UvA-DARE (Digital Academic Repository)

A Systematic Review of Human Resource Management Systems and Their Measurement

Boon, C.; Den Hartog, D.N.; Lepak, D.P.

Citation for published version (APA):

A Systematic Review of Human Resource Management Systems and Their Measurement

Corine Boon
Deanne N. Den Hartog
University of Amsterdam

David P. Lepak
University of Massachusetts Amherst

In the strategic human resource (HR) management literature, over the past three decades, a shared consensus has developed that the focus should be on HR systems rather than individual HR practices because the effects of HR practices are likely to depend on the other practices within the system. Despite this agreement, the extent to which the fundamental assumption in the field of interactions and synergy in the system holds true is unclear. We present a systematic review of 495 empirical studies on 516 HR systems in which we analyze the development of HR systems research over time and identify important trends, explicitly linking conceptualization and measurement of the HR system. Our findings suggest that the increasingly broad conceptualization and measurement of HR systems and the lack of clarity on the HR systems construct at different levels have hampered research progress. Much of the research to date does not align with the fundamental assumption of synergies between HR practices in a system, the measures have problems and increasingly confound HR systems with related concepts and outcomes, and insufficient attention is paid to the HR system construct at different levels. Overall, we thus still know little about the "systems" element and how synergies and interactions in an HR system operate. We offer actionable suggestions on how to advance HR systems research towards con-

ceptual clarity and construct refinement, focusing both on how to conceptualize, measure, and combine practices in systems and on studying such systems at different levels of analysis.

Strategic human resource management (SHRM) research increasingly focuses on the performance effects of human resource (HR) systems rather than individual HR practices (Combs, Liu, Hall, & Ketchen, 2006). Researchers tend to agree that the focus should be on systems because employees are simultaneously exposed to an interrelated set of HR practices rather than single practices one at a time, and the effects of HR practices are likely to depend on the other practices within the system (Delery, 1998; Jiang, Lepak, Han, Hong, Kim, & Winkler, 2012; Lepak, Liao, Chung, & Harden, 2006). Research indeed consistently shows a positive association between (broad) HR systems and performance (e.g., Boselie, Dietz, & Boon, 2005; Jiang, Lepak, Ju, & Baer, 2012), and the idea of complementarities or synergies between practices in an HR system is widely accepted as the conceptual logic behind the effectiveness of HR systems (e.g., Chadwick, 2010; Delery, 1998; Gerhart, 2007; Jiang, Lepak, Han et al., 2012). Despite this agreement, the extent to which this fundamental assumption in the field of interactions and synergy in the system holds true is unclear. In other words, our understanding of the "systems" element of HR systems seems more nascent than one might expect, given the sizable body of literature on HR systems.

In the past, several authors have noted fundamental problems in the research relating to how the "system" element of HR systems has been conceptualized. For example, over a decade ago, Lepak and colleagues (2006), in a review of HR systems research, highlighted that a wide variety of HR systems exist with labels such as high performance, commitment, and involvement HR systems but that how these systems are distinct in terms of the practices they include or exclude, how the selected practices help achieve the system's goal, and why these systems would have distinct effects on outcomes was not sufficiently clear. Our first aim is to review the available empirical studies on HR systems and compare studies over time to assess the extent to which the field has progressed in dealing with these issues. In addition, despite the agreement on the interactive nature of HR practices, no consensus has developed on how to combine HR practices into (synergistic) systems (e.g., Chadwick, 2010), and it remains unclear whether or how the field has progressed in terms of understanding how interactions within HR systems that are supposed to be complementary or synergistic work. Thus, going beyond previous reviews, our second aim is to assess the different ways to combine practices in HR systems studies to date to address whether and if so, how the field has progressed in assessing the synergistic effects of HR systems.

Construct development concerns the simultaneous process of validation of measures and theory, and because theory and measurement are inherently linked, both need to be considered in order to advance theory (Smith, Fischer, & Fister, 2003; Strauss & Smith, 2009). The HR field has paid relatively little attention to measurement of HR systems, and previous reviews have not yet focused in detail on these measures. While of course using different measures of the same underlying construct is of value to advance theory, if the same HR system is measured in vastly different ways without clarity as to why this is the case, the question does become whether

measures indeed still capture the same underlying construct and, thus, whether results of such studies are sufficiently comparable. Without good measurement and sound study design, empirical findings may reveal more about the measure than the construct, leading to inaccurate or misleading results (McGrath, 2005; Rossiter, 2008). Thus, our third aim is to review the development of study design and measurement of HR systems over the past three decades.

In sum, we present a systematic review of existing empirical studies on HR systems and analyze the development of the field over time. We take a comprehensive approach and focus on all choices researchers make when designing a study on HR systems, explicitly linking conceptualization and measurement of the HR system. We analyze developments in how HR systems have been conceptualized and measured, how practices are combined into systems, and how HR systems studies are designed. On the basis of this, we highlight conceptual and empirical problems in the current field and offer practical guidance on addressing some of the limitations undermining the current empirical literature, and we discuss theoretical and methodological advances needed to progress towards a better understanding of HR systems.

Our review extends previous work in several important ways. First, analyzing the development of HR systems research over time enables us to identify areas in which progress has been made and where such progress is lacking. In doing so, we identify the most pressing research needs and develop a future research agenda aimed at better understanding interrelationships between HR practices in a system. Second, we add to previous reviews through our focus on both the conceptualization and the measurement of HR systems. Beyond prior reviews, which were primarily conceptual with some addressing some aspects of study design, we also review HR system measures at the item level. As noted, jointly considering both theory and measurement is needed, and in doing so, we identify future research directions that can help establish correspondence between conceptualization and measurement and provide a stronger basis for further theory development on HR systems. Third, we focus specifically on the system element of HR systems by assessing every aspect of HR systems research. Most reviews focus either broadly on the field of SHRM and identify important themes such as human resource management (HRM) implementation or mediating mechanisms in the HRM–performance relationship (e.g., Jackson, Schuler, & Jiang, 2014; Jiang & Messersmith, 2018; Lengnick-Hall, Lengnick-Hall, Andrade, & Drake, 2009) or on specific issues (e.g., levels of analysis, Arthur & Boyles, 2007; Peccei & Van De Voorde, 2019; high performance work practices, Posthuma, Campion, Masimova, & Campion, 2013). Our review is broader in its coverage than those focused on specific issues that include the subset of articles related to that issue and is more exhaustive than those providing a broad thematic overview that focus on a selection of impactful articles (e.g., Lengnick-Hall et al., 2009; Wright & Ulrich, 2017).

Below, we first provide a brief overview of HR systems theory and then present our review showing how HR systems research has developed over the past three decades. Our findings suggest two main and interrelated issues that have hampered research progress: the increasingly broad conceptualization and measurement of HR systems and the lack of clarity on the HR systems construct at different levels. In addition, we see confounding of HR systems with related constructs and outcomes. Together, these problems imply that it is not always sufficiently clear what is responsible for found performance effects of HR systems, which suggests that some of the current evidence may be misleading, and that we lack knowledge about the "system" element of HR systems. We highlight areas of conceptual and empirical confusion in the composition and measurement of HR systems that have hindered theory building, and we offer actionable suggestions on how to advance HR systems research.

Literature Review

Conceptualizing HR Systems

SHRM can be defined as "the pattern of planned HR deployments and activities intended to enable an organization to achieve its goals" (Wright & McMahan, 1992: 298). Increasingly, the field has emphasized the importance of focusing on whether and how "systems" or "bundles" of HR practices jointly help organizations achieve strategic goals, rather than on single HR practices individually. An HR system can be defined as a combination of HR practices "that are espoused to be internally consistent and reinforcing to achieve some overarching results" (Lepak et al., 2006: 221). Conceptually, these *systems* of HR practices—as a whole—are proposed to affect performance-related outcomes (Delery, 1998; Wright & Boswell, 2002). Existing evidence provides some first meta-analytic support, as HR systems tend to be more strongly related to performance than individual HR practices (Combs et al., 2006). However, how this joint effect occurs seems less clear. Conceptually, all practices in a system are proposed to promote an overarching goal (e.g., Jiang, Lepak, Han, et al., 2012); however, it is not always clear what the overarching goal is, how HR systems are conceptualized, or how practices contribute to this goal.

Multiple conceptualizations of HR systems exist, including high performance (e.g., Huselid, 1995), commitment (e.g., Arthur, 1994), and involvement (e.g., Guthrie, 2001). Some scholars use general labels such as HR system or HR bundle without indicating a dominant strategic focus, while others study targeted HR systems focused, for example, on customer service or teamwork (Jackson et al., 2014). Different levels can be distinguished within HR systems: HR policies represent an organization's stated *intentions* about HR practices that should be implemented, whereas HR practices reflect the *actual* HR activities (Becker & Gerhart, 1996; Wright & Boswell, 2002). Techniques are methods used within practices, such as assessment centers in selection. One can also structure HR systems by focusing on broader types or subbundles of practices, such as those based on the ability-motivation-opportunity (AMO) model: ability-enhancing practices (e.g., selection, training), motivation-enhancing practices (e.g., performance management, rewards), and opportunity-enhancing practices (e.g., participation, job design; e.g., Jiang, Lepak, Ju, & Baer, 2012). The logic for this level of abstraction is that countless specific HR practices exist that at a broader policy level, form conceptually similar groupings of practices.

Already over a decade ago authors lamented that a precise and consistent definition of HR systems was lacking and that the variability across HR systems in terms of the included practices was considerable (e.g., Lepak et al., 2006). Here we review whether this has changed over time. We examine how systems are labeled and which practices and subbundles they contain to determine how HR systems that are labeled differently can be distinguished from each other and to what extent HR systems that are labeled similarly indeed are similar in terms of the practices they include. Ambiguity regarding the conceptual boundaries of a construct hinders knowledge accumulation, as it may be unclear what we are speaking about when we examine or compare (specific) HR systems (cf. Podsakoff, MacKenzie, & Podsakoff, 2016).

The System Element of HR Systems

The core assumption underlying HR systems research is that the effectiveness of an HR practice depends on the other practices in the system (Delery, 1998). When practices fit into a

coherent system (internal/horizontal fit), they reinforce one another and create synergies. When practices do not fit, they may detract from each other's effects. Thus, HR practices should be examined jointly rather than separately. Practices in a system can relate to one another in different ways. For example, an additive relationship assumes HR practices have independent effects and add up without influencing each other. In contrast, in an interactive relationship, the effectiveness of a practice depends on the presence or level of other practices. Practices may for instance be substitutes or show positive or negative synergies (e.g., Delery, 1998).

Assuming an additive relationship between practices typically implies calculating an HR system score by summing or averaging scores on individual practices into a scale score or index (Delery, 1998). This approach assumes that HRM is best viewed as a consistent system that has most impact if all practices send consistent signals about the organization's underlying intentions (Bowen & Ostroff, 2004). A suggested advantage of an additive index is that it allows for different ways (i.e., different combinations of practices) to achieve a high system score (e.g., Becker & Huselid, 1998). Yet many disagree with the use of additive indices, as these cannot capture the assumed synergies between practices, and advocate using methods that can capture these, such as cluster analysis or interactions (Becker & Gerhart, 1996; Chadwick, 2010). The few studies that compare different analytical techniques to test for synergies show that the different techniques yield different results and represent different underlying ideas about fit (Chadwick, 2010; Delery & Gupta, 2016). Overall, conceptual approaches to combining differ considerably, and disagreement exists on how to combine HR practices in a system. Knowing how the elements of an HR system interact is important in order to study whether "systems" indeed affect intended outcomes. How much empirical attention different ways of combining practices have received over time is not clear; thus, we review this and analyze trends in the field over time.

Study Design and Measurement

Theory and measurement are inherently linked, and the absence of rigorous study designs and valid measurement can hamper theoretical progress in the field. We thus also review this. We assess who is used as the source to provide information on the HR system. Early research relied mostly on a single (HR) manager to rate the system, which has problems, such as the potentially low reliability of such single-informant designs (e.g., Gerhart, Wright, McMahan, & Snell, 2000). However, even if multiple respondents are used, these sources may not be the most knowledgeable about specific practices or levels. For example, several studies focus on employee perceptions of HR systems (e.g., Den Hartog, Boon, Verburg, & Croon, 2013; Liao, Toya, Lepak, & Hong, 2009), which may not be suited for all research purposes, as employees might not be able to fully evaluate HR systems, especially practices that do not pertain to them personally or intended policies. The HR system may have different meanings at different levels, with different problems associated with each of the levels. Thus, we examine developments over time in the source used to rate the HR system and the levels at which the HR system is theorized and analyzed.

In addition, we review answer scales, as disagreement exists about appropriate rating or answer scales for capturing HR practices (Wright & Gardner, 2003). Answer scales can be more objective, such as the percentage of employees a practice covers, or more subjective, such as Likert-type scales indicating attitudes towards certain practices, and these can reflect different constructs. We assess the examined outcome, which is relevant as, for example,

when studies measure how employees feel about the HR system and relate this to attitudinal outcomes, overlap may occur between the HR system and outcome. Also, because HR system theory implicitly assumes that time is important, as HR systems are supposed to influence performance, the field needs study designs that allow testing for relationships over time and cannot rely on cross-sectional designs. Thus, we review whether longitudinal studies are done and what they focus on.

We review (changes in) the item types used to measure HR systems. Item content and wording can direct the respondents' attention to different aspects of the work environment (e.g., organization or manager), focus on individual experiences (individual referent) or on common experiences in the group (group referent), and describe objective or evaluate subjective characteristics (Klein, Conn, Smith, & Sorra, 2001). Different item types can reflect different underlying conceptual ideas, introduce different biases, and influence the variability between respondents (Klein et al., 2001), which can affect the construct that is actually measured (Clark & Watson, 1995). For example, research on referent-shift models shows that shifting the referent from the individual to the group or vice versa results in two conceptually distinct constructs (Chan, 1998). In general, more objective items tend to yield more agreement among raters than evaluative ones, and individual referents tend to evoke more idiosyncratic responses than group referents, as personal values or interpretations play a larger role in responses. Thus, item wording can alter the meaning of the captured construct and the extent to which respondents are likely to agree. Variation in types of items and their mixed use within one scale may lower validity and accuracy of measurement of HR systems and hamper comparability of results. Below, we present a systematic review focused on all aspects involved in studying HR systems (conceptualization, study design, measurement, and assessing systems) and the developments in this research over time.

Method

Literature Search

We conducted a search of the peer-reviewed academic literature on HR systems published before September 2017. We searched the Scopus and OVID PsycINFO databases, and cross-checked with the EBSCO Business Source Premier database. We searched for peer-reviewed articles containing the following keywords in the title or abstract: "human resource management system" (or human resource/HR/HRM system), "HR(M) bundle," "HR(M) configuration," "set of HR(M) practices," "human resource (management) practices," "high performance/ involvement/ commitment work system" (or high performance/ involvement/ commitment HR/HRM/work practices). In addition, we sent a message to the HR division listserver asking for in press articles. Our deletion of duplicates yielded 5,303 articles. To get a representative picture of the field, which is sufficiently comprehensive and manageable and of sufficient quality, we focused on journals with an impact factor over 1. Thus, we removed all articles published in journals without an impact factor (964 articles) or with an impact factor below one (451 articles), resulting in 3,888 articles. To be included, an empirical study had to meet the following criteria. First, it had to focus on multiple HR practices. Studies on a single practice were excluded. Next, it had to use a quantitative methodology and measure the HR system with a measurement scale. Third, it had to combine the HR practices in some way in a system in the analyses. We did not consider studies

in which HR practices were included individually in the analyses. In total, 495 articles met the criteria and were included in our review; these articles are listed in the online supplemental material.

Coding Procedure

Conceptualization

Some papers report multiple studies or use multiple HR systems; thus, the 495 articles included 516 HR systems. We coded these 516 systems using the following criteria.

HR system label. We coded the label that is used for the HR system, usually retrieved from the hypotheses, model, and tables. Categories were unspecified (for general labels, e.g., HR system, HR practices, HR configuration), high performance, high commitment, high involvement, (strategically) targeted (for labels that clearly specify the target of the HR system), and other.

HR practices or practice domains measured. On the basis of Lepak et al. (2006) and Combs et al. (2006), we coded the following HR practices: job analysis/job design, recruitment, selection, training and development, incentive compensation, other compensation, (self-managed) teams, participation/autonomy, (results-oriented) performance appraisal/management, job security, employee voice/grievance, promotion from within/career development/internal labor market, information sharing/communication, HR planning, flexible work/family-friendly practices, and other practices. We also coded how many practices were included.

Subbundles. We coded whether the study distinguishes between subsystems or subbundles. Categories were ability bundle, motivation bundle, and opportunity bundle (i.e., AMO model), as well as other and none. We coded only subbundles included in the analyses as separate bundles. When subbundles were mentioned only in theory or in discussing the overall HR system, but not included as variables in the analyses, subbundles were not coded.

The Type of Relationship Between the Practices and Bundles

We coded how individual HR practices were combined in systems. All studies that combined practices by averaging or summing scores of the individual practices or used subscale aggregation were coded as additive index, and a second category included studies that analyzed the HR system as a latent factor. All other approaches were first listed under the category other, and subsequently this group was further coded on how they combined practices (see the appendix). We included a category for unclear when no information was provided. We also coded whether and how subbundles were combined in analyses (included as separate bundles or other approaches).

Study Design

We coded all 495 articles in terms of their study design using the following criteria.

Levels. We coded the level of theory and level of analysis of the HR system. The level of theory was coded as organization when theory assumed differences between organizations or when employees were considered as one homogeneous group, as group/unit when assuming differences between units but units being homogeneous, and as individual when differences between individuals were assumed. Categories for level of analysis of the HR system were organization, group/unit, and individual. We also coded whether the study tested a multilevel model.

Data source. We coded who filled out the HR system measure: HR professionals, higher/middle-level managers (e.g., CEOs, unit/department managers), line (or team) managers, employees, others, or unclear. In addition, we coded the use of one or multiple sources.

Answer scale. Categories were presence (yes/no), coverage (the percentage of employees covered by a practice), Likert-type scale, other (for other answer scales), and unclear. We also coded whether one or multiple types of answer scales were used in one measure.

Outcomes. We coded which types of outcome(s) were examined in each study: attitudes, behaviors, performance (including different types of individual/organizational performance, e.g., productivity or task performance), other, or none (studies with the HR system as the outcome).

One or multiple time points. We coded whether studies were cross-sectional, used separate measurements in time, or were longitudinal in nature.

Measures

We coded whether the measure for the HR system was existing, adapted from existing measures, or newly developed. For the adapted ones, we listed references to the original measures up to three, and when four or more were used, we coded them as multiple. Of the 516 systems, 219 had (mostly) new measures, 193 adapted ones, and 100 an existing measure. For 4 of the systems, it was unclear. Part of our review focuses on the item level. For this, we needed full measures. For 209 studies, the measure was available in full in the article; of these, 29 were existing, 77 were (mostly) new, and 103 were adapted from existing ones. Of these, 34 were adapted from four or more measures. We coded the 77 newly developed ones and the 34 based on four or more existing ones (111 in total) for the following.

Policies, practices, or techniques. Items were coded as policies if they referred to organizational goals or objectives for managing HRs. We coded items referring to general practices, such as selection, as practices and as techniques if they referred to specific practice techniques used within a practice, such as selection interviews or assessment centers.

General vs. criterion focused. We coded whether items were general (e.g., referring to rigorous selection) or focused on a specific criterion (e.g., selection based on creativity).

Who offers HR practices. Different agents can offer HR practices, and we coded whether items referred to HR practices emanating from the organization, unit, or manager. We used unspecified when it was unclear who offered HR.

Item referent. We included the following categories when coding item referents: group (multiple individuals, such as employees, as the referent), job (a specific job or job category as the referent), individual (one individual as the referent), or unspecified/unclear.

Item focus. We coded whether items were descriptive or evaluative. When items refer to a practice in an objective way (e.g., how many hours of training), we coded them as descriptive, and when items contain a value judgement or refer to a feeling, we coded them as evaluative (e.g., communication is effective). We also used the category descriptive with Likert scale for descriptive items with a Likert scale, which includes more evaluation than percentages or coverage. We used the category descriptive and evaluative for mostly descriptive items that contain an evaluative element, using words such as "considerable" or "serious" (e.g., considerable importance is placed on staffing).

Results

Table 1 summarizes the coded data for the 516 systems on how HR systems are conceptualized and combined, and Table 2 summarizes the coded data on study design and measurement. To assess developments over time, we report results for five time periods (1991–2000, 2001–2005, 2006–2010, 2011–2015, 2016–2017) and the total period (Total).[1] When reporting changes, we report percentages for the first (1991–2000) and last (2016–2017) period.

Conceptualization of HR Systems

How Are HR Systems Labeled?

Table 1 shows that many different HR system labels are used. Unspecified labels such as HRM, HR practices, HR system, HR bundle, or HR configuration are widely used (34% overall), but their use has decreased over time (from 59% to 23%). With these generic labels, it is unclear what the goal of a system is. Labels such as high performance (35%), commitment (8%), or involvement (8%) HR systems are widely used with little change over time. Table 1 shows that targeted HR systems with more specific labels such as relationship-oriented HR system, knowledge-oriented HR system, and initiative-enhancing HRM system are less common (12% overall) but have increased over time (from 9% to 19%). The remaining studies (3%) mostly do not focus on (the extent to) which HR practices are offered but on preferences for, motivation for, satisfaction with, or effectiveness of HRM.

Problematically, different terms are often used for highly similar HR systems, which has not improved over time. For example, while the labels of high performance and high commitment HR systems suggest they are differentially strategically targeted HR systems (focused on increasing performance vs. commitment), they are used interchangeably in many studies, implying these labels have become more general than originally intended. The practices included in and the items used to measure these systems overlap strongly. For example, most practices are found in both

Table 1
Conceptualization of Human Resource (HR) Systems

	1991–2000 (34 Systems)	2001–2005 (61 Systems)	2006–2010 (114 Systems)	2011–2015 (202 Systems)	2016–2017 (105 Systems)	Total (516 Systems)
HR System Label						
Unspecified	59% (20)	52% (32)	37% (42)	29% (58)	23% (24)	34% (176)
High performance	21% (7)	28% (17)	33% (38)	41% (83)	36% (38)	35% (183)
High commitment	12% (4)	3% (2)	11% (12)	5% (10)	14% (15)	8% (43)
High involvement	6% (2)	8% (5)	7% (8)	8% (17)	10% (10)	8% (42)
(Strategically) targeted	9% (3)	7% (4)	8% (9)	14% (28)	19% (20)	12% (64)
Other	0% (0)	2% (1)	5% (6)	3% (6)	0% (0)	3% (13)
HR Practices						
Training/development	82% (28)	89% (54)	91% (104)	90% (182)	89% (93)	89% (461)
Participation/autonomy	85% (29)	70% (43)	74% (84)	68% (137)	70% (73)	71% (366)
Incentive compensation	76% (26)	77% (47)	75% (86)	66% (133)	59% (62)	69% (354)
Performance appraisal	50% (17)	56% (34)	74% (84)	66% (133)	68% (71)	66% (339)
Selection	62% (21)	57% (35)	62% (71)	59% (119)	52% (55)	58% (301)
Job analysis/design	71% (24)	51% (31)	59% (67)	45% (91)	43% (45)	50% (258)
Promotion from within/career development/internal labor market	47% (16)	59% (36)	50% (57)	46% (92)	45% (47)	48% (248)
Information sharing/communication	47% (16)	49% (30)	46% (53)	48% (96)	47% (49)	47% (244)
Other compensation	32% (11)	34% (21)	52% (59)	41% (82)	43% (45)	42% (218)
(Self-managed) teams	47% (16)	51% (31)	49% (56)	36% (73)	28% (29)	40% (205)
Job security	32% (11)	30% (18)	32% (37)	27% (55)	27% (28)	29% (149)
Recruitment	21% (7)	26% (16)	28% (32)	21% (43)	19% (20)	23% (118)
Employee voice/grievance	47% (16)	23% (14)	23% (26)	17% (34)	16% (17)	21% (107)
Flexible work/family-friendly practices	6% (2)	11% (7)	6% (7)	12% (25)	15% (16)	11% (57)
HR planning	9% (3)	5% (3)	4% (5)	3% (7)	0% (0)	3% (18)
Others	59% (20)	61% (37)	50% (57)	47% (94)	36% (38)	48% (246)
Number of Practices						
Average	8.1	8.5	8.4	7.4	7	7.7
Minimum	3	3	3	2	2	2
Maximum	16	15	16	16	15	16
Subbundles						
Ability	0% (0)	0% (0)	1% (1)	5% (10)	10% (10)	4% (21)
Motivation	3% (1)	0% (0)	1% (1)	4% (8)	10% (10)	4% (20)
Opportunity	0% (0)	0% (0)	1% (1)	5% (10)	10% (10)	4% (21)
Other	38% (13)	30% (18)	26% (30)	15% (31)	9% (9)	20% (101)
None	59% (20)	70% (43)	73% (83)	80% (161)	82% (86)	76% (393)
Relationship Between Practices						
Additive	47% (16)	66% (40)	76% (87)	72% (145)	63% (66)	69% (354)
Latent factor	6% (2)	11% (7)	11% (13)	21% (43)	29% (30)	18% (95)
Other	50% (17)	25% (15)	16% (18)	10% (20)	7% (7)	15% (77)
Unclear	0% (0)	3% (2)	2% (2)	1% (2)	2% (2)	2% (8)
Relationship Between Bundles						
Separate	80% (8)	92% (11)	86% (18)	83% (20)	92% (12)	86% (69)
Other	20% (2)	8% (1)	14% (3)	17% (4)	8% (1)	14% (11)

Table 2
Study Design and Measurement of Human Resource (HR) Systems

	1991–2000 (34 Systems/34 Studies, 6 Measures)	2001–2005 (61 Systems/58 Studies, 11 Measures)	2006–2010 (114 Systems/112 Studies, 26 Measures)	2011–2015 (202 Systems/192 Studies, 49 Measures)	2016–2017 (105 Systems/99 Studies, 19 Measures)	Total (516 Systems/495 Studies, 111 Measures)
Time Points						
One	91% (31)	92% (56)	89% (101)	89% (180)	84% (88)	88% (456)
Multiple	9% (3)	8% (5)	11% (13)	11% (22)	16% (17)	12% (60)
Outcomes						
Attitude	3% (1)	13% (8)	23% (26)	35% (71)	35% (37)	28% (143)
Behavior	3% (1)	3% (2)	11% (12)	20% (40)	24% (25)	16% (80)
Performance	59% (20)	57% (35)	55% (63)	52% (106)	46% (48)	53% (272)
Other	0% (0)	13% (8)	9% (10)	16% (32)	20% (21)	14% (71)
HRM as outcome	41% (14)	16% (10)	18% (21)	10% (20)	6% (6)	14% (71)
Level of Theory						
Organization	97% (33)	98% (60)	89% (102)	89% (179)	89% (93)	91% (467)
Group/unit	3% (1)	3% (2)	9% (10)	9% (18)	10% (11)	8% (42)
Individual	0% (0)	2% (1)	3% (3)	3% (6)	5% (5)	3% (15)
Level of Analysis						
Organization	94% (32)	77% (47)	69% (79)	60% (121)	63% (66)	67% (345)
Group/unit	3% (1)	7% (4)	7% (8)	9% (18)	10% (11)	8% (42)
Individual	3% (1)	11% (7)	23% (26)	30% (61)	33% (35)	25% (130)
Multilevel	0% (0)	2% (1)	10% (11)	14% (29)	27% (28)	13% (69)
Data Source						
HR professionals	41% (14)	56% (34)	40% (46)	33% (66)	26% (27)	36% (187)
High/middle manager	44% (15)	43% (26)	46% (53)	36% (72)	36% (38)	40% (204)
Line manager	6% (2)	11% (7)	11% (12)	11% (22)	10% (11)	10% (54)
Employee	6% (2)	20% (12)	26% (30)	39% (78)	50% (52)	34% (174)
Unclear	18% (6)	3% (2)	3% (3)	4% (9)	4% (4)	5% (24)
Other	3% (1)	0% (0)	2% (2)	0% (1)	0% (0)	1% (4)
How Many Sources						
One	68% (23)	69% (42)	74% (84)	76% (153)	76% (80)	74% (382)
Multiple	15% (5)	26% (16)	23% (26)	20% (41)	20% (21)	21% (109)
Answer Scale						
Presence (yes/no)	32% (11)	33% (20)	27% (31)	19% (39)	18% (19)	23% (120)
Coverage	29% (10)	26% (16)	20% (23)	14% (29)	10% (11)	17% (89)
Likert-type scale	53% (18)	59% (36)	69% (79)	68% (138)	81% (85)	69% (356)
Unclear	9% (3)	11% (7)	4% (5)	10% (20)	4% (4)	8% (39)
Other	32% (11)	21% (13)	9% (10)	8% (16)	6% (6)	11% (56)
Number of Answer Scales						
One	53% (18)	57% (35)	75% (86)	74% (150)	83% (87)	73% (376)
Multiple	38% (13)	33% (20)	20% (23)	16% (33)	14% (15)	20% (104)
Average Number of Items	21	20	18	15	20	19
Policies/Practices/Techniques?						
Policies	17% (1)	0% (0)	12% (3)	8% (4)	5% (1)	8% (9)
Practices	100% (6)	100% (11)	100% (26)	100% (49)	100% (19)	100% (111)
Techniques	17% (1)	27% (3)	19% (5)	20% (10)	32% (6)	23% (25)
Other	17% (1)	9% (1)	19% (5)	8% (4)	0% (0)	10% (11)
Policies/Practices/Techniques?						
One	67% (4)	64% (7)	58% (15)	67% (33)	68% (13)	65% (72)
Multiple	33% (2)	36% (4)	38% (10)	33% (16)	32% (6)	34% (38)
Criterion focused?						
Criterion focused	33% (2)	9% (1)	31% (8)	37% (18)	47% (9)	34% (38)
General	100% (6)	100% (11)	100% (26)	90% (44)	95% (18)	95% (105)
HR Practices Offered by						
Organization	33% (2)	36% (4)	65% (17)	51% (25)	68% (13)	55% (61)
Unit/team	17% (1)	18% (2)	8% (2)	8% (4)	11% (2)	10% (11)
Manager/management	17% (1)	18% (2)	15% (4)	31% (15)	42% (8)	27% (30)
Unspecified	83% (5)	100% (11)	92% (24)	94% (46)	95% (18)	94% (104)

(continued)

Table 2 (continued)

	1991–2000 (34 Systems/34 Studies, 6 Measures)	2001–2005 (61 Systems/58 Studies, 11 Measures)	2006–2010 (114 Systems/112 Studies, 26 Measures)	2011–2015 (202 Systems/192 Studies, 49 Measures)	2016–2017 (105 Systems/99 Studies, 19 Measures)	Total (516 Systems/495 Studies, 111 Measures)
HR Practices Offered by						
One	67% (4)	45% (5)	31% (8)	41% (20)	11% (2)	35% (39)
Multiple	33% (2)	55% (6)	69% (18)	59% (29)	89% (17)	65% (72)
Item Referent						
Group	83% (5)	100% (11)	88% (23)	84% (41)	89% (17)	87% (97)
Job	33% (2)	45% (5)	8% (2)	8% (4)	5% (1)	13% (14)
Individual	0% (0)	18% (2)	31% (8)	33% (16)	32% (6)	29% (32)
Unspecified	67% (4)	55% (6)	58% (15)	63% (31)	79% (15)	64% (71)
Item Referent						
One	33% (2)	9% (1)	27% (7)	29% (14)	21% (4)	25% (28)
Multiple	67% (4)	91% (10)	73% (19)	71% (35)	79% (15)	75% (83)
Item Focus						
Descriptive	50% (3)	45% (5)	31% (8)	29% (14)	26% (5)	32% (35)
Descriptive and Likert scale	67% (4)	64% (7)	65% (17)	63% (31)	84% (16)	68% (75)
Descriptive and evaluative	33% (2)	55% (6)	54% (14)	61% (30)	74% (14)	59% (66)
Evaluative	50% (3)	55% (6)	58% (15)	53% (26)	37% (7)	51% (57)
Item Focus						
One	17% (1)	18% (2)	31% (8)	27% (13)	21% (4)	25% (28)
Multiple	83% (5)	82% (9)	69% (18)	73% (36)	79% (15)	75% (83)

Note: HRM = human resource management.

types of systems, and several studies of high commitment HR systems (e.g., Kwon, Bae, & Lawler, 2010; Yamamoto, 2013) base the choice of practices in the system on work on high performance HR systems (e.g., on Becker & Huselid, 1998; Huselid, 1995). However, causal mechanisms linking different targeted combinations of practices to outcomes should at least to some extent differ; thus, the combinations should not be fully interchangeable. For example, practices in a system emphasizing enhancing worker efficiency should differ from those in a system focused on creating a highly able or innovative workforce. In addition, the system label used does not always reflect the original focus of the measure used. For example, Camelo-Ordaz, García-Cruz, Sousa-Ginel, and Valle-Cabrera (2011) use items from Lepak and Snell's (2002) commitment and collaboration HR measures but label the system high involvement. Also, unspecified labels are sometimes used for scales originally developed for targeted systems. These labeling issues can create confusion and ambiguity and may reflect misalignment between theory and measurement.

Which HR Practices Are Measured?

Studies vary strongly on the number of included HR practices, which reflects differences in the breadth of the conceptualization of the HR system. Surprisingly, many studies are not very specific in describing which practices they measured. If a measure was not provided, it was often unclear. The average number of practices in a system has slightly decreased (from 8.1 to 7.0), and the range has stayed relatively stable (between 2 and 16 practices). The combinations of practices included in HR systems, even in those with the same label, vary considerably. The most widely adopted practices are training/development (89%), participation/autonomy (71%), incentive compensation (69%), performance

appraisal (66%), selection (58%), and job design (50%), which is in line with earlier reviews (e.g., Boselie et al., 2005; Posthuma et al., 2013). The number of practices used in at least 50% of the studies has decreased (from 8 until 2010 to 5 thereafter), suggesting agreement about which practices should be included in HR systems has decreased rather than increased over time. Of the 516 systems, 24% include subbundles such as AMO or others, which has decreased over time (41% to 18%).

Studies also vary considerably on the inclusion of other practices, as 48% of HR systems overall include practices from the "other" category, including HR-related practices such as attitude surveys, mentoring, exit management, absence management, and diversity management, but also other constructs. The breadth of the "other" category content begs the question where the boundaries lie of what still constitutes an HR practice. For example, over time an increasing number of studies includes (transformational) leadership or supervisor support in the HR system (e.g., Zacharatos, Barling, & Iverson, 2005). In addition, concepts that are usually considered outcomes are included. For example, attitudes such as trust, fairness, and loyalty are increasingly included in HR systems (e.g., Chen, 2007; Prieto Pastor, Santana, & Sierra, 2010), and other elements such as skill level (e.g., De Grip & Sieben, 2009), climate, and organizational effectiveness (e.g., Ma, Silva, Callan, & Trigo, 2016) are sometimes included as well. Some studies include vertical alignment in the HR system, for example, the strategic importance of specific human capital (De Saá-Pérez & García-Falcón, 2002) or the strategic orientation of HRM (e.g., Jayaram, Droge, & Vickery, 1999). Thus, there is disagreement on which HR practices should be included in HR systems but more problematically, also on what is (or is not) an HR practice.

Besides the lack of agreement on what constitutes an HR practice to begin with, there is disagreement on the content some HR practice areas should cover. While at least some agreement is seen on what the most used practices, such as training, incentive compensation, or selection, typically entail, practices such as participation, job design, and communication are more ambiguous. The latter show a much larger variation in how they are conceptualized and measured. For example, the term "job design" is used for having job descriptions but also for challenging work. This conceptual disagreement at multiple levels raises the question whether we are capturing the same or different constructs in studies even when they are on similarly labeled systems. Lack of clarity on what is an HR practice, contamination of the system with outcomes, and lack of clarity in whether it is the combination of HR practices or the related variables, such as leadership, included in the system that yield an effect are all problems relating to this.

Assessing the System Element of HR Systems

Next, we assessed how authors combine HR practices into systems. Most studies (87% overall) use an additive index or a latent variable approach, and despite repeated calls for using other approaches that address the core theoretical assumption of interdependence of practices in systems, the use of these has decreased considerably over time (from 50% to 7%). Downsides of the additive approach include that practices are weighted equally and that it does not allow testing for the interactions and synergies proposed to underlie the effectiveness of HR systems. Using a latent factor allows for some weighting; however, it does not yet capture synergies. Overall, to date, only 15% of the studies combine practices into an HR system in other ways. The appendix lists studies using other ways to combine practices.

Some ways of combining practices are empirically based, such as cluster analysis (e.g., Arthur, 1994) or latent class analysis (e.g., De Menezes & Wood, 2006), which empirically derive sets of practices that are usually adopted together. One study uses sequential tree analysis (Guest, Conway, & Dewe, 2004), and one uses fuzzy set qualitative comparative analysis (Meuer, 2017)—techniques that can help identify which practices are most important for explaining the outcome.

Theoretically based methods to combine HR practices in a system include examining interactions between practices (16 studies). Studies vary from the examination of a specific interaction between two practices based on theoretical grounds, such as Frick, Goetzen, and Simmons (2013), who examine the interaction between teamwork and performance pay; to the examination of interactions between one specific practice (e.g., participation or teamwork) and all other practices included in the system (e.g., Gould-Williams & Gatenby, 2010); to the inclusion of all possible interactions between the practices included in the HR system (e.g., Darwish, Singh, & Mohamed, 2013). Also, 21 studies calculate a system score based on the presence, absence, or level of specific HR practices, for example, by scoring the HR system as 1 only if all practices (e.g., Kauhanen, 2009) or at least a certain number of practices (e.g., Laursen & Foss, 2003) are present or if the score on each of the practices is higher than a certain threshold, such as the median (e.g., Laroche & Salesina, 2017). Others indicate which/how many practices should be present. For example, Ichniowski and Shaw (1999) distinguish five HR systems based on presence/absence of specific practices.

Six studies use profile or pattern deviation and calculate the deviation of actual HR practices from an ideal type HR system. They differ in how they determine ideal types. Some use theoretically derived ideal types of HR systems (e.g., Delery & Doty, 1996), others combined these with expert ratings (e.g., Verburg, Den Hartog, & Koopman, 2007). Also, six studies use weighted measures, usually calculating an HR system index weighted on the basis of the proportion of workers covered by each practice (e.g., Galang, 1999), which takes differences in use of practices into account but does not capture synergies. Koster (2011) used the standard deviations of items to calculate internal fit to measure the inconsistency of experienced HR practices. Only six studies combine subbundles in nonadditive ways, such as interactions, profile deviation, or polynomial regression (e.g., Chenevert & Tremblay, 2009; Godard, 2007; Huselid, 1995). Bryson, Forth, and Kirby (2005) calculated a system score based on high scores on three subbundles. Overall, using other ways of assessing fit has decreased over time and they are seldom compared; thus, there is only limited systematic evidence on what "best" ways of combining practices in a system are.

Measurement and Study Design

Table 2 summarizes the coded data on measurement and study design.

Who Rates the System?

Variation in respondents providing the data on the HR system is increasing; most use HR professionals (36% overall), higher/middle-level managers (40%) and lower-level managers (10%), or employees (34%). Only 1% of studies use other sources (e.g., union reps, students). In 5% of studies, the respondent is unclear, and most of these use secondary data

(e.g., Kalleberg & Moody, 1994). Over time, the use of HR managers (from 41% to 26%) and higher/middle-level managers (44% to 36%) as respondents decreases and that of employees increases (6% to 50%). This shift toward more employee-rated HR systems is in and of itself not problematic, as different perspectives are of interest. However, what is measured as the "HR system" has different meanings and reflects different levels/constructs, including firm level–intended organizational policies, implemented practices valid for specific groups, and idiosyncratic perceptions of individual employees. This raises questions about whether results are always comparable. Also, despite the increase in studies that analyze data at the individual level (from 3% to 33%), the level of theory in most studies is still the organization, as individual-level theory has increased only from 0% to 5% over time. This mismatch is problematic, as individual-level data do not always capture meaningful organization-level characteristics.

Of the studies, 74% rely on one type of respondent (one source), such as HR managers or employees, to rate the HR system, while 21% use multiple sources, and this has not changed much over time. Most studies using multiple sources combine all responses into one HR system variable. As respondents from different organizational levels may have different perspectives, this is problematic. Combining ratings can imply combining different constructs that reflect different meanings of the HR system without taking these differences into account. The question is then what such combined measures capture. Using managers and employees as respondents and constructing a manager-rated and an employee-rated HR system is done in 14 studies, with the alignment between their views often being moderate at best (e.g., Den Hartog et al., 2013).

Answer Scales and Outcomes

Considerable variation exists in answer scales: presence, coverage, Likert-type scales, and other scales (usually a count, e.g., training hours) are found. Each answer scale reflects something different and sometimes even different constructs (e.g., coverage vs. attitudes). Also, quite a few measures use a mix of answer scales (20% overall). Variation is particularly high in older studies and when HR systems are rated by respondents other than employees (for employees, Likert scales are common). Over time, the use of descriptive answer scales such as presence, coverage, and counts has decreased, and the use of Likert-type scales has increased (from 53% to 81%). Particularly Likert-type scales that focus on agreement are criticized because it is unclear what a score actually means (Clark & Watson, 1995). Table 2 also shows that over time, employee attitudes are increasingly studied as outcomes (from 3% to 35%). Problematically, in several measures using Likert scales, HR system items are confounded with their outcomes not only because these are increasingly included in the system as noted above but also, for example, when employees rate perceptions of HR practices with evaluative items and the studied outcomes are their attitudes toward the job or organization.

Cross-Sectional or Over Time?

The number of studies using cross-sectional designs is slowly decreasing over time (91% to 84%), yet while using multiple time points has increased, most studies are not longitudinal but, rather, use two time points to separate independent from dependent

variables. A few studies measure the HR system once and the outcome multiple times (e.g., Wright, Gardner, Moynihan, & Allen, 2005) or measure both the HR system and the outcome at two time points. Only 2% of (recent) studies are truly longitudinal, using three time points and assessing change. Most longitudinal studies test causal (and reversed) relationships between HR systems and outcomes (e.g., Shin & Konrad, 2017); three studies go beyond this to explore how long it takes for the HR system to have an effect and how long these effects persist (e.g., Piening, Baluch, & Salge, 2013). This relative dearth of longitudinal studies is problematic for establishing causality and addressing the other roles time can play in HR systems.

HR Systems Measures

Our review shows considerable variation has existed in measures used in research on HR systems from the early research onwards. Many newly developed or (strongly) adapted measures are used in the reviewed studies (219 of the 516 HR systems were new; 193 were adapted). This implies most scales do not receive extensive scale validation through repeated use in multiple contexts. The number of items used varies from a very limited number (3; Litwin, 2013) to a much higher one (up to 60; Shin & Konrad, 2017). The average number of items is 19, which is relatively stable over time. All HR system measures contain items that measure HR practices; however, 34% (stable over time) use a mix of items tapping practices with items on policies and/or techniques. For example, in one scale, Huselid (1995) combines general practices (e.g., "What proportion of the workforce receives formal performance appraisals?") and techniques (e.g., "What proportion of the workforce is administered an employment test prior to hiring?"), and Ketkar and Sett (2009) combine practices (e.g., "We regularly involve our employees in decision making on job related matters") with policies (e.g., "Good performance is always recognized and rewarded in our firm"). Such combinations can confound multiple components of the HR system structure. For example, when combining policies and practices, it can be unclear whether respondents reported on intended or actual practices.

Measures also vary in whether items are general versus criterion-focused (e.g., aimed to enhance flexibility). Almost all measures (95%) contain general items (e.g., "Employees in this job are often asked by their supervisor to participate in decisions"; Delery & Doty, 1996), but an increasing number mix this with criterion-focused items. For example, F. Liu, Chow, Gong, and Wang (in press) mix general items (e.g., "Employees have various opportunities for upward mobility") and criterion-focused items (e.g., "My organization emphasizes training with focus on creativity"). A few measures are fully criterion focused, mostly for strategically targeted systems with criteria such as flexibility (e.g., S. Chang, Gong, Way, & Jia, 2013) or personal initiative (e.g., Hong, Liao, Raub, & Han, 2016). Some use a general label with a criterion-focused measure, such as Karatepe (2013), whose high performance work system measure consists of items focusing on customer service.

Types of Items

Table 3 shows example items for who offers HR, item referents, and item focus. Who offers HR in the items varies from the organization (e.g., the organization offers training),

Table 3

Item Wording: Sample Items

Issue		Example Items From the Same Scales
Human Resource Practices Offered by	—	"The organisation spends enough resources on EEO [equal employment opportunity] awareness and EEO-related training" *(organization)*
	—	"This unit is committed to the training and development of its employees" *(unit)*
	—	"Management in this unit is supportive of cultural difference in this organization" *(management)*
	—	"Interview panels are used during the recruitment and selection process in this unit" *(unspecified)* (four items from Ang, Bartram, McNeil, Leggat, & Stanton, 2013)
	—	"In this department we are provided with the training needed to achieve high standards of work" *(unit)*
	—	"Our line manager/supervisor consults us before making decisions" *(management)*
	—	"I am provided with sufficient training and development opportunities" *(unspecified)* (three items from Gould-Williams & Gatenby, 2010)
	—	"Our company provides employees with training assistance enabling them to upgrade their qualifications" *(organization)*
	—	"Supervisors keep open communications with employees" *(management)*
	—	"Performance assessments are conducted regularly" *(unspecified)* (three items from Zhang, Fan, & Zhu, 2014)
Item Referent	—	"We provide multiple career path opportunities for employees to move across multiple functional areas of the company" *(group)*
	—	"Performance appraisals are used primarily to set goals for personal development" *(unspecified)* (two items from Collins & Smith, 2006)
	—	"Goals that are applied to employees are usually challenging" *(group)*
	—	"Employees in this job will normally go through training programs every few years" *(job)*
	—	"I can decide how to get my job done" *(individual)*
	—	"A wild variety of training programs is provided in my company" *(unspecified)* (four items from Jaw & Liu, 2003)
	—	"Individuals' skill is important when recruiting new employees" *(group)*
	—	"How much influence do you have over what tasks you do?" *(individual)*
	—	"Grievance procedures cover pay issues" *(unspecified)* (three items from Ogbonnaya, Daniels, Connolly, & Van Veldhoven, 2017)
Item Focus	—	"Have you had an appraisal or Knowledge and Skills Framework development review in the last 12 months?" *(descriptive)*
	—	"Team members have a set of shared objectives" *(descriptive with Likert-type scale)*
	—	"Managers give clear feedback on my work" *(descriptive/evaluative)*
	—	"Different parts of the trust communicate effectively with each other" *(evaluative)* (four items from Ogbonnaya et al., 2017)
	—	"On average how many hours of formal training do employees in this job receive each year?" *(descriptive)*
	—	"Employees in this job have a reasonable and fair complaint process" *(evaluative)* (two items from Wright, Gardner, Moynihan, Park, Gerhart, & Delery, 2001)
	—	"Within the last 2 years, how many employees participated in formal job training?" *(descriptive)*
	—	"How important is job performance in determining the earnings of managers and administrators?" *(descriptive with Likert-type scale)*
	—	"Overall, how effective would you say your employee training is?" *(evaluative)* (three items from Delaney & Huselid, 1996)

to the unit/team, the manager, or unspecified. Of the measures, 94% included items that were unspecified, such as "I am provided with sufficient training and development opportunities" (Gould-Williams & Gatenby, 2010) or "Employee bonuses or incentive plans are based primarily on the performance of the organization" (Collins & Smith, 2006). A decreasing number of measures have items that consistently fall into one category (from 67% to 11%). These measures either have items that exclusively pertain to either the organization or the unit or all items are unspecified. The majority of the measures (65%) vary (having items on the organization, unit, managers, and unspecified in a mix), and such mixes increased over time (from 33% to 89%).

Item referents vary widely. We found that 87% of the measures include items with a group referent (e.g., employees), and 64% use items leaving the referent unspecified (e.g., "A wide variety of training programs is provided in my company"; Jaw & Liu, 2003). Using the job as item referent has decreased (33% to 5%), despite the job forming a specific and clear referent. Using an individual referent has increased (0% to 32%). Most measures (75%) use a mix of item referents in a single scale, which is relatively stable over time. Studies using two referents often mix a group referent with an unspecified referent (e.g., Collins & Smith, 2006). Others mix three (e.g., Ogbonnaya, Daniels, Connolly, & Van Veldhoven, 2017) or four (e.g., Jaw & Liu, 2003) referents in one measure.

Turning to item focus, most measures (68%) use descriptive items in combination with a Likert-type answer scale. An increasing number of scales (33% to 74%) have items that are a mix of descriptive and evaluative, combining a mostly descriptive statement with an adjective that asks for a value judgement (e.g., "Managers give *clear* feedback"), and the use of fully evaluative items has been relatively stable (51% overall; e.g., "Training is effective"). The use of purely descriptive items has decreased over time (50% to 26%). Similar to the item referents, 75% of the measures (relatively stable over time) combine items with different item content, up to all four (e.g., Ogbonnaya et al., 2017). Taken together, for all item-related criteria (who offers HR, the item referent, and item focus), most measures mix multiple item types. This has not improved and in part has even increased over time. These mixes raise questions on whether it is always clear what the overall scale is capturing and whether respondents can always fully judge item content or are always focused on the intended part of the work environment. Measures mixing different agents offering HR, group, and individual referents and including descriptive as well as evaluative items can be ambiguous. Ambiguous item wording can create confusion, change the meaning of the measured HR system, and negatively affect interrater agreement. At worst, it is unclear what is assessed. Our review suggests that these problems have increased in more recent work.

Discussion and Implications

We aimed to review three decades of HR systems research focusing on the "systems" element of HR systems to identify where the field has progressed and where it has not and to provide recommendations for moving this research forward. As noted, HR systems research overall suggests a positive relationship between HR systems and performance. However, the findings of this review show that the conclusion that research to date shows that HR *systems* are effective may be misleading. In most studies, conceptualization and measurement do not match the core theoretical assumption of complementarities or synergies between HR practices in a

system. Thus, while the empirical evidence so far may suggest that we can draw the broad conclusion that "investments in some broad set of HR practices yields returns," which practices this entails and whether and how practices jointly affect outcomes remains unclear. In addition, the measures used have problems and increasingly confound HR systems with related concepts and outcomes; thus, it is not always clear whether it is indeed the HR system causing effects. Finally, insufficient attention is paid to how differences between levels affect the meaning of the HR system construct. Overall, this makes it unclear exactly what is responsible for the found performance effects in HR systems research and shows we still know little about the theorized "systems" element or how synergies and interactions in an HR system operate.

Our review shows that despite earlier calls to study more specific and targeted systems (e.g., Lepak et al., 2006), approaches to measuring and combining HR practices in a system have moved even further towards a focus on broad undifferentiated HR systems. Our findings also show that over time, agreement in the field on how to measure HR systems has declined and confounding has increased, and it remains unclear which (sets of) practices drive the system's effect at different levels. Also, despite calls to address nonadditive effects (e.g., Chadwick, 2010), the use of additive approaches to combine HR practices in a system has increased rather than decreased recently. Research thus still provides only limited insight into the core theoretical assumption of complementarities or synergies between HR practices. In addition, theory on HR systems implicitly assumes that the HR system is influenced and shaped by time. Some first studies suggest that practices indeed vary in the timing of their effects and that effects of practices are likely to be nonlinear (e.g., Birdi et al., 2008; Piening et al., 2013), suggesting that cross-sectional studies may (at times) yield inaccurate results. While some progress has been made in showing causal effects of HR systems using additive indices, longitudinal studies have hardly examined the "system" element of HR systems over time. As very little explicit attention is paid to interrelationships between practices in a system over time, our understanding of how interrelationships between practices in HR systems develop and change is very limited.

The importance of (differences and differentiating between) levels in HR systems was noted earlier (Arthur & Boyles, 2007), and HR systems are increasingly studied at different levels, adding complexity to the conceptualization and measurement of HR systems. While this implies progress in terms of moving beyond considering only the organizational level, theorizing around HR systems at multiple levels has yet to follow suit, as even in studies measuring at the individual level, by far most theory (95%) is still focused exclusively on the organizational level. Misalignment between the level of the method and analyses and the level of theory can yield artefactual results, with found relationships being inaccurate because they do not capture meaningful variation at the right level (Klein, Dansereau, & Hall, 1994). Thus, more specificity in theory on the HR system at different levels is essential to move the field forward.

Over 80% of the studies use HR system measures that are new or are adapted from other scales and that have not received extensive scale validation, so empirical evidence that measures actually tap the intended constructs is limited (McGrath, 2005; Smith, 2005). The item types used are increasingly mixed, resulting in ambiguous scales with heterogeneous items that may not represent the same underlying construct (cf. Strauss & Smith, 2009). Also, there is a general trend over time towards the use of more perceptual and evaluative measurement: the use of individual employee respondents to rate the HR system and of individual item referents is increasing (focusing on the respondents'

individual experience rather than common experiences of the group), and more Likert scales and evaluative items are being used.

Overall, the broad and heterogeneous conceptualization and measurement of HR systems and lack of clarity in levels introduces theoretical and empirical imprecision because variation on the construct may represent variation in any or all of its levels or dimensions (Edwards, 2001; Smith et al., 2003; Strauss & Smith, 2009). This imprecision, which our review suggests is generally increasing rather than decreasing, hinders further theory development on HR systems. Theoretical progress in any field is typically characterized by construct refinement. Over time, distinctions between dimensions often become increasingly clear and constructs become more differentiated, and as a result, broader constructs become less useful (Edwards, 2001) and more rigorous empirical tests are necessary for scientific advancement (Schmidt & Pohler, 2018). In HR systems research, however, rather than a trend towards more specific theory development and related increasing precision in measurement, for example, by differentiation between different possible targeted systems, we see a trend towards even broader and less clear HR system constructs and operationalizations. From our analysis, we signal two main and interrelated areas that need specific attention in future work on HR systems to move the field forward in terms of construct refinement and building more knowledge on how HR practices combined in "systems" affect outcomes: measuring and combining practices in an HR system and conceptualizing and measuring the HR system at different levels. Below, on the basis of our review, we highlight problems in current empirical studies related to both of these areas, and for both, we offer a framework aimed to aid scholars in refining theory and matching conceptualization and measurement.

How to Measure and Combine Practices in an HR System

The first choice researchers need to make when designing a study on HR systems is which type of HR system to focus on. Despite earlier calls in the literature for more clarity and consistency in HR system labels and content (e.g., Lepak et al., 2006), our review shows that the terminology used to label HR systems has become increasingly unclear. Whether researchers study high performance, commitment, or involvement HR systems or focus on more strategically targeted HR systems, terms for these HR systems are not used consistently, and the definitions of such systems and differences between them are not clearly outlined. One can question whether different labels indeed always represent different systems or whether just as often, different labels are used for highly similar systems. Proliferation of different terms for the same concept is problematic because some researchers may see these as similar whereas others do not, and it raises questions about the cumulative understanding of the concept because the evidence is spread over research on concepts that are labeled differently, which inhibits conceptual progress of the field (see e.g., Podsakoff et al., 2016). Also, when systems with the same label are measured differently, the results of such studies may not be comparable. Our findings suggest that a clear label and definition, explaining the system's target and how the concept is similar and different from related constructs, is thus an important first step for researchers to take in theorizing and measuring the HR system.

What to Measure?

In contrast with the suggestion of some authors a decade ago that a growing consensus on the elements of an HR system existed (e.g., Lengnick-Hall et al., 2009), which would have

signaled construct refinement, our results show that the field has not progressed in terms of deciding which practices should be included in an HR system and why. Agreement on which practices should be included has even declined over time. If this were the case as the result of the development of multiple targeted HR systems that clearly include different practices, this would form progress; however, this type of increasing precision is not seen.

Our findings suggest that it is increasingly unclear what authors consider to be and not be an HR practice. If a measure includes items or dimensions that are not prototypic of the construct (e.g., a high performance HR system) but instead reflect a correlated construct (e.g., transformational leadership), the results may be misleading, as the measure reflects more than one construct (Smith et al., 2003). The results then may be driven by the related construct rather than the HR system. At the same time, if important dimensions of the constructs are not included in the measure, this can also lead to confusion and inaccurate prediction (Smith et al., 2003). For example, when a measure does not include training when in reality training has a large influence, the results may be misleading too. Interestingly, it is surprisingly difficult to find a clear definition of HR practices in the literature. Rather than defining HR practices, authors either take this for granted or generally refer to programs, policies, or actions of firms aimed at managing their HRs. Clarifying the boundaries of an HR system and of what is and is not an HR practice in it is needed and important to avoid contamination of the HR system concept (cf. Podsakoff et al., 2016), which our review suggests is a problem in many current studies.

Looking at the types of practices included in previous reviews (e.g., Lepak et al., 2006; Posthuma et al., 2013), most authors seem to agree that HR practices should refer to organizational actions or processes and job characteristics that focus on attracting, developing, and motivating employees and providing opportunities to contribute. However, contamination with other constructs is seen in several ways. First, our findings show an increased inclusion of individual leader behaviors (e.g., transformational leadership, supervisor support) in more recent HR systems measures, as well as increased inclusion of the (strategic) role of HR, such as the presence of an HR unit or HR integration with strategy. Both are generally not considered HR practices, as they do not represent organizational actions directed at employees. Second, constructs that can be seen as outcomes, such as trust, identification, skill level, loyalty, close ties, climate, and organizational effectiveness, are also increasingly considered part of HR systems. In the opportunity-enhancing bundle of the AMO model, and when HR systems are measured at the individual level, confounding HR practices with outcomes is particularly prevalent. For example, job rotation, teamwork, and participation structures such as suggestion systems form HR practices in this domain; however, other elements such as experienced work pressure and whether employees feel empowered are also included in HR systems measures, while these form outcomes rather than practices. Both forms of contamination are problematic, as in such studies, it can become unclear what exactly is responsible for the observed relationships.

To avoid concept proliferation and contamination, researchers need to first clearly define the HR system and whether the HR system is general or targeted to a specific outcome. Then, to measure the HR system, we suggest researchers select HR practices when they are organizational actions, processes, or job structures that directly affect employees and relate to system goals and leave out other broader structures and processes as well as attitudes, feelings, and behaviors of leaders and (groups of) employees. To choose which HR practices to include in the system, going forward, we suggest that studies would generally at least measure the six most widely adopted practices as shown by our review: *training and development, participation/ autonomy, incentive compensation, performance evaluation, selection*, and *job design*. These

include most of the core practices identified by Posthuma et al. (2013), who focused specifically on high performance HR systems. Inclusion of these common practices will enhance comparability of studies. This does not mean that all six practices are expected to relate to all possible outcomes or need to be at the core of all hypothesized systems. For example, in high involvement HR systems that aim to maximize current employees' involvement, selection may be less important, which can be hypothesized and tested. Such predictions and tests can help to build specific theory on the role of different HR practices in a system. Also, depending on the system's target, additional practices can be added, including a clear theory-based justification of why these are relevant. For example, when measuring a service-oriented HR system, Chuang and Liao (2010) add work-life balance–related practices because of the focus on employee and customer needs. Another way to focus the system measure on a specific target can be to use criterion-focused items (e.g., selection for creativity).

At the item level, it is important that all items should have a conceptual connection with the specific HR systems construct. The HR system originates at the organizational level, reflecting organizational actions towards employees. Thus, the most appropriate item types for such higher-level constructs are generally items that use the organization as the item source, that use the group as the item referent, and that are descriptive (e.g., the organization offers continuous training to employees), as these item characteristics have been shown to increase within-group agreement of higher-level constructs (Kozlowski & Klein, 2000).

How to Combine Practices in Systems?

A next important question is how HR practices are combined in a system. Despite aforementioned calls to use other ways to combine that allow, for example, for synergistic effects, only 15% of the studies we reviewed use alternative ways of combining HR practices into systems, and the use of these approaches has decreased over time. We suggest that in order to advance knowledge on HR systems, considering specific relationships between HR practices is important, and we should thus move away from an exclusive focus on the broad overall construct. The few available studies comparing multiple approaches to capturing synergies suggest that different ways of combining reflect different theoretical propositions and lead to different outcomes (Chadwick, 2010; Delery & Gupta, 2016). Different approaches can thus help to advance knowledge on specific relationships between HR practices in a system. However, so far, there has been limited attention for which analytical technique fits best with which underlying theoretical idea, and our review suggests different ways in which more specific theory on complementarities and synergies between HR practices in a system can be built. On the basis of our findings, we offer a framework with several key questions and describe associated areas for research aimed at building more specific understanding of interrelationships between HR practices in a system (see Table 4). In describing each question, we highlight the key assumptions and describe what to measure and how to combine practices in a system when using nonadditive approaches.

Weighting

Our review suggests that we lack knowledge on which HR practices in systems are relatively more important and why. Weighting assumes that some practices may be more important than others in explaining outcomes, depending, for example, on the context, type of employees, or the outcome. When explaining human capital, training may have a relatively strong effect,

Table 4
Four Approaches to Interrelationships Between Human Resource (HR) Practices in a System

	Research Question	Key Assumption	What to Measure?	How to Combine Practices in a System?
Weighting	Which HR practices are relatively more important?	HR practices vary in their importance in explaining outcomes.	Six common HR practices plus other practices that are relevant for the system's target	Multiple regression, HR system as a latent factor, relative weights analysis
Configurations	Which practices are typically combined in a system? What are the consequences of deviating from an ideal HR system?	There are different (equally effective) profiles of HR practices, and deviating from an ideal profile of HR practices is related to lower outcomes.	Six common HR practices plus other practices that are relevant for the system's target	Cluster analysis, latent class analysis, profile deviation
Interactions	Which (sets of) practices are substitutes or enhance or diminish each other's effects?	The effectiveness of one practice depends on the other practices in place.	Limited set of HR practices, theoretically driven	Interactions
Necessary and Sufficient Practices	Which practices need to be present for the system to be effective, and which practices make the difference between average and good performance?	Certain HR practices in a system may be essential, nonessential, or even counterproductive.	All available HR practices need to be included.	Fuzzy set qualitative comparative analysis, necessary condition analysis, sequential tree analysis, etc.

whereas when focusing on motivation as an outcome, rewards may have a stronger influence. To examine weighting, we suggest measuring the six most common HR practices, adding other practices that are theoretically relevant depending on the system's goal. Techniques such as multiple regression or modeling the HR system as a latent factor can be used to assess relative differences in the effects of practices within a system. Such research can yield more insight into how, why, and when HR practices vary in importance within the system. The relative influence of HR practices in a system may change over time, so timing of measurement is important. A cross-sectional design may suggest that certain practices do not have an effect, while this effect may not have occurred yet or is already gone or declining. Some sets of practices might have immediate effects on outcomes (e.g., reward for performance), whereas others might take more time (e.g., skill development), which has not yet received sufficient research attention.

Configurations

Configurations focus on which practices are typically combined in a system. The underlying assumptions are that there are different (equally effective) profiles of HR practices and that deviating from ideal HR systems is less effective. These assumptions originate from

configurational theories, which assume that relationships between HR practices in a system are nonlinear and synergistic (e.g., Meyer, Tsui, & Hinings, 1993). Although the configurational approach has long formed an important mode of theorizing in SHRM (e.g., Arthur, 1994; Delery & Doty, 1996), our review shows that it was not yet applied in many studies to date; thus, knowledge on configurations of HR practices and their consequences is limited. Going forward, research could extend previous work on HR system configurations by focusing on questions such as whether and how strategy influences the HR configuration, how HR configurations within organizations develop over time, or whether other characteristics such as the type of work or the phase of development of the company play a role. In addition, research can further examine what the consequences are of deviating from an "ideal" HR system. Does deviating from the ideal have more negative consequences for some HR practices than for others (i.e., does it matter more when less training is offered than when selection is less extensive)? To examine these types of issues, we suggest a broad set of HR practices—including the six most common HR practices and additional practices that are relevant for the system's goal—in order to capture any important combinations of HR practices. Cluster analysis or related techniques can be used to identify clusters or profiles of HR practices, and techniques such as profile deviation can be used to examine the consequences of deviating from the (ideal) profiles or clusters.

Interactions

Interaction approaches specify types of relationships between practices, testing the core assumption that the effectiveness of one practice depends on the other practices in place. Which (sets of) practices enhance or diminish each other's effects or form substitutes? Although interaction is a core assumption in research on HR systems, knowledge of such interactive relationships is limited, and studies that do examine interaction effects often do not have specific predictions regarding the type of interactions between practices. To move forward, future research can build on literature on interactions to create more specific predictions of the potential interactive relationships between the practices in an HR system. For example, in their taxonomy of interaction effects, Gardner, Harris, Li, Kirkman, and Mathieu (2017) distinguish between linear and quadratic interaction effects, and for each of these, they distinguish strengthening, weakening, and reversing effects, which may also hold for HR practices. Future work can examine interactions between (bundles of) practices or which practices weaken, reverse, substitute, or enhance each other's effects to enhance knowledge on how "powerful connections" and "deadly combinations" (Becker, Huselid, Pickus, & Spratt, 1997) operate.

Going beyond previous work, exploring potential nonlinear interactions between practices is also of interest. For example, do some synergies become stronger or weaker, or do they occur only at higher levels of certain practices? Or as a more concrete example, are the effects of performance-related pay strong only when performance appraisals are used intensively? Taking timing of the effects into account may also imply that several of the complementarities between practices can occur over time rather than in the same moment. How different HR practices with different timing of effects combine in a system forms an interesting new area for the field to address. For example, selection can have consequences for subsequent training needs. This means that besides complementarities at one point in time, causal complementarities should be considered. Also, HR practices themselves may change over time, which is in need of attention. Synergies between individual practices in a system

may suffer when one or more practices are changed, removed, or replaced by other practices, and for some types of interactions, a change may matter more than for others. Also, as with any change, learning may be needed, and when a practice is newly introduced or changed, it may take time before the desired effect occurs. Effectiveness may even decrease at first before going up. To capture this empirically, future research can examine how different combinations of HR practices and their effects change over time or how effects of earlier HR practices influence the effectiveness of subsequent practices.

Interactions are typically best examined by selecting a limited set of relevant HR practices, and the specific combinations tested should be driven by theory (cf. Gardner et al., 2017). Some studies, for example, identify one or more HR practices that are most important for achieving strategic goals or that are characteristic for the sector, which form the core practices of the system. For example, De Grip and Sieben (2009) identify performance evaluation as the core practice for the pharmacy sector. Then, interactions between the core practice and a narrow set of theoretically relevant noncore practices can be examined in order to capture synergies between practices.

Necessary and Sufficient Practices

Examining which practices need to be present for the system to be effective and which practices make the difference between average and good performance can also help to enhance knowledge about HR systems. Several studies develop a hierarchy of practices, which seems a promising area for future research, as this can help uncover which combinations drive a system's effects. Of these studies, a few empirically derive a hierarchy, for example, using fuzzy set qualitative comparative analysis to identify which practices are core practices and which are peripheral, based on the idea that some practices need to be present for achieving an outcome even if they may not be sufficient for achieving that result (cf. Dul, 2016). Another interesting area that has hardly been explored yet is whether some practices can hurt the effectiveness of others. For example, besides distinguishing between core and peripheral practices, Meuer (2017) also finds nonessential practices and practices that have to be absent because they lower the effect of the system. This too cannot be captured using additive indices and suggests that focusing all research only on a limited set of practices may be too narrow and may not reveal all complementarities between practices.

Knowing which practices drive the system's effect(s) and which are nonessential may also have important consequences for the efficiency and cost-effectiveness of HR systems. To date, the costs of HR practices have rarely been considered in studies on HR systems. However, very broad systems may be unnecessarily expensive when the same effects could be achieved using fewer or less costly practices. Knowing the costs (and returns) of different HR practices will help to make decisions about investments that are made in the HR system. For example, when two practices are substitutes, it may be valuable to be able to compare the costs of alternative practices. Weighting practices by cost in the HR system indices or utility analysis may form an interesting way to start addressing this issue empirically.

Future research can thus examine which practices are core practices and need to be present in order for the system to be effective and which are peripheral, nonessential, or even counterproductive. Here, too, time may play an important role. Individuals and organizations

at different times or in different stages may have different needs or preferences regarding HR practices. For example, individuals in different career or life stages may benefit more or less from certain practices, and a startup may have different core practices than an established company, so depending on such stages of development, different (sets of) practices may drive the system's effect. To date, theorizing and measuring a set of HR practices and their outcomes (correlates) is mostly done at one point in time, and we have not yet addressed many of the potential dynamics involved. To examine necessary and sufficient practices, we suggest including a broad set of HR practices in the HR system measure, the six common HR practices, and preferably all other practices that are implemented in the study's context to enable capturing all possible interrelationships. Statistical techniques such as fuzzy set qualitative comparative analysis, necessary condition analysis, or sequential tree analysis can be used to examine these types of questions.

Implications for Future Research

The framework presented above offers different ways of building a better knowledge base addressing specific interrelationships between HR practices in a system in order to better match HR systems research with the underlying assumption of complementarities between practices. Our framework extends previous work on synergies (e.g., Chadwick, 2010; Delery, 1998) by including additional ways of capturing synergies or complementarities such as necessary and sufficient practices, by calling for more work on temporal dynamics, and by suggesting which practices to focus on and how to measure these in order to enhance precision as well as comparability across studies. Examining the different questions suggested by the framework enables researchers to build more specific theory and evidence on how practices interact within HR systems, which practices are essential and which are not, and how time affects interrelationships between practices in a system. Building theory based on the idea that practices in a system may be essential, nonessential, or even counterproductive has consequences for the HR system concept structure. The additive approach to measure the HR system represents a family resemblance concept, where each item has attributes in common with one or more other items (Podsakoff et al., 2016), so it matters less which of the practices are present and which are absent. In contrast, necessary and sufficient concepts are defined by sets of individually necessary and collectively sufficient attributes. Thus, developing a hierarchy of practices fits with this concept structure. Shifting from a family resemblance to seeing HR systems as having a necessary and sufficient concept structure would provide opportunities to build better theory on which practices are most important for the effects of the HR system and which support these core practices.

Also, when developing more specific theory on the relationship between HR systems and outcomes, the outcome which the HR system intends to affect becomes more important. Our findings show an increased variation in outcome types over time. While most older studies used organizational performance as the outcome, recent studies address various outcomes, ranging from individual well-being to organizational innovation and flexibility. It is of course of interest to see how HR systems affect these different outcomes. However, these outcomes are very different in nature; thus, which outcome is considered when theorizing and testing the effects of HR systems matters, and including more than one outcome can be of interest. Studies could, for example, include two potentially competing outcomes (e.g., performance and well-being; efficiency and innovation) to examine differences in how a given set of

practices affects both. This, for example, allows asking whether the same or different practices drive performance and innovation or whether the positive impact of a system on performance comes at a well-being cost. For each of the research questions in Table 4, the specific target or outcome can inform theoretical predictions about which (sets of) practices should have most influence and why, how these practices interrelate, and which practices should not affect outcomes or are even counterproductive.

HR System Levels and Appropriate Measurement

A little over a decade ago, authors started to discuss the importance of how employees perceive HR systems and how such perceptions might differ from the organization's intentions (e.g., Bowen & Ostroff, 2004; Nishii & Wright, 2008; Purcell & Hutchinson, 2007). Since then, we see a strong increase in studies examining employee perceptions of HR systems and a simultaneous trend towards more evaluative measurement. However, the HR system has different meanings at different levels (Arthur & Boyles, 2007) and, thus, forms different constructs at these levels. Our review shows several issues that relate to this. For instance, asking key informants such as HR managers to report on HR practices (intended policies), asking managers about the practices they implement in their unit (implemented practices), or asking employees to report on their personal experience with these same HR practices (employee perceptions of practices) form three valid albeit noninterchangeable approaches. When and why these views are aligned or not also forms an interesting area of study. However, ratings of different sources are often combined in a single score even if they represent different levels, making it unclear at which level the system is conceptualized. Also, while studies increasingly focus on the individual level of measurement, our review shows a large (and increasing) variation in item types, particularly for employees as respondents. Theory building on levels other than organizational is still largely lacking. Going forward, levels need to be taken into account more explicitly in theorizing. Our review suggests that there are five distinct perspectives on HR systems at different levels that are currently mixed in empirical studies. Below, we offer a second framework to distinguish between these five perspectives and their measurement, using three variability assumptions proposed by Klein et al. (1994), namely, homogeneity, heterogeneity, and independence. Table 5 summarizes key differences between levels and highlights differences in assumptions, types of hypotheses, and measurement of the HR system at these different levels.

Assuming homogeneity implies the level of the theory is the group, group members are assumed to be sufficiently similar to characterize the group as a whole, and theory focuses on variation between groups. An example hypothesis is that an organization's HR system is positively related to organizational performance. When assuming heterogeneity, the level of the theory is individuals within a group (or groups in an organization), and theories focus on within-group effects, linking within-group variability in one construct to within-group variability in another. For example, employees who relative to their work group have more positive perceptions of the HR system are more committed. Assuming independence means that the value of the construct for one individual should be independent of its value for others. Between-individual variability in one construct is related to between-individual variability in another construct, for example, individual satisfaction with the HR system is positively related to organizational commitment. The underlying assumption has consequences for how constructs are conceptualized and which propositions can be derived, as well as for

Table 5
Five Perspectives on Human Resource (HR) Systems at Different Levels

Concept	HR System Level				Outcome
	Intended HR System	Manager-Rated HR System	Collective Employee Perceptions of the HR System	Employee Perceptions of the HR System	Employee Attitudes Towards the HR System
Variability Assumption	Homogeneity	Homogeneity or heterogeneity (groups within organizations)	Homogeneity	Heterogeneity (individuals within groups)	Independence
Example Hypothesis	An organization's HR system is positively related to organizational performance.	Manager-rated HR systems are positively related to team performance (homogeneity). The relationship between intended HR systems and manager-rated HR systems is moderated by manager's implementation quality (heterogeneity).	An organization's HR system is positively related to organizational performance.	The relationship between manager-rated HR systems and employee perceptions of HR systems is moderated by manager communication.	Employee satisfaction with the HR system is positively related to employee motivation.
Level	Organization/unit	Work group	Organization/unit, work group	Individual	Individual
Proposed Data Source	HR/higher-level managers	Line manager/team manager	Employees	Employees	Employees
Proposed Answer Scale	Presence/coverage, Likert-type scale (frequency)	Presence/coverage, Likert-type scale (frequency)	Presence, Likert-type scale (frequency)	Presence, Likert-type scale (frequency)	Likert-type scale (frequency or agreement)
Item Source	Organization/unit	Organization/unit, management	Organization/unit, management	Organization/unit, management	Organization/unit, management
Item Referent	Group, job	Group, job	Group, job	Group, job, individual	Individual
Item Focus	Descriptive	Descriptive	Descriptive	Descriptive	Evaluative

measurement and analysis, because if these are not consistent with the variability assumption, one may draw erroneous conclusions from the data (Klein et al., 1994). For example, while most studies using employees as raters of the HR system adopt a homogeneity assumption focused on common experiences in the group, HR system measures in such studies often focus on individual experiences, use many evaluative items, and often do analyses at the individual level, thus capturing between-individual rather than between-group differences. To clarify this, we propose five different perspectives on HR systems: intended HR system, manager-rated HR system, collective employee perceptions, individual perceptions, and (as the most proximal outcome) employee attitudes towards the HR system.

Intended HR System

This is the designed organizational-level HR system (Nishii & Wright, 2008) rated by key informants such as HR managers or higher-level managers. When studying the intended HR system, homogeneity is assumed, linking organizational-level HR systems to organizational-level outcomes. Measurement of higher-level constructs should be mostly descriptive to decrease within-group variability in responses. Thus, we suggest using the organization as the item source,

a group referent, descriptive items (Kozlowski & Klein, 2000), and more descriptive answer scales (presence, coverage, or Likert-type scales focusing on frequency rather than agreement).

Manager-Rated HR System

When the HR system is rated by line managers, studies may take two theoretical approaches. When adopting a homogeneity assumption, group-level HR systems are related to group-level outcomes, and when adopting a heterogeneity assumption, variability in manager-rated HR systems relative to each other is explained. Such within-group effects are relevant in light of the increase of multilevel studies (Peccei & Van De Voorde, 2019). Theoretical work on intended, implemented, and employee perceptions of HR systems typically uses cross- or multilevel theory (e.g., Nishii & Wright, 2008). Such models link the HR system—as a homogeneous organizational-level construct—to HR systems implemented by managers and individual-level perceptions, attitudes, and behaviors and suggest that variability increases when HR systems are implemented in the organization. Conceptually and statistically, a homogeneous group-level construct cannot explain within-group variance. Yet individual-level sources of within-group variation could be identified (moderators), explaining how and why individuals perceive and respond differently to group-level characteristics, thus adopting a heterogeneity assumption, for example, by explaining which factors affect the relationship between intended and manager-rated HR systems. Measuring the manager-rated HR system is—other than the data source—similar to the intended HR system, as it concerns a higher (group) level construct.

Collective Employee Perceptions of the HR System

Our framework distinguishes between three types of employee-rated HR systems. The first is collective employee perceptions, at the organization/group level. Although several studies acknowledge that employee perceptions of HR systems may differ between individuals, most studies treat employees as homogeneous, for example, using social exchange theory to suggest that an HR system represents a long-term investment in employees, which employees reciprocate by showing more effort. When assuming homogeneity, the focus is on common experiences of the group, making it appropriate to use a group referent, the organization/unit as the item source, and more descriptive items. A referent-shift composition model can be used (Chan, 1998), orienting the HR system measure to the group level to test whether employees in a job group share consistent perceptions of HR systems. Kehoe and Wright (2013) and Wu and Chaturvedi (2009), for example, use such a model and aggregate individual HR system perceptions to the group or organizational level.

Employee Perceptions of the HR System

When using employees as raters of the HR system measure while assuming heterogeneity, the focus shifts to individual perceptions of the HR system relative to the group. Here again, multilevel models are relevant that explain differences between manager ratings to employee perceptions of the same HR system. For example, manager communication and demographic similarity were shown to strengthen the relationship between implemented and perceived HR systems (Den Hartog et al., 2013; Jiang, Hu, Liu, & Lepak, 2017), yet research that examines sources of variation at the different levels within

the HR system is still scarce. Future research could examine which factors enhance the similarity between manager-rated and perceived HR systems to further develop cross-level theory. Here too, measurement typically will focus on the organization/unit, using descriptive items. Depending on the research question, the item referent could be the group/job or the individual.

Employee Attitudes Towards the HR System

Using employees as raters of the HR system measure, adopting an independence assumption implies that the focus is on differences between individuals, independent of the group. However, as theory assumes that the HR systems construct originates at the organizational level, and is implemented and communicated within the organization, the group- or organization-level HR system should always play a role. Our review shows that only 15 studies explicitly use an independence assumption, and these all focus on employee attitudes towards or evaluations of the HR system. We propose to explicitly distinguish between employees' perceptions of and their attitudes towards the HR system. Employee attitudes towards the HR system should not form part of the HR system construct, not even at the perceived level, but should be treated separately as a proximal attitudinal outcome of (employee perceptions of) the HR system. To measure this outcome, researchers can use individual item referents and evaluative items. For example, some studies asked respondents how satisfied with or motivated they feel by a set of HR practices (e.g., Runhaar, Sanders, & Konermann, 2013).

Implications for Future Research

So far, researchers lack a shared terminology of HR systems at different levels. Our framework offers a common language for studying HR systems at different levels and based on the findings of our review, extends previous work on HR system levels (e.g., Arthur & Boyles, 2007) by also proposing distinct approaches to employee-rated HR systems that have received increasing attention the past decade and by including specific suggestions for measurement at the item level. We hope this can facilitate construct-valid measurement at each level, allow integration of research findings, and suggest new research avenues. The five perspectives are distinct in terms of the appropriate theorizing and measurement. Each perspective can help answer different questions, each tests different core assumptions, and all are potentially important in HR systems research. Moving forward, it is important that researchers clearly specify which type and level they are focusing on. Doing so can help build knowledge on each of the different types, driving concept refinement and theory building at each of these levels as well as enhancing understanding of effects across levels.

An interesting question related to the differences between HR systems at different levels is the relative importance of practices in HR systems at different levels. Our review shows that employee-rated (perceived) HR systems typically contain similar practices as the (intended) systems rated by (HR) managers. However, due to differences in interpretation or salience of the items included in the measure, the measure may have a different meaning and a different structure across groups (Furr, 2011). The relative importance of the dimensions may differ, or the dimensions may be different altogether (Tay, Woo, & Vermunt, 2014). As noted, not all HR

practices may be equally relevant or salient for (all) employees, for example, HR planning or selection may be visible only to some. Thus, when studies show that some practices within a system have a weaker or stronger effect, this may be due to differences in visibility and relevance for different rater groups rather than differences in effectiveness. Such differences may actually be an important source of information that requires further exploration rather than being evidence of measurement error (McGrath, 2005). Future work can build theory and examine differences in the relative importance of (sets of) practices for key informants (e.g., employees, managers).

Another interesting area for future research relating to levels is whether informal practices are offered. While theory assumes that HR systems are designed at the organizational level, not all practices that are received or perceived by employees need to form part of the formal HR system (as intended by the organization). For example, there may be practices that are not part of the intended HR system but that are offered by line managers on their own initiative or negotiated by employees to fit their specific needs and wishes (e.g., I-deals). For example, Yanadori and van Jaarsveld (2014) distinguished between "formal" HR practices, which are present (reported by managers) and employees participate in (reported by employees), and "informal" HR practices, which are not present in the organization but employees do participate in. Both formal and informal HR practices were similarly positively associated with satisfaction and profitability, suggesting that the formality and informality of practices might both have positive effects. Future work is needed on when this might not be the case and more generally, on what the role is of practices that are not part of the intended HR system. For example, which practices do managers offer beyond formal HR systems? How do I-deals influence wider HR system perceptions?

Examining the HR system at different levels also matters for the timing of the effects. Differences in time scales affect the nature of links among levels. Lower-level constructs and processes tend to have more rapid dynamics than higher-level ones, which makes it easier to capture change in lower-level relationships (Kozlowski & Klein, 2000). Thus, the effects of employee attitudes towards the HR system on individual outcomes are expected to occur sooner than the effects of organizational-level HR systems on (organizational) outcomes. Thus, in designing HR systems studies, differences in time lag at the different levels is of interest.

Conclusion

We reviewed the empirical research on HR systems to date to identify trends and progress over time and to pinpoint areas where progress is lacking. We used the findings to identify directions for future research aimed toward further understanding of how interrelationships between practices in an HR system affect outcomes (summarized in Table 6). Most research to date does not align with the fundamental assumption of synergies between HR practices in a system. The problems our review highlighted in conceptualization at different levels, measurement, and combining practices into systems hamper progress of the field in terms of understanding the "system" element of HR systems. We offered two frameworks aimed at enhancing conceptual clarity and construct refinement. The suggestions for future research from these frameworks can help to develop less ambiguous and more rigorously developed measures and build more specific theory and evidence on how practices interact within HR

Table 6
Directions for Future Research

	Research Needs and Sample Research Questions
Measuring and Combining Practices in a Human Resource (HR) System	– Research needs to avoid concept proliferation by using clear conceptual definitions, explaining the system's target and how the concept is similar and different from related constructs. – Future research should limit the focus on HR practices to organizational actions or processes and job characteristics that focus on attracting, developing, and motivating employees. – Future research needs to separate HR practices and systems from related concepts, such as leadership and the role of HR, and outcomes. – Future studies could at least measure the six core practices as shown by our review: training and development, participation/autonomy, incentive compensation, performance evaluation, selection, and job design. – Depending on the system's target, additional practices can be selected, including a clear justification of why they are expected to be relevant. – Future studies should measure the (organization-level) HR system using items that use the organization as the item source, use the group as the item referent, and are descriptive. – Future work could build more specific theory on interactions of practices within HR systems by examining weighting, configurations, interactions, and necessary and sufficient practices (see Table 4). – Time needs to be incorporated in research in different ways. Which (combinations of) HR practices have short-term and which have long-term effects? Are effects of HR practices nonlinear over time? How do the effects of different combinations of HR practices change over time? What is the impact of a change in one practice on the effectiveness of other practices and of the whole system? – Future research should start to take costs of the HR practices into consideration. Can similar system effects be achieved with fewer or less costly practices? – Future research could weigh practices by cost in the HR system indices or use utility analysis. – Theory building is needed to enhance understanding of why HR systems affect outcomes, with increasing use of a necessary and sufficient concept structure. – More understanding of the effect on different types of outcomes is needed. Why do HR systems affect different types of (potentially contradictory) outcomes?
HR System Levels and Appropriate Measurement	– Future research should align the level of theory, measurement, and analysis. – Research should not mix item types in one measure. – Future research could distinguish between five different perspectives on HR systems: intended HR system, manager-rated HR system, collective employee perceptions, individual perceptions, and employee attitudes towards the HR system (see Table 5). – Researchers should clearly specify which type and level they are focusing on and how this relates to their specific research question in order to drive concept refinement and theory building at each of these levels. – Future research could examine which factors enhance the similarity between intended, manager-rated, and perceived HR systems in order to further develop cross-level theories on HR systems. – Future work should differentiate between practices that are relevant and visible for (HR) managers and employees in order to build meaningful concepts on each level. – Future research should study the nature, antecedents, and consequences of informal practices. – Future work should investigate when, why, and how practices that are not part of the intended HR system are offered and perceived. How do I-deals influence wider HR system perceptions? Which factors affect what practices managers offer beyond the formal HR system? – Future studies should take differences in time lag between the different levels into account.

2529

Appendix
Types of Relationships Between Practices

	1991–2000	2001–2005	2006–2010	2011–2015	2016–2017
Cluster Analysis	(Arthur, 1992, 1994; Cutcher-Gershenfeld, 1991; Doeringer, Evans-Klock, & Terkla, 1998; Kalleberg & Moody, 1994; Nyambegera, Sparrow, & Daniels, 2000; Osterman, 1994)	(Bae & Yu, 2005; Camelo, Martín, Romero, & Valle, 2004; De Saá-Pérez & García-Falcón, 2002; De Saá Pérez & García-Falcón, 2004; Doeringer, Lorenz, & Terkla, 2003; Ferratt, Agarwal, Brown, & Moore, 2005)	(De Saá-Pérez & Díaz-Díaz, 2010; Horgan & Mühlau, 2006; De Menezes & Wood, 2006; Toh, Morgeson, & Campion, 2008; Zheng, O'Neill, & Morrison, 2009)	(Kato & Owan, 2011; Leffakis & Dwyer, 2014; Marti-Audí, Valverde, & Heraty, 2013; Sgobbi & Cainarca, 2015; Yoon & Chae, 2012)	(Ogbonnaya, Daniels, Connolly, & Van Veldhoven, 2017, two studies)
Profile Deviation	(Delery & Doty, 1996)	(Ahmad & Schroeder, 2003)	(Jimenez-Jimenez & Martinez-Costa, 2009; Jimenez-Jimenez & Sanz-Valle, 2008; Verburg, Den Hartog, & Koopman, 2007)	(Arthur, 2011)	—
Interaction	(Delaney & Huselid, 1996; Pérotin & Robinson, 2000; Whitfield, 2000; Wright, McCormick, Sherman, & McMahan, 1999)	(Cappelli & Neumark, 2001)	(Birdi et al., 2008; E. Chang, 2006; Frenkel & Lee, 2010; Gould-Williams & Gatenby, 2010)	(Arthur, 2011; Darwish, Singh, & Mohamed, 2013; Frick, Goetzen, & Simmons, 2013)	(Andreeva, Vanhala, Sergeeva, Ritala, & Kianto, 2017; Conway, Fu, Monks, Alfes, & Bailey, 2016; Pavlov, Mura, Franco-Santos, & Bourne, 2017; Singh, Darwish, Wood, & Mohamed, 2017)
System Score Based on Presence, Absence, or Level of Specific HR Practices	(Delaney & Huselid, 1996; Hoque, 1999; Ichniowski & Shaw, 1999; Michie & Sheehan, 1999; Roche, 1999)	(Bryson, Forth, & Kirby, 2005; Forth & Millward, 2004; Guest, Michie, Conway, & Sheehan, 2003; Laursen & Foss, 2003; Michie & Sheehan-Quinn, 2001)	(Brown, Forde, Spencer, & Charlwood, 2008; Chenevert & Tremblay, 2009; De Grip & Sieben, 2009; Kauhanen, 2009; Mohr & Zoghi, 2008)	(Böckerman, Bryson, & Ilmakunnas, 2012, 2013; Bonet, 2014; Davies, McNabb, & Whitfield, 2015; Kroon, Van De Voorde, & Timmers, 2013)	(Laroche & Salesina, 2017)
Weighted Measure	(Galang, 1999)	(Guthrie, 2001)	(W. Liu, Guthrie, Flood, & MacCurtain, 2009)	(S. Chang, Gong, Way, & Jia, 2013; Guthrie, Flood, Liu, MacCurtain, & Armstrong, 2011; Marti-Audí et al., 2013)	
Latent Class Analysis			(De Menezes & Wood, 2006)	(Hauff, Alewell, & Hansen, 2014; Kim, Wehbi, Dellifraine, & Brannon, 2014)	
Other	Sequential tree analysis (Guest, Conway, & Dewe, 2004)			Standard deviation of the items (Koster, 2011)	Fuzzy set qualitative comparative analysis (Meuer, 2017)

systems, which practices are essential and which are not, how time plays a role, and how HR systems operate at different levels.

Note

1. The results using 2-year time periods and the full coding scheme are available from the authors.

References

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: 19-43.

Andreeva, T., Vanhala, M., Sergeeva, A., Ritala, P., & Kianto, A. 2017. When the fit between HR practices backfires: Exploring the interaction effects between rewards for and appraisal of knowledge behaviours on innovation. *Human Resource Management Journal*, 27: 209-227.

Ang, S. H., Bartram, T., McNeil, N., Leggat, S. G., & Stanton, P. 2013. The effects of high-performance work systems on hospital employees' work attitudes and intention to leave: A multi-level and occupational group analysis. *International Journal of Human Resource Management*, 24: 3086-3114.

Arthur, J. B. 1992. The link between business strategy and industrial relations systems in American steel minimills. *Industrial and Labor Relations Review*, 45: 488-506.

Arthur, J. B. 1994. Effects of human resource systems on manufacturing performance and turnover. *Academy of Management Journal*, 37: 670-687.

Arthur, J. B. 2011. Do HR system characteristics affect the frequency of interpersonal deviance in organizations? The role of team autonomy and internal labor market practices. *Industrial Relations*, 50: 30-56.

Arthur, J. B., & Boyles, T. 2007. Validating the human resource system structure: A levels-based strategic HRM approach. *Human Resource Management Review*, 17: 77-92.

Bae, J., & Yu, G.-C. 2005. HRM configurations in Korean venture firms: Resource availability, institutional force and strategic choice perspectives. *International Journal of Human Resource Management*, 16: 1759-1782.

Becker, B., & Gerhart, B. 1996. The impact of human resource management on organizational performance: Progress and prospects. *Academy of Management Journal*, 39: 779-801.

Becker, B., & Huselid, M. A. 1998. High performance work systems and firm performance: A synthesis of research and managerial implications. *Research in Personnel and Human Resource Management*, 16: 53-101.

Becker, B., Huselid, M. A., Pickus, P. S., & Spratt, M. F. 1997. HR as a source of shareholder value: Research and recommendations. *Human Resource Management*, 36: 39-47.

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.

Böckerman, P., Bryson, A., & Ilmakunnas, P. 2012. Does high involvement management improve worker wellbeing? *Journal of Economic Behavior & Organization*, 84: 660-680.

Böckerman, P., Bryson, A., & Ilmakunnas, P. 2013. Does high involvement management lead to higher pay? *Journal of the Royal Statistical Society: Series A*, 176: 861-885.

Bonet, R. 2014. High-involvement work practices and the opportunities for promotion in the organization. *Industrial Relations*, 53: 295-324.

Boselie, P., Dietz, G., & Boon, C. 2005. Commonalities and contradictions in HRM and performance research. *Human Resource Management Journal*, 15: 67-94.

Bowen, D. E., & Ostroff, C. 2004. Understanding HRM-firm performance linkages: The role of the "strength" of the HRM system. *Academy of Management Review*, 29: 203-221.

Brown, A., Forde, C., Spencer, D., & Charlwood, A. 2008. Changes in HRM and job satisfaction, 1998–2004: Evidence from the Workplace Employment Relations Survey. *Human Resource Management Journal*, 18: 237-256.

Bryson, A., Forth, J., & Kirby, S. 2005. High-involvement management practices, trade union representation and workplace performance in Britain. *Scottish Journal of Political Economy*, 52: 451-491.

Camelo, C., Martín, F., Romero, P. M., & Valle, R. 2004. Human resources management in Spain: Is it possible to speak of a typical model? *International Journal of Human Resource Management*, 15: 935-958.

Camelo-Ordaz, C., García-Cruz, J., Sousa-Ginel, E., & Valle-Cabrera, R. 2011. The influence of human resource management on knowledge sharing and innovation in Spain. *International Journal of Human Resource Management*, 22: 1442-1463.

Cappelli, P., & Neumark, D. 2001. Do "high-performance work practices" improve establishment-level outcomes? *Industrial and Labor Relations Review*, 54: 737-775.

Chadwick, C. 2010. Theoretic insights on the nature of performance synergies in human resource systems: Toward greater precision. *Human Resource Management Review*, 20: 85-101.

Chan, D. 1998. Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83: 234-246.

Chang, E. 2006. Individual pay for performance and commitment HR practices in South Korea. *Journal of World Business*, 41: 368-381.

Chang, S., Gong, Y., Way, S. A., & Jia, L. 2013. Flexibility-oriented HRM systems, absorptive capacity, and market responsiveness and firm innovativeness. *Journal of Management*, 39: 1924-1951.

Chen, S.-J. 2007. Human resource strategy and unionization: Evidence from Taiwan. *International Journal of Human Resource Management*, 18: 1116-1131.

Chenevert, D., & Tremblay, M. 2009. Fits in strategic human resource management and methodological challenge: Empirical evidence of influence of empowerment and compensation practices on human resource performance in Canadian firms. *International Journal of Human Resource Management*, 20: 738-770.

Chuang, C.-H., & Liao, H. 2010. Strategic human resource management in service context: Taking care of business by taking care of employees and customers. *Personnel Psychology*, 63: 153-196.

Clark, L. A., & Watson, D. 1995. Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7: 309-319.

Collins, C. J., & Smith, K. G. 2006. Knowledge exchange and combination: The role of human resource practices in the performance of high-technology firms. *Academy of Management Journal*, 49: 544-560.

Combs, J., Liu, Y., Hall, A., & Ketchen, D. 2006. How much do high-performance work practices matter? A meta-analysis of their effects on organizational performance. *Personnel Psychology*, 59: 501-528.

Conway, E., Fu, N., Monks, K., Alfes, K., & Bailey, C. 2016. Demands or resources? The relationship between HR practices, employee engagement, and emotional exhaustion within a hybrid model of employment relations. *Human Resource Management*, 55: 901-917.

Cutcher-Gershenfeld, J. 1991. The impact on economic performance of a transformation in workplace relations. *Industrial and Labor Relations Review*, 44: 241-260.

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. *International Journal of Human Resource Management*, 24: 3343-3362.

Davies, R., McNabb, R., & Whitfield, K. 2015. Do high-performance work practices exacerbate or mitigate the gender pay gap? *Cambridge Journal of Economics*, 39: 537-564.

De Grip, A., & Sieben, I. 2009. The effectiveness of more advanced human resource systems in small firms. *International Journal of Human Resource Management*, 20: 1914-1928.

Delaney, J. T., & Huselid, M. A. 1996. The impact of human resource management practices on perceptions of organizational performance. *Academy of Management Journal*, 39: 949-969.

Delery, J. E. 1998. Issues of fit in strategic human resource management: Implications for research. *Human Resource Management Review*, 8: 289-309.

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: 802-835.

Delery, J. E., & Gupta, N. 2016. Human resource management practices and organizational effectiveness: Internal fit matters. *Journal of Organizational Effectiveness*, 3: 139-163.

De Menezes, L. M., & Wood, S. 2006. The reality of flexible work systems in Britain. *International Journal of Human Resource Management*, 17: 106-138.

Den Hartog, D. N., Boon, C., Verburg, R. M., & Croon, M. A. 2013. HRM, communication, satisfaction, and perceived performance: A cross-level test. *Journal of Management*, 39: 1637-1665.

De Saá-Pérez, P., & Díaz-Díaz, N. L. 2010. Human resource management and innovation in the Canary Islands: An ultra-peripheral region of the European Union. *International Journal of Human Resource Management*, 21: 1649-1666.

De Saá-Pérez, P., & García-Falcón, J. M. 2002. A resource-based view of human resource management and organizational capabilities development. *International Journal of Human Resource Management*, 13: 123-140.

De Saá Pérez, P., & García-Falcón, J. M. 2004. The influence of human resource management in savings bank performance. *Service Industries Journal*, 24(2): 51-66.

Doeringer, P. B., Evans-Klock, C., & Terkla, D. G. 1998. Hybrids or hodgepodges? Workplace practices of Japanese and domestic startups in the United States. *Industrial and Labor Relations Review*, 51: 171-186.

Doeringer, P., Lorenz, E., & Terkla, D. G. 2003. The adoption and diffusion of high-performance management: Lessons from Japanese multinationals in the West. *Cambridge Journal of Economics*, 27: 265-286.

Dul, J. 2016. Necessary condition analysis (NCA): Logic and methodology of "necessary but not sufficient" causality. *Organizational Research Methods*, 19: 10-52.

Edwards, J. R. 2001. Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods*, 4: 144-192.

Ferratt, T. W., Agarwal, R., Brown, C. V., & Moore, J. E. 2005. IT human resource management configurations and IT turnover: Theoretical synthesis and empirical analysis. *Information Systems Research*, 16: 237-255.

Forth, J., & Millward, N. 2004. High-involvement management and pay in Britain. *Industrial Relations*, 43: 98-119.

Frenkel, S. J., & Lee, B.-H. 2010. Do high performance work practices work in South Korea? *Industrial Relations Journal*, 41: 479-504.

Frick, B. J., Goetzen, U., & Simmons, R. 2013. The hidden costs of high-performance work practices: Evidence from a large German steel company. *Industrial and Labor Relations Review*, 66: 198-224.

Furr, M. 2011. *Scale construction and psychometrics for social and personality psychology*. Thousand Oaks, CA: Sage.

Galang, M. C. 1999. Stakeholders in high-performance work systems. *International Journal of Human Resource Management*, 10: 287-305.

Gardner, R. G., Harris, T. B., Li, N., Kirkman, B. L., & Mathieu, J. E. 2017. Understanding "it depends" in organizational research: A theory-based taxonomy, review, and future research agenda concerning interactive and quadratic relationships. *Organizational Research Methods*, 20: 610-638.

Gerhart, B. 2007. Horizontal and vertical fit in human resource systems. In C. Ostroff & T. A. Judge (Eds.), *Perspectives on organizational fit*: 317-348. New York: Erlbaum.

Gerhart, B., Wright, P. M., McMahan, G. C., & Snell, S. A. 2000. Measurement error in research on human resources and firm performance: How much error is there and how does it influence effect size estimates. *Personnel Psychology*, 53: 803-834.

Godard, J. 2007. Unions, work practices, and wages under different institutional environments: The case of Canada and England. *Industrial and Labor Relations Review*, 60: 457-476.

Gould-Williams, J. S., & Gatenby, M. 2010. The effects of organizational context and teamworking activities on performance outcomes. *Public Management Review*, 12: 759-787.

Guest, D. E., Conway, N., & Dewe, P. 2004. Using sequential tree analysis to search for "bundles" of HR practices. *Human Resource Management Journal*, 14: 79-96.

Guest, D. E., Michie, J., Conway, N., & Sheehan, M. 2003. Human resource management and corporate performance in the UK. *British Journal of Industrial Relations*, 41: 291-314.

Guthrie, J. P. 2001. High-involvement work practices, turnover, and productivity: Evidence from New-Zealand. *Academy of Management Journal*, 44: 180-190.

Guthrie, J. P., Flood, P. C., Liu, W., MacCurtain, S., & Armstrong, C. 2011. Big hat, no cattle? The relationship between use of high-performance work systems and managerial perceptions of HR departments. *International Journal of Human Resource Management*, 22: 1672-1685.

Hauff, S., Alewell, D., & Hansen, N. 2014. HRM systems between control and commitment: Occurrence, characteristics and effects on HRM outcomes and firm performance. *Human Resource Management Journal*, 24: 424-441.

Hong, Y., Liao, H., Raub, S., & Han, J. H. 2016. What it takes to get proactive: An integrative multilevel model of the antecedents of personal initiative. *Journal of Applied Psychology*, 101: 687-701.

Hoque, K. 1999. Human resource management and performance in the UK hotel industry. *British Journal of Industrial Relations*, 37: 419-443.

Horgan, J., & Mühlau, P. 2006. Human resource systems and employee performance in Ireland and the Netherlands: A test of the complementarity hypothesis. *International Journal of Human Resource Management*, 17: 414-439.

Huselid, M. A. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, 38: 635-672.

Ichniowski, C., & Shaw, K. 1999. The effects of human resource management systems on economic performance: An international comparison of U.S. and Japanese plants. *Management Science*, 45: 704-721.

Jackson, S. E., Schuler, R. S., & Jiang, K. 2014. An aspirational framework for strategic human resource management. *The Academy of Management Annals*, 8: 1-56.

Jaw, B.-S., & Liu, W. 2003. Promoting organizational learning and self-renewal in Taiwanese companies: The role of HRM. *Human Resource Management*, 42: 223-241.

Jayaram, J., Droge, C., & Vickery, S. K. 1999. The impact of human resource management practices on manufacturing performance. *Journal of Operations Management*, 18: 1-20.

Jiang, K., Hu, J., Liu, S., & Lepak, D. P. 2017. Understanding employees' perceptions of human resource practices: Effects of demographic dissimilarity to managers and coworkers. *Human Resource Management*, 56: 69-91.

Jiang, K., Lepak, D. P., Han, K., Hong, Y., Kim, A., & Winkler, A.-L. 2012. Clarifying the construct of human resource systems: Relating human resource management to employee performance. *Human Resource Management Review*, 22: 73-85.

Jiang, K., Lepak, D. P., Ju, J., & Baer, J. C. 2012. How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms. *Academy of Management Journal*, 55: 1264-1294.

Jiang, K., & Messersmith, J. 2018. On the shoulders of giants: A meta-review of strategic human resource management. *International Journal of Human Resource Management*, 29: 6-33.

Jimenez-Jimenez, D., & Martinez-Costa, M. 2009. The performance effect of HRM and TQM: A study in Spanish organizations. *International Journal of Operations & Production Management*, 29: 1266-1289.

Jimenez-Jimenez, D., & Sanz-Valle, R. 2008. Could HRM support organizational innovation? *International Journal of Human Resource Management*, 19: 1208-1221.

Kalleberg, A. L., & Moody, J. W. 1994. Human resource management and organizational performance. *American Behavioral Scientist*, 37: 948-962.

Karatepe, O. M. 2013. High-performance work practices, work social support and their effects on job embeddedness and turnover intentions. *International Journal of Contemporary Hospitality Management*, 25: 903-921.

Kato, T., & Owan, H. 2011. Market characteristics, intra-firm coordination, and the choice of human resource management systems: Theory and evidence. *Journal of Economic Behavior & Organization*, 80: 375-396.

Kauhanen, A. 2009. The incidence of high-performance work systems: Evidence from a nationally representative employee survey. *Economic and Industrial Democracy*, 30: 454-480.

Kehoe, R. R., & Wright, P. M. 2013. The impact of high-performance human resource practices on employees' attitudes and behaviors. *Journal of Management*, 39: 366-391.

Ketkar, S., & Sett, P. K. 2009. HR flexibility and firm performance: Analysis of a multi-level causal model. *International Journal of Human Resource Management*, 20: 1009-1038.

Kim, J., Wehbi, N., Dellifraine, J. L., & Brannon, D. 2014. The joint relationship between organizational design factors and HR practice factors on direct care workers' job satisfaction and turnover intent. *Health Care Management Review*, 39: 174-184.

Klein, K. J., Conn, A. B., Smith, D. B., & Sorra, J. S. 2001. Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *Journal of Applied Psychology*, 86: 3-16.

Klein, K. J., Dansereau, F., & Hall, R. J. 1994. Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, 19: 195-229.

Koster, F. 2011. Able, willing, and knowing: The effects of HR practices on commitment and effort in 26 European countries. *International Journal of Human Resource Management*, 22: 2835-2851.

Kozlowski, S. W. J., & Klein, K. J. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations*: 3-90. San Francisco: Jossey-Bass.

Kroon, B., Van De Voorde, K., & Timmers, J. 2013. High performance work practices in small firms: A resource-poverty and strategic decision-making perspective. *Small Business Economics*, 41: 71-91.

Kwon, K., Bae, J., & Lawler, J. J. 2010. High commitment HR practices and top performers: Impacts on organizational commitment. *Management International Review*, 50: 57-80.

Laroche, P., & Salesina, M. 2017. The effects of union and nonunion forms of employee representation on high-performance work systems: New evidence from French microdata. *Human Resource Management*, 56: 173-189.

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.

Leffakis, Z. M., & Dwyer, D. J. 2014. The effects of human resource systems on operational performance in mass customisation manufacturing environments. *Production Planning and Control*, 25: 1213-1230.

Lengnick-Hall, M. L., Lengnick-Hall, C. A., Andrade, L. S., & Drake, B. 2009. Strategic human resource management: The evolution of the field. *Human Resource Management Review*, 19: 64-85.

Lepak, D. P., Liao, H., Chung, Y., & Harden, E. E. 2006. A conceptual review of human resource management systems in strategic human resource management research. *Research in Personnel and Human Resources Management*, 25: 217-271.

Lepak, D. P., & Snell, S. A. 2002. Examining the human resource architecture: The relationships among human capital, employment, and human resource configurations. *Journal of Management*, 28: 517-543.

Liao, H., Toya, K., Lepak, D. P., & Hong, Y. 2009. Do they see eye to eye? Management and employee perspectives of high-performance work systems and influence processes on service quality. *Journal of Applied Psychology*, 94: 371-391.

Litwin, A. S. 2013. Not featherbedding, but feathering the nest: Human resource management and investments in information technology. *Industrial Relations*, 52: 22-52.

Liu, F., Chow, I. H.-S., Gong, Y., & Wang, H. in press. Mediating links between HRM bundle and individual innovative behavior. *Journal of Management and Organization*. doi:10.1017/jmo.2016.47

Liu, W., Guthrie, J. P., Flood, P. C., & MacCurtain, S. 2009. Unions and the adoption of high performance work systems: Does employment security play a role? *Industrial and Labor Relations Review*, 63: 109-127.

Ma, S., Silva, M. G., Callan, V. J., & Trigo, V. 2016. Control and commitment HR practices, job satisfaction and turnover intentions: A comparison between local and multinational firms in China. *International Journal of Human Resource Management*, 27: 974-990.

Martí-Audí, N., Valverde, M., & Heraty, N. 2013. Human resource management in the Spanish call centre sector: The bird cage model of call centre work. *International Journal of Human Resource Management*, 24: 308-329.

McGrath, R. E. 2005. Conceptual complexity and construct validity. *Journal of Personality Assessment*, 85: 112-124.

Meuer, J. 2017. Exploring the complementarities within high-performance work systems: A set-theoretic analysis of UK firms. *Human Resource Management*, 56: 651-672.

Meyer, A. D., Tsui, A. S., & Hinings, C. R. 1993. Configurational approaches to organizational analysis. *Academy of Management Journal*, 36: 1175-1195.

Michie, J., & Sheehan, M. 1999. HRM practices, R&D expenditure and innovative investment: Evidence from the UKs 1990 Workplace Industrial Relations Survey (WIRS). *Industrial and Corporate Change*, 8: 211-234.

Michie, J., & Sheehan-Quinn, M. 2001. Labour market flexibility, human resource management and corporate performance. *British Journal of Management*, 12: 287-306.

Mohr, R. D., & Zoghi, C. 2008. High-involvement work design and job satisfaction. *Industrial and Labor Relations Review*, 61: 275-296.

Nishii, L. H., & Wright, P. M. 2008. Variability within organizations: Implications for strategic human resource management. In D. B. Smith (Ed.), *The people make the place: Dynamic linkages between individuals and organizations*: 225-248. Mahwah, NJ: Erlbaum.

Nyambegera, S. M., Sparrow, P., & Daniels, K. 2000. The impact of cultural value orientation on individual HRM preferences in developing countries: Lessons from Kenyan organizations. *International Journal of Human Resource Management*, 11: 639-663.

Ogbonnaya, C., Daniels, K., Connolly, S., & Van Veldhoven, M. 2017. Integrated and isolated impact of high-performance work practices on employee health and well-being: A comparative study. *Journal of Occupational Health Psychology*, 22: 98-114.

Osterman, P. 1994. How common is workplace transformation and who adopts it? *Industrial and Labor Relations Review*, 47: 173-188.

Pavlov, A., Mura, M., Franco-Santos, M., & Bourne, M. 2017. Modelling the impact of performance management practices on firm performance: Interaction with human resource management practices. *Production Planning and Control*, 28: 431-443.

Peccei, R., & Van De Voorde, K. 2019. The application of the multilevel paradigm in human resource management–outcomes research: Taking stock and going forward. *Journal of Management*, 45: 786-818.

Pérotin, V., & Robinson, A. 2000. Employee participation and equal opportunities practices: Productivity effect and potential complementarities. *British Journal of Industrial Relations*, 38: 557-583.

Piening, E. P., Baluch, A. M., & Salge, T. O. 2013. The relationship between employees' perceptions of human resource systems and organizational performance: Examining mediating mechanisms and temporal dynamics. *Journal of Applied Psychology*, 98: 926-947.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. 2016. Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. *Organizational Research Methods*, 19: 159-203.

Posthuma, R. A., Campion, M. C., Masimova, M., & Campion, M. A. 2013. A high performance work practices taxonomy: Integrating the literature and directing future research. *Journal of Management*, 39: 1184-1220.

Prieto Pastor, I. M., Santana, M. P. P., & Sierra, C. M. 2010. Managing knowledge through human resource practices: Empirical examination on the Spanish automotive industry. *International Journal of Human Resource Management*, 21: 2452-2467.

Purcell, J., & Hutchinson, S. 2007. Front-line managers as agents in the HRM-performance causal chain: Theory, analysis and evidence. *Human Resource Management Journal*, 17: 3-20.

Roche, W. K. 1999. In search of commitment-oriented human resource management practices and the conditions that sustain them. *Journal of Management Studies*, 36: 653-678.

Rossiter, J. R. 2008. Content validity of measures of abstract constructs in management and organizational research. *British Journal of Management*, 19: 380-388.

Runhaar, P., Sanders, K., & Konermann, J. 2013. Teachers' work engagement: Considering interaction with pupils and human resources practices as job resources. *Journal of Applied Social Psychology*, 43: 2017-2030.

Schmidt, J. A., & Pohler, D. M. 2018. Making stronger causal inferences: Accounting for selection bias in associations between high performance work systems, leadership, and employee and customer satisfaction. *Journal of Applied Psychology*, 103: 1001-1018.

Sgobbi, F., & Cainarca, G. C. 2015. High-performance work practices and core employee wages: Evidence from Italian manufacturing plants. *Industrial and Labor Relations Review*, 68: 426-456.

Shin, D., & Konrad, A. M. 2017. Causality between high-performance work systems and organizational performance. *Journal of Management*, 43: 973-997.

Singh, S., Darwish, T. K., Wood, G., & Mohamed, A. F. 2017. Institutions, complementarity, human resource management and performance in a South-East Asian petrostate: The case of Brunei. *International Journal of Human Resource Management*, 28: 2538-2569.

Smith, G. T. 2005. On construct validity: Issues of method and measurement. *Psychological Assessment*, 17: 396-408.

Smith, G. T., Fischer, S., & Fister, S. M. 2003. Incremental validity principles in test construction. *Psychological Assessment*, 15: 467-477.

Strauss, M. E., & Smith, G. T. 2009. Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5: 1-25.

Tay, L., Woo, S. E., & Vermunt, J. K. 2014. A conceptual and methodological framework for psychometric isomorphism: Validation of multilevel construct measures. *Organizational Research Methods*, 17: 77-106.

Toh, S. M., Morgeson, F. P., & Campion, M. A. 2008. Human resource configurations: Investigating fit with the organizational context. *Journal of Applied Psychology*, 93: 864-882.

Verburg, R. M., Den Hartog, D. N., & Koopman, P. L. 2007. Configurations of human resource management practices: A model and test of internal fit. *International Journal of Human Resource Management*, 18: 184-208.

Whitfield, K. 2000. High-performance workplaces, training, and the distribution of skills. *Industrial Relations*, 39: 1-25.

Wright, P. M., & Boswell, W. R. 2002. Desegregating HRM: A review and synthesis of micro and macro human resource management research. *Journal of Management*, 28: 247-276.

Wright, P. M., & Gardner, T. M. 2003. The human resource–firm performance relationship: Methodological and theoretical challenges. In D. Holman, T. D. Wall, P. Clegg, P. Sparrow, & A. Howard (Eds.), *The new workplace: A guide to the human impact of modern work practices*: 311-330. London: John Wiley & Sons.

Wright, P. M., Gardner, T. M., Moynihan, L. M., & Allen, M. R. 2005. The relationship between HR practices and firm performance: Examining causal order. *Personnel Psychology*, 58: 409-446.

Wright, P. M., Gardner, T. M., Moynihan, L. M., Park, H. J., Gerhart, B., & Delery, J. E. 2001. Measurement error in research on human resources and firm performance: Additional data and suggestions for future research. *Personnel Psychology*, 54: 875-901.

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: 551-571.

Wright, P. M., & McMahan, G. C. 1992. Theoretical perspectives for strategic human resource management. *Journal of Management*, 18: 295-320.

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.

Wu, P.-C., & Chaturvedi, S. 2009. The role of procedural justice and power distance in the relationship between high performance work systems and employee attitudes: A multilevel perspective. *Journal of Management*, 35: 1228-1247.

Yamamoto, H. 2013. The relationship between employees' perceptions of human resource management and their retention: From the viewpoint of attitudes toward job-specialties. *International Journal of Human Resource Management*, 24: 747-767.

Yanadori, Y., & van Jaarsveld, D. D. 2014. The relationships of informal high performance work practices to job satisfaction and workplace profitability. *Industrial Relations*, 53: 501-534.

Yoon, S. J., & Chae, Y. J. 2012. Management of paradox: A comparative study of managerial practices in Korean and Japanese firms. *International Journal of Human Resource Management*, 23: 3501-3521.

Zacharatos, A., Barling, J., & Iverson, R. D. 2005. High-performance work systems and occupational safety. *Journal of Applied Psychology*, 90: 77-93.

Zhang, M., Fan, D., & Zhu, C. J. 2014. High-performance work systems, corporate social performance and employee outcomes: Exploring the missing links. *Journal of Business Ethics*, 120: 423-435.

Zheng, C., O'Neill, G., & Morrison, M. 2009. Enhancing Chinese SME performance through innovative HR practices. *Personnel Review*, 38: 175-194.