11-2020

Personnel selection: a longstanding story of impact at the individual, firm, and societal level

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Paul R. SACKETT

Charlene ZHANG

European Journal of Work and Organizational Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pewo20

Personnel selection: a longstanding story of impact at the individual, firm, and societal level

Filip Lievens , Paul R. Sackett & Charlene Zhang

To cite this article: Filip Lievens , Paul R. Sackett & Charlene Zhang (2020): Personnel selection: a longstanding story of impact at the individual, firm, and societal level, European Journal of Work and Organizational Psychology, DOI: 10.1080/1359432X.2020.1849386

To link to this article: https://doi.Org/10.1080/1359432X.2020.1849386

Published online: 26 Nov 2020.

EUROPEAN JOURNAL OF WORK AND ORGANIZATIONAL PSYCHOLOGY
https://doi.org/! 0.1080/1359432X.2020.1849386

Personnel selection: a longstanding story of impact at the individual, firm, and societal level

Filip Lievens[3], Paul R. Sackett[b] and Charlene Zhang[b]

[a]Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore; [b]Depazrtment of Psychology, University of Minnesota-Twin Cities, Minneapolis, MN, USA

ABSTRACT

This paper discusses how and why the field of personnel selection has made a long-lasting mark in work and organizational psychology. We start by outlining the importance and relevance of the well-established analytical framework (criterion-related validity, incremental validity, utility) for examining the impact of selection at the individual (job performance) level. We also document the substantive criterion-related validities of most common selection procedures on the basis of cumulative meta-analytic research. Next, we review more recent research that investigated the impact of selection at the more macro organizational (firm performance) level. We show that the positive relationship between selection and performance at the individual-level translates to the organizational-level. Finally, we draw upon a longstanding project on situational judgement tests to exemplify the tradition of implementing interventions for improving the way selection is done in specific settings. We reflect on the reasons for this programme's impact on the selection process and its decision makers. We end with recommendations to researchers in personnel selection and other fields for increasing the impact of their research projects.

ARTICLE HISTORY
Received 23 January 2020
Accepted 6 November 2020

KEYWORDS
Personnel Selection; validity; job Performance; firm Performance; situational Judgement Tests; impact; human Capital; utility

"Applied psychology grew from a desire to make the world a better place, and selection was the field in which the movement gained traction." ('Ployhart et al., 2017, p. 291)

This quote comes from a recent review of selected research. It illustrates that selection is generally considered to be a field where academics are not only influential by virtue of their publication record and academic citations but can also have a profound and sustainable impact on individuals, team, and organizations. In the field of personnel selection, impact is typically broadly defined (Pulakos, 2005). One key criterion is related to the ability to adequately estimate the quality of the future hires. This is then evidenced by high performance and low turnover of the hires but also by better performance of the teams/departments they work in and the organization as a whole. As another criterion, personnel selection serves as an important lever in establishing a diverse organization. That explains why selection procedures are no longer rank ordered only in terms of how valid they are but also in terms of the subgroup differences they exhibit (e.g., gender, ethnicity, etc.). Apart from the quality and diversity of the selected applicants, the applicant experience during the selection process (e.g., their perceptions of these selection process) has also received growing attention to gauge the impact of selection. Due to the rise of employer branding, it has increasingly become important for hiring organizations that applicants are not only taking part in a job-related and fair process but that they are also engaged and have "fun" in the often technologically driven selection process so that it reflects on the organizations' image as an employer. Finally, the impact of selection can be assessed via a host of pragmatic criteria such as the ability to reduce costs.

We were pleased to be given the opportunity to write a paper about the impact of selection research. As it is an impossible task to be exhaustive and comprehensive on this account, we had to make some hard decisions. One such decision is that we mainly address the first criterion (validity), although we acknowledge that all other criteria are also of key importance. Second, we decided to tackle only three main themes that exemplify the impact of selection research. First, we focus on the most traditional theme: we outline the well-documented and established impact of selection (in terms of criterion-related validity, incremental validity, and utility) at the level of individual job performance. The second section reviews the strand of studies that have taken a macro multilevel angle in examining the effects of selection beyond individual job performance to firm performance. Third, the last major section shifts the focus to a concrete example, namely our longstanding cumulative line of work on SJTs as a series of interventions to change selection procedures so that they exert more impact on a variety of outcomes.

Impact of selection on individual job performance}

The o^ntrepiece **df LiXdl** [illegible] **SUlbdion** at the indi/dual level consists of estimating the relationship between *scares* on a predictor (e.g., cognitive ability test, personality inventory) and a wide array of key outcomes (e.g., indices of job performance, turnover, training performance, organizational citizenship, various forms of counterproductive work behaviour, such as theft, various forms of withdrawal, such as attrition and absence, customer service or other outcomes). Historically, the

CONTACT Filip Lievens © filiplievens@smu.edu.sg

correlation coefficient has been used to document the strength of this relationship, with such correlations labelled as indices of *"criterion-related validity"*. Importantly, a finding of non-zero criterion-related validity reveals the impact of using a given selection predictor because it shows it is possible to raise the mean performance of selected individuals by using this predictor in the selection process. Criterion-related validity research has been common practice for over a century (for reviews, see Ployhart et al., 2017; Ryan & Ployhart, 2014; Sackett & Lievens, 2008; Sackett et al., 2017).

As many studies are done with sample sizes smaller than 100 (Cascio & Aguinis, 2008), criterion-related validity is often estimated with substantial uncertainty. For example, even a sample of 100 results in a correlation with a 95% confidence interval of ±.20. Pooling data across studies via meta-analysis emerged as a key solution to the sampling error problem. Since the 1970s (Schmidt & Hunter, 1977), there have been hundreds of meta-analyses of different predictor-criterion relationships. One highly influential paper is Schmidt and Hunter (1998) compilation of meta-analyses of various predictors of job performance. It contains a table giving the mean correlation with performance for all of these predictors (corrected for range restriction and unreliability in the performance measure). Table 1 updates Schmidt and Hunter's table. Generally, it attests to the substantive validities obtained in selection.

The above focuses on the validity of individual predictors. However, to examine *incremental validity* multiple predictors can be used, either sequentially, or combined into a composite. Schmidt and Hunter (1998) focus was on cognitive ability, and for each predictor they computed the validity of a regression-weighted composite pairing that predictor and cognitive ability. As such, the highest composite validities to be obtained can rise to about .65. In short, we have strong evidence regarding the availability of a wide rangeof predictors that show substantial relationships with job performance, either alone (criterion-related validity) or in combination with other predictors (incremental validity).

Closely linked to the issue of validity is the pragmatic issue of *utility,* namely the value to the organization of using a selection system. As an illustration, we focus here on a straightforward model for estimating utility in the case of continuous criteria, put forward by Naylor and Shine (1965). In their formulation, the criterion is presented in standard score form (i.e., z scores with mean = 0 and standard deviation = 1). The improvement in mean criterion performance due to testing is the product of two variables: (a) the validity of the predictor, and (b) a value reflecting the selection ratio (i.e. the percentage of candidates screened in), specifically, the mean predictor score, in z score form, among those selected by the test. Table 2 shows the product of validity and mean test score among those selected for various selection ratios and levels of validity. The highest mean criterion performance is observed when higher validity is paired with lower selection ratios. Even a highly valid predictor will have only a very small effect on criterion performance when the organization faces a high selection ratio.

An appealing feature of the Naylor-Shine approach is that it expresses the value of selection in a standard score metric, permitting a direct comparison with the effect of other organizational interventions that use a Cohen's *d* metric (e.g., the effect of a training intervention, for example). For a treatment of less technical alternate metrics and visualization approaches, we refer to Kuncel and Rigdon (2012).

Table 1. Summary of corrected correlations between predictors and job performance.

Predictor	Validity corrected	Source (k)
Cognitive ability test	.51	Hunter (1980; *k* = , p. 151)
Job knowledge test	.48	Hunter and Hunter (1984; *k* = , p. 10)
C (generic personality test)	.27	Mount and Barrick (1995; *k* = , p. 172)
C (contextualized personality test)	.30	Shaffer and Postlethwaite (2012; *k* = ,*p.* 22)
Biographical data inventory	.33	Rothstein et al. (1990[c])
Integrity test	.18	Van Iddekinge et al. (2012, *k* = ,*p.* 74)
Structured interview	.25 (level 3)[a].44 (level 4)[a]	Huffcutt et al. (2014; *k* = , p. 2) Huffcutt et al. (2014; *k* = , p. 8)
Unstructured interview	.16	Huffcutt et al. (2014; *k* = , p. 9)
SJT (knowledge)	.26	McDaniel et al. (2011; *k* = ,*p.* 96)
SJT (behavioural tendency)	.26	McDaniel et al. (2011; *k* = ,*p.* 22)
Work sample	.33	Roth et al. (2005; *k* = , *p.* 54)
AC (overall assessment rating)	.44	Sackett et al. (2017; *k* = , p. 17)
Interests	.16	Nyeetal. (2017; *k* = , p. 92)
Interest congruence	.32	Nyeetal. (2017; *k* = , p. 92)
Reference check	.26	Hunter and Hunter (1984; *k* = , p. 10)
Work experience	.06	C. Van Iddekinge et al. (2019; *k* = ,*p.* 44)
Educational level	.09	Ng and Feldman (2009; *k* = , p. 85)
Person-organization fit	.15	Arthur et al. (2006; *k* = ,*p.* 36)
Job tryout	.44	Hunter and Hunter (1984; *k* = , p. 20)
Peer ratings	.49	Hunter and Hunter (1984; *k* = , p. 31)
T&E point method	.11	McDaniel et al. (1988; *k* = , p. 91)
T&E behavioural consistency method	.11	McDaniel et al. (1988; *k* = , p. 15)
Graphology	.02	Bar-Hillel and Ben-Shakhar (1986); Ben-Shakhar (1989); Ben-Shakhar et al. (1986); Neter and Ben-Shakhar (1989)[b]
Age	.03	Sturman (2003; *k* = , p. 115)

Notes, k = number of effect sizes, C = conscientiousness, T&E = training & experience, SJT = situational judgement test, AC = assessment centre.

[a]Level of standardization in interviews, with higher values denoting higher standardization.

[b]Obtained from Schmidt and Hunter (1998).

[c]As the.33 was aggregated from multiple meta-analyses, there was no exact *k*.

Table 2. Product of validity and mean test scores at various hypothetical selection ratios and levels of validity.

Selection Ratio	Mean Predictor Score Among Those Selected	Mean Criterion Score Among Those Selected When Validity = .50	Validity = .30	Validity = .10
0.01	2.52	1.26	0.76	0.25
0.05	2.05	1.03	0.62	0.21
0.10	1.76	0.88	0.53	0.18
0.20	1.40	0.70	0.42	0.14
0.30	1.16	0.58	0.35	0.12
0.40	0.97	0.49	0.29	0.10
0.50	0.80	0.40	0.24	0.08
0.60	0.64	0.32	0.19	0.06
0.70	0.50	0.25	0.15	0.05
0.80	0.35	0.18	0.11	0.04
0.90	0.20	0.10	0.06	0.02
0.95	0.11	0.06	0.03	0.01
0.99	0.03	0.02	0.01	0.00
1.00	0.00	0.00	0.00	0.00

We end this section with the observation that the documentation of criterion-related validity is now a routine part of the work of industrial/organizational psychologists. Many consulting firms will conduct a local study, sample size permitting, as part of their work with a client organization. Such studies do not appear in the peer-reviewed literature: documenting the value of a selection system fora specific organization is of value to that organization, but not seen as a general contribution to knowledge. For instance, Ones et al.'s (1993) meta-analysis of integrity tests located 695 validity coefficients. Only 67 came from the published literature; others were provided by test publishers from their work with organizations.

*Impact of selection on firm performance""^*

The/>rior section convincingly demonstrates the validity and uti/ty of selection for predicting performance at the individual lafel (e.g., Schmidt & Hunter, 1998). Much of this research is /ased on the fundamental assumption that the effectiveness of selection practices ultimately translates to higher levels, namely team success and especially organizational success. Therefore, researchers have asserted the importance of incorporating higher-level outcomes or using multilevel perspectives in investigating the effect of individual-level selection on organizational performance (e.g., Fulmer & Ployhart, 2014; Lepak et al., 2012; Ployhart, 2012b, 2012a; Ployhart & Moliterno, 2011; Renkema et al., 2016; Schneider et al., 2012, 2000). Below we therefore review this important stream of research for demonstrating the impact of selection.

Empirical support for the selection-firm performance relationship

Empirical examination of the selection-performance link has taken several different approaches. In one approach, selection practices are examined either as an individual HR practice (so that the effect of selection can be isolated) or under the umbrella term "staffing", wherein selection is combined with other activities such as recruitment. Terpstra and Rozell (1993) conducted one of the first studies using this approach. They surveyed the HR department heads about staffing practices among 201 companies, including the follow-up studies of recruitment sources, validation studies of selection predictors, and the use of structured, standardized interviews, cognitive ability tests, and biographical data. They found that the number of distinct staffing practices implemented was positively related to higher annual profit and profit growth, and the relationship was moderated by industry and firm size. Since then, the correlational links between selection and organizational financial performance (Delery & Gupta, 2016; Skaggs & Youndt, 2004), labour productivity and quality (Koch & McGrath, 1996; Macduffie, 1995), operational performance (Ahmad & Schroeder, 2003), and perceived performance (Singh, 2004; Vlachos, 2008) have been repeatedly established. Aggregating over 15 effect sizes, Combs et al. (2006) reported a corrected correlation of .14 between staffing selectivity and firm performance.

Another stream of studies takes a systems-approach and scrutinizes selection practices as embedded within HR systems alongside non-selection practices (e.g., training, compensation, job design) and how they relate to firm-level performance as a whole. HR systems that have been studied the most are high-performance work systems (HPWS; Huselid, 1995; Zacharatos et al., 2005), which can consist of any number of effective HR practices, including selective staffing, training and development, compensation and incentives, performance appraisals, employee involvement, and information sharing (Lepak et al., 2006). This body of research has largely supported the positive relationship between some selection-included HR systems and organizational performance. A seminal study by Huselid (1995) cross-sectionally surveyed 968 firms on their use of high-performance work practices (HPWPs). They factor analysed 13 practices and obtained two factors: 1) practices intended to enhance employee skills, such as information sharing programmes, job analysis, attitude surveys, and training, and 2) organizational structures which included selection procedures and those intended to enhance employee motivation, such as performance appraisals and performance-based promotions. Results showed a positive relationship between HPWS as a whole and organizational financial performance. Combs et al. (2006) meta-analytically reviewed this research and reported that a one standard deviation increase in use of HPWPs was associated with a .20 standard deviation increase in firm performance. Similarly, Rabi et al. (2014) found an overall positive relationship between HPWS and business performance (p = .28), a link that varied in magnitude across nations but remained positive in all 29 countries included.

Longitudinal evidence for selection-firm performance link

There is no doubt that some type of relationship exists between selection practices and firm-level performance. However, over-reliance on cross-sectional designs makes it difficult to understand the nature of this relationship (Godard & Delaney, 2000; Wall & Wood, 2005). In fact, some studies found that the relationship between staffing activities and organizational performance disappeared once other non-selection HR practices (e.g., Absar et al., 2012; Singh, 2004) or prior performance was accounted for (Guest et al., 2003; Wright et al., 2005). So, do more effective selection practices cause better organizational

outcomes or are they simply an indication of overall firm success (e.g., Becker & Gerhart, 1996; Huselid & Becker, 1996)?

Three larger studies shed light on this. While Wright et al. (2005) found that past and concurrent performance largely accounted for the relationship between HR practices and future performance, Saridakis et al. (2017) conducted a meta-analysis of eight longitudinal studies and reported a corrected correlation of .29 between the HPWS at time 1 and firm performance at time 2, supporting the causal effect of HR practices. Kim and Ployhart (2014) further explored the selection-performance link and examined whether it was dependent on the economic context. Using the Great Recession as a naturalistic experiment, they demonstrated that selective staffing, operationalized as the selection ratio, was predictive of profit growth indirectly through productivity. Furthermore, results showed that training was more beneficial for prerecession profitability, whereas staffing was more beneficial for postrecession recovery. So, selective staffing seems to accumulate generic human capital which aids organizations in adapting to environmental changes. Other longitudinal studies also supported the effects of selection and training practices on unit-level performance (Ployhart et al., 2011; Van Iddekinge et al., 2009).

Empirical investigation of the "Black box"

To examine the "black box" of how selection practices exert influence on firm-level criteria, several theoretical frameworks have identified a number of mediators of the selection-performance relationship. The ability, motivation, and opportunity (AMO) model specifies three possible pathways through which HR practices can result in organizational success, namely ability-enhancing (e.g., recruitment, selection, training), motivation-enhancing (e.g., compensation, performance appraisal), and opportunity-enhancing (e.g., job design, involvement) (Appelbaum et al., 2000; Boxall & Purcell, 2003). Consistent with the AMO model, selection activities have been found to contribute to organizational success by increasing workforce's knowledge, skills, and abilities (KSAs). Hatch and Dyer (2004) showed that manufacturing firms that screened applicants on education level and technical skills performed better because employees were able to learn by doing or reduce the amount of manufacturing defects more quickly. Similarly, Katou and Budhwar (2010) found that the positive association between selection and development policies and organizational performance was fully mediated by employee skill (see also Jiang et al., 2012)

Besides enhancing employee KSAs, selection practices can also affect employee attitudes. Studies found that organizational commitment mediated the link between selection and firm performance (Ahmad & Schroeder, 2003; Wright et al., 2003). Messersmith et al. (2011) further showed that the mediating effect of organizational commitment was explained by its effect on organizational citizenship behaviour.

Other than exploring mediators of the selection-performance relationship, a separate body of research investigated employee attributes that are typical results of selection practices as predictors of organizational performance. Consistent with Schneider's (1987) attraction-selection-attrition (ASA) model, these individual attributes were found to emerge as more homogeneous within than between groups or organizations, such as personality (Giberson et al., 2005; Jordan et al., 1991; Oh et al., 2015; Ployhart et al., 2006; Schaubroeck et al., 1998; Schneider et al., 1998; Slaughter et al., 2004) and aptitude. Often, such attributes are referred to as human capital (conceptualized as unit-level KSAOs), capturing the collective characteristics of an organization's workforce (Ployhart et al., 2006), and used as explanatory mechanisms for the selection-performance relationship. This is in line with the resource-based view of the firm (Barney, 1991), which attributes inter-organization performance differences to the heterogeneity in their resources, a central type of which is human capital. Taking the human capital perspective, Cabello-Medina et al. (2011) found that selection based on learning potential and interpersonal skills enhanced the uniqueness of human capital or employee skills, and in turn improved innovation and overall firm performance.

Indirectly demonstrating the importance of human capital, a study conducted by Li et al. (2015) in China found that human capital-enhancing HR practices, including selective staffing, were especially beneficial for firms in regions with a lower proportion of highly educated people and education investment. In such cases, HR practices are an alternative approach of enhancing an organization's human capital. Chowdhury et al. (2014) reported that the relationship between job-related experience and revenue productivity was stronger in firms in which employees held higher levels of formal education. Similar results were found in a longitudinal study by Ployhart et al. (2009). Changes in human capital, operationalized as unit-level service orientation, resulted in subsequent changes in unit-level effectiveness. Furthermore, the effect of human capital was greater among units with consistently high unit service orientation flows. To sum up this literature, a meta-analysis (Crook et al., 2011) found an overall positive relationship between human capital and firm performance (p = .21). The link was stronger for more proximal measures: Firm-specific human capital measures (p = .24) were more predictive than general human capital ones (p = .14), and human capital predicted operational performance measures (p = .26) more strongly than global performance measures (p = .15).

Situational judgement tests and high-stakes selection: a 10-year research Programme

This last section focuses on a third theme that shows the impact of selection. Specifically, interventions are often implemented to alter the way selection is done in a specific setting. To this end, we review a large-scale project about the use of Situational Judgement Tests (SJTs) in high-stakes selection contexts (in this case the admissions process to college/university). SJTs are selection procedures that present candidates with a hypothetical situation and ask what to do in that situation (Motowidlo et al., 1990).

We chose this longstanding project because it serves as a concrete example of interventions to make modifications to selection procedures in light of various considerations (e.g., validity improvement, diversity, cost reduction, candidate experience). Finally, results obtained in this large-scale project had direct ramifications for the actual design of the admissions process. In turn, the outcomes of this admissions process were

high-stakes: they had substantial impact at the individual (e.g., people's careers), school (e.g., inflow and demographic make-up of candidates), and societal level (e.g., diversity of future physician population). So, many stakeholders were involved and critically evaluated the admissions process and the tests used. Obviously, this project is just one possible illustration about how some (but not all) aspects from the two prior reviews can generate real-world impact. That is, we will mainly discuss the impact of selection at the individual level, although some broader level implications will also be mentioned.

Background

In the late 90s, practitioners and researchers had embraced SJTs because they seemed "psychometric alchemy" (Landy, 2007, p. 418). That is, meta-analytic criterion-related validities were acceptable, subgroup differences seemed to be lower than for cognitive ability tests, and the candidate experience was favourable. Yet, most studies had been conducted in employment settings and especially with incumbents in organizations. A key unresolved question was whether SJTs could also be used as supplements to more traditional cognitive ability tests in high-stakes contexts (like college admissions). After all, a major challenge in admissions settings was to uncover reliable and valid selection procedures for assessing interpersonal skills (aka "soft" skills or "21st-century" skills) among large candidate pools. Assessing such interpersonal skills among medical students early on (in the admissions stage) could also reduce the gap that in those years often existed between the objectives underlying the curricula of medical schools (i.e., highlighting both cognitive and interpersonal competencies) and their actual medical selection and training practices (i.e., mostly focused on cognitive and clinical skills).

When the board overseeing the Flemish admissions exam challenged us with this very question, we posited that we hypothesized video-based SJTs to be useful approaches for capturing interpersonal skills for at least three reasons. First, assessing people's interpersonal procedural knowledge (i.e., knowing how to act in interpersonal situations) via video-based situations might serve as precursor of their interpersonal behaviour in actual interactions (e.g., during internships, on the job). Second, selecting students on their interpersonal procedural knowledge at the time of admission might facilitate future interpersonal skills training. Third, we expected the realistic job-related situations inherent in video-based SJTs to receive favourable candidate reactions.

Validating the basic set up

We received the funding to develop and implement SJTs. The first results were disappointing, though. The interpersonal SJT scores did not predict GPA. The picture changed when we used interpersonal courses as criterion measures (Lievens, Buyse, & Sackett, 2005a, 2005b; This was confirmed in a longitudinal and multiple-cohort study (Lievens, 2013; Lievens & Sackett, 2012): Whereas cognitive tests emerged as the best predictors of GPA, the interpersonal skills assessment via SJTs had significant added value over cognitive tests for predicting interpersonal GPA. In fact, the SJT scores became more valid over the years and were able to predict physicians' job performance nine years later. The interpersonal skills training received during medical education did not seem to negate selecting students on interpersonal skills in the first place. Moreover, the candidate experience was a good one. Students perceived the interpersonal SJT as significantly more face valid than the cognitive tests (Lievens, 2013). This was also critical in view of the high-stakes and the high visibility of the admission exam. The most direct impact of these results was that SJTs remained an integral part of the admissions process.

Generating evidence-based knowledge via field experiments

Once these validity and candidate experience results had removed initial scepticism regarding the role of interpersonal SJTs in the admissions process, the attention shifted towards cutting costs and streamlining the SJTs, as demanded by the board. Yet, at the same time, it also opened a window of opportunity for us to set up interventions via field experiments and gather evidence-based information in support of potential changes to the SJTs. Below, we discuss two such interventions. One pressing question of the board was whether a video format was really necessary. If a traditional written version had equal validity, then the costly investments in the design and logistics related to video-based SJTs could be avoided. At that time, a laboratory experiment of Chan and Schmitt (1997) had shown that students perceived video-based SJTs to be more face valid than the written SJT (see also Richman-Hirsch et al., 2000). In addition, subgroup differences were lower for the video-based SJT. However, the effects of presentation format (video vs. written) on criterion-related validity were still unknown. Therefore, we set up a field experiment. We transformed the video-based SJT that was used in a prior year in the admissions exam in a written format, holding SJT verbal content constant (Lievens & Sackett, 2006). The cohort of students completing the SJTs across the years was similar. Results showed that the interpersonal video-based SJT had a significantly lower cognitive loading and significantly higher predictive validity for predicting interpersonally oriented criteria than the written SJT (see also the later meta-analysis of Christian et al., 2010). This result had direct impact on how the SJT was administered to thousands of candidates because it provided the evidence needed to continue the video-based SJT.

Another lingering debate related to which response instruction format to use in SJTs. Generally, two broad response instruction formats are distinguished: (a) a knowledge instruction (aka "should do" instruction), asking candidates to show whether they know what the most effective answer is, and (b) a behavioural tendency instruction (aka "would do" instructions) that requires candidates to report how they typically behave. In prior research, the response instruction format affected the amount of response distortion in SJTs (M.A. McDaniel et al., 2007; Nguyen et al., 2005). SJTs with behavioural tendency instructions were found to be more susceptible to response distortion than SJTs with knowledge instructions. Yet almost all research on these two response instructions had been conducted with incumbents. Clearly, the motivation to self-present is quite different in low-stakes

incumbent settings than in high-stakes applicant settings. As the response instruction issue was a long-standing debate in the board overseeing the admission exam, the board agreed to let us randomly assign candidates to one of two formats (with the SJT content being exactly the same across them). A contingency plan was in place to equate scores in case substantial mean score differences were found. In line with prior research, the knowledge instruction SJT correlated more highly with cognitive ability than did the behavioural tendency SJT and there was no difference between the SJT criterion-related validity under the response instruction sets. Yet, contrary to prior research in low-stakes settings, there were no meaningfully important differences between mean scores for the response instruction sets. In essence, candidates presented with "would do" instructions ignored the instructions and gave socially desirable "should do" responses. So, the type of response instructions did not matter much in a high-stakes context for reducing response distortion (Lievens et al., 2009). So, our advice to the board was to use knowledge instructions because they make faking a non-issue.

Ensuring sustained impact over time

Once the use of SJTs became established and their design was systematically improved in this high-stakes setting, new challenges emerged. The SJT development and implementation process had to stand the test of time. One challenge related to the possibility of people retaking the SJTs. Therefore, it was important to ascertain the size of the retest effects. We found that retest effects on SJTs were in the same range as with cognitive tests (Lievens et al., 2005a). Due to retesting, it also became crucial to develop alternate forms of SJTs. Although there existed standard alternate test development procedures, this was not the case for SJTs. Moreover, we faced the challenge that the law precluded us from pretesting the items. This is because the law did not allow putting pilot items (that would ultimately not be used in computing people's test scores) into the actual SJT. Therefore, we started experimenting with various SJT alternate-form development approaches (i.e., random assignment, incident isomorphism, and item isomorphism). Generally, results revealed that even small changes in the context of the situations presented resulted in significantly lower alternate-form consistency. Conversely, placing more constraints on the alternate-form development process (as in cloning approaches) proved beneficial (Lievens & Sackett, 2007). This was another evidence-based recommendation that was adopted in the design of SJTs in high-stakes settings.

Another challenge over time was that (just like other tests used in high-stakes situations) SJTs became the subject of commercially available coaching programmes. This invoked a whole set of new questions. First, to assess the size of the effects of commercial coaching we compared the pre- and post-coaching SJT scores of individuals who had failed the overall exams once, sought commercial coaching, and then retook the exams later (Lievens et al., 2012): With coaching, SJT scores improved sizably (d = 0.59). A second issue is that coaching is commonly viewed as a source of unfairness. People with more financial resources are more able to access these coaching programmes and thus obtain a competitive advantage. So, we examined the effects of organizationally provided coaching. Such coaching is made freely available to all, thus addressing the disparity in access to coaching (Stemig et al., 2015). We found that the most effective coaching procedures were freely available, officially sanctioned, and involved practice with items similar to those on the actual SJT. In addition, SJT scores' criterion-related validity was not degraded by coaching. So, it then became important to make less advantaged students aware of the value of this free coaching.

Finally, technological innovations have to be taken into account. In the past, many organizations adhered to the classic multiple-choice (MC) format given its ease of administration and scoring. Yet, the availability of text analytics makes it nowadays possible to quickly score constructed (aka open-ended) responses. In addition, constructed response format might leverage diversity effects due to their lower cognitive load. Therefore, in a related research project outside of academia (at that time logistics precluded implementing this in the medical admission process), we randomly assigned people to two constructed response formats (Lievens et al., 2015,2019). In one format, people typed their answers to multimedia interpersonal situations in a text box, whereas in another format they reacted via a webcam. In terms of promoting a more diverse inflow, the two constructed response formats outperformed the MC format because the MC format led to unwanted cognitive demands. Among the two constructed response formats, the webcam format was the winner because it led to a more diverse candidate intake. There were no differences among the response formats in predictive power. So, constructed response formats and especially webcam formats benefit organizations that strive to increase diversity inflow while maintaining valid predictions. This is important because success in predicting outcomes of interest is no longer the only issue in designing and evaluating selection. Many organizations also value diversity as an outcome, and there is a large literature on balancing the sometimes-competing values of maximizing the performance vs. maximizing the diversity of those selected (see De Corte et al., 2011 for techniques for examining trade-offs between such competing objectives).

Reflections

In sum, the impact of this longstanding research programme on the effectiveness of Situational Judgement Tests as measures of interpersonal skills is multifold. First, the results were not only published in top-tier journals but were also covered by the media. This informed the general public about the validity of the process. Second, the impact of the use of SJTs went beyond the level of individual performance. As the SJT focused on interpersonal skills, it made clear to medical universities that students should also further sharpen their interpersonal and communication skills once in medical school. So, a couple of years after SJTs became part of the admission exam, we noticed that medical schools started changing their curricula, thereby increasing the number and weight of interpersonal courses. Third, the results had a direct impact on generating scientific evidence and best practices on how SJTs in an admissions process were to be developed, implemented, scored, and maintained. Partially on the basis of this pioneering work in Flanders,

universities across the world started (or considered) using SJTs for assessing interpersonal skills in admissions (USA, UK, Australia, Switzerland, Germany, Austria, the Netherlands, etc.).Fourth, the results of our field experiments exemplified how working together with the board in tweaking existing SJT design led to direct actionable implications for selection practice. These field experiments also inspired various other studies that systematically manipulated SJT building blocks (although mostly in lab experiments, for a review of these studies, see Lievens & Sackett, 2017).

We conclude with some reflections about why we think this project affected how SJTs are used and developed in high-stakes settings. First, there was pressure from the society at large (general public) and media to prove that the admissions process (including SJTs) really worked given the ramifications it had for people's future lives and careers. Second, it was the right time to conduct this kind of research because scientific evidence was urgently needed. Prior research was conducted mostly in employment settings and there were a lot of unknowns (e.g., the validity to be expected and the effects of presentation format, response format, instruction format, retesting, coaching). The urge to provide answers to these unresolved questions stemmed not only from a local need but also from a global need because many colleges around the world were struggling with them. In fact, at that time, across the world, there was a striking discrepancy between medical school objectives and the admissions process. Although the objectives of curricula acknowledged the importance of interpersonal skills, most formal medical school admissions tended to primarily assess academic achievement in science domains and cognitive abilities. Therefore, an editorial in the leading medical journal Lancet suggested that, "a more holistic and sophisticated approach to selection - based on predictors of care that are both valid and patient-relevant - needs to be developed and applied" (Barr, 2010). Our project showed that SJTs (in their various formats) implement part of such an approach.

In hindsight, the impact might also have stemmed from the strategy we used for involving the board and communicating with it. At the outset, the board that consisted of 90% of professors in a specific medical field was mostly ignorant with respect to tests and psychometric theory. As an example, they often mentioned that a test was essentially similar to an exam ("it is something you design in one evening, right"). This serves as a good illustration of the acknowledged gap between our scientific knowledge base and the beliefs of applied human resource practitioners. More generally, Rynes et al. (2002) documented a large gap between what HR practitioners believe and the actual findings from research. For example, many HR practitioners seem to believe that ability testing is less effective than the research base shows them to be. Thus, potentially valuable selection tools do not become part of the selection systems in some organizations due to lack of knowledge or disbelief about selection practice effectiveness.

To address this science-practice gap, each year, we presented our test design approach and research results, thereby taking ample time for discussion with the board. Along the years, the board's interest in test and psychometric issues grew. Moreover, at the end of our presentations, we also signalled key unresolved issues to the board, thereby triggering questions on how to address them in the future. This strategy worked well because we were talking to medical professors, physicians, and specialists who were positively inclined to evidence-based research and hence open to interventions and field experiments. This was best exemplified by their endorsement to implement the instruction format manipulation. Generally, our strategy created an awareness for pressing issues to be investigated, a shared commitment to address them via field experiments, and a curiosity for the outcomes. This fits well with strategies of Muchinsky (2004) to successfully implement psychometrically straightforward test development principles in organizational contexts. He referred to his strategy as a balancing act wherein he combined strategies of education, shared responsibility, negotiation, respect, and recognition of available knowledge of all stakeholders.

We want to end with several cautionary notes, though. We were not always successful in persuading the board. For example, we explained above that our scientific evidence favoured a video-based format over the traditional written format. That was also the reason why the board supported to keep investing in a video-based SJT for several years. However, the video-based SJT was ultimately replaced by a written one due to electricity breakdowns caused by the number of large screens needed to project the video clips to over 7,000 candidates. So, eventually logistics were regarded as more important than scientific evidence. As another example, we also encountered situations where our message to the board was too complex. For instance, we could not convince them of the importance of range restriction issues so that the predictor composites in the admission exam were not optimally computed (Sackett et al., 2007). This bears resemblance to the difficulty of persuading managers of utility calculations (Latham & Whyte, 1994). Finally, although the SJT was still to be used as part of the admissions process, the funding to examine the SJT suddenly stopped after about ten years. Apart from budget constraints imposed by the government, a key reason was that the board and the Ministry felt that "they had confidence in SJTs and they knew SJTs worked". This was a pity because test construction is a never-ending process.

Discussion

Impact outside of academia in selection: main conclusions

This paper reviewed why selection is one of the success stories in work and organizational psychology. As one reason, there exist well-established paradigms and analytics for investigating the impact of selection. Estimating the criterion-related validity of selection has become common practice and has cemented the view that carefully selecting people adds above random selection. Moreover, incremental validity examinations document the added value of selection procedures over already existing ones, whereas utility frameworks enable converting the benefits of selection into financial metrics and comparing them to other HR interventions.

As a second reason for the impact of selection, a myriad of meta-analyses has documented the criterion-related validity of selection procedures. In fact, almost each selection procedure

has its own meta-analysis in terms of the predictive power of that procedure for various criteria and settings. We should be proud of this rich evidence base and promote the validities of selection procedures as key testaments of our ability to predict human performance at work (also known as the "supreme problem;" Hall, 1917; Ployhart et al., 2017). Along these lines, Schmitt (2014) concluded: *"Some authors are critical of the personnel selection field, given the magnitude of the validities reported. However, if one considers the complexity of the job performance phenomena and the organizational constraints on performance and our ability to define and measure performance, the size of the coefficients actually represents one of the most remarkable achievements of psychology."* (pp. 58). Others compared the criterion-related validities of selection procedures to the validities of medical procedures. Generally, the criterion-related validities of selection procedures had the same or higher validities than those of common medical procedures (e.g., histamine, ibuprofen, Viagra; Meyer et al., 2001).

Third, besides support of the validity/utility of selection practices at the micro (individual) level and the interventions to improve selection procedures, empirical evidence at the macro (organizational) level demonstrated that selection also contributes to firm success. A large body of empirical work suggests that the positive relationship between HR practices (selection) and performance at the individual-level translates to the organizational-level, contrary to the groundless claims that HR is irrelevant and worthless (e.g., Hammonds, 2005). As noted by Kaufman (2014), *"it would be difficult to overstate how influential and energizing this line of research has been to modern HRM"* (p. 211).

Fourth, in selection, researchers often work closely together with practitioners (also referred to as the scientist-practitioner model) to further improve the way selection is done in specific settings. We exemplified this intervention-based approach with our longstanding project on developing SJTs for admissions. The opportunity to set up experiments in the field enabled us to conduct rigorous selection research, while at the same time addressing problems that selection practitioners care about and struggle with.

Undoubtedly, our ability to establish the impact of selection at these different levels is quintessential and valuable as the world of selection evolves to embrace new developments in gamification, mobile technology, statistical methods, machine learning, artificial intelligence, and Big Data (Landers et al., 2018; Putka et al., 2018; Sajjadiani et al., 2019). So, we are convinced that personnel selection will remain an exciting and impactful field of study.

Impact outside of academia in domains other than selection: recommendations

We hope that this paper serves as a strong illustration of the practical and sustained impact that academic work can have "outside of academia". We do not see specific reasons why the field of personnel selection would be better suited for academics than other fields to make a sustained, real-world impact outside of academia. Generally, we believe that applied, actionable impact will be most direct and most feasible for topic research areas where there is a clear link to an area of professional practice (e.g., selection, training, compensation, employee attitudes), where new work and organizational psychology graduates can find employment, and where target consumers of the research can be clearly identified. Academics from such organizational psychology areas should be able to meet academic/publication criteria and practical client needs (e.g., have an impact on organizations). Interestingly, in his role as SIOP president, Frank Landy, wrote a letter to former US president Barrack Obama: This letter mentioned all the different fields in work and organizational psychology that academics could have an impact on.

On the basis of our longstanding experiences in the selection field, we end with painting some broad-brush recommendations for academics in fields other than selection to conduct rigorous, publication-worthy research as well as to make an impact outside of academia. First of all, building one's publication record and one's reputation of expertise in a domain increases the frequency with which one is approached with invitations to participate in (or lead) important projects with organizations (like developing SJTs for admissions). So, having established research expertise in a specific domain and communicating about it seems like an important precondition. Second, there is the adage of "choosing wisely" both in terms of the type of practical project and the partners involved. Regarding the latter, it is important for the longevity of the project to establish a network of "stable" practitioners with whom one works. One of the problems we have experienced is that when a practical project is finally under way our linking pin in the organization moves jobs or even leaves the organization. Relatedly, the decision to invest in a practical project should be based on the ability of the project to generate a large and relevant sample size for testing the key research questions. If a project will not generate sufficient "power" to draw conclusive results, it might not be the one that will lead to a publication. A final criterion is that one should be wary of investing in practical projects that are essentially nothing more than "applied problem solving" (Hollenbeck, 2008). This means that one implements an already well-established solution in an organization. Clearly, such projects are worthwhile for organizations but they often fail to generate new knowledge. It is better to start with "interesting" research questions that for example, pit various competing perspectives against each other or that deviate from current practices. We thus suggest to get especially involved in practical projects that enable to set up field experiments. Along these lines, our longstanding SJT admissions project gave some excellent suggestions on how to generate curiosity among the organization involved and create a spirit of experimentation and evidence-based management.

Acknowledgments

We would like to thank Philipp Schäpers and Christoph N. Herde for their insightful comments to a prior version of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Charlene Zhang http://orcid.org/0000-0001-6975-5653

References

Absar, N., Nimalathasan, B., & Mahmood, M. (2012). HRM-market performance relationship: Evidence from Bangladeshi organizations. *South Asian Journal of Global Business Research,* 1(2), 238-255. https://doi.org/10.1108/20454451211252750

Ahmad, S., & Schroeder, R. G. (2003). The impact of human resource management practices on operational performance: Recognizing country and industry differences. *Journal of Operations Management, 21(1),* 19-43. https://doi.Org/10.1016/S0272-6963(02)00056-6

Appelbaum, E., Bailey, T., Berg, P., Kalleberg, A. L., & Bailey, T. A. (2000). *Manufacturing advantage: Why high-performance work systems pay off.* Cornell University Press.

Arthur, W., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology, 91(4),* 786-801. https://d0i.0rg/l 0.1037/0021-9010.91.4.786

Bar-Hillel, M., & Ben-Shakhar, G. (1986). Scientific aspects of graphology. In H. R. Arkes, & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 263-279). Cambridge University Press.

Barney, J. (1991). Special theory forum the resource-based model of the firm: Origins, implications, and prospects. *Journal of Management, 17*(1), 97-98. https://d0i.0rg/l 0.1177/014920639101700107

Barr, D. A. (2010). Science as superstition: Selecting medical students. *The Lancet,* 376(9742), 678-679. https://doi.org/10.1016/S0140-6736d0)61325-6

Becker, B., & Gerhart, B. (1996). The impact of human resource management on organizational performance: Progress and prospects. *Academy of Management Journal,* 39(4), 779-801. https://doi.org/10.5465/256712

Ben-Shakhar, G. (1989). Non-conventional methods in personnel selection. In P. Herriot (Ed.), *Handbook of assessment in organizations: Methods and practice for recruitment and appraisal* (pp. 469-485). Wiley.

Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y, Ben-Abba, E., & Flug, A. (1986). Can graphology predict occupational success? Two empirical studies and some methodological ruminations. *Journal of Applied Psychology, 71(4),* 645-653. https://d0i.0rg/l 0.1037/0021-9010.71.4.645

Boxall, P., & Purcell, J. (2003). *Strategy and human resource management: Management, work and organisations.* Palgrave MacMillan.

Cabello-Medina, C., López-Cabrales, A., & Valle-Cabrera, R. (2011). Leveraging the innovative performance of human capital through HRM and social capital in Spanish firms. *The International Journal of Human Resource Management,* 22(4), 807-828. https://doi.org/10.1080/09585192.2011.555125

Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *The Journal of Applied Psychology,* 93(5), 1062-1081. https://doi.org/10.1037/0021-9010.93.5.1062

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology,* 82(1), 143-159. https://doi.org/10.1037/0021-9010.82.1.143

Chowdhury, S., Schulz, E., Milner, M., & Van De Voort, D. (2014). Core employee based human capital and revenue productivity in small firms: An empirical investigation. *Journal of Business Research, 67(11),* 2473-2479. https://d0i.0rg/l 0.1016/j.jbusres.2O14.03.007

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgement tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology,* 63(1), 83-117. https://doi.0rg/lO.l 111/j.1744-6570.2009.01163.x

Combs, J., Liu, Y., Hall, A., & Ketchen, D. (2006). How much do high-performance work practices matter? A meta-analysis of their effects on organizational performance. *Personnel Psychology, 59(3),* 501-528. https://d0i.0rg/l 0.1111/j.l 744-6570.2006.00045.X

Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen, D. J., Jr. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology,* 96 (3), 443-456. https://doi.org/10.1037/a0022147

De Code, W., Sackett, P. R., & Lievens, F. (2011). Designing Pareto-optimal selection systems: Formalizing the decisions required for selection system development. *Journal of Applied Psychology, 96,* 907-926. https://d0i.0rg/l 01037/a0023298

Delery, J., & Gupta, N. (2016). Human resource management practices and organizational effectiveness: Internal fit matters. *Journal of Organizational Effectiveness: People and Performance, 3(2),* 139-163. https://d0i.0rg/l 0.1108/JOEPP-03-2016-0028

Fulmer, I. S., & Ployhart, R. E. (2014). "Our most important asset": A multidisciplinary/multilevel review of human capital valuation for research and practice. *Journal of Management, 40(1),* 161-192. https://doi.org/10.1177/0149206313511271

Giberson, T. R., Resick, C. J., & Dickson, M. W. (2005). Embedding leader characteristics: An examination of homogeneity of personality and values in organizations. *Journal of Applied Psychology,* 90(5), 1002-1010. https://d0i.0rg/l 0.1037/0021 -9010.90.5.1002

Godard, J., & Delaney, J. T. (2000). Reflections on the "high performance" paradigm's implications for industrial relations as a field. *ILR Review, 53* (3), 482-502. https://d0i.0rg/l 0.1177/001979390005300307

Guest, D. E., Michie, J., Conway, N., & Sheehan, M. (2003). Human resource management and corporate performance in the UK. *British Journal of Industrial Relations, 41(2),* 291-314. https://doi.org/10.1111/1467-8543.00273

Hall, G. S. (1917). Practical relations between psychology and the war. *Journal of Applied Psychology,* 1(1), 9-16. https://doi.org/10.1037/h0070238

Hammonds, K. (2005). Why we hate HR. *Fast Company, 97,* 41-47. https://www.fastcompany.com/53319/why-we-hate-hr

Hatch, N. W., & Dyer, J. H. (2004). Human capital and learning as a source of sustainable competitive advantage. *Strategic Management Journal, 25* (12), 1155-1178. https://doi.org/10.1002/smj.421

Hollenbeck, J. R. (2008). The role of editing in knowledge development: Consensus shifting and consensus creation. In Y. Baruch, A. M. Konrad, H. Aguinis, & W. H. Starbuck (Eds.), *Journal editing: Opening the Black box* (pp. 16-26). Jossey Bass.

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology: Employment interview validity. *International Journal of Selection and Assessment, 22(3),* 297-309. https://d0i.0rg/l 0.1111 /ijsa.12078

Hunter, J. E. (1980). *Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization to the General aptitude test battery (GATE). Washington.* Department of Labor, Employment Service.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin,* 96(1), 72-98. https://doi.Org/10.1037/0033-2909.96.1.72

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal, 38(3),* 635-672. https://d0i.0rg/l 0.5465/256741

Huselid, M. A., & Becker, B. E. (1996). Methodological issues in cross-sectional and panel estimates of the human resource-firm performance link. *Industrial Relations: A Journal of Economy and Society, 35(3),* 400-422. https://doi.Org/10.1111/j.1468-232X.1996.tb00413.x

Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms. *Academy of Management Journal, 55* (6), 1264-1294. https://doi.org/10.5465/amj.2011.0088

Jordan, M., Herriot, P., & Chalmers, C. (1991). Testing Schneider's ASA theory. *Applied Psychology, 40(1),* 47-53. https://d0i.0rg/l 0.1111 /j. 1464-0597.1991.tbOl 357.x

Katou, A. A., & Budhwar, P. S. (2010). Causal relationship between HRM policies and organisational performance: Evidence from the Greek manufacturing sector. *European Management Journal,* 28(1), 25-39. https://doi.org/10.1016/j.emj.2009.06.001

Kaufman, B. E. (2014). The historical development of American HRM broadly viewed. *Human Resource Management Review, 24(3),* 196-218. https://doi.org/10.1016/j.hrmr.2O14.03.003

Kim, Y, & Ployhart, R. E. (2014). The effects of staffing and training on firm productivity and profit growth before, during, and after the Great Recession. *Journal of Applied Psychology, 99,* 361-389. https://doi.org/10.1037/a0035408

Koch, M. J., & McGrath, R. G. (1996). Improving labor productivity: Human resource management policies do matter. *Strategic Management Journal,* 17(5), 335-354. https://doi.Org/10.1002/(SICI)1 097-0266 (199605) 17:5<335::AID-SMJ814>3.0.CO;2-R

Kuncel, N. R., & Rigdon, J. 2012. Communicating research findings. In *Handbook of psychology* (2nd ed., Vol. 12, pp. 43-58). Wiley.

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science.

Landy, F. J. (2007). The validation of personnel decisions in the twenty- first century: Back to the future. In S. M. McPhail (Ed.), *Alternate validation strategies: Developing and leveraging existing validity evidence* (pp. 409-426). Jossey-Bass.

Latham, G. P., & Whyte, G. (1994). The futility of utility analysis. *Personnel Psychology, 47(1),* 31-46. https://doi.Org/10.1111/j.1744-6570.1994. tb02408.x

Lepak, D. P., Jiang, K., Han, K., & Castellano, W. G. 2012. Strategic HRM moving forward: What can we learn from micro perspectives? In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology* (Vol. 27, pp. 231-259). John Wiley & Sons Ltd.

Lepak, D. P., Liao, H., Chung, Y, & Harden, E. E. (2006). A conceptual review of human resource management systems in strategic human resource management research. In J. J. Martocchio (Ed.), *Research in personnel and human resources management* (pp. 217-271). Emerald Group Publishing Limited.

Li, X., Qin, X., Jiang, K., Zhang, S., & Gao, F.-Y. (2015). Human resource practices and firm performance in China: The moderating roles of regional human capital quality and firm innovation strategy. *Management and Organization Review, 11(2),* 237-261. https://doi.org/10.1017/mor. 2015.12

Lievens, F. (2013). Adjusting medical admission: Assessing interpersonal skills via situational judgment tests. *Medical Education, 47(2),* 182-189. https://doi.0rg/IO.I 111/medu.12089

Lievens, F., Buyse, T, & Sackett, P. R. (2005a). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology,* 58(4), 981-1007. https://doi.Org/10.1111/j.1744-6570.2005. 00713.x

Lievens, F., Buyse, T„ & Sackett, P. R. (2005b). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90(3), 442-452.* https://doi.org/ 10.1037/0021-9010.90.3.442

Lievens, F., Buyse, T., Sackett, P. R., & Connelly, B. S. (2012). The effects of coaching on situational judgment tests in high-stakes selection. *International Journal of Selection and Assessment, 20(3),* 272-282. https://doi.Org/10.1111/j.I 468-2389.2012.00599.x

Lievens, F., De Cotte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management, 41(6),* 1604-1627. https:// doi.org/10.1177/0149206312463941

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology,* 91(5), 1181-1188. https://doi.org/10.1037/0021-9010.91.5.1181

Lievens, F., & Sackett, P. R. (2007). Situational judgment tests in high stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology,* 92(4), 1043-1055. https://doi.org/10.1037/0021-9010.92.4.1043

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology,* 97(2), 460-468. https:// doi.org/10.1037/a0025741

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102(1),* 43-66. https://doi. org/10.1037/apl0000160

Lievens, F., Sackett, P. R., Dahlke, J. A., Oostrom, J. K., & De Soete, B. (2019). Constructed response formats and their effects on minority-majority differences and validity. *Journal of Applied Psychology, 104(5),* 715-726. https://doi.org/! 0.1037/apl0000367

Lievens, F., Sackett, P. R., & Sackett, P. R. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *Journal of Applied Psychology, 94(4),* 1095-1101. https://doi.org/10.1037/a0014628

Macduffie, J. P. (1995). human resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry. *Industrial & Labor Relations Review, 48(2),* 197-221. https://d0i.0rg/I 0.1177/001979399504800201

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology,* 60(1), 63-91. https://doi.org/10. 1111/j.I 744-6570.2007.00065.X

McDaniel, M. A., Psotka, J., Legree, P. J., Yost, A. P., & Weekley, J. A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology, 96(2),* 327-336. https://d0i.0rg/I 0.1037/a0021983

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology, 41(2),* 283-309. https://doi.org/10. 1111/j.I 744-6570.1988.tb02386.x

Messersmith, J. G., Patel, P. C., Lepak, D. P„ & Gould-Williams, J. S. (2011). Unlocking the black box: Exploring the link between high-performance work systems and performance. *Journal of Applied Psychology,* 96(6), 1105-1118. https://doi.org/10.1037/a0024710

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist,* 56(2), 128-165. https://doi.org/10.1037/0003-066X.56.2.128

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75(6),* 640-647. https://doi.Org/10.1037/0021-9010.75.6.640

Mount, M. K., & Barrick, M. R. (1995). The big five personality dimensions: Implications for research and practice in human resources management. In G. R. Ferris & K. M. Rowland (Eds.), *Research in personnel and human resources management* (Vol. 13, pp. 153-200). Greenwich, CT: JAI Press.

Muchinsky, P. M. (2004). When the psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples, and recommendations. *Personnel Psychology,* 57(1), 175-209. https://d0i.0rg/I 0.1111/j.I744-6570.2004.tb02488.x

Naylor, J. C., & Shine, L. C. (1965). A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology, 3(2),* 33-42.

Neter, E., & Ben-Shakhar, G. (1989). The predictive validity of graphological inferences: A meta-analytic approach. *Personality and Individual Differences, 10(7),* 737-745. https://d0i.0rg/I 0.1016/0191 -8869(89)90120-7

Ng, T. W. H., & Feldman, D. (2009). How broadly does education contribute to job performance. *Personnel Psychology, 62(1),* 89-134. https://doi.org/ 10.1111/j.I 744-6570.2008.01130.x

Nguyen, N. T„ Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment, 13(4),* 250-260. https://doi.0rg/IO.llll/j. 1468-2389.2005.00322.X

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *Journal of Vocational Behavior, 98,*138-151. https://d0i.0rg/I 0.1016/j.jvb.2O16.11.002

Oh, I.-S, Kim, S., & Van Iddekinge, C. H. (2015). Taking it to another level: Do personality-based human capital resources matter to firm performance? *Journal of Applied Psychology, 100(3),* 935-947. https://doi.org/10.1037/ a0039052

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78(4),* 679-703. https://doi.org/10.1037/0021-9010.78.4.679

Ployhart, R. E. (2012a). Multilevel selection and the paradox of sustained competitive advantage. In *Oxford library of psychology.* In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection.* Oxford University Press.

Ployhart, R. E. (2012b). The psychology of competitive advantage: An adjacent possibility. *Industrial and Organizational Psychology,* 5(1), 62-81. https://doi.Org/10.1111 /j.1754-9434.2011.01407.x

Ployhart, R. E., & Moliterno, T. P. (2011). Emergence of the human capital resource: A multilevel model. *Academy of Management Review,* 36(1), 127-150. https://d0i.0rg/I 0.5465/amr.2009.0318

Ployhart, R. E., Schmitt, N., & Tippins, N. T. (2017). Solving the supreme problem: 100 years of selection and recruitment at the journal of applied psychology. *The Journal of Applied Psychology, 102(3),* 291-304. https://doi.org/10.1037/apl0000081

Ployhart, R. E., Van Iddekinge, C. H„ & Mackenzie Jr, W. I. (2011). Acquiring and developing human capital in service contexts: The interconnectedness of human capital resources. *Academy of Management Journal,* 54(2), 353-368. https://doi.org/10.5465/amj.2011.60263097

Ployhart, R. E., Weekley, J. A., & Baughman, K. (2006). The structure and function of human capital emergence: A multilevel examination of the attraction-selection-attrition model. *Academy of Management Journal,* 49 (4), 661-677. https://doi.org/10.5465/amj.2006.22083023

Ployhart, R. E., Weekley, J. A., & Ramsey, J. (2009). The consequences of human resource stocks and flows: A longitudinal examination of unit service orientation and unit effectiveness. *Academy of Management Journal,* 52(5), 996-1015. https://doi.org/10.5465/amj.2009.44635041

Pulakos, E. D. (2005). *Selection assessment methods: A guide to implementing formal assessments to build a high-quality workforce.* Society for Human Resource Management.

Putka, D. J., Beatty, A. S., & Reeder, M. C. (2018). Modern prediction methods: New perspectives on a common problem. *Organizational Research Methods, 21(3),* 689-732. https://d0i.0rg/I 0.1177/1094428117697041

Rabi, T., Jayasinghe, M., Gerhart, B., & Kühlmann, T. M. (2014). A meta-analysis of country differences in the high-performance work system-business performance relationship: The roles of national culture and managerial discretion. *Journal of Applied Psychology,* 99(6), 1011-1041. https://d0i.0rg/I 0.1037/a0037712

Renkema, M., Meijerink, J., & Bondarouk, T. (2016). Advancing multilevel thinking and methods in HRM research. *Journal of Organizational Effectiveness: People and Performance,* 3(2), 204-218. https://doi.org/10.1108/JOEPP-03-2016-0027

Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology,* 85(6), 880-887. https://d0i.0rg/I 0.1037/0021-9010.85.6.880

Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58(4),* 1009-1037. https://doi.Org/10.1111/j.1744-6570.2005.00714.x

Rothstein, H. R., Schimdt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology,* 75(2), 175-184. https://d0i.0rg/I 0.1037/0021-9010.75.2.175

Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology, 65,* 693-717. https://doi.org/10.1146/annurev-psych-010213-115134

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41(2),* 149-174. https://d0i.0rg/I 0.1002/hrm.I 0029

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology,* 59(1), 419-450. https://doi.org/10.1146/annurev.psych.59.103006.093716

Sackett, P. R., Lievens, F., Berry, C. M., & Landers, R. N. (2007). A cautionary note on the effects of range restriction on predictor intercorrelations. *Journal of Applied Psychology, 92(2),* 538-544. https://doi.org/10.1037/0021-9010.92.2.538

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: 100 years of research in the *journal of applied psychology. Journal of Applied Psychology, 102(3),* 254-273. https://doi.org/10.1037/apl0000151

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology, 104(10),* 1207-1225. https://doi.org/10.1037/apl0000405

Saridakis, G., Lai, Y., & Cooper, C. L. (2017). Exploring the relationship between HRM and firm performance: A meta-analysis of longitudinal studies. *Human Resource Management Review, 27(1),* 87-96. https://doi.org/10.1016/j.h rm r.2016.09.005

Schaubroeck, J., Ganster, D. C., & Jones, J. R. (1998). Organization and occupation influences in the attraction-selection-attrition process. *Journal of Applied Psychology,* 83(6), 869-891. https://doi.org/10.1037/0021-9010.83.6.869

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62* (5), 529-540. https://doi.Org/10.1037/0021-9010.62.5.529

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(2),* 262-274. https://d0i.0rg/I 0.1037/0033-2909.124.2.262

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior,* 1(1), 45-65. https://doi.org/10.1146/annurev-orgpsych-031413-091255

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40 (3),* 437-453. https://doi.Org/10.1111/j.1744-6570.1987.tb00609.x

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2012). A funny thing happened on the way to the future: The focus on organizational competitive advantage lost out. *Industrial and Organizational Psychology, 5(1),* 96-101. https://d0i.0rg/I 0.1111 /j.1754-9434.2011.01411 .X

Schneider, B., Smith, D. B., & Sipe, W. P. (2000). *Personnel selection psychology: Multilevel considerations.*

Schneider, B., Smith, D. B., Taylor, S., & Fleenor, J. (1998). Personality and organizations: A test of the homogeneity of personality hypothesis. *Journal of Applied Psychology, 83(3),* 462-470. https://doi.org/10.1037/0021-9010.83.3.462

Shaffer, *J. A.,* & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65 (3),* 445-494. https://d0i.0rg/I 0.1111 /j.1744-6570.2012.01250.x

Singh, K. (2004). Impact of HR practices on perceived firm performance in India. *Asia Pacific Journal of Human Resources, 42(3),* 301 -317. https://doi.org/10.1177/1038411104048170

Skaggs, B. C., & Youndt, M. (2004). Strategic positioning, human capital, and performance in service organizations: A customer interaction approach. *Strategic Management Journal,* 25(1), 85-99. https://doi.org/10.1002/smj.365

Slaughter, J. E., Zickar, M. J., Highhouse, S., & Mohr, D. C. (2004). Personality trait inferences about organizations: Development of a measure and assessment of construct validity. *Journal of Applied Psychology,* 89(1), 85-103. https://d0i.0rg/I 0.1037/0021-9010.89.1.85

Stemig, M., Sackett, P. R., & Lievens, F. (2015). Effects of organizationally-endorsed coaching on performance and validity of situational judgment tests. *International Journal of Selection of Assessment, 23(2),* 175-182. https://doi.org/10.1111/ijsa.12105

Sturman, M. C. (2003). Searching for the inverted u-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/ performance, and age/performance relationships. *Journal of Management, 29(5),* 609-640. https://doi.org/10.1016/S0149-2063(03)00028-X

Terpstra, D. E., & Rozell, E. J. (1993). The relationship of staffing practices to organizational level measures of performance. *Personnel Psychology, 46* (1), 27-48. https://doi.Org/10.1111/j.1744-6570.1993.tb00866.x

Van Iddekinge, C., Arnold, J. D., Frieder, R. E., & Roth, P. L. (2019). It's required, but is it job-related? A meta-analysis of the validity of prior work experience. *Academy of Management Proceedings, 72(1),* 571-598. https://d0i.0rg/I 0.5465/AMBPP.2018.278

Van Iddekinge, C. H., Ferris, G. R., Perrewe, P. L., Perryman, A. A., Blass, F. R., & Heetderks, T. D. (2009). Effects of selection and training on unit-level performance over time: A latent growth modeling approach. *Journal of Applied Psychology, 94(4),* 829-843. https://doi.org/10.1037/a0014453

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., &Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology,* 97(3), 499-530. https://doi.org/10.1037/a0021196

Vlachos, I. (2008). The effect of human resource practices on organizational performance: Evidence from Greece. *The International Journal of Human Resource Management,* 7 9(1), 74-97. https://doi.org/10.1080/09585190701763933

Wall, T. D., & Wood, S. J. (2005). The romance of human resource management and business performance, and the case for big science. *Human Relations, 58(4),* 429-462. https://doi.org/! 0.1177/0018726705055032

Wright, P. M., Gardner, T. M., & Moynihan, L. M. (2003). The impact of HR practices on the performance of business units. *Human Resource Management Journal, 13(3},* 21-36. https://doi.Org/10.1111/j.1748-8583.2003.tb00096.x

Wright, P. M., Gardner, T. M., Moynihan, L. M., & Allen, M. R. (2005). The relationship between HR practices and firm performance: Examining causal order. *Personnel Psychology, 58(2},* 409-446. https://doi.org/10.1111/j.l 744-6570.2005.00487.X

Zacharatos, A., Barling, J., & Iverson, R. D. (2005). High-performance work systems and occupational safety. *Journal of Applied Psychology,* 90(1), 77-93. https://doi.org/! 0.1037/0021-9010.90.1.77