: Macmillan.

Sternberg, R. J. (1985): *Beyond IQ: A triarchic theory of human intelligence.* New York, NY: Cambridge University Press.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology, 23*, 565–578.

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48*, 227–236.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York, NY: Wiley.

Thorndike, R. L,. & Hagen, E. (1961). *Measurement and evaluation in psychology and education (2nd ed.).* Oxford, UK: John Wileys & Sons, Inc.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review, 38,* 406–427.

Thurstone, L. L. (1944). Second-order factors. *Psychometrika, 9,* 71–100.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. *Journal of Personality, 60*, 225–251.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Venter, A., & Maxwell, S. E. (2000). Issues in the use of multiple regression analysis. In H. E. A. Tinsley & S. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling: A Comprehensive Guide for Applied researchers in the Biological Sciences, Social Sciences, and Humanities* (pp. 151–182). San Diego, CA: Academic Press.

Watson, E. D., & Clark, L. A. (1997). Extraversion and its positive emotional core. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality. Theory and research* (pp. 767–794). New York, NY: Guilford press.

Wiggins, J. S., & Trapnell, P. D. (1997). Personality structure: The return of the Big Five. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality. Theory and research* (pp. 737–766). New York, NY: Guilford press.

Vinchur, A. J. (2007). A history of psychology applied to employee selection. In L. L. Koppes (Ed.), *Historical perspectives in industrial* *and organizational psychology* (pp. 193–218). Mahwah, NJ: Lawrence Erlbaum.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8,* 216–226.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90,* 108–131.

Viteles, M. S. (1925). *The clinical viewpoint in vocational selection. Journal of Applied Psychology, 2*, 131–138.

Viteles, M. S. (1932). *Industrial psychology*. New York, NY: Norton.

Vrieze, S. I., & Grove, W. M. (2009). Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Professional Psychology: Research and practice, 40*, 525–531.