

The impact of research synthesis methods on industrial–organizational psychology: The road from pessimism to optimism about cumulative knowledge

David S. DeGeest^{*†} and Frank L. Schmidt

This paper presents an account of the impact that research synthesis methods, in the form of psychometric meta-analysis, has had on industrial/organizational (I/O) psychology. This paper outlines the central contributions of psychometric meta-analysis in providing a method for developing cumulative knowledge. First, this paper describes the concerns and the state of the field before the development of meta-analytic methods. Second, the paper explains how meta-analysis addressed these concerns. Third, the paper details the development of psychometric meta-analysis through VG research and describes how the use of psychometric meta-analysis spread to other topic areas in the field. Finally, the paper presents illustrative example literatures, such as training and leadership, where meta-analysis had crucial impacts. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: meta-analysis; industrial–organizational psychology; validity generalization; research synthesis

Introduction

Research synthesis methods, first in the form of validity generalization (VG) and later psychometric meta-analysis, have had a broad impact in the field of industrial–organizational (I/O) psychology and the related disciplines of human resources management (HRM) and organizational behavior (OB). The development of research synthesis methods that simultaneously correct for the distorting effects of both sampling error and measurement error, and sometimes other research artifacts, radically altered these research fields. Major changes in theoretical paradigms, theory development and testing, and the empirical practices of researchers and practitioners have come about as a result.

VG is a particular type of psychometric meta-analysis conducted to determine whether a particular psychological construct, test, or measure has validity in predicting job performance regardless of situation or setting. This question is important to employers for economic and legal reasons (Schmidt and Hunter, 1998). Psychometric meta-analysis grew out of VG research and has proven to be the most effective tool developed in the I/O field to conduct research synthesis. Most forms of meta-analysis take into account, and correct for, only the effects of sampling error on research findings. Psychometric meta-analysis is a form of meta-analysis that simultaneously corrects both for sampling error and for measurement error. Sampling error is the random deviation of a sample from the properties of the population it is intended to represent; as such it creates random variation in results. Measurement error is the random deviation of the observed score or measure from the true value it is intended to measure. The main effect of measurement error is to bias the size of observed effects and relationships downward. Sampling error and measurement error are present in all real-world data

Department of Management and Organizations, Tippie College of Business, University of Iowa, IA, U.S.A.

**Correspondence to: David S. DeGeest, W252 John Pappajohn Business Building, University of Iowa, Iowa City, IA 52242-1994, U.S.A.*

†E-mail: david-degeest@uiowa.edu

sets and both create distortions in research results that need to be corrected to reveal the real meaning of the data.

Psychometric meta-analysis methods are based on the random effects model rather than the fixed-effects model (cf. Hunter and Schmidt, 2004; Schmidt *et al.*, 2009). The procedures estimate the amount of between-study variability of effect sizes that is due to sampling error variance and variance due to study differences in (a) measurement error, (b) restrictions in the range of the observed values, and (c) other research artifacts (e.g. dichotomization of continuous measures). (A research artifact is any methodological property of the study that distorts research results.) Any remaining variability in study outcomes is viewed as being potentially due to moderators (interactions between study properties and study results) and there are attempts to identify these moderators (usually via sub-grouping of studies into groups homogeneous on hypothesized moderators). At this point, the model becomes a mixed effects model if the moderators account for all the remaining variance (Overton, 1998). Otherwise, the model remains a random effects model.

A number of statistics are usually reported; those on which the most important interpretations are based are: (a) the mean of the correlational or *d* effect size values (after correction for measurement error and other research artifacts) that estimate the mean population parameter; (b) the estimated standard deviation (SD) of these population parameters (independent of sampling error, measurement error, etc.); (c) credibility intervals about these means based on the SD in (b); and (d) confidence intervals around the mean corrected values. Credibility intervals are most relevant when researchers are interested in the distribution of population parameters. For example, an 80% credibility interval around the estimated population mean indicates the bounds within which 80% of the (variable) study population effect size values (population parameters) are estimated to lie. A set of computer programs is widely used to implement these procedures (Schmidt and Le, 2004).

In the remaining sections of this paper, we first describe the pessimism about prospects for progress that characterized the field in the years before the introduction of psychometric meta-analysis. The second section then details the development of psychometric meta-analysis and its first application in VG. The final section examines the changes in research areas beyond VG of employee selection methods in the wider field since the introduction and adoption of psychometric meta-analysis as the primary method of research synthesis.

The state of the field before meta-analysis

The two major conceptual components in I/O psychology in the 1970s that would change radically with the development of VG and the subsequent adoption of meta-analysis in the field were the theory of situational specificity and pessimism regarding the hope for cumulative knowledge. Researchers mourned their seemingly fundamental inability to create replicable results: different studies produced different results, both in terms of statistical significance and the size of relationships. It was difficult for researchers in I/O psychology to answer basic questions important to social programs and policy, such as whether employee training programs were effective or whether various tests were useful in classifying and placing individuals for employment (Burke, 1984; Schmidt, 1992).

Situational specificity was a theory or belief held in I/O psychology from about 1925 and into the 1970s that postulated that the predictive validity of personnel selection procedures was situation-specific (Ghiselli, 1966; Schmidt and Hunter, 1984). Predictive validity is defined as the ability of a personnel selection test or other selection procedure to predict later performance on the job and is assessed via the correlation between the test and later job performance. The situational specificity theory held that variation in observed validities across settings was caused by subtle differences in the nature of job performance in different settings. This meant that apparently identical jobs in two different businesses or two different organizations would be characterized by location-based differences that altered the validity of employment tests used in those locations. Moreover, these differences in jobs were believed to be difficult or impossible to detect via job analysis or other systematic means.

Belief in situational specificity created a fundamental instability in selection and hiring practices and required expensive local validation studies in each setting or organization. As Murphy and Newman (2003) noted, 'Prior to 1977, personnel selection systems were something like a vintage automobile. They were custom built from the ground up, relying on procedures that consumed lots of time and energy, and nobody was sure they would really work' (p. 409). There was little or no recognition of the role of sampling error in producing variable results across small studies. Because of limited statistical sophistication, researchers at the time lacked an understanding of how large sampling errors could be in small studies. They also failed to appreciate the variability in results produced by variation across studies in measurement error and range restrictions on the measures used to estimate the validity coefficients.

The problem posed by situational specificity in employment testing was related to a larger trend in social science research: studies commissioned to examine the efficacy of various social programs, situational interventions, and other applications of economics, sociology, and psychology showed unexpected variability in effectiveness; they were effective in some studies and settings but not others. This instability had multiple

effects on I/O psychology and social science research in general. First, and most pervasively, it led to a cycle of pessimism. On any particular question or topic, individual researchers would conduct studies with relatively small samples and find conflicting answers about effect sizes and statistical significance for similar studies. Qualitative narrative reviews that examined the studies would fail to come to meaningful conclusions and would usually suggest that a search for moderators might explain the variability in findings. No study or review was able to create a simple explanation or cumulative body of knowledge describing the effects and findings of a particular research program. At this point, researchers would often move on to a new topic, only to repeat the cycle of conflicting findings for effect sizes and statistical significance. As described by Schmidt (1992), many in the social sciences came to believe that cumulative knowledge was hopeless. Cronbach (1975), a famous research methodologist, explicitly stated that he believed that cumulative knowledge was impossible in psychology and the social sciences. The variation in findings also led to a search for moderator effects both in validity studies and across a variety of topics in applied psychology (Berdie, 1961; Frederiksen and Gilbert, 1960; Frederiksen and Melville, 1954; Ghiselli, 1956; Grooms and Endler, 1960; Saunders, 1956). The failure of such moderator research to clarify matters added to the pessimism in the field, creating a sense that social science may lack the capacity to carve nature at its joints.

A major contributor to the apparent variability in studies was the widespread use of statistical significance tests. Starting with Fisher's (1923) development of the technique, significance testing had been used as the dominant data analysis procedure in the field. As an example of one major problem created by such usage, Oakes (1986) showed that researchers endorsed the false decision rule that if the statistic, such as the *t*-statistic or *F*-statistic, is not significant, there is no relationship between the variables being studied. The predominant mode of significance testing ignored Type II or beta errors (the failure to find a true effect when one exists). Studies based on small samples are prone to low statistical power (Cohen, 1962, 1988, 1990), making Type II or beta errors common. In many research literatures, statistical power averaged about 50%, resulting in maximally conflicting findings: half the studies reported finding a relationship and half reported finding no relationship (Schmidt, 1996; Schmidt and Hunter, 1997).

Meta-analysis in VG research

The adoption of research synthesis in the form of psychometric meta-analysis through the channel of VG research had major effects on the field of I/O psychology from the late 1970s onward and produced important ramifications for how future research was to be conducted and how individual studies were viewed. Meta-analysis has allowed researchers to demonstrate generalizable results across situations for relationships between variables and to identify replicable moderators, and has revealed other information that was obscured, distorted or unclear in the previous primary studies.

Schmidt and Hunter (2003) described the development of the methods that demonstrated that differences in the validities of occupational aptitude tests across occupations and jobs were largely due to sampling error and other research artifacts. Studies such as Schmidt *et al.* (1979), Hunter and Hunter (1984), Pearlman *et al.* (1980), and Pearlman (1982) applied these procedures to tests used in hiring for a wide variety of occupations. Hunter and Hunter's (1984) research program on validity coefficients of the U.S. Department of Labor's General Aptitude Test Battery (GATB) provides an example. The 425 validity studies used in this analysis showed substantial variation in validity coefficients, but meta-analysis of these studies demonstrated that most of this variation was due to sampling error. In addition, corrections for measurement error in the job performance measure and for range restriction on test scores showed that mean levels of validity were much higher than previously realized. Schmidt (1992) provides a list of other meta-analyses where large amounts of observed variability were mostly or completely attributable to sampling error. This research provided powerful disconfirming evidence for the theory of situational specificity. The impact of psychometric meta-analysis continued into the 1980s, the 1990s and later, creating important restructurings of basic theories and beliefs in I/O psychology and related disciplines in a variety of topic areas, as discussed later.

The illustration presented in Schmidt (2010) is a simple concrete example of psychometric meta-analysis. This data set consisted of 16 validity studies of a clerical test conducted in different organizations. Sample sizes varied from 30 to 323. Observed validities varied from 0.02 to 0.39; the mean was 0.20 and the SD of validities was 0.08. Only half were statistically significant ($p < 0.05$). Taken at face value, these data were highly variable and defied any simple interpretation. Application of psychometric meta-analysis showed that 98% of this variability was due to sampling error. After corrections for sampling error and measurement error, the final meta-analytic results showed a distribution with a mean of 0.32 and a SD of only 0.01. These results showed that the apparent complexity of results was artifactual and that the reality underlying the apparent complexity was quite simple and parsimonious.

VG research was soon extended to other procedures for predicting job performance such as employment interviews and assessment centers (a standardized evaluation of employees based on multiple raters or assessments such as role playing, case studies, and psychological test batteries). The predictive validity of these procedures was also shown to be consistent across times and situations rather than situation specific. Also,

other researchers contributed to the credibility of VG measures when they developed alternate statistical and mathematical models to estimate meta-analytic correlation coefficients and produced the same results and conclusions (Callender and Osburn, 1980; Raju and Burke, 1983). Outside of the I/O psychology field, Hedges and Olkin (1985) developed a model for meta-analytic analysis of correlations, although these methods typically did not include corrections for measurement error and other artifacts that distort research findings.

Hunter and Hirsh (1987) reviewed meta-analyses completed from the time of the development of the original VG procedure by Schmidt and Hunter (1977) in 1977 up to 1987, and noted that over 200 meta-analyses examining the relationship between selection procedures and job performance had been completed by that time. These meta-analyses examined a wide variety of job performance predictors, including biographical data, personality, reference checks, education, interviews, traditional training and experience evaluations, college GPA, interest, and age. In many of these meta-analyses, there was variability that sampling error, measurement error, and range restriction could not account for, suggesting the possibility of moderators of effect (variables that affect the size of the relationship between job performance and the independent variables studied). In the case of employment interviews, this finding led to some of the first studies of interview structure (degree of standardization of questions asked and scoring of answers) as a moderator for the effectiveness of interviews (McDaniel *et al.*, 1984; Wiesner and Cronshaw, 1988). This is an example of what has become a common theme: meta-analytic results opened doors for new directions in research.

The first examples of VG also demonstrated the power of general cognitive ability (GCA; intelligence in lay terms) as a predictor of job performance across situations and job types (Schmidt *et al.*, 1979, 1986), a finding that would have a major impact on the field and establish GCA as perhaps the central construct in I/O psychology. GCA, a construct originally defined by Spearman (1904), has been shown to correlate with a broad variety of behaviors and performances, from matters such as the likelihood of committing a crime, to maintenance of one's physical health to more mundane tasks such as the ability to read a bus schedule (Gottfredson, 1997; Lubinski and Humphreys, 1997). Hunter and Hunter's (1984) work with the GATB test provided the first meta-analytic evidence that GCA was a highly valid predictor of job performance for all jobs. The latest and most technically accurate meta-analysis indicates that GCA validity varies from 0.74 for the most mentally demanding jobs to 0.39 for unskilled jobs (Hunter *et al.*, 2006). A number of later primary and meta-analytic studies demonstrated that GCA was also a predictor of occupational status, income, job performance, and rate of career advancement (cf. Hunter and Schmidt, 2004). GCA continues to be studied as a predictor of educational and occupational attainment (Judge *et al.*, 2010a), health outcomes (Judge *et al.*, 2010b), and turnover (Maltarich *et al.*, 2010),[‡] among other life outcomes.

Expansions in VG research

Work on VG research continued to expand in the 1990s. The Hunter and Schmidt (1990) meta-analysis book followed their earlier book (Hunter *et al.*, 1982) and explicated the principles and methods of psychometric meta-analysis in more detail for researchers. Two major meta-analytic findings from this decade were Barrick and Mount (1991), which demonstrated the generalized predictive validity of personality assessments, and Ones, Viswesvaran, & Schmidt (Ones *et al.*, 1993), which demonstrated the generalizable predictive validity of integrity tests, a type of psychological assessment designed to measure a potential employee's honesty, reliability, trustworthiness, and dependability. Barrick and Mount's (1991) study was important in a number of ways. Rothstein (2003) cited the Barrick and Mount (1991) study as reinvigorating personality research in I/O psychology and creating a 'systematic program of research designed to further examine the role of personality in job performance' (p.129). As of this writing, the Barrick and Mount (1991) study has been cited 2823 times. Both meta-analyses overturned a fundamental and widely accepted conclusion based on traditional narrative reviews—namely, that personality measures did not predict job performance (Ghiselli, 1973; Guion and Gottier, 1965). The development of the Five-factor model (FFM) of personality (Goldberg, 1993) provided a framework that allowed Barrick and Mount (1991) to use meta-analysis to demonstrate a generalizable relationship between the trait of conscientiousness and job performance across all types of jobs and to demonstrate the utility of the traits of agreeableness and extraversion in work with crucial interpersonal components such as sales and management.

Importantly, Barrick and Mount (2003) noted that their 1991 meta-analysis stimulated later research on the proximal motivational mechanisms that connected personality to distal performance (Barrick *et al.*, 1993). For example, they showed that the impact of conscientiousness on job performance was mediated by work-related goal setting. This typifies a fundamental impact of meta-analytic research: the creation of new, theoretically driven research advances based on the big-picture conclusions that can be drawn from meta-analysis. Because of these developments, Guion (1998) observed that 'Meta-analyses have provided

[‡]Note: This is a negative relationship such that individuals with higher GCA are less likely to engage in voluntary turnover.

grounds for optimism' (p. 145). Combining the advances of psychometric meta-analysis with the development of the FFM for personality variables allowed for important theoretical gains to be made and created new avenues of research for I/O psychology.

The Ones *et al.* (1993) meta-analysis of integrity tests not only provided evidence that personality and character matter in personnel selection but also demonstrated how psychometric meta-analysis can identify unexpected utility for a selection method. While integrity tests were initially designed to predict counterproductive work behaviors (such as violence on the job and employee theft), this meta-analysis revealed that integrity tests had the ability to predict overall job performance as well as counterproductive behaviors. This led to a later conclusion in Schmidt and Hunter (1998) that a test battery consisting of an integrity test and an intelligence (GCA) test given at the time of hire provided the highest validity based on any combination of two tests for predicting later job performance. Ones *et al.* (2003) later also used meta-analysis to demonstrate the utility of integrity tests for predicting employee absenteeism.

VG research was conducted on other important personnel selection methods. These methods included: situational judgment tests (McDaniel *et al.*, 2001), a type of employment test designed to predict how individuals will react to specific work situations or problems; work sample tests (Roth *et al.*, 2008); and assessment centers (Woehr and Arthur, 2003). There has also been meta-analytic research on organizational citizenship behaviors (such as helping others at work; Organ and Ryan (1995)), counter-productive work behaviors (such as theft or violence on the job; Salgado (2002)), the relationship between organizational citizenship behaviors and counterproductive behaviors (Dalal, 2005), and absenteeism (Scott and Taylor, 1985).

Overall, a pattern developed in this time period. As new techniques, constructs, predictors, or criteria were developed or gained popularity, or as new measurement issues arose in the pre-existing selection methods, or as new theories of job performance changed the conclusions drawn about the predictive validity of selection instruments, new primary studies were conducted and meta-analysis was then used to assess the overall meaning and importance of the resulting literatures.

The practical applications of VG

The largest single practical application of VG was built around the U.S. Department of Labor's GATB. This research was sponsored by the U.S. Department of Labor in the 1980s. The GATB program was designed as a large-scale VG project meant to demonstrate the generalized validity of the GATB in predicting job performance across the approximately 12 000 jobs in the Department of Labor's *Dictionary of Occupational Titles* (DOT; 1977). John Hunter of Michigan State began work on this project in 1980 and produced four technical reports for the Department of Labor (Hunter, 1983a,b,c,d), as well as a journal publication of the basic GATB VG results (Hunter and Hunter, 1984). The GATB applicant referral program, which allowed job seekers to take the GATB and send their results to potential employers, was initially introduced in North Carolina but was quickly offered nationwide and ultimately adopted in 42 states with hundreds of thousands of applicants tested and referred to employers every year (Schmidt and Hunter, 2003). Both small firms and Fortune 500 companies took advantage of the program (Madigan *et al.*, 1986). Some companies conducted their own studies and found that employees selected through the VG program demonstrated better job performance, higher training performance, less absenteeism, and fewer accidents compared with employees selected through other means (McKinney, 1984). Because the GATB employment referral program was only a referral service, organizations could use additional selection procedures beyond those in the GATB program in making final hiring decisions. The program was free to employers and applicants alike and proved that research results from VG could have wide applicability and practical, useful results for employers and workers.

After several years, the GATB program's use of separate test norms for whites, blacks, and Hispanics became controversial. This technique had been introduced by the Department of Labor to prevent adverse impact on minorities, and this use of 'race norming' was openly announced in materials distributed to employers and states using the program. During the Reagan Administration, the U.S. Department of Justice challenged this score adjustment component of the GATB program, claiming that it was a form of reverse racial discrimination (Schmidt and Hunter, 2003). However, a May 1989 report (Hartigan and Wigdor, 1989) from the National Research Council (an arm of the National Science Foundation) supported the entire program, both the VG aspects and the score adjustments. In 1990, Secretary of Labor Elizabeth Dole pushed for a suspension of the GATB program but was forced to withdraw her proposal due to overwhelmingly negative feedback from employers and other users of the program (Schmidt and Hunter, 2003).

The Civil Rights Act of 1991 became law in early 1992 and banned any score adjustments based on race, sex, or ethnicity. In February 1992, the Department of Labor announced that it would continue to offer the program but without score adjustments. However, Labor Secretary Elizabeth Dole unilaterally suspended the GATB VG program in the summer of 1992 because of her concern about test score differences between minorities and the majority. Nonetheless, the program had demonstrated on a national scale the utility and efficacy that

VG research, and consequently meta-analysis, could have. Employing organizations around the country found VG research to be useful in their hiring practices and the challenge from the Department of Justice resulted in powerful scientific support from the National Research Council. Ultimately, the GATB referral program was discontinued, not due to its foundation in meta-analysis, but due to the difficult and complex social problem of racial and ethnic group test score differences (cf. Sackett *et al.*, 2001).

The changing professional standards that have evolved for VG in personnel selection are indicative of the impact that meta-analysis has had on the field of I/O psychology and its associated disciplines. As Sackett (2003) observed, the 1985 *Standards for Educational and Psychological Testing* (American Educational Research Association, 1985), a document produced jointly by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education, acknowledges the value of meta-analysis. The *Standards* identifies meta-analysis both as a way to draw quantitative conclusions about a body of research as well as a tool for transporting validity to a new test situation. The 1985 standards (Standard 1.16, p. 16) gave priority to local validity studies, with meta-analytic evidence receiving a conditional status as a strategy to be used when a local validity study is not available or cannot be conducted. However, in the 1999 revision of the Standards, this position changed and meta-analytical validity evidence was given full credibility.

The 1987 *Principles for the Validation and Use of Personnel Selection Tests* (Society for Industrial and Organizational Psychology, Inc., 1987), published by the Society for Industrial and Organizational Psychology, focused more attention on VG than did the 1985 *Standards*. The *Principles* do not frame VG as a conditional strategy; rather, they state that researchers can rely on VG tools to the extent that adequate VG data are available. The 1987 *Principles* presaged the developments in the 1999 *APA Standards* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), which endorsed the use of multiple lines of support, both local validity studies and meta-analytic validity research, as a strategy to support test use. By 1999, the *Standards* gave primacy neither to local nor to meta-analytic validity studies. They simply indicated that the use of a local or meta-analytic validity study was acceptable, provided that the data used are appropriate to the job in question and there are no known moderators of the relationship between the selection test and job performance that could threaten conclusions about the general validity of the test.

Sackett (2003) made the current state of meta-analysis and the impact it had on standards in personnel research quite clear. The treatment of VG in the *Standards* and the *Principles*, the endorsement of the fundamental concept of VG by the National Academy of Sciences Committee on the GATB, and agreement on the fundamental underpinnings of VG in the 'Forty Questions'⁵ Document all attest to the high degree of acceptance of VG. (p.109). Sackett (2003) also concluded that the psychometric meta-analysis model used in the VG research and later, general research synthesis, revolutionized thinking in I/O psychology and integrated the concepts of sampling error, measurement error, and range restriction into a compact framework that could be used to develop cumulative knowledge in the field.

Applications of meta-analysis beyond VG

In addition to the influence and success of VG, nearly every major research area in I/O psychology and related fields has undergone dramatic changes due to the application of meta-analysis. Education practices in the field have also changed. Textbooks in I/O psychology and OB now routinely cite meta-analyses rather than primary studies to summarize research results. Graduate students in these areas now typically take a course on meta-analysis or have a section of their research methods' coursework devoted to meta-analytic procedures. Meta-analyses routinely appear in top-tier applied psychology journals such as the *Journal of Applied Psychology*, *Personnel Psychology*, and the *Academy of Management Journal*.

Recent meta-analyses in organizational psychology have addressed a broad range of topics across different levels of analysis. At the levels of the organization or business unit, Harter *et al.* (2002; 2010) and Whitman *et al.* (2010) demonstrated that unit-level job satisfaction and employee engagement has positive, generalized effects on business unit financial performance and customer satisfaction. Another meta-analysis showed, across 83 different organizations, the generalized positive effects on job performance of the Productivity Measurement and Enhancement System (ProMES; Pritchard *et al.*, 2008), a performance management system designed by organizational psychologists to provide workers and employees with quick and effective feedback on their performance. Meta-analyses of team research continue to be popular, with one meta-analysis summarizing how different team work processes impact team effectiveness (LePine *et al.*, 2008).

⁵The *Forty Questions Document* noted above is a reference to Schmidt *et al.* (1985), an paper in a major I/O psychology journal that describes the uses of psychometric meta-analysis and VG. The title of the work is 'Forty questions about VG and meta-analysis'.

Other meta-analyses focus on individuals as the unit of analysis. One such study examined the relationship between job turnover and the FFM model of personality and found that Emotional Stability, one of the FFM personality traits that showed generalized predictive validity for job performance (Hurtz and Donovan, 2000), is also a generalized negative predictor of turnover (Zimmerman, 2008). Another study attempted to untangle the ambiguous causal relationship between job attitudes and job performance (Ricketta, 2008). Other studies focus on multicultural and international issues. Dean *et al.* (2008) meta-analytically examined ethnic and gender subgroup differences in assessment center ratings to show that gender differences in ratings from these evaluations are smaller than previously thought, but that some ethnic subgroup differences are larger than previously believed. Taras *et al.* (2010) showed that Hofstede's (1980) cultural value dimensions had validity in predicting (in decreasing order) individual emotions, attitudes, behaviors, and job performance.

Older examples also span a variety of topics and units of analysis. Meta-analysis of team research was popular: an example of one such meta-analysis examined the negative and positive correlates of conflict and ambiguity for members' roles in teams (Fisher and Gitelson, 1983). Meta-analyses of leadership were also popular. An example is a meta-analytic test of Fiedler's contingency theory of leadership, a dominant theory of leadership at the time (Peters *et al.*, 1985). Other meta-analyses studied questions about attitudes and beliefs, such as the relatively low accuracy of self-ratings of ability and skill (Mabe and West, 1982) and the negative relationship between job satisfaction and absenteeism (Terborg *et al.*, 1982). Other studies focused on more specific interventions and assessments, such as the small but positive effect of realistic job previews in reducing subsequent employee turnover (Premack and Wanous, 1985), the positive, generalizable validity of LSAT for predicting performance in law school (Linn and Hastings, 1983), and the limited abilities of financial analysts to predict stock growth (Coggin and Hunter, 1983). In short, researchers in organizational psychology have pursued and continue to pursue meta-analytic studies across a wide variety of subjects and continue to recognize psychometric meta-analysis as an important research tool.

Additional examples of influential meta-analyses are found in the literature on managerial training. Burke and Day's (1986) meta-analysis on the effectiveness of managerial training prompted a subsequent stream of meta-analytic research on management training including Collins and Holton (2004), Taylor *et al.* (2009), and Powell and Yalcin (2010). The conclusions from these studies have repeatedly been that management training programs can be effective in nearly all situations at changing particular behaviors and for the acquisition of knowledge, particularly in the areas of time management and human relations skills. Other meta-analyses assess the results of training across a variety of organizational contexts. For example, a recent meta-analysis examined how combinations of training content, trainee attributes, and trainees' affective reactions to a training influence the outcomes of a training program (Sitzmann *et al.*, 2008). Other meta-analytic studies investigated the relationships among different training criteria such as behavior, learning, and performance (Alliger *et al.*, 1997) or how trainees use their training in applied settings and share knowledge from training with others (Arthur *et al.*, 2003). All three of these meta-analyses had powerful impacts on traditional models of learning and training in I/O psychology, resulting in updates to Kirkpatrick's (2000) widely used model of learning, a demonstration on the effectiveness of lectures as a form of training, and reconsideration of the value of affective reactions to training.

Colquitt *et al.* (2000) used meta-analysis to further expand on Alliger's (1997) work and examine how individual differences, situational factors, and job career factors influence an individual's motivation during training and the subsequent outcomes of this motivation. Colquitt, LePine, and Noe (2000) demonstrated the importance of motivation as a critical factor in determining the efficacy of training, how individuals transfer learning from training to their work performance, and how effective they are in sharing training knowledge with others. In a more recent study on the impact of motivation, Payne *et al.* (2007) used meta-analytic methods to calibrate the impact of goal orientation (a psychological variable reflecting the motivation of individuals to learn versus the motivation to perform well in front of others) on training outcomes. These two meta-analyses have influenced how researchers conduct subsequent primary studies and created new opportunities for researchers to examine the role of motivation in training. In short, meta-analysis has impacted research on management training, motivation in training, and training in general by demonstrating the validity and value of training to organizations across the board.

Meta-analysis has also been used extensively in the leadership literature, a popular topic of study in I/O psychology and OB. Meta-analysis has provided some clarity on a difficult subject, as well as changing how studies were conducted in light of the findings of the meta-analyses. Judge *et al.* (2004) reported a relatively low relationship between GCA and leadership, noting that this relationship is weaker than expected based on earlier qualitative reviews (Bass, 1990; Kirkpatrick and Locke, 1991; Mann, 1959; Stogdill, 1948). Similar work was done demonstrating that leadership cannot be explained solely as a result of personality traits in the popular FFM framework (Judge *et al.*, 2002). Several years later, this line of research inspired a group of researchers to consider conceptual models of positive, 'bright-side' and negative, 'dark-side' personality traits in leaders linked to leadership emergence and efficacy (Judge *et al.*, 2009). Finally, multiple meta-analyses have been used to show the validity, uniqueness, and importance of transformational (charismatic) leadership, a specific set of leadership behaviors highly motivating to employees and a central topic of research in the leadership literature (Bono and Judge, 2004; Eagly *et al.*, 1995, 2003; Judge *et al.*, 2004b). The impact of meta-analysis on

the research literature is the same in I/O psychology and OB: important changes in research conclusions based on the quantitative analysis of previous studies, fundamental changes to the existing theoretical paradigms, and the development of new lines of research.

Meta-analysis, construct proliferation, and cumulative knowledge

Meta-analysis has also allowed researchers to address issues related to construct proliferation and construct validity, issues that have confounded the development of I/O psychology and its related fields, and perhaps all of psychology and the social sciences (Becker, 2009; Le *et al.*, 2010). Construct validity refers to the meaning and theoretical and/or practical usefulness of a construct or variable. Construct proliferation exists when measures of supposedly distinct constructs correlate perfectly or nearly perfectly after the appropriate corrections are made for the downward biases created by measurement error. For example, Le *et al.* (2010) found that the appropriately corrected correlation between measures of organizational commitment and job satisfaction was 0.91, demonstrating that these two constructs are probably not distinguishable in the minds of respondents.

Two examples of construct validity issues come from the leadership literature. Judge *et al.* (2004b) used meta-analysis to overturn the previous dismissal of the Ohio State leadership behavior constructs known as Consideration and Initiating Structure, demonstrating that the measures were in fact valid and useful in leadership research. Their paper notes that the dismissal of these constructs was premature: 'Perhaps the conclusions in past qualitative reviews were not tested meta-analytically because the literature had already been pronounced as dead before the advent of meta-analysis in I/O psychology. Why meta-analyze a literature, researchers may have asked, that had long been obsolete?' (p. 44). More recently, DeRue *et al.* (in press) used meta-analytic techniques to address issues of construct proliferation in the leadership literature: 'Our hope is that this study begins to reverse the trend of construct proliferation in the leadership literature, and thus provides some clarity to leadership studies... with the goal of developing an integrative understanding of leadership in organizations' (p. 38). In short, meta-analysis can help to clarify past research on a particular psychological variable and allows researchers to avoid both the problem of creating too many theoretical variables in explanations of phenomena and the problem of prematurely dismissing valid and useful variables.

Schmidt (1984, 1992, 1996, 2010) described the production of cumulative knowledge as the most important problem in psychology and the social sciences, and stated that if cumulative knowledge and general principles were impossible to establish, then 'psychology can never be a science—only *at best* a technology producing answers limited in their applicability to specific situations'. This is why Guion (1998) expressed optimism regarding the development and adoption of the meta-analytic tools presented in Schmidt and Hunter (1977), Callender and Osburn (1980), Raju and Burke (1983), and Hedges and Olkin (1985). These tools represented the opportunity to develop cumulative knowledge and allow the social sciences the ability to be scientific in a new and fundamentally different way. This position is generally consistent with other opinions and reviews of the value of meta-analysis in behavioral research and developing cumulative knowledge (cf. Green and Hall, 1984; Lipsey and Wilson, 1993), although debate continues (cf. Curran, 2009).

Finally, it is important to point out that no one meta-analysis is necessarily the final word. Any given meta-analysis may have methodological or other deficiencies and so may be superseded by later meta-analyses. One example of this is the series of psychometric meta-analyses examining the relationship between job satisfaction and job performance. The first meta-analysis on this relationship was that of Petty *et al.* (1984), based on a limited number of studies. This was followed by a more comprehensive meta-analysis by Iaffaldano and Mchinsky (1985). However, primary studies continued to appear on this relationship and Judge *et al.* (2001) produced a meta-analysis based on both a more thorough search for studies and more complete corrections for measurement error biases. Their meta-analysis has also encouraged further meta-analytic research on the direction of the causal relationship between job attitudes and job performance (Harrison *et al.*, 2006).

Conclusion

In conclusion, research synthesis in the form of psychometric meta-analysis had a dramatic impact on developments in I/O psychology and related disciplines. The central purpose of social sciences, the development of cumulative knowledge, was made possible through the advancement of meta-analytic techniques. The ability to develop cumulative knowledge in I/O psychology lifted the state of the field from one of epistemological pessimism to general optimism that cumulative knowledge is possible. The spread of meta-analytic methods from personnel selection into other areas of the field of I/O psychology created a number of powerful results that changed researchers' conceptions of how to conduct research, led to the revision of basic theories founded on erroneous interpretations of data, and created an opportunity for I/O psychology to create cumulative knowledge. Future use of psychometric meta-analysis will no doubt lead to further discoveries and increases

in the development of knowledge as it relates to behavior and performance in workplace and organizational settings.

References

- Alliger GM, Tannenbaum SI, Bennett W, Traver H, Shotland A. 1997. A meta-analysis of the relations among training criteria. *Personnel Psychol* **50**(2): 341–358.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 1985. *Standards for Educational and Psychological Testing*. American Psychological Association, Washington, DC.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 1999. *Standards for Educational and Psychological Testing*. American Educational Research Association, Washington, DC.
- Arthur W, Bennett W, Edens PS, Bell ST. 2003. Effectiveness of training in organizations: a meta-analysis of design and evaluation features. *J Appl Psychol* **88**(2): 234–243.
- Barrick MR, Mount MK. 1991. The Big Five dimensions and job performance: a meta-analysis. *Personnel Psychol* **44**: 1–26.
- Barrick MR, Mount MK. 2003. Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychol* **56**: 45–74.
- Barrick MR, Mount MK, Strauss JP. 1993. Conscientiousness and performance of sales representatives: test of the mediating effects of goal setting. *J Appl Psychol* **78**(5): 715–722.
- Bass BM. 1990. *Bass and Stogdill's Handbook of Leadership*. Free Press, New York.
- Becker BJ. 2009. Model-based meta-analysis, in *The Handbook of Research Synthesis and Meta-analysis* (Cooper H, Hedges LV, Valentine JC eds), pp 377–398. 2nd ed. Russell Sage Foundation, New York.
- Berdie RF. 1961. Intra-individual variability and predictability. *Educ Psychol Meas* **21**: 663–676.
- Bono JE, Judge TA. 2004. Personality and transformational and transactional leadership: a meta-analysis. *J Appl Psychol* **89**(5): 901–910.
- Burke MJ. 1984. Validity generalization: a review and critique of the correlation model. *Personnel Psychol* **37**: 93–113.
- Burke R, Day R. 1986. A cumulative study of the effectiveness of management training. *J Appl Psychol* **71**(2): 232–245.
- Callender JC, Osburn Hg. 1980. Development and test of a new model for validity generalization. *J Appl Psychol* **65**: 543–558.
- Coggin TD, Hunter JE. 1983. Problems in measuring the quality of investment information: the perils of the information coefficient. *Financ Analysts J* **39**(3): 25–33.
- Cohen J. 1962. The statistical power of abnormal–social psychological research: a review. *J Abnormal Soc Psychol* **65**: 145–153.
- Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Erlbaum, Hillsdale, NJ.
- Cohen J. 1990. Things I have learned (so far). *Amer Psychol* **45**(12): 1304–1312.
- Collins DB, Holton EF. 2004. The effectiveness of managerial leadership development programs: a meta-analysis of studies from 1982 to 2001. *Hum Resour Dev Q* **15**(2): 217–248.
- Colquitt JA, LePine JA, Noe RA. 2000. Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *J Appl Psychol* **85**(5): 678–707.
- Cronbach LJ. 1975. Beyond the two disciplines of scientific psychology revisited. *Amer Psychol* **30**: 116–127.
- Curran PJ. 2009. The seemingly quixotic pursuit of a cumulative psychological science: introduction to the special issue. *Psychol Meth* **14**: 77–80.
- Dalal RS. 2005. A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *J Appl Psychol* **90**: 1241–1255.
- Dean MA, Roth PL, Bobko P. 2008. Ethnic and gender subgroup differences in assessment center ratings: a meta-analysis. *J Appl Psychol* **93**(3): 685–691.
- DeRue S, Nahrgang J, Wellman N, Humphrey SE. 2011. Trait and behavioral theories of leadership: an integration and meta-analytic test of their validity. *Personnel Psychol* **64**: 7–52.
- U.S. Department of Labor, Employment Division. 1977. *Dictionary of Occupational Titles*. 4th ed. U.S. Government Printing Office, Washington, DC.
- Eagly AH, Johnson BT. 1990. Gender and leadership style: a meta-analysis. *Psychol Bull* **108**(2): 233–256.
- Eagly AH, Johannsen-Schmidt MC, van Engen ML. 2003. Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychol Bull* **129**: 569–591.
- Eagly AH, Karau SJ, Makhijani MG. 1995. Gender and the effectiveness of leaders: a meta-analysis. *Psychol Bull* **117**(1): 125–145.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. 1991. Publication bias in clinical research. *Lancet* **337**: 867–872.

- Fisher RA. 1923. Statistical test of agreement between observation and hypothesis. *Economica* **8**: 39–147.
- Fisher CD, Gitelson R. 1983. A meta-analysis of the correlates of role conflict and role ambiguity. *J Appl Psychol* **68**(2): 320–333.
- Frederiksen N, Gilbert AC. 1960. Replication of a study of differential predictability. *Educ Psychol Meas* **20**: 759–767.
- Frederiksen N, Melville SD. 1954. Differential predictability in the use of test scores. *Educ Psychol Meas* **14**: 647–656.
- Ghiselli EE. 1956. Differentiation of individuals in terms of their predictability. *J Appl Psychol* **40**: 374–377.
- Ghiselli EE. 1966. *The Validity of Occupational Aptitude Tests*. Wiley, New York.
- Ghiselli EE. 1969. *The Validity of Occupational Aptitude Tests*. Wiley, New York.
- Ghiselli EE. 1973. The validity of aptitude tests in personnel selection. *Personnel Psychol* **26**(4): 461–477.
- Goldberg LR. 1993. The structure of phenotypic personality traits. *Amer Psychol* **48**(1): 26–34.
- Gottfredson LS. 1997. Why g matters: the complexity of everyday life. *Intell* **24**: 79–132.
- Green BF, Hall JA. 1984. Quantitative methods for literature reviews. *Annu Rev Psychol* **35**: 37–53.
- Grooms and Endler. 1960.
- Guion RM. 1998. *Assessment, Measurement, and Prediction for Personnel Decisions*. Erlbaum, New York.
- Guion RM, Gottier RF. 1965. Validity of personality measures in personnel selection. *Personnel Psychol* **18**(2): 135–164.
- Harrison DA, Newman DA, Roth PL. 2006. How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Acad Manage J* **49**: 305–325.
- Harter JK, Schmidt FL, Asplund J, Killham EA. 2010. Causal impact of employee work perceptions on the bottom line of organizations. *Perspect Psychol Sci* **5**: 378–389.
- Harter JK, Schmidt FL, Hayes TL. 2002. Business unit level relationships between employee satisfaction/engagement and business outcomes: a meta-analysis. *J Appl Psychol* **87**: 268–279.
- Hartigan JA, Wigdor AK eds. 1989. *Fairness in Employment Testing: Validity Generalization, Minority Issues and the General Aptitude Test Battery*. National Academy of Sciences Press, Washington, DC.
- Hedges LV, Olkin I. 1985. *Statistical Methods for Meta-analysis*. Academic Press, New York.
- Hofstede G. 1980. *Culture's Consequences: International Differences in Work-related Values*. Sage, Beverly Hills, CA.
- Hunter JE. 1983a. *The Dimensionality of the General Aptitude Test Battery (GATB) and the Dominance of the General Factors over Specific Factors in the Prediction of Job Performance*. U.S. Employment Service, USDOL, Washington, DC, USES Test Research Report No. 44.
- Hunter JE. 1983b. *Test Validation for 12,000 Jobs: An Application of Job Classification and Validity Generalization Analysis to the General Aptitude Test Battery*. U.S. Employment Service, USDOL, Washington, DC, USES Test Research Report No. 45.
- Hunter JE. 1983c. *Fairness of the General Aptitude Test Battery: Ability Differences and Their Impact on Minority Hiring Rates*. U.S. Employment Service, USDOL, Washington, DC, USES Test Research Report No. 46.
- Hunter JE. 1983d. *The Economic Benefits of Personnel Selection Using Ability Tests: A State of the Art Review Including a Detailed Analysis of the Dollar Benefit of U.S. Employment Service Placements and a Critique of the Low-cutoff Method of Test Use*. U.S. Employment Service, USDOL, Washington, DC, USES Test Research Report No. 47.
- Hunter JE, Hirsh RH. 1987. Application of meta-analysis, in *International Review of Industrial and Organizational Psychology* (Cooper CL, Robertson IT eds), pp 321–357. Wiley, New York.
- Hunter JE, Hunter RF. 1984. Validity and utility of alternative predictors of job performance. *Psychol Bull* **96**(1): 72–98.
- Hunter JE, Schmidt FL. 1990. *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*. 1st ed. Sage, Beverly Hills, CA.
- Hunter JE, Schmidt FL. 2004. *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*. 2nd ed. Sage, Thousand Oaks, CA.
- Hunter JE, Schmidt FL, Jackson GB. 1982. *Meta-analysis: Cumulating Research Findings Across Studies*. Sage, Beverly Hills, CA.
- Hunter JE, Schmidt FL, Le H. 2006. Implications of direct and indirect range restriction for meta-analysis. *J Appl Psychol* **91**(3): 594–612.
- Hurtz GM, Donovan JJ. 2000. Personality and job performance: the big five revisited. *J Appl Psychol* **85**: 869–879.
- Iaffaldano MT, Mchinsky PM. 1985. Job satisfaction and job performance: a meta-analysis. *Psychol Bull* **97**: 251–273.
- Judge TA, Bono JE, Ilies R, Gerhardt MW. 2002. Personality and leadership: a qualitative and quantitative review. *J Appl Psychol* **87**(4): 765–780.
- Judge TA, Colbert AE, Ilies R. 2004a. Intelligence and leadership: a quantitative review and test of theoretical propositions. *J Appl Psychol* **89**(3): 542–552.
- Judge TA, Ilies R, Dimotakis N. 2010a. Are health and happiness the product of wisdom? The relationship of general mental ability to educational and occupational attainment, health, and well-being. *J Appl Psychol* **95**: 454–468.

- Judge TA, Klinger RL, Simon LS. 2010b. Time is on my side: Time, general mental ability, human capital, and extrinsic career success. *J Appl Psychol* **95**: 92–107.
- Judge TA, Piccolo RF, Ilies R. 2004b. The forgotten ones? The validity of consideration and initiating structure in leadership research. *J Appl Psychol* **89**(1): 36–51.
- Judge TA, Piccolo RF, Kosalka T. 2009. The bright and dark sides of leader traits: a review theoretical extension to the leader trait paradigm. *Leadersh Q* **20**(6): 855–875.
- Judge TA, Thoresen CJ, Bono JE, Patton GK. 2001. The job satisfaction-job performance relationship: a qualitative and quantitative review. *Psychol Bull* **127**(3): 376–407.
- Kirkpatrick DL. 2000. Evaluating training programs: the four levels, in *The ASTD Handbook of Training Design and Delivery* (Piskurich GM, Beckschi P, Hall B eds), pp 133–146. McGraw Hill, New York.
- Kirkpatrick SA, Locke EA. 1991. Leadership: do traits matter? *Acad Manage Exec* **5**: 48–59.
- Le H, Schmidt FL, Harter JK, Lauver KJ. 2010. The problem of empirical redundancy of constructs in organizational research: an empirical investigation. *Organ Behav Hum Decis Making* **112**: 112–125.
- LePine JA, Piccolo RF, Jackson CL, Mathieu JE, Saul JR. 2008. A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychol* **61**(2): 273–307.
- Linn RL, Hastings CN. 1983. A meta analysis of the validity of predictors of performance in law school (Report Number LSAC-83-1), *Law School Admission Council, Reports of LSAC sponsored research: Volume IV 1978–1983*, pp 507–544. Law School Admission Council, Newtown, PA.
- Lipsey MW, Wilson DB. 1993. The efficacy of psychological, educational, and behavioral treatment. *Amer Psychol* **48**: 1181–1209.
- Lubinski D, Humphreys LG. 1997. Incorporating general intelligence into epidemiology and the social sciences. *Intell* **24**: 159–201.
- Mabe PA, West SG. 1982. Validity of self-evaluation of ability: a review and meta-analysis. *J Appl Psychol* **67**(3): 280–296.
- Madigan RM, Scott KQ, Deadrick DL, Stoddard JA. 1986. Employment testing: the U.S. Job Service is spearheading a revolution. *Personnel Admin* **57**: 102–107.
- Maltarich et al. 2010.
- Mann RD. 1959. A review of the relationships between personality and performance in small groups. *Psychol Bull* **56**: 241–270.
- McDaniel MA, Morgeson FP, Finnegan EB, Campion MA. 2001. Use of situational judgment tests to predict job performance: a clarification of the literature. *J Appl Psychol* **86**(4): 730–740.
- McDaniel MA, Whetzel DL, Schmidt FL, Mauser SD. 1994. The validity of employment interviews: a comprehensive review and meta-analysis. *J Appl Psychol* **79**: 599–616.
- McKinney MW. 1984. *Final Report: Validity Generalization Pilot Study*. U.S. Employment Service, USDOL, Southern Test Development Field Center, Raleigh, NC.
- McNatt D. 2000. Ancient Pygmalion joins contemporary management: a meta-analysis of the result. *J Appl Psychol* **70**: 342–353.
- Morgeson FP, Campion MA, Dipboye RL, Hollenbeck JR, Murphy K, Schmitt N. 2007. Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychol* **60**: 1029–1049.
- Murphy KR, Newman DA. 2003. The past, present, and future of validity generalization, in *Validity Generalization: A Critical Review* (Murphy KR ed.). Lawrence Erlbaum Associates, Mahwah, pp 403–424.
- Oakes M. 1986. *Statistical Inference: A Commentary for the Social and Behavioral Sciences*. Wiley: New York.
- Oh IS. 2009. The Five-Factor Model of personality and job performance in East Asia: a cross-cultural validity generalization study. Unpublished doctoral dissertation, University of Iowa.
- Ones DS, Viswesvaran C, Schmidt FL. 1993. Comprehensive meta-analysis of integrity test validities: finding and implications for personnel selection and theories of job performance. *J Appl Psychol Monograph* **78**: 679–703.
- Ones DS, Viswesvaran C, Schmidt FL. 2003. Personality and absenteeism: a meta-analysis of integrity tests. *Eur J Personal* **17**: S19–S38.
- Organ DW, Ryan K. 1995. A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychol* **48**(4): 775–802.
- Overton RC. 1998. A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychol Meth* **3**(3): 354–379.
- Payne SC, Youngcourt SS, Beaubien JM. 2007. A meta-analytic examination of the goal orientation nomological net. *J Appl Psychol* **92**(1): 128–150.
- Pearlman K. 1982. The Bayesian approach to validity generalization: a systematic examination of the robustness of procedures and conclusions. Unpublished doctoral dissertation, Department of Psychology, George Washington University.
- Pearlman K, Schmidt FL, Hunter JE. 1980. Validity generalization results for tests used to predict job proficiency and training criteria in clerical occupations. *J Appl Psychol* **65**: 373–407.
- Peters LH, Hartke DD, Pohlmann JT. 1985. Fiedler's contingency theory of leadership: an application of the meta-analysis procedures of Schmidt and Hunter. *Psychol Bull* **97**(2): 274–285.

- Petty MM, McGee GW, Cavender JW. 1984. A meta-analysis of the relationships between individual job satisfaction and individual performance. *Acad Manage Rev* **9**(4): 712–721.
- Powell KS, Yalcin S. 2010. Managerial training effectiveness: a meta-analysis 1952–2002. *Personnel Rev* **39**(2): 227–241.
- Premack SL, Wanous JP. 1985. A meta-analysis of realistic job preview experiments. *J Appl Psychol* **70**(4): 706–719.
- Pritchard RD, Harrell MM, DiazGranadaos D, Guzman MJ. 2008. The productivity measurement and enhancement system: a meta-analysis. *J Appl Psychol* **93**(3): 540–567.
- Raju NS, Burke MJ. 1983. Two new procedures for studying validity generalization. *J Appl Psychol* **65**: 373–407.
- Riketta M. 2008. The causal relation between job attitudes and job performance: a meta-analysis of panel studies. *J Appl Psychol* **93**(2): 472–481.
- Roth P, Bobko P, McFarland L, Buster M. 2008. Work sample tests in personnel selection: a meta-analysis of black–white differences in overall and exercise scores. *Personnel Psychol* **61**(3): 637–661.
- Rothstein H. 2003. Progress is our most important product: contributions of validity generalization and meta-analysis to the development and communication of knowledge in I/O psychology, in *Validity Generalization: A Critical Review* (Murphy KR ed.). Lawrence Erlbaum Associates, Mahwah.
- Sackett PR. 2003. The status of validity generalization research: key issues in drawing inferences from cumulative research findings, in *Validity Generalization: A Critical Review* (Murphy KR ed.). Lawrence Erlbaum Associates, Mahwah, pp 114–912.
- Sackett PR, Schmitt N, Elingson JE, Kabin MB. 2001. High-stakes testing in employment, credentialing, and higher education: prospects in a post-affirmative action world. *Amer Psychol* **56**: 302–318.
- Salgado JF. 2002. The Big Five personality dimensions and counterproductive behaviors. *Int J Sel Assess* **10**: 117–125.
- Saunders DR. 1956. Moderator variables in prediction. *Educ Psychol Meas* **16**: 209–222.
- Schmidt FL. 1984. Meta-analysis: implications for cumulative knowledge in the behavioral and social sciences. Invited Division 5 Address, 92nd annual convention of the American Psychological Association, Toronto, August 25, 1984. (Available through ERIC/CAPS Clearinghouse; ED No. ED 251-722.)
- Schmidt FL. 1992. What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *Amer Psychol* **47**: 1173–1181.
- Schmidt FL. 1996. Statistical significance testing and cumulative knowledge in psychology: implications for the training of researchers. *Psychol Meth* **1**: 115–129.
- Schmidt FL. 2010. Detecting and correcting the lies that data tell. *Perspect Psychol Sci* **5**(3): 233–242.
- Schmidt FL, Hunter JE. 1977. Development of a general solution to the problem of validity generalization. *J Appl Psychol* **62**: 529–540.
- Schmidt FL, Hunter JE. 1984. A within-setting test of the situational specificity hypothesis in personnel selection. *Personnel Psychol* **37**: 317–326.
- Schmidt FL, Hunter JE. 1988. The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychol Bull* **124**(2): 262–274.
- Schmidt FL, Hunter JE. 1997. Eight common but false objections to the discontinuation of statistical significance tests, in *What If There Were No Significance Tests* (Harlow L, Mulaid S, Steiger J eds). Erlbaum, Hillsdale, NJ.
- Schmidt, Hunter. 1998.
- Schmidt FL, Hunter JE. 2003. History, development, evolution, and impact of validity generalization and meta-analysis methods, 1975–2001, in *Validity Generalization: A Critical Review* (Murphy KR ed.). Lawrence Erlbaum Associates, Mahwah.
- Schmidt FL, Hunter JE, Outerbridge AN. 1986. Impact of job experience and ability on job knowledge, work sample performance, and supervisory rating of job performance. *J Appl Psychol* **71**(3): 432–439.
- Schmidt FL, Hunter JE, Pearlman K, Hirsh HR. 1985. Forty questions about validity generalization and meta-analysis. *Personnel Psychol* **38**: 697–798.
- Schmidt FL, Le H. 2004. Software for the Hunter–Schmidt meta-analysis methods. Department of Management and Organizations, University of Iowa, Iowa City, U.S.A.
- Schmidt FL, Pearlman K, Hunter JE, Shane GS. 1979. Further tests of the Schmidt–Hunter Bayesian validity generalization procedure. *Personnel Psychol* **32**: 257–281.
- Schmidt FL, Oh IS, Hayes TL. 2009. Fixed vs. random effects models in meta-analysis: model properties and comparison of differences in results. *Br J Math Stat Psychol* **62**: 97–128.
- Scott KD, Taylor SG. 1985. An examination of conflicting findings on the relationship between job satisfaction and absenteeism: a meta-analysis. *Acad Manage J* **28**(3): 599–612.
- Sitzmann T, Brown KG, Casper WJ, Ely K, Zimmerman RD. 2008. A review and meta-analysis of the nomological network of trainee reactions. *J Appl Psychol* **93**(2): 280–295.
- Society for Industrial and Organizational Psychology, Inc. 1987. *Principles for the Validation and Use of Personnel Selection Procedures*. 3rd ed. Society for Industrial and Organizational Psychology, Inc., College Park, MD.
- Society for Industrial and Organizational Psychology, Inc. 2003. *Principles for the Validation and Use of Personnel Selection Procedures (Third Edition)*. 4th ed. Society for Industrial and Organizational Psychology, Inc., College Park, MD.

- Spearman C. 1904. 'General intelligence,' objectively determined and measured. *Am J Psychol* **15**: 201–293.
- Stogdill RM. 1948. Personal factors associated with leadership: a survey of the literature. *J Psychol* **25**: 35–71.
- Taras V, Kirkman BL, Steel P. 2010. Examining the impact of *Culture's Consequences*: a three-decade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *J Appl Psychol* **95**(3): 405–439.
- Taylor PJ, Russ-Eft DF, Taylor H. 2009. Transfer of management training from alternative perspectives. *J Appl Psychol* **94**(1): 104–121.
- Terborg JR, Lee TW, Smith FJ, Davis GA, Tubrin M. 1982. Extension of the Schmidt and Hunter validity generalization procedure to the prediction of absenteeism behavior from knowledge of job satisfaction and organizational commitment. *J Appl Psychol* **67**(4): 440–449.
- Whitman DS, Van Rooy DL, Viswesvaran C. 2010. Satisfaction, citizenship behaviors, and performance in work units: a meta-analysis of collective construct relations. *Personnel Psychol* **63**: 41–81.
- Wiesner WH, Cronshaw SF. 1988. The moderating impact of interview format and degree of structure on the validity of the employment interview. *J Occup Psychol* **61**: 275–290.
- Woehr DJ, Arthur W. 2003. The construct-related validity of assessment center rating: a review and meta-analysis of the role of methodological factors. *J Manage* **29**(2): 231–258.
- Zimmerman RD. 2008. Understanding the impact of personality traits on individuals' turnover decisions: a meta-analytic path model. *Personnel Psychol* **61**(2): 309–348.