

Measuring the organisational impact of training: The need for greater methodological rigor

Keywords: training and organizational performance, methodological rigor, validity

Abstract

We review the methodological rigor of empirical quantitative studies that have investigated the training and organisational performance relationship. Through a content analysis of 219 studies published in quality journals, we reveal significant validity threats (internal, external construct and statistical conclusion validity) that raise questions about the methodological rigor of the field. Our findings suggest that the time is appropriate for a renewed methodological endeavour to understanding the relationship between training and organisational performance. We make specific recommendations to enhance methodological rigor and generate research finding will enhance operationalisation of theory, help researchers to make inferences about causality and inform the decision making of HRD practitioners.

INTRODUCTION

In this paper, we review 40 years of quantitative empirical studies that have investigated the training-organisational performance relationship to identify the methodological features of these studies and the extent to which they are subject to validity threats. Training is an important construct in the HRD and learning and development (L&D) disciplines (Bell, Tannenbaum, Ford, Noe, & Kraiger, 2017) and numerous industry-based reports document the considerable investment made by organizations in employee training and development (e.g. Bersin by Deloitte, 2016). In addition, scholars have argued that training enhances organisational performance including productivity, innovation, customer service quality and financial performance (Aguinis & Kraiger, 2009; Kim & Ployhart, 2014; Noe, Clark & Klein 2014) yet the evidence base to make these claims is based on a preponderance of cross-sectional research designs that shed little light on causality. Since 1979 when Miron published the first study on this relationship and McClelland, (1979) the past four decades has witnessed a sustained increase in empirical studies investigate the training-organisational performance relationship with major growth in published studies since 2010. The extensiveness of past research highlights the importance of training in organisations and the need for researchers to provide practitioners with robust findings on the strength of the relationship, the linking mechanisms and the boundary conditions explaining the relationship.

While there are many published reviews and syntheses on the topic of training in organisations (eg. Bell et al, 2017; Noe et al, 2014; Salas, Tannenbaum, Kraiger & Smith-Jentsch, 2012) these reviews have primarily focused on identifying and reporting key themes and knowledge accumulation on training to date. However, existing reviews seldom engage with the methodological features of studies on training-organisational performance relationship and the rigor with which research is undertaken. In contrast to prior reviews, our primary aim in this study

is to evaluate the methodological characteristics of existing research investigating training and its organisational performance outcomes and specifically to identify the threats to validity that exist in these studies. Given that the training-organisational performance relationship is extensively studied and is central to the arguments that HRD and L&D specialists make to justify investment in training major question arises as to the quality of the evidence available on this relationship to date.

Three sets of reasons arise that a need to focus on methodological rigor. First, from the perspective of theory, scholars to date have not always used research designs that reflect the key assumptions of the theories they use to study the relationship. For example, many studies make use of human capital theory (Riley et al 2017; Becker, 1964) and the resource based view (Barney 1991) however, these theories envisage a long-term contribution of investment in human resources to organisational performance. Yet the majority of studies use cross-sectional designs and post-predictive designs (i.e. where respondents provide information on both assessments of current training and their firm's performance at the same time) and therefore do not provide a robust testing of the propositions of the theories used. Wright, Gardner, Moynihan and Allen (2005) describes these designs as post-predictive because they are actually predicting past performance or performance up to the point of the survey. Similar arguments are made for studies that utilise social exchange theory (Blau, 1964) and behavioural theories (Jackson & Schuler, 1975). Therefore, it is reasonable to conclude that existing studies do not provide a robust operationalisation of the theoretical foundations of these studies.

Second, from an empirical perspective two important issues arise. First, there is the problem of contextual validity. The majority of studies have been conducted in an Anglo-American context (USA, UK, Canada, Australia and New Zealand) therefore our current

understanding of the relationship may not be completely valid given the emergence of Asia-Pacific, Middle Eastern and African economies. In addition, the majority of studies focus on professional full time employees yet the world of employment has changed significantly with the emergence of international workers and the gig economy. This suggests that the context of the training-organisational performance relationship has changed in significant ways thus suggesting a need to understand the complexities of the relationship. A second empirical reason for analysing the way in which the training-organisational performance relationship has been investigated concerns the issue of establishing causality. This represents the empirical gold standard of science however many existing studies make use of research designs (typically surveys) that do not enable inferences to be made about causality. Wright et al (2005) highlights that survey designs can never ultimately 'prove' cause and many of what are considered well designed studies have paid little attention to temporal precedence and /or alternative explanations for the relationship. This issue has also received prominence in the HRD and training literature. For example, both Sitzmann and Weinhardt (2018) and Bainbridge, Sanders, Cugin and Lin (2017) have drawn attention to the needs for greater methodological rigor in understanding how training and other HRM practices contribute to organisational performance. In the HRD context, Brown and Latham (2018) highlighted the need for both rigor and relevance in HRD research.

Third, from managerial and HRD practice perspectives it is important to generate valid insights and robust research findings concerning the strength and direction of the relationship between training and organisational performance. Given that the field of HRD focuses on the investigation of learning and development processes in workplace settings it is important that research findings within the field should inform practice in these settings. Thus an important motivation for this study speaks to recent debates concerning the role of research in generating

evidence that is of value in the real world (Brown, & Latham, 2018; Gubbins, Harney, Van der Werff & Rousseau, 2018). The import of this discussion suggests that academic HRD research is moving further away from addressing ‘real world ‘problems that have interest and relevance to practitioners. For research to be relevant to practitioners, it must also be rigorously conducted. Paterson, Harms and Tuggle (2018) proposed that greater methodological rigor should lead to greater relevance to practitioners. Aguinis et al. (2010) highlighted the concept of customer-centric science and emphasize that careful and rigorous reporting of research results should serve the needs of both academics and practitioners. HRD and L&D scholars are positioned at the theory-practice interface. On the one hand they generate evidence that can be used by practitioners to make a case for investment in training (Rousseau & Barends, 2011) and on the other hand, they are concerned to develop a body of knowledge that is robust and answers key theoretical and empirical questions concerning the training-organisational performance relationship (Tharenou, Saks & Moore, 2007).

Our overarching goal in this paper is therefore to review prior research on the training-organisational performance relationship to illuminate the extent of validity problem in existing studies and to use the outputs of our analysis to make methodological suggestions to address identified validity threats in future research. In doing so we seek to enthuse scholars within HRD and L&D to conduct research that achieves the following outcomes. First scholars should conduct research that provides a strong operationalisation of the theoretical perspectives used to formulate hypotheses; second, they should provide a more fine-grained understanding of the training-organisational performance relationship and go further in answering the question of causality and third they need to generate findings that will help HRD and L& D practitioners to make evidence-based decisions about investment by organisations in training. For the purposes of this paper validity is defined as the essential trustworthiness of study findings and scholars have highlighted

four categories of validity that are central to methodological rigor (Cook & Campbell, 1976; Casper, Eby, Bordeaux, Lockwood & Lambert 2007; Brutus, Aguinis & Wassmer, 2013). *Internal validity* is concerned with the causality and accuracy of conclusions and is something that plagues a lot of research in the HRD/HRM fields in establishing a relationship between training practices and organizational performance (Tharenou et al., 2007; Bainbridge et al., 2017). *External validity* focuses on the extent to which findings on that relationship are generalizable to different locations, research settings, organizations, employee groups and across time. *Construct validity* is concerned with the types of measures that are used to operationalize both training and organizational performance and *statistical conclusion validity* focuses on the extent to which it is possible to make inferences about the training-organizational performance relationship. In quantitative investigations, these dimensions are central to the legitimacy of the field (Bacon, 2016; MacCarthy, Lewis, Voss & Narsimhan, 2013) of research findings amongst academics and the quality of evidence generated for practitioners (Gelade, 2006).

We make two contributions to the field of HRD and specifically to understanding the training-organisational performance relationship. First, we provide an original overview of existing research on the training –organisational performance relationship in that we discuss key issues related to the validity of the research base. In doing so we identify methodological issues that have received relatively less attention to date. Second, we advance understanding of the priority validity threats that future researchers should focus on in order to enhance the quality of research findings. For each area of validity, we discuss the research implications of the threats identified and make suggest on methodological approaches that will decrease or eliminate some of these threats. We structure our paper as follows. We first define the core concepts that underpin the research in this paper. Second, we describe in detail the methodology we used to conduct this

study and then present our findings. Finally, we discuss the implications of our findings for methodological rigor and suggests a number of priority recommendations to address causality, contextual validity of studies, the construct validity of the training measure and greater understanding of linking mechanisms and boundary conditions explaining the relationship.

DEFINING TRAINING AND ORGANISATIONAL PERFORMANCE.

Training. Training is defined in different ways in the literature (Dipboye, 2018; Bell et al 2017) with some definitions emphasising current knowledge, skill and ability needs and others focusing on future needs. Training however can be defined as consisting of both ‘training and development’ with the former focused on knowledge, skills and abilities (KSAs) required for the current job role and the latter focusing on KSAs required for a future role (Garavan, 1995; Kraiger, Passmore, dos Santos & Malvezzi, 2015). The future component is conceptualised as development. Training in its narrower sense is sponsored by the organisation because it is assumed to have immediate organisational benefits whereas development may be sponsored by the organisation however, it may also be initiated by employees and without recognition or awareness by the organisation. Sitzmann & Weinhardt (2018) argue that the vast majority of training in organisations focuses on what they describe as hard skills or the development of KSAs that are directly applicable to the job. Tharenou et al (2007) in their meta-analysis of training focused primarily on these hard skills components and excluded soft skill or development programmes. They defined training as “the systematic acquisition and development of the knowledge, skills and attitudes required by employees to adequately perform a job or task and to improve performance” (Tharenou et al, 2007, p6). Recent studies of the training-organisational performance relationship have included training focused on enhancing employees’ soft skills (Kim & Ployhart, 2014).

Therefore, we include in this review studies that reported findings related to training that enhances both current and future KSAs (Kim & Ployhart, 2014; Berk & Kase, 2010). This definition incorporates training that focuses on the development of generic or soft skills as well as training that take place in the classroom and on-the job (Salas et al, 2012) focused on developing hard or skills that are immediately applicable to the job. We selected studies that reported on formal training rather than informal training or training that occurs as part of day-to-day on-the-job experiences, trial an error and learning by doing (Nelson & Winter, 1982; Nikolova, Van Ruyseveldt, De Witte, & Syroit, 2014). In addition, we only included studies of training conducted in workplace settings.

Organisational Performance. Organisational performance is conceptualised as a multi-dimensional construct (Paauwe, 2004) with studies measuring it in different ways. It is the ultimate dependent variable that researchers can use to justify investment in training (Richard, Devinney, Yip, & Johnson, 2009) and includes human resource, operational and financial performance dimensions (Dyer & Reeves, 1995; Tharenou et al., 2007). However, some studies use the term ‘organizational effectiveness’ which Richard et al. (2009) conceptualize as a broader and more general construct that focuses on internal organizational performance in comparison to external organizational performance measures focused on accounting and financial metrics.

Scholars operationalize organizational performance using objective and subjective measures or a combination of both. The majority of studies utilize subjective measures including in some cases a composite index or as a single organizational performance item. We define organizational performance to include the three categories proposed by Tharenou et al. (2007): HR-related, operational and financial. We define human resource outcomes as proximal outcomes such as collective KSAs, motivation, employee turnover, job satisfaction and organizational commitment

(Dyer & Reeves, 1995). We define operational outcomes as distal outcomes comprising labour productivity, innovation, customer service and customer retention (Jiang, Wang & Zhao 2012; Rauch & Hatak, 2016). Finally, we define financial outcomes to comprise three categories: (a) financial performance, (b) product market performance and (c) shareholder return (Richard, Devinney, Yip, & Johnson, 2009). The financial performance category comprises measures of profit, return on assets and return on investment. Product market performance comprises measures such as sales and market share and shareholder return includes measures such as total shareholder returns and economic value added. We acknowledge the different approaches taken by scholars concerning this categorization. Rauch and Hatak (2016) for example, did not include HR outcomes as organizational performance outcomes; however, Jiang et al. (2012) in their meta-analysis included HR outcomes in their definition of organizational performance.

METHOD

Sample and Procedure

We draw on studies published in quality training, HRD, organizational behaviour, industrial/organizational psychology and HRM journals. We examined studies published between 1979-2018 to assess the field and we confined our analysis to articles published in quality journals and specialist journals in the training and HRD fields. We defined a quality journal as those rated 1-4 stars in the Academic Journal Guide, Chartered Association of Business Studies, UK listing (2018). This is an authoritative listing of journal quality. Our starting point for the review was 1979. We utilised this starting point because Tharenou et al (2007) in the one meta-analysis published to date on the training-organisational performance relationship identified that year as the starting point for they meta-analysis. We checked to ascertain whether ant earlier studies have been published given that the criteria for inclusion in a meta-analysis are more restrictive than is

the case for a methodological review of the We searched Business Source Premier, Social Citation Index and Google Scholar using the following terms: ‘training and individual outcomes’, ‘training and organizational outcomes’ or variants ‘training and HR outcomes’, ‘training and organizational performance outcomes’, ‘training and organizational effectiveness outcomes’ and ‘training and financial outcomes’ to identify relevant articles. We used Google Scholar to search for the most cited articles. We also conducted manual searches of journals that typically publish empirical investigations on the training –organisational performance relationship to ensure that we had captured the relevant articles. Our initial search led to 2455 articles. To be included in the review, each article was analysed using three criteria. First, we only included articles that reported empirical findings. We, therefore, excluded papers that were theoretical, conceptual or literature reviews. This reduced our sample of studies to 1105 papers. Second, we only included studies conducted in workplace settings and this further reduced our sample to 756 papers. Third, each study needed to investigate the effects of training on one or more of the three categories of outcome specified by Tharenou et al (2007), human resource, organisational and financial and to use quantitative methods. This reduced our sample of papers to 219. We reviewed the title, abstract and content of each study against these criteria to determine suitability for inclusion in this review. Our final sample of studies were published in 36 journals of which the following are examples; *Journal of Organizational Behavior*, *Personnel Psychology*, *The International Journal of Human Resource Management*, *Human Resource Management*, *Human Resource Management Journal*, *Human Resource Development International* and *Human Resource Development Quarterly*.

Coding Process

To investigate the four categories of validity we utilised content analysis (Krippendorff, 2013; Hoobler & Johnson, 2004). Content analysis helps researchers to identify and elaborate on

different validity characteristics (Durla, Rigor & Pfarrer, 2007). We followed the hierarchically system of codes proposed by Aguinis, Pierce, Bosco and Muslin (2009) to identify the dimensions to be included in each category of validity.

Internal validity. We assessed three dimensions of internal validity: a) the structure of the data (cross-sectional or longitudinal), b) the research designs used to investigate the training-organisational performance relationship: post-predictive (the measurement of training after the performance period), retrospective (where respondents are asked to recall training practices that existed prior to performance period), contemporaneous (the gathering of concurrent data on training and organisational performance), predictive (the gathering of data on training at one point in time that is related to subsequent organisational performance) or multiple research designs and c) the types of relationship investigated (direct, mediated, moderated, moderated mediation).

External validity. We assessed seven dimension of external validity: a) level of analysis of organisational performance (firm, establishment, business unit, multilevel); b) sample location (North America, Europe, Asia, Africa, Australia/New Zealand, not specified); c) industry (single industry, multi-industry, not specified), d) sector (private, public, both, not specified); e) organization size (specified, not specified) f) firm/workplace/business unit characteristics (past performance, geographic location, industry or sector, size, age, ownership, competition, number of hierarchical levels, export orientation, diversification, innovation, HR strategy, asset/investment /capital, single or multiple establishment, employee groups, business status, restructuring, level of unionization) and g) subject-level characteristics (gender, job tenure, education, contract type, working hours, wage levels, age, occupation, race, number of dependents, marital status).

Construct validity. We assessed the construct validity of both the predictor and dependent variables.

Training. We coded for eight dimensions of the predictor or independent variable: a) operationalization of the training construct: absolute (the amount of training employees received), proportional, (the percentage of workers within an organisation trained). content, (the type of training provided); emphasise (the perceived importance of the training provided by the organisation) and effectiveness, (the perceived effectiveness of the training provided) or the use of combined measures) training measurement development (existing measure without adaptation, existing measure with adaptation, idiosyncratic (one specifically developed for use in the study) , single item measure, multiple item measure, binary measure); c) type of training measure (subjective measures only, objective measures only, subjective and objective measures); d) number of informants for training measure (single informant, multiple informant, not specified); e) measurement: reliability evidence for training measure (alpha, inter-rater, test-retest) ,f) measurement: validity evidence of training measure (any content validity evidence, any construct validity evidence, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), discriminant validity, convergent validity); g) procedural remedies to reduce common method variance (CMV) (where data for both the predictor and dependent variable are obtained from the same person in the same measurement context using the same item context) for the training measure (used, not used); and h) statistical methods used for CMV for the training measure (used, not used).

Organisational Performance. We coded for eight dimensions of the dependent variable: a) the type of organisational performance measure used (subjective measure only, objective measure only, combined measures); b) measurement development of the organisational performance variable (existing measures used without adaptation, existing measures used with adaptation, idiosyncratic, single item measure, multiple item measure); c) organisational

performance domain measured (human resource, organisational performance financial outcomes , multiple organisational performance outcomes); d) source of organisational performance measures (same source as training measure , multiple sources, not specified); e) measurement: reliability evidence of organisational performance measures (alpha, inter-rater, test-retest); f) measurement: validity evidence of organisational performance measures (any content validity evidence reported, any construct validity evidence reported, EFA, CFA, discriminant validity, convergent validity); g) procedural remedies to reduce common method variance (CMV) for the organisational performance variable (used, not used); and h) statistical methods used for CMV for the organizational performance variable (used, not used).

Statistical conclusion validity. We coded for nine dimensions of statistical conclusion validity: a) simple inferential statistics (correlation, t-test, chi-square); b) analysis of statistical relationships (multiple regression, ANOVA and ANCOVA, logistic regression, MANOVA and MANCOVA, canonical correlation, HLM, panel analysis, SEM and path analysis; c) tests for mediation (Baron and Kenny and alternative models); d) tests for moderation (MMR); e) reporting of effect sizes and the magnitude of effect sizes ; f) the reporting of statistical assumption (randomization, independence, measurement level of variable, normality, linearity and variance); g) statistical software used to assess relationships (SPSS, Amos, M plus, LISREL, Stata, not specified); h) response rate reported (yes, no); and i) sample size (mean).

Inter-rater reliability and validity. Three of the paper authors were provided with a detailed coding taxonomy developed by the first two authors accompanied by an explanation of each category of validity. Each coder independently coded the data utilizing these coding categories. Our approach is similar to that used by Casper et al. (2007), Hiller, De Church, Murase, and Doly (2011) and Bainbridge et al. (2017). First, the three coders independently coded an initial

sample (25) of studies to check for the reliability of coding. Second, we computed the reliability of our coding, made appropriate adjustments and tightened up where necessary the coding taxonomy. The key challenges we encountered related to the categorisation of the training and the organisational performance variables, the categorisation of the research design and the identification of the statically assumptions reported in the paper. Third, following the issuing of new instructions to each coder, we asked a fourth author to code the first set of 25 papers. The first three coders met with the fourth coder to compare coding decisions. We discussed areas of disagreement, explored alternative classification possibilities and when we reached agreement, we adjusted the coding taxonomy. The adjustments primarily related to clearly defining the emphasis and effectiveness training variables and broadening our definition of organisational performance to include customer related outcomes. Where coders had made identical classifications, these consensus codes were recorded in the taxonomy. Each coder then proceeded to code the full set of studies. We calculated agreement between coders for the final coding process using a Cohen's kappa level of .70 (Brutus et al, 2013). We found the following: Cohen's kappa for each of the four categories in the taxonomy internal validity (0.90), external validity (0.87), construct validity (0.77) and statistical conclusion validity (0.87).

FINDINGS

Internal Validity

The key trends that emerge from the analysis on internal validity are summarized as follows.

Use of cross-sectional designs. Ninety-one percent of studies used a cross-sectional research design. Cross-sectional designs do contribute to the literature where they used in the initial phase of investigating novel research questions and potential moderator and mediator hypotheses

not previously tested in the literature. They are also useful to help researchers develop new scales and represent a cost-effective way of demonstrating that two or more variables are related to each other. However, cross-sectional designs have limitations in terms of establishing causation, which as we pointed out earlier represents the gold standard in terms of research designs. Researchers have expressed concerns about the value of cross sectional designs to address the fundamental question that underpins organizational investment in training, which is whether training makes a difference to the bottom line. Cross-sectional designs are particularly ineffective when measuring organizational and financial performance outcomes as these types of outcomes require significant time lags to be realized. Only 9% of studies use a longitudinal research design and they typically measured the training construct at one point in time and used this measure to predict subsequent performance while also controlling for prior or concurrent performance. We encountered significant difficulties in studies in making judgments about the type of research design used. For example, studies were frequently not precise in describing the timing of training implementation and subsequent measures of performance were taken. Studies varied considerably in the time lag between training and organizational performance. The average time span between the measurement of the training construct and performance was 4.66 years. The longest time was 14 years and the shortest was 0.5 years. Examples of longitudinal research studies include Kim and Ployhart, (2015) and Choi and Yoon (2015). The use of longitudinal designs can help researchers can show that changes in training are associated with subsequent changes in organisational performance. This type of design allows a causal type of interpretation to be drawn however unless they are an experimental design the inferences that can be drawn about causality are limited. The limited use of longitudinal designs and the lack of use of experimental designs is a significant limitation of current training-organisational performance research.

Use of post-predictive designs. The majority (54%) of studies utilize a post-predictive design, which involves the use of organizational performance measurements collected prior to the measurement of the training variable. Wright et al. (2005: 412) draw attention to the limitation of post-predictive studies arguing that they “measure HR practices after the performance period, resulting in actually predicting *past* performance”. Therefore, while a significant number of studies reported a positive relationship between training and outcomes, it is not possible to make claims about a causal relationship between training and organisational performance due to the over-reliance on post-predictive designs. Post-predictive research design involves a single point in time collection of both training and organisational performance data. Researchers typically asked respondents to report current training practices, but ask about organizational outcomes up to the point of measurement of the training variable. Examples include Ahmad and Schroeder (2003), Gurbuz and Mert (2011) and Fletcher (2016). A small number of studies use survey methods to gather data on training and archival data to measure outcomes related to past performance (e.g. Beugelsdijk, 2008; Chen & Huang, 2009). This latter type of study, while interesting, falls into the post-predictive category because the measures of outcomes occurred prior to the measurement of the training variable.

A small number of studies that use ‘retrospective’ designs (5%). These involve asking participants to recall training programs that were in existence prior to the performance period. Examples of studies that use these types of design are Kampkotter and Marggraf (2015) and Zwick (2006). Retrospective research designs are subject to inaccuracy of recall (Wright et al., 2005) and make it difficult to draw conclusions related to causality. Contemporaneous designs (3%) involve researchers gathering data on training practices and organisational performance data using the same timeframe. Wright et al. (2005) point out that this design is problematic from a causality

perspective, because the performance data may be gathered both prior to and concurrent with the training practices measure. Predictive designs (13%) investigate whether training implemented at one point in time are related to future organizational performance. Examples of predictive designs include Barrett and O'Connell (2001) and Park and Jacobs (2011). These studies are the most robust in helping researchers to draw inferences about causality. Overall, studies reveal a positive link between training and organizational performance however, we can only draw limited conclusions about causality and for that matter reverse causality.

Investigation of direct relationships. The initial stages of the development of a research field typically focus on the measurement of a direct relationship and it progresses there is a focus on understanding indirect paths and contingencies that affect the direct relationship. The majority of studies (51%) investigated a direct relationship between training and organisational performance and researchers continue to investigate a direct relationship however the analysis indicates that researchers increasingly investigate linking mechanisms that potentially better explain the link between training and organisational performance and investigated what if or contingency type questions. Eighteen percent of the total studies included in our review studies reported partially mediated relationships, 14% reported fully mediated relationships, 13% reported moderated relationships and 4% reported moderated mediation relationships. Therefore, researchers increasingly pay more attention to understanding the processes connecting training to organisational performance and the boundary conditions that affect the generalizability of direct relationships. The investigation of moderated-mediated relationships is a relatively new statistical method and we found a number of recent studies utilised this type of analysis to understand the interaction of linking mechanisms with boundary conditions. However, the use of moderated mediation requires careful operationalisation of both the training and organisational performance

measures. We found an absence of replication type studies despite calls for this type of investigation in the HRM, international management and OB literature (Harzing, 2016).

External Validity

The following findings emerge on threats to external validity.

Level of measurement of organizational performance outcomes. We found that the bulk of studies investigated organizational performance at the firm level (74%) with 17% of studies investigated the relationship at the establishment level and 9% at the business unit level. This is an interesting finding because studies that are conducted at the firm level assumes that there is litter heterogeneity across the firm where as studies that utilise a business unit or establishment level of analysis are more likely to capture heterogeneity. This is most likely to be the case in large multi-nationals and multi-unit organisations.

An ethno-centric Anglo-American focus on sample location. We found significant bias in terms of the countries and regions in which data on training and organizational performance is collected. Studies derived samples from five regions with more than one-quarter from North America and more than one third from European countries. Twenty-seven percent of studies derived samples from Asia with the majority of these from China. We found a small number of studies that generated samples from Africa and Australia. There is a significant under-representation of samples from Eastern Europe, Latin America and the Middle East. Therefore, studies have, to date, rely on a small number of countries in which to generate samples, which is a significant threat to external validity and the potential to generalize findings across different countries, cultures and regions.

Industry sector and size of firm. The majority of studies report information on industry context. Forty-one percent of studies were undertaken using single industry samples and 52% of

studies used multi-industry samples. While multi-industry samples help researchers to enhance the generalizability of findings, single industry samples help increase measurement precision and allow researchers to capture dimensions of context more effectively. The analysis reveals that researchers have not paid attention to the reporting of firm size in empirical investigations. This is not unique to quantitative investigations with Saunders and Townsend, (2016) highlighting that it is also a problem with qualitative studies in general. Forty percent of studies did not specify the size of the organization when reporting findings or describing the methods used to conduct the study. The lack of attention to the reporting of organization sector and size is particularly problematic and studies are inconsistent in the way they report organization size: some studies report the mean; others the median; and in other studies, organization size is reported as a log in relation to assets or revenue. These deficiencies in reporting or sector, size and industry make it difficult for researchers to conduct moderated meta-analysis.

Organization, Individual and subject-level characteristics. Organization- and subject-level characteristics in published studies are not reflective of the diversity of organizations in which training is implemented and the nature of the global workforce in general. There is a major underreporting of both sets of characteristics in existing studies. We found the following trends for the reporting of organization age (20%), ownership (11%), the competitive context (6%), the organisations asset base or level of capital investment (6%) and the level of unionization (12%). There is very poor reporting of individual or subject-level characteristics. Only 11.5% of studies reported gender, 10% reported job tenure of study respondents, and 9% report education level. There is a very low level of reporting of employee age (6%), occupation (4%) and race (2%). The majority of studies do not report essential sample characteristics and therefore make it difficult to draw inferences about the generalizability of findings. Even based on the limited reporting of

organization and subject level-characteristics the samples used by studies do not reflect the diversity of organizations in which training is undertaken and the changing nature of organizations, workforces and work itself.

Construct Validity

The following findings emerge on threats to construct validity.

Operationalization and measurement of the training construct. Clearly defined operationalization of the training construct (independent variable) is a major research design issue. We found four distinct operationalisations of the training construct. Thirty-one percent of studies operationalize training as a content measure, 7% as an effectiveness measure, 7% as an absolute measure and 9% as a proportional measure. Twenty percent of studies use a combination of measures. Some of these operationalisations are complex because they involve personal judgements and respondent recall about effectiveness and are therefore potentially subject to random measurement error. Furthermore, measures that focus on effectiveness may be rated more favourably by different categories of study respondents. These errors may lead to the finding of spurious relationships between training and outcomes. Thirty percent of studies utilized idiosyncratic measures exclusively to measure the training construct, 4% used a binary measure and 13% of studies used a single item measure. Twenty-one percent of studies used an existing measure with adaptation and 13% used an existing measure without adaptation. Overall, many studies create a measure of training that is unique to the study and the use of single item measures is controversial and raises important questions about how the rigor of measurement of the predictor variable (Fuchs & Diamantopoulos, 2009)

Use of subjective measures of training and single informants. The use of subjective measures of training and single informants to measure the training construct represents a weakness

of published studies. Wall and Wood (2005) highlight the need to secure assessments from two or more persons and the use of the same raters across different organizations. This problem is compounded in multi-organization studies where researchers rely on single informants (e.g. training or HR specialists) who are expected to have knowledge of the training construct. Seventy-four percent of studies relied on a single informant to provide data on the training construct and 7% of studies used multiple informants. The majority of studies utilized a subjective measure of training (71%) with 23% of studies utilizing an objective measure such as archival data and 6% used a combination of objective and subjective measures. Researchers criticise studies that rely on single informants due to measurement error issues, low reliability and statistical inference problems (Sanders & Frenkel, 2011).

Assessment of reliability, validity and CMV of training measures. Given the use of both self-reports of training and single item measures there is a low incidence of reporting of reliability. The average α for the training measure was 0.81. A significant number of studies do not pay attention to validity issues. The same issue arises in respect of the reporting of validity evidence due to the use of single item measures of training. Twenty-eight percent of studies used EFA, 16% used CFA, and 18% report discriminant and 14% convergent validity. Forty-one percent of studies did not use procedural remedies to reduce CMV and 91% of studies did not make use of statistical remedies to address training measure CMV.

Measurement of organisational performance. Strong research design requires that measurement of organisational performance variable(s) should be from a different source than that used to measure the training construct. Furthermore, researchers highlight the value of objective measures of organisational performance (Richards et al, 2009). The measurement of organisational performance is more rigorously measured than is the case for the measurement of training.

However, research that is more recent highlights the value of subjective measures (Singh, Darwish & Potočnik 2016). Fifty-eight percent of studies measured organisational performance using a subjective measure; 32% used an objective measure and 10% used a combination of subjective and objective measures. The use of objective measures therefore helps ensure that data on organisational performance comes from a different source to that of the training measure. Sixteen percent of studies used an existing organisational performance measures without adaptation, 36% used a measure of organisational performance with adaptation, 16% used an idiosyncratic measure of organisational performance, 46% of studies used a multiple item measure of organisational performance and 26% of studies used a single item measure of organisational performance. Forty-three percent of studies used measures of organizational performance, 24% used measures of financial performance, 23% used measures of human resource outcomes and 29% of studies use multiple measures of organisational performance.

The collection of data on both training and organisational performance from the same source is problematic (Donaldson & Grant-Vallone, 2002; Podsakoff, MacKenzie, Lee & Podsakoff, 2003). The use of single source data can have the effect of both inflating and deflating correlations reported. Sixty-one percent of studies utilize the same source to measure both the training and organisational performance variables and in 18% of studies, this dimension was not specified. Therefore, measures of both training and organisational performance are subject to common method bias. These features hamper the extent to which it is possible to infer a relationship between training and outcomes and can result in correlation errors leading to spurious associations.

Given the increased use of multiple items to measure, organisational performance there is a higher incidence of reporting of reliability data (57%). The average α for measures of

organisational performance was 0.83. Studies paid less attention to providing evidence of content and construct validity of organisational performance measures. Sixteen percent of studies reported evidence of construct validity and 4% reported evidence of content validity. The reporting of EFA (18%), CFA (12%), discriminant validity (18%) and convergent validity (21%) is low considering that researchers make greater significantly greater use of multiple item measures of organisational performance. Finally, studies pay little attention to addressing common method variance in respect of organisational performance measures. Forth two percent of studies did not report procedural remedies and 91% of studies do not report statistical remedies to address CMV.

Statistical Conclusion Validity

The following findings emerge on threats to statistical conclusion validity.

Sample size and response rates. A large sample size helps researchers to minimize sampling error. It also affects the extent to which one can generalize. The mean sample size varied depending on the level of analysis of outcomes investigated. The average sample size for firm-level studies is 627 employee's workplace level was 84 employees; business unit level is 150 employees. Overall the mean sample size is effective however its adequacy depends on how respondents were selected (randomly or convenience), the study purpose and the data analysis procedures used. In reality, the resources available or the sample size in previous studies frequently determines sample size. However, a variety of data analysis packages such as Mplus (Muthen & Muthen, 2002), R (Kabacoff, 2017) and Stats (StataCorp, 2013) can be used to determine the sample size.

The response rate ranged from 22% to 53% and the average response rate is 43%. We found a lack of clarity and inconsistency in the reporting of response rates. Some studies reported response rates as a percentage of the number sent out, some as a percentage of usable responses

and others as a percentage of those sent out but not deliverable. Studies that use convenience or purposeful samples reported higher response rates than studies using random samples, which reported lower response rates.

Reporting of Effect Sizes. We investigated whether studies reported effect sizes and we analyzed the magnitude of effect sizes found in studies. Both Pek and Flora (2016) and Wilkinson (1999) highlight the importance of the reporting of effect sizes as an important feature of well-conducted research. Overall, we found that many of the earlier studies did not report effect size however, an analysis of articles from 2010 reveals that greater attention is paid to the reporting of effect sizes and the level of significance of effect sizes reported. Effect size was not reported in 48% of studies. In terms of the magnitude of effect sizes reported, we found that the majority of effect sizes reported were small. The distribution of effect sizes using Cohen's (1988) categorization was 42% small (0.20 or more), 33% medium (0.50 or more) and 5% were large (0.80 or more). Twenty percent of studies reported affect size of less than 0.20. Additional analysis of effect sizes indicates that they are significantly lower for the measurement of financial performance compared to operational performance. In addition, they are significantly higher for cross sectional rather than longitudinal studies and for studies that utilized subjective rather than objective measures of organizational performance.

Reporting of Statistical Assumptions. Nimon (2012) highlighted the importance of reporting of statistical assumptions as central to the rigor of quantitative research. We utilised the categorisation provided by Nimon (2012) to inform this analysis. Overall, we found very low levels of reporting of statistical significantly since 2010. Twenty-seven percent of studies reported on the randomization of the sample data, 14% reported on the independence of data, 26% reported on the measurement level of the training variable and 33% reported on the measurement level of

the organisational performance measure. A slightly larger percentage of studies provided comments or data demonstrating the normality of the data (34%) however only a small percentage of studies made explicit comments on the linearity of the data (14%) and the issues related to variance (including homogeneity of variance, homogeneity of regression, sphericity and homoscedasticity) (5%). We did however find that these issues were more likely to be reported in studies published in high ranked journals (4 and 4* journals in the ABS list) and in recent times the level of reporting of statistical assumptions has improved

Use of statistical analysis techniques. The majority of studies reported correlations (78%) followed by t-tests (14%) and chi-square tests (1%). To conduct analysis of statistical relationships, studies typically employed multiple regression techniques (59%), SEM and path analysis (18%), panel analysis (10%), and AVOVA and ANCOVA (10%). In the case of studies that investigated moderation, the majority use MMR; whereas for studies testing mediation, the most common method used was Baron and Kenny's (1986) approach or tests for moderation conducted using SEM. In most cases, the software used to conduct analysis is not reported the most frequently used packages were SPSS, MPlus, AMOS, and LISREL.

DISCUSSION

This research study set out to investigate the extent of methodological rigor within a very homogeneous field of investigation related to the relationship between training and organisational performance. We specifically focused on the extent to which this body of research was subject to internal, external, construct and statistical conclusion validity threats. Our area of investigation is therefore a very narrow one with clear or distinct boundaries. So what does our review tell us about the state of methodological rigor in training and organisational performance research? Five key

trends are apparent: (a) empirical research on the relationship is growing and becoming more international, (b) quantitative methods are the predominant empirical approach (c) the majority of empirical investigations draws on a very small selection of research methods, and (d) major threats to validity persist within the field. The latter problem is notable in relatively new field however, there are also debates concerning more mature field such as that reviewed here about the lack of precision of measures and methods used in empirical investigations. Rost and Ehrmann (2017) for example revealed that within the area of management research there is reporting bias towards win-win results and Chatterji, Durand, Levine and Touboul, (2016) highlighted significant validity problems with self-report data. Therefore, validity threats are not unique to the training-organisational performance field of investigation.

Overall, the field on investigation is characterised by a high degree of methodological conservatism relative to the broader area of management and psychology. Researchers continue to use the same methods that are pervasive within the field despite the significant validity threat problems related to these approaches. In addition, researchers do not often acknowledge these problems and there is a hesitancy to utilise methods that are innovative or more rigorous. These problems highlight a clear need for greater methodological rigor to be a key priority for future research. We suggest that attention to some of the validity threats identified here will help researchers address three core issues: (a) the utilisation of methods that will help researchers make inferences about the casual nature of the relationship between training and organisational performance and better operationalisation of theories used to generate hypotheses, (b) the generation of samples from unique country and institutional contexts and categories of workers that will help address external or contextual validity issues , (c) greater precision in the measurement of the predictor variable and (d) the use of more sophisticated research designs to

understand boundary conditions and micro-level mechanisms linking training to organisational performance.

The pursuit of the Gold standard: Demonstrating a Causal Relationship

To date researchers have not made sufficient use of research designs that will allow inferences to be drawn about causality. Our analysis highlighted significant threats to internal validity that undermine efforts to achieve this goal. This is however a problem that is not unique to training and HRD research with both Wright et al (2005) and Bainbridge et al (2016) highlighting that it is also a problem within strategic human resource management research. However, our analysis highlights that there is a need to utilise research methods that will generate evidence to make a better case for the impact of training on organizational performance. Therefore, there is a case to be made to make greater efforts to utilise longitudinal designs (Ployhart, Weekley & Ramsey, 2009). They provide an important opportunity but also significant challenges for training and organizational performance researchers. The challenge is to collect data on organizational performance sometime after the collection of data on training (Van de Voorde, Paauwe, & Van Veldhoven, 2010) and to collect measures of training and organizational performance at Times 1, 2 and 3. This will allow researchers to make inferences about causality and reverse causality. Training-organizational performance research will be significantly enhanced if researchers track the training investment over time and identify its impacts on organisational performance when training levels are altered or changed. The issue of temporal ordering is central to making inferences about causality therefore to do this effectively researchers need to have a minimum of three measurements of both predictor and criterion variables (Ployhart & Vandenberg, 2010). In terms of statistical conclusion validity this will require the analysis measurement invariance (Vandenberg, 2002) given that it is difficult to say whether respondents are using the

same conceptual frame of reference as they respond to the survey at multiple time periods. It is also important to acknowledge that the use of longitudinal research designs is not without difficulty. Zhu (2012) for example highlights that longitudinal research designs may suffer from omitted variable bias (Beck, 2011) and endogenous regressors (Hamilton & Nickerson, 2003) and Stritch (2017) highlights the need to investigate variation in data.

However, the use of survey methodologies will only go so far in addressing the causality issue. Experimental designs may be the only effective method in terms of eliminating other alternative explanations for the relationship between training and organisational performance. Studies that field experiments may be better suited to infer causality. Field experiments are potentially valuable in answering relevant questions about training and outcomes that may be difficult to investigate using other methods (Shadish, Cook & Campbell, 2002). They can be used to investigate the effects of multiple training conditions. For example, researchers could investigate the performance of high training versus low training business units or to investigate a strategic training investment choice and its impact on specific outcome metrics. This type of design could help researchers capture the effects of strategic training choices. Field experiments are, however, not the complete answer. They are not particularly useful when researchers wish to understand the mechanisms that explain why training impacted organisational performance. However, they are a significant step in helping researchers to explain causality. Field experiments allow researchers to gather data on outcomes as data that naturally occurs in organizations and allows the independent variable to be manipulated. This situation allows causal inferences to be drawn about the impact of training on organisational performance. Researchers point out that the implementation of field experiments is complex due to the difficulty of finding an equivalent control group. Dehejia and Wahba (2002) proposed propensity score matching (PSM) which helps

researchers to match on observed characteristics. In the case of training and organizational performance research, the matching can be on issues such as firm size, sector, and industry and technology intensity. Khandker, Koolwal and Samad (2009) proposed the double differences approach, which unlike PSM allows for selection bias on unobserved characteristics but assumes that these characteristics do not change over time.

Greater attention to External or Contextual Validity

To date research on the training-organisational performance relationship is subject to external validity threats or what Ahuja and Novelli (2017) call the problem of contextual validity. This is manifest in a situation where the majority of the research it conducted in Western or developed institutional contexts and is focused on a narrow category of workers. Therefore, much of the research suffers from a generalizability problem. Therefore, researchers need to conduct research in a broader range of countries and generate samples in underrepresented country and institutional contexts such as the Middle Eastern, Eastern Europe, African and Latin American countries. We also recommend that researchers need to generate samples in different industry and sectoral contexts and with firms across micro, SME and large organizations. For example, there is scope to generate samples in public sector and not-for-profit organizations and we need more studies within unique industry contexts. There is also a need to study the relationship with different categories of employees. Current research has a strong bias towards investigate white-collar professionals, those who hold full-time jobs and who have significant job security working in high-income countries. Bergman and Jean (2016) for example highlighted the poor representation of low to medium skilled employee, temporary workers and wage earners in industrial-organisational psychology research

Greater Precision of Measurement of the Training Variable

Our analysis highlighted significant issues related to construct validity in respect of both the predictor and criterion variables. This problem is demonstrated in respect the training measure is the over use of idiosyncratic measures, the use of single item measures and the lack of replication of measures in different studies. In the context of training, we found only five studies that used measures of training that were used in two or more previous studies. Researchers have therefore not sufficiently established the construct validity and reliability of published measures across multiple studies. What is also surprising that well established measures such as those found in Fields (2002), the developmental experiences measures developed by Wayne, Shore and Linden, (1997) and components of the learning transfer system (Bastes, Holton & Hatala, 2012) are less frequently used in studies investigating the training-organisational performance relationship. An important challenge in the context of measuring the training construct is the distinction between individual and organisational level measures of training. There is a strong trend towards the use of individual level perception measures of training related to issues such as effectiveness, importance and the content of the training with fewer studies utilising true organisational level measures of training such as the amount of training or the proportion of employees trained.

We recommend the use of archival data to enhance the construct validity of the training measure. Using archival data to measure the training construct may prove valuable because it consists of data gathered in the ordinary course of business without any involvement of a researcher (Spector, Liu & Sanchez, 2015). Organizations are likely to retain training data for compliance, regulatory and grant funding purposes. We do, however, acknowledge problems with archival data on training. SMEs and not-for-profit organizations may not gather and maintain accurate, up-to-date training records (Nolan & Garavan, 2016). Further, the training records will not have been created with the particular research question in mind. The lack of match between

the data and the question potentially present internal validity problems. The use of multiple sources for the training construct will provide researchers with better insights into the coverage of the training within an organization.

Enhanced Understanding of Boundary Conditions and Micro-Level Mechanisms Linking Training to Organisational Performance

An important feature of the growth of a field methodologically is the shift away from the investigation of direct relationships to investigation of indirect relationships and boundary conditions. We noted that the core mechanisms underlying the training-organisational performance relationship are only beginning to be researched. These linking mechanisms may relate to individual characteristics, leadership, team and organisational and external contextual processes through which training impacts organisational performance. Much of the existing research does not account for the precise mechanisms that link training to organisational performance and there is a need to jumpstart this line of research by focusing on specific micro linking mechanisms and researching organisational performance outcomes that are proximate to that mechanism and seeking out a sample where it may be found. There is a major need to utilise research designs to engage with both contingency and configurational perspectives to investigate the complexities of the training-organisational performance relationship. Scholars in HRM for example have highlighted the 'black box' problem and this is equally applicable to the training-organisational performance relationship (Messersmith, Patel, Lepak & Gould-Williams, 2011). This 'black box' is particularly acute in the context of the training-organisational performance relationship where the investigation of boundary conditions is embryonic.

CONCLUSIONS

In this paper, we have conducted a methodological review of the training-organisational performance literature to identify the extent to which it has rigor. We specifically analysed existing studies to identify threats to internal, external, construct and statistical conclusion validity. Our analysis of methodological rigor will help researchers to make decisions about research designs that more effectively operationalise theories used to investigate the training-organisational performance relationship, utilise methods that enable inferences to be made about causality and reverse causality and generate a body of research evidence that can be used by practitioners to make decisions about investment in training. We call for renewed vigour and enthusiasm for a significant shift in the way we research that relationship and we are saying that old approaches have not served us well in generating evidence that training makes a difference to organisational performance. Rather than simply continue as to fore we need to jumpstart the research area by utilising longitudinal research designs and field experiments, by paying greater attention to the generalisability of research findings by seeking out new contexts in which to conduct research, by paying greater retention to the way we measure training and finally by researching mediated and moderated relationships. We acknowledge however that our review has a number of limitations. First, we focused solely on studies published in the English language and on studies that investigated training as an independent variable. We therefore omitted studies that considered training as a moderator, mediator or dependent variable. We only included quantitative studies and therefore omitted studies that used qualitative designs. We are however confident that enhanced rigor of research on the training-organisational performance relationship will be of benefit to both practitioners and researchers.

References

- Aguinis, H., & Kraiger, K. (2009). Benefits of Training and Development for Individuals and Teams, Organizations and Society. *Annual Review of Psychology*, *60*, 451-474.
- Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. (2009). First decade of Organizational Research Methods: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods*, *12*(1), 69-112.
- Aguinis, H., Werner, S., Lanza Abbott, J., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, *13*(3), 515-539.
- 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. DOI:10.1016/S0272-6963(02)00056-6.*
- Ahuja, G., & Novelli, E. (2017). Redirecting research efforts on the diversification–performance linkage: The search for synergy. *Academy of Management Annals*, *11*(1), 342-390.
- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.
- Akrofi, S. (2016). Evaluating the effects of executive learning and development on organizational performance: Implications for developing senior manager and executive capabilities. *International Journal of Training and Development*, *20*(3), 177-199.*
- Allen, D., Hancock, J., Vardaman, J., & McKee, D. (2014). Analytical mindsets in turnover research. *Journal of Organizational Behavior*, *35*(1), 61-86.
- Aragon, I. B., & Valle, R. S. (2013). Does training managers pay off? *The International Journal of Human Resource Management*, *24*(8), 1671-1684.*

- Aragon-Sanchez, A., Barba-Aragón, I., & Sanz-Valle, R. (2003). Effects of training on business results¹. *The International Journal of Human Resource Management*, 14(6), 956-980.*
- Aragón, M. I. B., Jiménez, D. J., & Valle, R. S. (2014). Training and performance: The mediating role of organizational learning. *BRQ Business Research Quarterly*, 17(3), 161-173.
- Audea, T., Teo, S. T., & Crawford, J. (2005). HRM professionals and their perceptions of HRM and firm performance in the Philippines. *The International Journal of Human Resource Management*, 16, 532-552. *
- Bacon, D. R. (2016). Progress in the Legitimacy of Business and Management Education Research: Rejoinder to “Identifying Research Topic Development in Business and Management Education Research Using Legitimation Code Theory”. *Journal of Management Education*, 40(6), 700-704.
- Bainbridge, H. T. J., Sanders, K., Cogan, J. A. & Lin, C.-H. (2017). The Pervasiveness and Trajectory of Methodological Choices: A 20-Year Review of Human Resource Management Research. *Human Resource Management*. DOI: 10.1002/hrm.21807.
- Bal, P.M. & Dorenbosch, L. (2015). Age-related differences in the relations between individualized HRM and organizational performance: A large-scale employer survey. *Human Resource Management Journal*, 25(1), 41-61.*
- Ballot, G., FakhFakh, F., & Taymaz, E. (2001). Firms' human capital, R&D and performance: a study on French and Swedish firms. *Labour Economics*, 8(4), 443-462.*
- Ballot, G., Fakhfakh, F., & Taymaz, E. (2006). Who benefits from training and R&D, the firm or the workers?. *British Journal of Industrial Relations*, 44(3), 473-495.*

- Barling, J., Weber, T., & Kelloway, E. K. (1996). Effects of transformational leadership training on attitudinal and financial outcomes: A field experiment. *Journal of Applied Psychology, 81*(6), 827.*
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management, 17*(1), 99-120.
- Baron, R. & Kenny, D. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.
- Barrett, A., & O'Connell, P. J. (2001). Does training generally work? The returns to in-company training. *Industrial & Labor Relations Review, 54*(3), 647-662.*
- Bartel, A.P.(1994). Productivity gains from the implementation of employee training programs. *Industrial Relations, 33*, 411-425.*
- Bates, R., Holton III, E. F., & Hatala, J. P. (2012). A revised learning transfer system inventory: factorial replication and validation. *Human Resource Development International, 15*(5), 549-569.
- Beck, N. (2011). Of fixed-effects and time-invariant variables. *Political Analysis, 19*(2), 119-122.
- Becker, G. S. (1964). *Human Capital: A theoretical and empirical analysis*. Chicago: University of Chicago Press.
- Bell, B. S., Tannenbaum, S. I., Ford, J. K., Noe, R. A., & Kraiger, K. (2017). 100 Years of Training and Development Research: What We Know and Where We Should Go. *Journal of Applied Psychology*. Advance online publication.
<http://dx.doi.org/10.1037/apl0000142>.

- Bell, J. L., & Grushecky, S. T. (2006). Evaluating the effectiveness of a logger safety training program. *Journal of Safety Research*, 37(1), 53-61.*
- Berglund, H., Dimov, D., & Wennberg, K. (2018). Beyond bridging rigor and relevance: the three-body problem in entrepreneurship. DOI: <https://doi.org/10.1016/j.jbvi.2018.02.001>
- Bergman, M. E., & Jean, V. A. (2016). Where have all the “workers” gone? A critical analysis of the unrepresentativeness of our samples relative to the labor market in the industrial–organizational psychology literature. *Industrial and Organizational Psychology*, 9(1), 84-113.
- Berk, A., & Kaše, R. (2010). Establishing the value of flexibility created by training: Applying real options methodology to a single HR practice. *Organization Science*, 21(3), 765-780.
- Bersin by Deloitte (2016). UK Corporate Learning Factbook 2015: Benchmarks, Trends, and Analysis of the UK Training Market. Available at: www.bersin.com
- Beugelsdijck, S. (2008). Strategic human resource practices and product innovation. *Organization Studies*, 29(6), 821-847.*
- Birdi, K., Clegg, C. Patterson, M., Robinson, A., Stride, C.B., Wall, T.D. & Wood, S.J. (2008). The impact of human resource and operational management practices on company productivity: A longitudinal study. *Personnel Psychology*, 61, 467-501.*
- Birley, S., & Westhead, P. (1990). Growth and performance contrasts between ‘types’ of small firms. *Strategic Management Journal*, 11(7), 535-557.*
- Black, S. E., & Lynch, L. M. (1996). Human-capital investments and productivity. *The American Economic Review*, 86, 263-267.*
- Black, S. E., & Lynch, L. M. (2001). How to compete: the impact of workplace practices and information technology on productivity. *Review of Economics and Statistics*, 83(3), 434-445.*

- Blau, P. M. (1964). *Exchange and power in social life*. New York: John Wiley.
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, *36*, 1065-1105.
- Brown, T. C., & Latham, G. P. (2018). Maintaining relevance and rigor: How we bridge the practitioner–scholar divide within human resource development. *Human Resource Development Quarterly*, *29*(2), 99-105.
- Brutus, S., Aguinis, H., & Wassmer, U. (2013). Self-reported limitations and future directions in scholarly reports: Analysis and recommendations. *Journal of Management*, *39*(1), 48-75.
- Buch, R., Dysvik, A., Kuvaas, B & Nerstad, C.G.L (2015). Exploring the interplay among training intensity, job autonomy, and supervisor support in predicting knowledge sharing. *Human Resource Management*, *54*(4), 623-635.*
- Cappelli, P., & Neumark, D. (2001). Do “high-performance” work practices improve establishment-level outcomes? *Industrial & Labor Relations Review*, *54*(4), 737-775.*
- Casper, W. J., Eby, L. T., Bordeaux, C., Lockwood, A., & Lambert, D. (2007). A review of research methods in IO/OB work-family research. *Journal of Applied Psychology*, *92*(1), 28-43.
- April Chang, W. J., & Chun Huang, T. (2005). Relationship between strategic human resource management and firm performance: A contingency perspective. *International Journal of Manpower*, *26*(5), 434-449.
- Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, *37*(8), 1597-1614.

- Chen, C. & Huang, J. (2009). Strategic human resource practices and innovation performance – the mediating role of knowledge management capacity. *Journal of Business Research*, 62, 104-114.*
- Chi, N. W., Wu, C. Y., & Lin, C. Y. Y. (2008). Does training facilitate SME's performance? *The International Journal of Human Resource Management*, 19(10), 1962-1975.*
- Choi, M. & Yoon, H.J. (2015). Training Investment and Organisational Outcomes: A Moderated Mediation Model of Employee Outcomes and Strategic Orientation of the HR Function. *International Journal of Human Resource Management*, 26 (20) 2632-2651.*
- Chowhan, J. (2016). Unpacking the black box: understanding the relationship between strategy, HRM practices, innovation and organizational performance. *Human Resource Management Journal*, 26(2), 112-133.*
- Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*, 12(4), 425-434.
- Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five decade study of Academy of Management Journal. *Academy of Management Journal*, 50, 1281-1303.
- Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally.
- Croon M. A., van Veldhoven M., Peccei R., Wood S. J. (2014). Researching individual well-being and performance in context: Multilevel mediational analysis for bathtub models. In van Veldhoven M., Peccei R. (Eds.), *Well-being and performance at work: The role of context*. Hove, England: Psychology Press, 129-154.

- d'Arcimoles, C.-H. (1997). Human Resource Policies and Company Performance: A Quantitative Approach Using Longitudinal Data. *Organization Studies*, 18(5), 857–874. DOI:10.1177/017084069701800508. *
- Darwish, T. K., Singh, S., & Mohamed, A. F. (2013). The role of strategic HR practices in organisational effectiveness: an empirical investigation in the country of Jordan. *The International Journal of Human Resource Management*, 24(17), 3343–3362. DOI:10.1080/09585192.2013.775174.*
- Darwish, T.K., Singh, S. & Wood, G. (2015). The impact of human resource practices on actual, and perceived organisational performance in a Middle Eastern emerging market. *Human Resource Management*, 55(2), 261-281.*
- Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1), 151-161.
- Delaney, J. T., & Huselid, M. A. (1996). The impact of human resource management practices on perceptions of organizational performance. *Academy of Management journal*, 39(4), 949-969.*
- Delery, J. E., & Doty, D. H. (1996). Modes of theorizing in strategic human resource management: Tests of universalistic, contingency, and configurational performance predictions. *Academy of Management Journal*, 39(4), 802-835.*
- Dello Russo S., Mascia D., Morandi F. (2016). Individual perceptions of HR practices, HRM strength and appropriateness of care: A meso, multilevel approach. *The International Journal of Human Resource Management*. doi:10.1080/09585192.2016.
- Dhar, R.L. (2015). Service quality and the training of employees: The mediating role of organizational commitment. *Tourism Management*, 46, 419-430.*

- Dipboye, R. L. (2018). Work Analysis. In *The Emerald Review of Industrial and Organizational Psychology*, 495-534. Emerald Publishing Limited.
- Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, 17(2), 245-260.
- Duriau, V.J., Rigor, R.K. & Pfarrer, M.D. (2007). A Content Analysis of the Content Analysis Literature in Organization Studies – Research Themes, Data Sources and Methodological Refinements. *Organizational Research Methods*, 10(1) 5-34.
- Dyer, L., & Reeves, T. (1995). Human resource strategies and firm performance: what do we know and where do we need to go?. *International Journal of Human Resource Management*, 6(3), 656-670.
- Dysvik, A., Kuvaas, B. & Buch, R. (2014). Perceived training intensity and work effort: The moderating role of perceived supervisor support. *European Journal of Work and Organizational Psychology*, 23(5), 729-738.*
- Edwards, J. R. (2008). To prosper, organizational psychology should . . . overcome methodological barriers to progress. *Journal of Organizational Behavior*, 29, 469–491.
- Ely, R. J. (2004). A field study of group diversity, participation in diversity education programs, and performance. *Journal of Organizational Behavior*, 25(6), 755-780.*
- Esteban-Lloret, N.N., Aragon-Sanchez, A. & Carrasco-Hernandez, A. (2016). Determinants of employee training: Impact on organizational legitimacy and organizational performance. *The International Journal of Human Resource Management*, DOI: 10.1080/09585192.2016.1256337.*

- Faems, D., Sels, L., De Winne, S., & Maes, J. (2005). The effect of individual HR domains on financial performance: evidence from Belgian small businesses. *The International Journal of Human Resource Management*, *16*(5), 676-700.*
- Fey, C. F., Björkman, I., & Pavlovskaya, A. (2000). The effect of human resource management practices on firm performance in Russia. *International Journal of Human Resource Management*, *11*(1), 1-18.*
- Fields, D (2002). *Taking the measure of work: A guide to validated scales for organizational research and diagnosis*. Thousand Oaks: SAGE Publications
- Fletcher, L. (2016). Training perceptions, engagement, and performance: comparing work engagement and personal role engagement. *Human Resource Development International*, *19*(1), 4-26.*
- Fraser, S., Storey, D., Frankish, J. & Roberts, R. (2002). The relationship between training and small business performance: an analysis of the Barclays Bank small firms training loans scheme. *Environment and Planning C: Government and Policy*, *20*, 211-233. *
- Frenkal, S.J. & Bednall, T. (2016). How training and promotion opportunities, career expectations, and two dimensions of organizational justice explain discretionary work effort. *Human Performance*, *29*(1), 16-32. *
- Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft*, *69*(2), 195.
- Garavan, T. N. (1995). HRD stakeholders: Their philosophies, values, expectations and evaluation criteria. *Journal of European Industrial Training*, *19*(10), 17-30.

- Garavan, T. N. (2007). A strategic perspective on human resource development. *Advances in Developing Human Resources*, 9(1), 11-30.
- García, M. Ú. (2005). Training and business performance: the Spanish case. *The International Journal of Human Resource Management*, 16(9), 1691–1710.
doi:10.1080/09585190500239341. *
- Gardner, T. M., Wright, P. M., & Moynihan, L. M. (2011). The impact of motivation, empowerment, and skill-enhancing practices on aggregate voluntary turnover: The mediating effect of collective affective commitment. *Personnel Psychology*, 64(2), 315-350.*
- Gelade, G. A. (2006). But what does it mean in practice? The Journal of Occupational and Organizational Psychology from a practitioner perspective. *Journal of Occupational and Organizational Psychology*, 79(2), 153-160.
- Ghebreorgis, F., & Karsten, L. (2007). Human resource management and performance in a developing country: the case of Eritrea. *The International Journal of Human Resource Management*, 18(2), 321-332.*
- Glaveli, N., & Karassavidou, E. (2011). Exploring a possible route through which training affects organizational performance: the case of a Greek bank. *The International Journal of Human Resource Management*, 22(14), 2892–2923. DOI:10.1080/09585192.2011.606113. *
- Godard J. (2014). The psychologisation of employment relations? *Human Resource Management Journal*, 24(1), 1–18.
- Gooderham, P., Parry, E., & Ringdal, K. (2008). The impact of bundles of strategic human resource management practices on the performance of European firms. *The International Journal of Human Resource Management*, 19(11), 2041-2056.*

- Gubbins, C., Harney, B., van de Werff, L., & Rousseau, D. M. (2018). Enhancing the trustworthiness and credibility of human resource development: Evidence-based management to the rescue? *Human Resource Development Quarterly*.
- Guerrazzi, M. (2016). The effect of training on Italian firms' productivity: Microeconomic and macroeconomic perspectives. *International Journal of Training and Development*, 20(1), 38-57.*
- Guerrero, S., & Barraud-Didier, V. (2004). High-involvement practices and performance of French firms. *The International Journal of Human Resource Management*, 15(8), 1408–1423. doi:10.1080/0958519042000258002. *
- Gurbuz, S., & Mert, I. S. (2011). Impact of the strategic human resource management on organizational performance: evidence from Turkey. *International Journal of Human Resource Management*, 22(8), 1803–1822. doi:10.1080/09585192.2011.565669. *
- Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic organization*, 1(1), 51-78.
- Harel, G. H., & Tzafrir, S. S. (1999). The effect of human resource management practices on the perceptions of organizational and market performance of the firm. *Human Resource Management*, 38(3), 185-199.*
- Harzing, A. W. (2016). Why replication studies are essential: learning from failure and success. *Cross Cultural & Strategic Management*, 23(4), 563-568.
- Hatch, N. W., & Hdyer, J. H. (2004). Human capital and learning as a source of sustainable competitive advantage. *Strategic Management Journal*, 25(12), 1155-1178.*
- Hiller, N.J., De Church, L.A., Murase, T. & Doly, D. (2011) Searching for Outcomes of Leadership: A 25-Year Review. *Journal of Management*, 37(4), 1137-1177.

- Holzer, H. J., Block, R. N., Cheatham, M., & Knott, J. H. (1993). Are training subsidies for firms effective? The Michigan experience. *Industrial & Labor Relations Review*, 46(4), 625-636.*
- Hoobler, J. & Johnson, N. (2004). An analysis of current human resource management publications. *Personnel Review*, 33, 665-676.
- Horgan, J., & Mühlau, P. (2006). Human resource systems and employee performance in Ireland and the Netherlands: A test of the complementarity hypothesis. *The International Journal of Human Resource Management*, 17(3), 414-439.*
- Huang, T. C. (2000). Are the human resource practices of effective firms distinctly different from those of poorly performing ones? Evidence from Taiwanese enterprises. *International Journal of Human Resource Management*, 11(2), 436-451.*
- Ichniowski, C., Shaw, K., & Prennushi, G. (1997). The effects of human resource practices on manufacturing performance: A study of steel finishing lines. *American Economic Review*, 87(3), 291-313.*
- Ismail, H.N. (2016). Training and organizational commitment: Exploring the moderating role of goal orientation in the Lebanese context. *Human Resource Development International*, 19(2), 152-177. *
- Jackson, S. E., & Schuler, R. S. (1995). Understanding human resource management in the context of organizations and their environments. *Annual Review of Psychology*, 46(1), 237-264.
- Ji, L., Huang, J., Liu, Z., Zhu, H., & Cai, Z. (2012). The effects of employee training on the relationship between environmental attitude and firms' performance in sustainable development. *The International Journal of Human Resource Management*, 23(14), 2995-3008.*

- Jiang, J., Wang, S. & Zhao, S. (2012). Does HRM facilitate employee creativity and organizational innovation? A study of Chinese firms. *The International Journal of Human Resource Management*, 23(19), 5024-4047.*
- Jiang, K., Takeuchi, R., & Lepak, D. P. (2013). Where Do We Go From Here? New Perspectives On The Black Box In Strategic Human Resource Management Research. *Journal of Management Studies*, 50(8), 1448-1480.
- Johnson, J. W. (1996). Linking employee perceptions of service climate to customer satisfaction. *Personnel Psychology*, 49(4), 831-851.*
- Kabacoff, R. (2017). Bar Plots. Quick-R Website. . Retrieved from <https://www.statmethods.net/graphs/bar.html>.
- Kampkotter, P. & Marggraf, K. (2015). Do employees reciprocate to intra-firm trainings? An analysis of absenteeism and turnover rates. *The International Journal of Human Resource Management*, 26(22), 2888-2907.*
- Katou, A, Budhwar, P.S. & Patel, C. (2014). Content vs Process in the HRM-performance relationship: An empirical examination. *Human Resource Management*, 53(4), 527-544.*
- Katou, A. A., & Budhwar, P. S. (2006). Human resource management systems and organizational performance: a test of a mediating model in the Greek manufacturing context. *The International Journal of Human Resource Management*, 17(7), 1223–1253. DOI:10.1080/09585190600756525. *
- Khandker, S., B. Koolwal, G., & Samad, H. (2009). *Handbook on impact evaluation: quantitative methods and practices*. The World Bank.
- Khatri, N. (2000). Managing human resource for competitive advantage: a study of companies in Singapore. *International Journal of Human Resource Management*, 11(2), 336-365.*

- 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*(3), 361-389. DOI: 10.1037/a0035408361.*
- Koch, M. J., & McGrath, R. G. (1996). Improving labor productivity: Human resource management policies do matter. *Strategic Management Journal, 17*(5), 335-354.*
- Köhler, T., Landis, R. S., & Cortina, J. M. (2017). From the Editors: Establishing Methodological Rigor in Quantitative Management Learning and Education Research: The Role of Design, Statistical Methods, and Reporting Standards.
- Kooij, D. T., Guest, D. E., Clinton, M., Knight, T., Jansen, P. G., & Dijkers, J. S. (2013). How the impact of HR practices on employee well-being and performance changes with age. *Human Resource Management Journal, 23*(1), 18-35.*
- Kraiger, K., Passmore, J., Dos Santos, N. R., & Malvezzi, S. (Eds.). (2014). *The Wiley Blackwell handbook of the psychology of training, development, and performance improvement*. John Wiley & Sons.
- Krippendorff, K. (2013). *Content Analysis: An Introduction to its Methodology* (3rd Ed). Chicago IL: University of Chicago Press.
- Kroll, A. & Moynihan, D.P. (2015). Does training matter? Evidence from performance management reforms. *Public Administration Review, 75*(3), 411-420.*
- Kwon, K., & Rupp, D. E. (2013). High-performer turnover and firm performance: The moderating role of human capital investment and firm reputation. *Journal of Organizational Behavior, 34*(1), 129–150. DOI:10.1002/job.1804. *
- Lau, C. & Ngo, H. (2004). The HR system, organizational culture, and product innovation. *International Business Review, 13*, 685-703.*

- Laursen, K. & Foss, N.J. (2003). New human resource management practices, complementarities and the impact on innovation performance. *Cambridge Journal of Economics*, 27, 243-263.*
- Lee, G.J. (2012). Firm size and the effectiveness of training for customer service. The *International Journal of Human Resource Management*, 23(12), 2597-2613.*
- Lee, G.J. (2015). Training match and mismatch as a driver of key employee behaviors. *Human Resource Management Journal*, 25(4), 478-495.*
- Liao, T. S., Rice, J., & Martin, N. (2011). The role of the market in transforming training and knowledge to superior performance: evidence from the Australian manufacturing sector. *The International Journal of Human Resource Management*, 22(2), 376-394.*
- Liao, Y. (2006). Task characteristics as a moderator of the relationship between human resource management control and product innovation. *International Journal of Management*, 23(2), 348-355.*
- Lin, L. (2011). Electronic human resource management and organizational innovation: the roles of information technology and virtual organizational structure. *The International Journal of Human Resource Management*, 22(2), 235-257.*
- Litz, R.A. & Stewart, A.C. (2000). Trade name franchise membership as a human resource management strategy: Does buying group training deliver 'true value' for small retailers? *Entrepreneurship, Theory & Practices*, Fall, 125-135.*
- Lopez, S.P., Peon, J.M.M. & Ordas, C.J.V.(2005). Human resource practices, organizational learning and business performance. *Human Resource Development International*, 8(2), 147-164.*
- Lui, S.S., Lau, C. & Ngo, H. (2004). Global convergence, human resources best practices, and firm performance: A paradox. *Management International Review*, 44(2), 67-86.*

- Mabey, C. & Ramirez, M. (2005). Does management development improve organizational productivity? A six-country analysis of European firms. *The International Journal of Human Resource Management*, 16(7), 1067-1082.*
- Mabey, C., & Gooderham, P.N. (2005). The impact of management development on perceptions of organizational performance in European firms. *European Management Review*, 2, 131-142. *
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84.
- MacCarthy, B. L., Lewis, M., Voss, C., & Narasimhan, R. (2013). The same old methodologies? Perspectives on OM research in the post-lean age. *International Journal of Operations & Production Management*, 33(7), 934-956.
- MacDuffie, J. P. & Kochan, T.A. (1995). Do U.S. Firms Invest Less in Human Resources? Training in the World Auto Industry. *Industrial Relations*, 34, 147–168.
- Martell, K., & Carroll, S. (1995). Which executive human resource management practices for the top management team are associated with higher firm performance, *Human Resource Management*, 34(4), 497-512.*
- Mathieu, J. E., & Chen, G. (2011). The etiology of the multilevel paradigm in management research. *Journal of Management*, 37, 610-641.
- McNamara, T.K., Parry, E., Lee, J. & Pitt-Castouphe, M. (2012). The effect of training on organizational performance: Differences by age composition and cultural context. *The International Journal of Human Resource Management*, 23(6), 1226-1244.*
- Menard, S. (2002). Longitudinal Research. 2nd ed. California: Sage.

- Meschi, P. X., & Metais, E. (1998). A socio-economic study of companies through their training policies: New empirical considerations in the French context. *MIR: Management International Review*, 31(1), 25-48.*
- 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.
- Milia, L.D. & Birdi, K. (2010). The relationship between multiple levels of learning practices and objective and subjective organizational financial performance. *Journal of Organizational Behavior*, 31, 481-498.*
- Miron, D., & McClelland, D. C. (1985). The Impact of Achievement Motivation Training of Small Businesses. *California Management Review*, 21(4), 13–28. *
- Mitchell, T.R. & James, L.R. (2001). Building Better Theory: Time and the Specification of when Things Happen. *Academy of Management Review*, 26(4), 530-547.
- Munoz Castellanos, R. M., & Salinero Martín, M. Y. (2011). Training as a source of competitive advantage: performance impact and the role of firm strategy, the Spanish case. *The International Journal of Human Resource Management*, 22(03), 574-594.*
- Murray, B., & Raffaele, G. C. (1997). Single-site, results-level evaluation of quality awareness training. *Human Resource Development quarterly*, 8(3), 229-245.*
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599-620.
- Neall & Tuckey (2014). A methodological review of research on the antecedents and consequences of workplace harassment. *Journal of Occupational and Organizational Psychology*, 87, 225-257.

- Nelson, R. R., & Winter, S. G. (1982). The Schumpeterian tradeoff revisited. *The American Economic Review*, 72(1), 114-132.
- Nembhard, I. M., & Tucker, A. L. (2011). Deliberate learning to improve performance in dynamic service settings: Evidence from hospital intensive care units. *Organization Science*, 22, 907-922.
- Newkirk-Moore, S., & Bracker, J. S. (1998). Strategic management training and commitment to planning: critical partners in stimulating firm performance. *International Journal of Training and Development*, 2(2), 82-90.*
- Ng, Y. C., & Siu, N. Y. (2004). Training and enterprise performance in transition: evidence from China. *The International Journal of Human Resource Management*, 15(4-5), 878-894.*
- Ngo, H. Y., Turban, D., Lau, C. M., & Lui, S. Y. (1998). Human resource practices and firm performance of multinational corporations: influences of country origin. *International Journal of Human Resource Management*, 9(4), 632-652.*
- Nikolova, I., Van Ruysseveldt, J., De Witte, H., & Syroit, J. (2014). Work-based learning: Development and validation of a scale measuring the learning potential of the workplace (LPW). *Journal of Vocational Behavior*, 84(1), 1-10.
- Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: a mini-review. *Frontiers in Psychology*, 3, 322.
- Noe, R.A., Clarke, A.D.M., & Klein, H.J. (2014). Learning in the twenty-first-century workplace. *Annual Review of Organizational Psychology and Organizational Behavior*, 1, 245-275. DOI: 10.1146/annurev-orgpsych-031413-091321.
- Nolan, C. T., & Garavan, T. N. (2016). Human Resource Development in SMEs: A Systematic Review of the Literature. *International Journal of Management Reviews*, 18(1), 85-107.

- Ogunyomi, P. & Bruning, N.S. (2016). Human resource management and organizational performance of small and medium enterprises (SMEs) in Nigeria. *The International Journal of Human Resource Management*, 27(6), 612-634.*
- Otten, S. (2010). "Research in Education: rigor and relevance", Working Paper, Michigan State University.
- Park, Y., & Jacobs, R. L. (2011). The influence of investment in workplace learning on learning outcomes and organizational performance. *Human Resource Development Quarterly*, 22(4), 437-458.*
- Paterson, T. A., Harms, P. D., & Tuggle, C. S (2018). Revisiting the rigor–relevance relationship: An institutional logics perspective. *Human Resource Management*. 1-13.
- Paul, A. K., & Anantharaman, R. N. (2003). Impact of people management practices on organizational performance: analysis of a causal model. *International Journal of Human Resource Management*, 14(7), 1246-1266.*
- Paauwe, J. (2004). *HRM and performance: Achieving long-term viability*. Oxford University Press on Demand.
- Pecci, R. & Van de Voorde, K. (2016). The Application of the Multilevel Paradigm in Human Resource Management–Outcomes Research: Taking Stock and Going Forward. *Journal of Management*. DOI: 10.1177/0149206316673720.
- Pek, J., & Flora, D. B. (2017). Reporting Effect Sizes in Original Psychological Research: A Discussion and Tutorial.
- Ployhart, R.E. & Vandenberg, R.J. (2010) Longitudinal Research: The Theory, Design and Analysis of Change. *Journal of Management*. 30(1), 94-120.

- 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.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879.
- Podsakoff, P.M., MacKenzie, S.B., & Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 65, 539-569.
- Rahman, A.A., Ng, S.I., Sambasivan, M. & Wong, F. (2013). Training and organizational effectiveness: Moderating role of knowledge management process. *European Journal of Training and Development*, 37(5), 472-488.*
- Randensbusch, S.W. & Byrk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd Ed.) Thousand Oaks, Sage.
- Rauch, A., & Hatak, I. (2016). A meta-analysis of different HR-enhancing practices and performance of small and medium sized firms. *Journal of business venturing*, 31(5), 485-504.
- Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring organizational performance: Towards methodological best practice. *Journal of management*, 35(3), 718-804.
- Riley, S. M., Michael, S. C., & Mahoney, J. T. (2017). Human capital matters: Market valuation of firm investments in training and the role of complementary assets. *Strategic Management Journal*, 38(9), 1895-1914.
- Roe, R. A. (2008). Time in applied psychology: The study of “what happens” rather than “what is”. *European Psychologist*, 13(1), 37-52.

- Rost, K., & Ehrmann, T. (2017). Reporting biases in empirical management research: The example of win-win corporate social responsibility. *Business & Society, 56*(6), 840-888.
- Rousseau, D. M., & Barends, E. G. R. (2011). Becoming an evidence-based manager. *Human Resource Management Journal, 21*, 221-235.
- Russell, J. S., Terborg, J. R., & Powers, M. L. (1985). Organizational performance and organizational level training and support. *Personnel Psychology, 38*(4), 849-863.*
- Rynes, S. L., & Bartunek, J. M. (2017). Evidence-based management: Foundations, development, controversies and future. *Annual Review of Organizational Psychology and Organizational Behavior, 4*, 235-261.
- Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science In The Public Interest, 13*(2), 74-101.
- Sanders, K., & Frenkel, S. (2011). HR-line management relations: characteristics and effects. *International Journal of Human Resource Management, 22*, 1611–1617.
- Sanders, K., Cogin, J. A., & Lin, C. H. V. (2017). Methodological choices of HR research conducted in Asia. *Frontiers of Business Research in China, 11*(1), 11.
- Santos, J.P., Caetano, A. & Tavares, S.M. (2015). Is training leaders in functional leadership a useful tool for improving the performance of leadership functions and team effectiveness? *The Leadership Quarterly, 26*, 470-484.*
- Saunders, M.N.K. & Townsend, K. (2016). Reporting and Justifying the Number of Interview Participants in Organization and Workplace Research. *British Journal of Management, 27*(4), 836–852.
- Scheel, T., Rigotti, T., & Mohr, G. (2014). Training and Performance of a Diverse Workforce. *Human Resource Management, 53*(5), 749-772.*

- Schmelter, R., Mauer, R., Borsch, C. & Brettel, M. (2010). Boosting corporate entrepreneurship through HRM practices: Evidence from German SMEs. *Human Resource Management*, 49(4), 715-741.*
- Seidle, B., Fernandez, S. & Perry, J.L. (2016). Do leadership training and development make a difference in the public sector? A panel study. *Public Administration Review*, 76(4), 603-613.*
- Shadish, W.R., Cook, T. & Campbell, D.T. (2002) The Renaissance of Field Experimentation in Evaluating Interventions. *Annual Review of Psychology*, 60, 607-629.
- Shaw, J. D., Delery, J. E., Jenkins, G. D., & Gupta, N. (1998). An organization-level analysis of voluntary and involuntary turnover. *Academy of Management Journal*, 41(5), 511-525.*
- Sheehan, M. (2012). Investing in management development in turbulent times and perceived organisational performance: a study of UK MNCs and their subsidiaries. *The International Journal of Human Resource Management*, 23(12), 2491-2513.*
- Sheldon, P., & Sanders, K. (2016). Contextualizing HRM in China: differences within the country. <https://doi.org/10.1080/09585192.2016.1192323>
- Singh, S., Darwish, T. K., & Potočnik, K. (2016). Measuring organizational performance: A case for subjective measures. *British Journal of Management*, 27(1), 214-224.
- Sitzmann, T., & Weinhardt, J. M. (2018). Training engagement theory: A multilevel perspective on the effectiveness of work-related training. *Journal of Management*, 44(2), 732-756.
- Spector P. E., Liu C., Sanchez J.I. (2015). Methodological and substantive issues in conducting multinational and cross cultural-research. *Annual Review of Organizational Psychology and Organizational Behavior*, 2, 101–131.

- Srivastava, S. P., MacLean, B., Macnab, R. F., & Jackson, H. R. (1982). Davis Strait: structure and evolution as obtained from a systematic geophysical survey.
- StataCorp, L. P. (2013). Stata multilevel mixed-effects reference manual. *College Station, TX: StataCorp LP.*
- Stritch, J. M. (2017). Minding the time: A critical look at longitudinal design and data analysis in quantitative public management research. *Review of Public Personnel Administration, 37*(2), 219-244.
- Sung, S. & Choi, J.N. (2014). Multiple dimensions of human resource development and organizational performance. *Journal of Organizational Behavior, 35*, 851-870.*
- Tabvuma, V., Georgellis, Y. & Lange, T. (2015). Orientation training and job satisfaction: A sector and gender analysis. *Human Resource Management, 54*(2), 303–321.*
- Taylor, P. J., Russ-Eft, D. F., & Chan, D. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology, 90*, 692-709.
- Thang, L. C., & Quang, T. (2005). Antecedents and consequences of dimensions of human resource management practices in Vietnam. *The International Journal of Human Resource Management, 16*(10), 1830-1846.*
- Tharenou, P., Saks, A. M., & Moore, C. (2007). A review and critique of research on training and organizational-level outcomes. *Human Resource Management Review, 17*(3), 251–273. DOI: <http://dx.doi.org/10.1016/j.hrmr.2007.07.004>
- Tzafir, S. S. (2005). The relationship between trust, HRM practices and firm performance. *The International Journal of Human Resource Management, 16*(9), 1600-1622.*

- Tzafrir, S. S. (2006). A universalistic perspective for explaining the relationship between HRM practices and firm performance at different points in time. *Journal of Managerial Psychology, 21*(2), 109-130.*
- Úbeda-García, M., Claver Cortés, E., Marco-Lajara, B., & Zaragoza-Sáez. (2014). Strategy, training and performance fit. *International Journal of Hospitality Management, 42*, 100-116. DOI: 10.1016/j.ijhm.2014.06.001.*
- Ubeda-García, M., Marco-Lajara, B., Sabater-Sempere, V., & García-Lillo, F. (2013). Does training influence organisational performance? Analysis of the Spanish hotel sector. *European Journal of Training and Development, 37*(4), 380-413.*
- Úbeda-García, M., Cortés, E. C., Marco-Lajara, B., & Zaragoza-Sáez, P. (2014). Strategy, training and performance fit. *International Journal of Hospitality Management, 42*, 100-116.
- Van De Voorde, K., Paauwe, J., & Van Veldhoven, M. (2010). Predicting business unit performance using employee surveys: monitoring HRM- related changes. *Human resource management journal, 20*(1), 44-63.
- Van Eerde, W., Tang, K., & Talbot, G. (2008). The mediating role of training utility in the relationship between training needs assessment and organizational effectiveness. *International Journal of Human Resource Management, 19*(1), 63- 73.
DOI:10.1080/09585190701763917.*
- Vandenberg, R. J., Richardson, H. A., & Eastman, L. J. (1999). The impact of high involvement work processes on organizational effectiveness a second-order latent variable approach. *Group & Organization Management, 24*(3), 300-339.*

- Vandenberg, R.J. (2002) Towards a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures. *Organizational Research Methods*, 5(2):139-158 .
- Voegtlin, C., Boehm, S. A., & Bruch, H. (2015). How to empower employees: using training to enhance work units' collective empowerment. *International Journal of Manpower*, 36(3), 354-373.
- 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.
- Walsworth, S. & Verma, A. (2007). Globalization, human resource practices and innovation: Recent evidence from the Canadian Workplace and Employee Survey. *Industrial Relations*, 46(2), 222-240.*
- Wayne, W. J., Shore, L. M., & Linden, R. C. (1997). Perceived Organizational Behaviors and their Effects on Organizational Effectiveness in Limited-Menu Restaurants. *Academy of Management Journal*, 40(1), 82-111.
- Weber, R.P. (1990) Basic Control Analysis (2nd Ed). Newbury Rank, CA. Sage.
- Welbourne, P. (2011). Twenty-first century social work: the influence of political context on public service provision in social work education and service delivery. *European Journal of Social Work*, 14(3), 403-420.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594.
- Wright, P. M., McCormick, B., Sherman, W. S., & McMahan, G. C. (1999). The role of human resource practices in petro-chemical refinery performance. *International Journal of Human Resource Management*, 10(4), 551-571.*

- Wright, P.M., Gardner, T.M., Moynihan, L.M. and Allen, M.R. (2005). The relationship between HR practices and firm performance: Examining causal order. *Personnel Psychology*, 58, 409–446.
- Wright, T.A. & Sweeney, D.A. (2015). The Call for Increased Replication, Extension and Mixed Methods Study Designs in Organizational Research. *Journal of Organizational Behavior*, 37(3), 480-486. DOI: 10.1002/JOB.2059.
- Wright, P. M., & Ulrich, M. D. (2017). A road well traveled: The past, present, and future journey of strategic human resource management. *Annual Review of Organizational Psychology and Organizational Behavior*, 4, 45-65.
- Yang, Y., Chen, Y. & Yang, L. (2013). Gender gap, training and financial performance: Evidence from public accounting industry. *The International Journal of Human Resource Management*, 24(19), 3697-3718. *
- Zheng, C., Morrison, M., & O'Neill, G. (2006). An empirical study of high performance HRM practices in Chinese SMEs. *The International Journal of Human Resource Management*, 17(10), 1772-1803.*
- Zhu, L. (2012). Panel data analysis in public administration: Substantive and statistical considerations. *Journal of Public Administration Research and Theory*, 23(2), 395-428.
- Zwick, T. (2006). The impact of training intensity on establishment productivity. *Industrial Relations: A Journal of Economy and Society*, 45(1), 26-46.*