1. Corrections should be marked with the Adobe Annotation & Comment Tools below:

2. To save your proof corrections, click the ‘Publish Comments’ button. Publishing your comments saves the marked up version of your proof to a centralized location in Wiley’s Online Proofing System. Corrections don’t have to be marked in one sitting – you can publish corrections and log back in at a later time to add more.

3. When your proof review is complete we recommend you download a copy of your annotated proof for reference in any future correspondence concerning the article before publication. You can do this by clicking on the icon to the right of the ‘Publish Comments’ button and selecting ‘Save as Archive Copy...’.

4. When your proof review is complete and you are ready to send corrections to the publisher click the ‘Complete Proof Review’ button that appears above the proof in your web browser window. Do not click the ‘Complete Proof Review’ button without replying to any author queries found on the last page of your proof. Incomplete proof reviews will cause a delay in publication. Note: Once you click ‘Complete Proof Review’ you will not be able to mark any further comments or corrections.

Firefox, Chrome, Safari Users

If your PDF article proof opens in any PDF viewer other than Adobe Reader or Adobe Acrobat, you will not be able to mark corrections and query responses, nor save them. To mark and save corrections, please follow these instructions to disable the built-in browser PDF viewers in Firefox, Chrome, and Safari so the PDF article proof opens in Adobe within a Firefox or Chrome browser window.
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 9.0 or above). (Note that this document uses screenshots from Adobe Reader X)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:

This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We’ve picked out some of these tools below:

1. Replace (Ins) Tool – for replacing text.
   - Strikethrough (Del) Tool – for deleting text.

   **How to use it**
   - Highlight a word or sentence.
   - Insert on the Replace (Ins) icon in the Annotations section.
   - The replacement text is into the blue box that appears.

   **How to use it**
   - Highlight a word or sentence.
   - Click on the Strikethrough (Del) icon in the Annotations section.

2. Add note to text Tool – for highlighting a section to be changed to bold or italic.

   **How to use it**
   - Highlight the relevant section of text.
   - Click on the Add note to text icon in the Annotations section.
   - Type in instructions on what should be changed regarding the text into the yellow box that appears.

3. Add sticky note Tool – for making notes at specific points in the text.

   **How to use it**
   - Click on the Add sticky note icon in the Annotations section.
   - Click at the point in the proof where the comment should be inserted.
   - Type the comment into the yellow box that appears.

4. Add note to text Tool – for highlighting a section to be changed to bold or italic.

   **How to use it**
   - Highlight the relevant section of text.
   - Click on the Add note to text icon in the Annotations section.
   - Type in instructions on what should be changed regarding the text into the yellow box that appears.

   **How to use it**
   - Click on the Add sticky note icon in the Annotations section.
   - Click at the point in the proof where the comment should be inserted.
   - Type the comment into the yellow box that appears.
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

5. Attach File Tool – for inserting large amounts of text or replacement figures.
   Inserts an icon linking to the attached file in the appropriate place in the text.
   **How to use it**
   - Click on the Attach File icon in the Annotations section.
   - Click on the proof to where you'd like the attached file to be linked.
   - Select the file to be attached from your computer or network.
   - Select the colour and type of icon that will appear in the proof. Click OK.

6. Add stamp Tool – for approving a proof if no corrections are required.
   Inserts a selected stamp onto an appropriate place in the proof.
   **How to use it**
   - Click on the Add stamp icon in the Annotations section.
   - Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears).
   - Click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as if it is, this would normally be on the first page).

7. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.
   Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks.
   **How to use it**
   - Click on one of the shapes in the Drawing Markups section.
   - Click on the proof at the relevant point and draw the selected shape with the cursor.
   - To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
   - Double click on the shape and type any text in the red box that appears.

For further information on how to annotate proofs, click on the Help menu to reveal a list of further options:
Single- and Multiple-Informant Research Designs to Examine the Human Resource Management—Performance Relationship

Juan Carlos Bou-Llusar, Inmaculada Beltrán-Martín, Vicente Roca-Puig and Ana Belén Escrig-Tena

Department of Business Management and Marketing at Universitat Jaume I, Facultad de Ciencias Jurídicas y Económicas, 12071 Castellón, Spain
Corresponding author email: ibeltran@emp.uji.es

During the last decades, many empirical studies have analysed the relationship between human resource management and firm performance. Despite the call for multiple-rater designs, a relatively large number of researchers still rely on survey responses provided by a single informant in each organization. Single-informant designs suffer from a number of problems, especially when the responses provided by different types of raters across firms are pooled into a single dataset prior to assessing their equivalence across raters. Using an illustration of the relationship between high performance work systems and firm performance, in this paper we observe that responses provided by managers holding different positions (human resource managers and sales managers) differ significantly and therefore pooling their responses into a single dataset may result in confusing conclusions. Furthermore, we demonstrate that differences arise in the estimated parameters when a multiple-key-informant approach, compared to a single-informant design, is adopted. For these reasons, data collection using multiple key informants is recommended, based on the assumption that some raters in the firm will be more knowledgeable about the variables of interest than others.

Introduction

A crucial question in the human resource management (HRM) literature is the analysis of the relationship between the human resource strategy of the firm and organizational performance (Appelbaum et al., 2000; MacDuffie, 1995; Youndt et al., 1996). Although most of the empirical studies have found a positive association between the two measures (Combs et al., 2006), the methodological limitations raised by several authors (e.g. Gerhart, Wright and McMahan, 2000; Gerhart et al., 2000; Huselid and Becker, 2000) suggest that the conclusions they draw are premature to say the least. Our study aims to contribute to the methodological debate in the HRM field by exploring the following research questions.

First, over the past years several studies in the HRM field have used data collected through questionnaires administered to informants (one informant per firm) whose positions vary across organizations. For instance, this is the case of studies that use data provided, without distinction, by HR managers or senior managers (e.g. Park et al., 2003), by various staff members from CEO to junior managers (Guthrie, 2001), by HR managers, owners and senior managers (Harel and Tzafrir, 1999), or even by unspecified respondents (Delaney and Huselid, 1996). These studies then pool responses to create a single dataset for statistical analyses (Rungtusanatham et al., 2008) so the information about HRM and performance is provided by the HR managers in some cases and by the senior or other managers in other cases. However, it is important to take into account that
when there are systematic differences in survey ratings depending on who provides the answers (or, in the words of Huselid and Becker, 2000), when the ‘respondents matter’) the rater chosen to respond to the survey becomes a critical issue. Conclusions drawn from studies that pool responses obtained from different types of raters should be analysed with caution when measurement invariance is not examined prior to pooling data (Rungtusanatham et al., 2008). It is not our intention to call into question the results of this research stream but rather, in our first research question, to empirically analyse the consequence of pooling responses provided by different types of informants and to examine whether the proposed relationships between HRM and performance vary depending on the respondent chosen to assess the variables.

Second, the traditional data collection strategy employed in the HRM field assumes that a single person is able to provide accurate information about all the variables that refer to the whole organization (Gerhart, Wright and McMahan, 2000). This approach increases the probability that the relationships between HRM and performance will be affected by common method variance (CMV) (Podsakoff et al., 2003). A recommended strategy in the HRM field to avoid the risk of CMV is to collect data from multiple sources, i.e. using various informants in each organization to provide responses to different questions in the same questionnaire. Multiple-source studies assume that some raters are more knowledgeable than others in assessing the measures of interest (i.e. differential accuracy assumption) (Huselid and Becker, 2000). Consistent with Huselid and Becker (2000) and Wright et al. (2001), we believe that more attention should be paid to ensuring that the most knowledgeable informants are used in order to increase the validity of the measures and to reduce the potential CMV. In our second research question we analyse whether the relationship between HRM and performance varies when multiple informants in a single company assess the variables about which they have more information compared to a single-respondent survey design to evaluate the same variables.

An overview of research approaches in the HRM–performance literature

Many of the empirical studies in the HRM field analysing the relationship between HRM and performance use survey research. According to Rungtusanatham et al. (2008), data in these studies are collected through different survey research approaches, including those using either a single or multiple informants. To provide a parsimonious and systematic consideration of how published articles on the HRM–performance relationship rely on single or multiple informants, Appendix 1 shows a classification of 97 studies on this relationship included in Jiang et al.’s (2012) meta-analysis, all of which used surveys to collect all or part of the research data. A direct content analysis (Hsieh and Shannon, 2005) was conducted to classify the papers. Inspired by the Weber protocol (Weber, 1990), the coding categories were defined based on Rungtusanatham et al.’s (2008) classification approaches, which were refined after coding a sample of the selected articles. The reliability of the allocation of articles in each category was addressed following the two-step procedure used in other content analyses such as Furrer, Thomas and Goussevskaia (2008), Jeung et al. (2011) and Clark et al. (2014). First, two of the authors of this paper independently reviewed all 97 selected articles and coded them in one of the four categories (based on a detailed examination of the methodology used); the two authors coded 79% of the articles in the same category. The percentage of agreement rose to above 90% when it was considered that some divergences came from papers that shared features from more than one approach, which the coders had allocated in different categories. In a second step, vagueness or discrepancies between the two coders were resolved through research team discussions and assigned to the category considered to be the best fit.

The papers were classified into four distinct approaches. In the first approach (approach 1) a single informant in each organization provides answers to all the questions in the questionnaire, but differs in that the position of the informants varies across the organizations. For example, Audea, Teo and Crawford (2005) used data obtained by aggregating responses from HR managers, labour or union representatives and general managers from different organizations and performed the statistical analyses using information from the pooled dataset.

Approach 2 involves sending a questionnaire to a single informant in each organization but the informants hold a similar recognized position in all the surveyed organizations and provide...
Single- and Multiple-Informant Research Designs

Q3

answers to all the questions included in the questionnaire. For instance, Batt and Colvin (2011) administered a questionnaire to the senior managers of a sample of US call centres, who answered questions related to both the independent variable (HR practices) and the dependent variables (quits, dismissals, turnover and customer satisfaction).

The remaining two approaches involve using multiple informants from each organization. Approach 3 corresponds to studies in which all the informants in each organization provide answers to all the questions included in the questionnaire. For instance, Youndt and Snell (2004) distributed a questionnaire containing items related to different HR configurations, the three components of the firm’s intellectual capital and firm performance to various managers in each participating firm: the two highest ranking executives (CEO and president) and the vice-president of HR. For 71 firms these authors received responses from two or three respondents who provided information about all the questions included in the survey which was then used to calculate the interrater agreement for each measure.

Approach 4 entails distributing different sections of the questionnaire to different ‘key’ informants. This is the case of Akhtar, Ding and Ge’s (2008) study about the influence of strategic human resource practices on firm performance in a sample of Chinese firms. In this research, the general manager of the surveyed firms responded to questions about company performance, while the HR manager responded to questions related to the HR practices.

From this review, we can conclude that the majority of studies in the HRM field still rely on single-informant data to test their hypotheses. Appendix 1 shows that 60 of the studies we reviewed (62%) collect data from a single informant in the firm. Wall and Wood (2005) drew similar conclusions, observing that 21 of the 25 studies they analysed use single respondents. Of the 60 studies, 32 adopt approach 1 and 28 adopt approach 2 in their research designs. In other words, more than half of the single-informant studies did not know the position held by the person who had answered the questionnaire. In addition, the pre-eminence of single-informant designs in the HRM literature indicates that concerns with CMV are still not properly addressed in this field. Regarding multiple-informant designs, studies in approach 3 are not very common in the HRM literature (see Appendix 1), probably because, although they improve the reliability of the measures, CMV may still exist and the cost of collecting data from multiple informants does not pay off. Finally, 32 studies (33%) collect data following approach 4. That is, only one-third of the analysed studies have a strategy to deal with CMV. In the following sections, we shall address some of the features of the above-mentioned approaches in more detail.

Research question 1: Does the respondent matter in HRM research?

The overview of the research approaches in the HRM–performance literature suggests that many of the empirical studies pooled data provided by different types of informants (approach 1). Implicit in such a procedure is the ‘parallel raters’ assumption, which ignores the existence of systematic differences in responses provided by informants holding different positions in the firm. However, Huselid and Becker (2000) highlight the relevance of using informants whose position enables them to provide relevant data, giving support to the ‘differential accuracy’ assumption. Under this assumption some raters (key informants) will give more reliable descriptions of the measures studied than others, and therefore these types of informants should be preferred ‘because they are supposedly knowledgeable about the issues being researched and are willing to communicate about them’ (Kumar, Stern and Anderson, 1993, p. 1634). If the differential accuracy assumption is realistic, data collection following approach 1 would lead to biased results that may threaten the validity of conclusions about HRM–performance relationships.

In the HRM–performance field, there are several reasons why pooling data provided by different types of respondents in different firms can lead to confusing conclusions. First, not all managers in the organization have the same knowledge and information about HR practices and performance; for instance, senior managers (particularly in large organizations) may not always know precisely which HR practices are implemented in the organization (Huselid and Becker, 2000), or HR managers may not have detailed information on how line managers implement the practices. If the influence of HRM
Research question 2: Common method variance and multiple-key-informant research designs in HRM research

A second inherent problem in studies that use data obtained through a single-informant strategy is that their results may be affected by CMV when a single rater evaluates both the predictor and the criterion variables (Wall and Wood, 2005). According to Podsakoff et al. (2003), this type of CMV may result in inflated observed correlations among the variables assessed by the same rater.

CMV can be partly explained by the sources of systematic differences across raters described in the previous section. Other potential sources of CMV should also be taken into account, however. For instance, Podsakoff et al. (2003) consider the consistency motif as one of the sources of CMV, defined as the tendency of raters to maintain consistency in their responses to the different questions included in the survey. In addition, acquiescence, or the ‘tendency to agree with attitude statements regardless of content’ (Winkler, Kanouse and Ware, 1982, p. 555), may lead to higher correlations among the items that are described in similar terms in the survey, even when they are not conceptually related. Finally, respondents' enduring or transient mood states can also produce artifactual covariance in measures, depending on the positive or negative mood of the respondent when he or she answers the survey questions.

A strategy suggested to avoid the problems of CMV consists of collecting data from multiple sources within the same firm (Wall and Wood, 2005), in particular selecting 'key' informants (one or several) who provide responses to questions about which they are more knowledgeable or that are more closely linked to their areas of expertise (Huselid and Becker, 2000) (approach 4). Regardless of the difficulties inherent in multiple-respondent research (e.g. missing data, low survey response rates, higher cost of data collection etc.) the use of a multiple-key-informant strategy to collect data seems a promising approach to advance knowledge about the relationship between HRM and firm performance. Research adopting a multiple-informant strategy should obtain information about HR practices from at least one rater and information about the performance variables from a different rater (or set of different raters).
Single- and Multiple-Informant Research Designs

This approach allows the researcher to deal with the above-mentioned problems associated with the single-informant research design. In our illustration, results obtained through approach 4 will be the baseline model against which we compare the results of single-informant designs. Through this comparison we shall examine our second research question, namely, are there differences in the proposed HRM—performance relationships when a multiple-key-informant design (approach 4) as opposed to a single-informant design is adopted?

Illustration: Influence of high performance work systems on firm performance through HR flexibility

Theoretical model

To illustrate the issues presented in the previous sections, we propose a theoretical model that includes the relationships between high performance work systems (HPWS), human resource flexibility (HRF) and firm performance (FP) (Figure 1). Our analyses are based on the relationships specified in this model.

HRM aims to increase productivity and effectiveness, and relies on conditions that help employees to identify the firm’s goals and to work hard to accomplish them (Whitener, 2001). In recent decades HPWS have predominated in the HR literature. HPWS are made up of four dimensions, namely comprehensive staffing, extensive training, developmental performance appraisal and equitable reward systems (Snell and Dean, 1992). The resource-based view of the firm argues that HPWS can lead to competitive advantages by developing a unique and valuable human capital pool (Delery, 1998).

Over the past decade, the increasing dynamism of competitive environments and the emergence of new principles to manage firms point to HRF as a potential mediator of the relationship between HRM and organizational outcomes (Bhattacharya, Gibson and Doty, 2005). From a resource-based view perspective, HRF is made up of three dimensions: functional flexibility, skill malleability and behaviour flexibility (Beltrán-Martín et al., 2008; Bhattacharya, Gibson and Doty, 2005).

Several studies have demonstrated that HRF significantly affects FP through different paths, such as developing more efficient means of accomplishing task requirements (Boxall, 1999), reducing the number of line managers and costs of administrative levels (Valverde, Tregaskis and Brewster, 2000), contributing to the adoption of innovative solutions in the firm, and increasing productivity (Lado and Wilson, 1994). In addition, HPWS are likely to influence HRF. For instance, training and staffing activities favour the abilities needed to perform a variety of tasks effectively (Friedrich et al., 1998), or the provision of equitable rewards encourages employees’ willingness to move and reallocate as the need arises (Dyer and Shafer, 2002). The above reasoning leads us to propose that a relevant mediating process by which HPWS affect FP is through improvements in the flexibility levels of human resources.

Measures of the variables

The unit of analysis in our illustration is the commercial departments in the companies surveyed as we are interested in the linkages between HPWS, HRF and FP related to salespeople in our sample of firms. Employees in the commercial department are increasingly important to competition in current environments (Slater and Olson, 2000). To measure HPWS, we used Snell and Dean’s (1992) scale, which covers items related to selective staffing, comprehensive training, developmental performance appraisal and equitable reward systems applicable to employees in the commercial department. Due to the small sample sizes used in our illustration, we simplified the model by constructing a single indicator as the mean value of the items corresponding to each HPWS dimension (Bagozzi and Edwards, 1998). The resulting four composite variables were used as observable indicators of a first-order factor corresponding to HPWS. With this procedure we provide an adequate representation of the underlying dimensionality of the HPWS scale (Landis, Beal and Tesluk, 2000) and achieve a reasonable ratio of cases to observed items (Bagozzi and Edwards, 1998).

We measured HRF with the measurement scale developed by Beltrán-Martín et al. (2008). The HRF items assess the extent to which employees in the commercial department currently possess the capabilities and attributes listed, on a scale ranging from 1 (applies to very few employees) to 7 (applies to most of the employees). Concerning FP and in accordance with the description of the unit of analysis in our study, we endeavoured to...
use measures that are more directly related to the commercial department, rather than FP indicators where other variables (e.g. market trends or competitors’ movements) may contaminate the results (Challis, Samson and Lawson, 2005). This measure refers to the extent to which relationships with customers are efficient and customers’ needs and expectations are fulfilled, and it has been used in previous studies in the HRM field (e.g. Liao and Chuang, 2004). We used a relative performance measure which asked informants to assess their performance over the past three years compared with that of their competitors (Delaney and Huselid, 1996). Appendix 2 provides a detailed description of the measures used. The fieldwork for this study took place during the period May–October 2005 on a sample of Spanish industrial and service companies with 100 or more employees, using the data collection strategies described below.

Data collection and results for research question 1

For research question 1 we analyse data obtained through a single informant in each organization and we are aware that the positions of the key informants differ across organizations (approach 1). We administered a survey in such a way that a single informant in each firm, either the HR manager or the sales manager, responded to all the questions on HPWS, HRF and FP. Of the usable responses received, for 108 firms the informant was the HR manager and for 49 firms, the sales manager. A key feature of approach 1 is that the responses from the various groups of informants are pooled into a single dataset. Thus, a total of 157 valid responses (108 + 49) were used in the subsequent analyses.

We estimated the model in Figure 1 for the pooled data group (n = 157) and also for the sales manager (n = 49) and HR manager (n = 108) groups separately. We used the multiple-group structural equation modelling (SEM) approach (see Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000) to compare the existence of invariance across groups of informants. Specifically, we conducted an omnibus test of the equality of sample covariance matrices and means across groups, and a series of tests of invariance (e.g. configural, metric, scalar or factor invariance) by constraining sets of specific model parameters in a series of nested models.

Table 1 shows the (robust) goodness-of-fit indices of the models for the pooled data, the sales manager group and the HR manager group. The pooled data show only a marginal fit, with goodness-of-fit indices satisfying the recommended values (see for example Bentler and Bonett, 1980). In the sales manager group we find an excellent fit to the data with a non-significant chi-squared test and values of fit indices within the recommended values,\(^1\) while in the case of the

\(^1\)Due to the small sample size of the sales manager group, the non-significance of the chi-squared statistic might be
Single- and Multiple-Informant Research Designs

Table 1. Goodness-of-fit indices for the single-informant design

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$ value</th>
<th>TLI</th>
<th>CFI</th>
<th>SRMR</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled data</td>
<td>157</td>
<td>89.272</td>
<td>62</td>
<td>0.013</td>
<td>0.903</td>
<td>0.923</td>
<td>0.063</td>
<td>0.053</td>
</tr>
<tr>
<td>Sales managers</td>
<td>49</td>
<td>62.463</td>
<td>62</td>
<td>0.459</td>
<td>0.992</td>
<td>0.994</td>
<td>0.102</td>
<td>0.012</td>
</tr>
<tr>
<td>HR managers</td>
<td>108</td>
<td>101.827</td>
<td>62</td>
<td>0.001</td>
<td>0.835</td>
<td>0.869</td>
<td>0.079</td>
<td>0.077</td>
</tr>
</tbody>
</table>

TLI, Tucker–Lewis index; CFI, comparative fit index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation.

*Satorra–Bentler scaled chi-squared (Satorra and Bentler, 1994).

HR manager group the fit of the model is inadequate.

Table 2 shows standardized item factor loadings and regression coefficients for the three groups. The item factor loadings for the FP scale are high in general and comparable in size across groups. In the case of HPWS and HRF, factor loadings differ across groups, mainly for ‘equitable reward systems’ ($\lambda_{10}$), in which the loading is not statistically significant in the sales manager group. Moreover, for items ‘selective staffing’ ($V_1$) and ‘skill malleability’ ($V_{12}$) factor loadings are empirically undefined (Rindskopf, 1984), resulting in negative error variance estimates and standardized loadings larger than one.

These results, together with the lack of model fit in the case of the HR manager group, raise doubts about the adequacy of the proposed model when it is applied to different types of informants, and eventually when data are pooled into a single group. Differences between groups suggest that sales and HR managers do not use equivalent interpretive frames of reference in the evaluating process (Vandenberg, Lance and Taylor, 2005). In particular, differences in factor loadings across groups suggest the existence of possible ‘conceptual’ disagreement (Cheung, 1999) between sales and HR managers. This implies that the two types of informants cannot be considered as equivalent.

The sales manager group and the HR manager group also differ in terms of the relationships between HRM, HRF and FP. Attending to the standardized parameters shown in Table 2, we find that in the pooled data group HPWS has a statistically significant influence on HRF (0.498), HRF positively influences FP (0.285) and HRF partially mediates the HPWS–FP relationship, with significant total and indirect effects (0.382 and 0.142, respectively). In the sales manager group, HRF has a significant influence on FP (0.405), but we find no statistically significant effect of HPWS on HRF or on FP (neither direct nor indirect). A clearly different pattern of relationships is found in the HR manager group. Here the effect of HPWS on HRF is statistically significant (0.636), as is the total effect of HPWS on FP (0.409). However, HRF does not significantly affect FP. These results suggest that the model is not invariant as regards the type of informant and points to ‘psychological’ disagreement (Cheung, 1999) between sales and HR managers. It also suggests that drawing conclusions based on pooled data from sales and HR managers showing distinct response patterns would provide ambiguous interpretations of the relationships between HPWS, HRF and FP.

Additional analyses are needed to more formally test the significance of the differences between types of informants. A series of tests of invariance was performed to assess the equivalence...
Table 2. Standardized parameter estimates and robust standard errors (in parentheses) for the single-informant design

<table>
<thead>
<tr>
<th>Item factor loadings</th>
<th>Pooled data ((N = 157))</th>
<th>Sales manager ((N = 49))</th>
<th>HR manager ((N = 108))</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP ((\text{improvements in customer satisfaction}))</td>
<td>0.800 ((_________)*)</td>
<td>0.825 ((________)*)</td>
<td>0.797 ((________)*)</td>
</tr>
<tr>
<td>FP ((\text{customer retention}))</td>
<td>0.795 (0.091)*</td>
<td>0.704 (0.166)*</td>
<td>0.833 (0.098)*</td>
</tr>
<tr>
<td>FP ((\text{improvements in communication with customers}))</td>
<td>0.813 (0.077)*</td>
<td>0.691 (0.126)*</td>
<td>0.861 (0.089)*</td>
</tr>
<tr>
<td>FP ((\text{reduction of complaints from customers}))</td>
<td>0.731 (0.094)*</td>
<td>0.611 (0.186)*</td>
<td>0.780 (0.106)*</td>
</tr>
<tr>
<td>FP ((\text{adjusting to customers’ behaviour patterns}))</td>
<td>0.612 (0.100)*</td>
<td>0.497 (0.181)*</td>
<td>0.700 (0.104)*</td>
</tr>
<tr>
<td>FP ((\text{approaching customers quickly}))</td>
<td>0.606 (0.116)*</td>
<td>0.617 (0.160)*</td>
<td>0.619 (0.129)*</td>
</tr>
<tr>
<td>HPWS ((\text{selective staffing}))</td>
<td>0.405 (0.285)*</td>
<td>1.555 (4.356)</td>
<td>0.242 (0.299)*</td>
</tr>
<tr>
<td>HPWS ((\text{comprehensive training}))</td>
<td>0.360 ((________))</td>
<td>0.292 ((________))</td>
<td>0.310 ((________))</td>
</tr>
<tr>
<td>HPWS ((\text{developmental performance appraisal}))</td>
<td>0.639 (0.381)*</td>
<td>0.306 (0.176)*</td>
<td>0.533 (0.453)*</td>
</tr>
<tr>
<td>HPWS ((\text{equitable reward systems}))</td>
<td>0.336 (0.294)*</td>
<td>0.082 (0.125)</td>
<td>0.559 (0.596)*</td>
</tr>
<tr>
<td>HRF ((\text{functional flexibility}))</td>
<td>0.524 ((________))</td>
<td>0.627 ((________))</td>
<td>0.504 ((________))</td>
</tr>
<tr>
<td>HRF ((\text{skill malleability}))</td>
<td>0.993 (0.265)*</td>
<td>0.698 (0.195)*</td>
<td>1.087 (0.350)*</td>
</tr>
<tr>
<td>HRF ((\text{behaviour flexibility}))</td>
<td>0.856 (0.128)*</td>
<td>0.857 (0.158)</td>
<td>0.626 (0.148)*</td>
</tr>
<tr>
<td>Direct effects ((\text{HPWS} \rightarrow \text{HRF}))</td>
<td>0.498 (0.291)*</td>
<td>0.070 (0.116)</td>
<td>0.636 (0.629)*</td>
</tr>
<tr>
<td>Direct effects ((\text{HPWS} \rightarrow \text{FP}))</td>
<td>0.241 (0.328)</td>
<td>0.038 (0.247)</td>
<td>0.266 (0.548)</td>
</tr>
<tr>
<td>Direct effects ((\text{HRF} \rightarrow \text{FP}))</td>
<td>0.285 (0.107)*</td>
<td>0.405 (0.166)*</td>
<td>0.265 (0.145)</td>
</tr>
<tr>
<td>Indirect effect ((\text{HPWS} \rightarrow \text{FP}))</td>
<td>0.142 (0.131)*</td>
<td>0.004 (0.071)</td>
<td>0.144 (0.282)</td>
</tr>
<tr>
<td>Total effect ((\text{HPWS} \rightarrow \text{FP}))</td>
<td>0.382 (0.334)*</td>
<td>0.035 (0.245)</td>
<td>0.409 (0.536)*</td>
</tr>
</tbody>
</table>

FP, firm performance; HPWS, high performance work systems; HRF, human resource flexibility.

*p < 0.05.

Table 3. Goodness-of-fit indices for single-informant designs: Sales versus HR managers

<table>
<thead>
<tr>
<th></th>
<th>(\chi^2)</th>
<th>df</th>
<th>p value</th>
<th>(\Delta \chi^2)</th>
<th>(\Delta df)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omnibus tests of invariance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equality of variance–covariance</td>
<td>141.589</td>
<td>91</td>
<td>&lt;0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equality of means</td>
<td>22.580</td>
<td>13</td>
<td>0.047</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equality of variance–covariance and means</td>
<td>166.599</td>
<td>104</td>
<td>&lt;0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invariance analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Configural invariance</td>
<td>187.450</td>
<td>124</td>
<td>&lt;0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metric invariance</td>
<td>214.206</td>
<td>134</td>
<td>&lt;0.05</td>
<td>26.493</td>
<td>10</td>
<td>0.003</td>
</tr>
<tr>
<td>Scalar invariance</td>
<td>237.326</td>
<td>147</td>
<td>&lt;0.05</td>
<td>23.269</td>
<td>13</td>
<td>0.039</td>
</tr>
<tr>
<td>Factor invariance</td>
<td>240.633</td>
<td>150</td>
<td>&lt;0.05</td>
<td>2.938</td>
<td>3</td>
<td>0.401</td>
</tr>
</tbody>
</table>

*Satorra–Bentler scaled chi-squared (Satorra and Bentler, 1994).

*Chi-squared scaled difference test statistics were computed using the Satorra and Bentler (2001) procedure.

of the information reported by the two groups of informants. First, we conducted an omnibus test of the equality of sample covariance matrices across groups (Jöreskog and Sörbom, 1989). The goodness-of-fit indices shown in the first row of Table 3 reveal that the null hypothesis is not supported by the data \((\chi^2 = 141.589, \text{df} = 91, p < 0.05)\). Therefore, the two types of informants cannot be considered as subsamples of the same population, since the measurement and the relationships between HPWS, HRF and FP differ between sales and HR managers. A test of equality of sample means reports a chi squared in the vicinity of the critical value \((\chi^2 = 22.580, \text{df} = 91, p < 0.05)\).
Single- and Multiple-Informant Research Designs

df = 13, p = 0.047), suggesting that the null hypothesis of equal means across groups may be acceptable. These results provide evidence that the respondent matters in analysing the relationship between the variables included in the model and advise against pooling data from sales and HR managers.

Although the rejection of the equality of sample covariance matrices indicates some kind of differences between groups of informants, the omnibus test provides no information on the particular source of invariance (Vandenberg and Lance, 2000) and it is an extremely rigorous test (Cheung and Rensvold, 1999). Following the recommendations of several authors on the procedure to perform analyses of invariance (e.g. Byrne, 1994; Steenkamp and Baumgartner, 1998) we estimated a series of nested models with increasingly restrictive tests to determine what kind of invariance exists between the two types of informants. A multiple-group model with no across-group restrictions was taken as a reference (or less restricted) model. This model represents a test of ‘configural’ invariance (i.e. equivalence of factor structure across groups), a hypothesis that is rejected, as the goodness-of-fit indices attest ($\chi^2 = 187.450$, df = 124, $p < 0.05$). Note that if the configural invariance test is not satisfied, the subsequent tests of invariance may not make sense from a substantive point of view (Vandenberg, Lance and Taylor, 2005), since an adequate reference model does not exist (Cheung and Rensvold, 1999). Nevertheless, in the lower part of Table 3 we report the fit indices for additional tests of metric ($\Delta \chi^2 = 26.493$, $\Delta$ df = 10, $p = 0.003$), scalar ($\Delta \chi^2 = 23.269$, $\Delta$ df = 13, $p = 0.039$) and factor covariance invariance ($\Delta \chi^2 = 2.938$, $\Delta$ df = 3, $p = 0.401$). Overall, the invariance analysis suggests that there are significant differences between sales managers and HR managers and lends support to the assumption that the respondent matters both in how the constructs of interest are measured and in the analysis of their relationships. In other words, the results are (highly) dependent on the type of respondent and their relative numbers in the pooled dataset.

Data collection and results for research question 2

For our second research question, we analyse data collected through multiple key informants from each organization who respond to different sections of the survey (approach 4). The questionnaire for our study was split into two sections in line with this data collection approach. The first section corresponds to questions about HPWS and the second section includes questions about HRF and FP only. Each of these sub-questionnaires was sent directly to the corresponding key informant. Wright et al. (2001) suggest the respondent’s title is a suitable test of accuracy when collecting data, and empirical studies usually choose key informants in terms of their formal roles in the organization. We refer to previous empirical studies in the field to identify the key informants for our variables.

Regarding the key informant for the HPWS section of the survey, Arthur and Boyles (2007) state that many of the empirical studies in the HRM field rely on information provided in responses from the HR manager to evaluate HR practices. However, those authors note that such responses may not be as reliable in relatively large and complex firms because HR managers may not have enough information about the specific practices used for all the employees and positions in the firm. One of the solutions several authors have suggested to improve informant reliability when using the HR manager as the key informant is to focus on a specific ‘core’ group of employees; this reduces ambiguity and the information processing requirements for this rater (Lepak et al., 2006). Our focus on employees in the commercial department, and therefore on HPWS applicable to these employees, contributes to reducing the risk of low reliability.

Regarding the key informant for the HRF questions, in accordance with previous literature assessing employee skills and behaviours, we consider that the supervisor is the most appropriate rater of employee flexibility. This section of the survey was therefore addressed to the sales managers of the surveyed companies. Viswesvaran, Ones and Schmidt (1996) state that supervisors’ ratings are the most commonly used measure of employee performance (Cascio, 1991). In the HRM literature, a number of empirical studies have used supervisors’ ratings to assess several employee outcomes, such as job performance (e.g. Yam, Fehr and Barnes, 2014), creativity (de Stobbeleir, Ashford and Buyens, 2011), customer service behaviours (Liden et al., 2014) and organizational citizenship behaviours (e.g. Bolino et al., 2006), among others.

Finally, in our illustration the key informant for the FP scale was the firms’ sales manager.
Several studies have demonstrated that subjective measures of FP provided by managers are positively associated with objective measures of performance (Baer and Frese, 2003). However, Wall et al. (2004) point out that when choosing the appropriate manager to assess subjective performance, and in order to avoid the risk of CMV, he or she should not be the same person as the manager assessing HR practices. In addition, given that the performance measure refers to the commercial department, the company’s sales manager may have a more proximal knowledge of department performance than higher-level managers. Prior studies analysing the performance in sales and commercial departments have chosen sales managers as the key informants of this department’s performance (e.g. Matsuo, 2006).

Regarding the sample size, we received information from 73 firms; 31 companies reported complete information from both key informants (HR manager and sales manager); 42 companies did not respond to both questionnaires, reporting only partial information. Specifically, 23 companies only reported the information from the sales managers, and 19 companies only provided information from the HR managers. To deal with the problem of missing data that arises when only one of the two key informants report survey responses, we use direct (or full information) maximum likelihood (ML) estimation of SEM with missing data (Arbuckle, 1996; Neale, 2000; Newman, 2003). This is a model-based procedure for missing data (Kline, 1998) currently implemented in most standard SEM software (e.g. Mplus, EQS, among others). In contrast to traditional approaches to missing data such as listwise deletion, which would imply a loss of 57% of the final sample, direct ML estimation utilizes all the available information, even that for firms where only one manager responded. It increases the statistical power of the analysis and provides more efficient and less biased estimates under the missing completely at random or missing at random assumptions (Little and Rubin, 1987). Robust standard errors as well as (robust) scaled chi-squared goodness-of-fit tests (Satorra and Bentler, 1994) were also reported in the multiple-informant analysis. Therefore the sample used in the statistical analyses was made up of a total of 73 companies that reported either total (n = 31) or partial (n = 42) information about the variables.

To examine research question 2, we first estimated the model in Figure 1 for the multiple-key-informant group (n = 73) and then compared these estimates with those reported for the single-informant analysis in order to test for significant differences between the two research designs in the proposed relationships between HPWS, HRF and FP.

The goodness-of-fit for the model in Figure 1 using data from multiple key informants yields a χ² test statistic value of 107.782 with 62 degrees of freedom (other goodness-of-fit indices are Tucker–Lewis index 0.722; comparative fit index 0.779; standardized root mean square residual 0.141, root mean square error of approximation 0.101), indicating a poor fit to the data. Standardized parameter estimates are shown in Table 4. Most of the factor item loadings are statistically significant and present reasonably high values (except for λ6, which is below 0.5, and λ10, which is not statistically significant) and there are no problems of empirical identification. We find no significant causal relationships between HPWS, HRF and FP, results that clearly depart from those obtained in the single-informant analysis shown in Table 2.

Due to the small sample size using the multiple-key-informant research design, the conflicting results with the single-informant analysis, and the importance of rejecting the existence of significant relationships for a meaningful interpretation of the findings, we assess the statistical power of the analysis to detect significant effects. Using the procedure of Satorra and Saris (1985) and Saris, Satorra and van der Veld (2009), we find low power for all the substantive parameters. The likelihood of detecting an effect, if it truly exists, is 0.208, 0.187 and 0.238 for the effects of HPWS on HRF, HPWS on FP and HRF on FP, respectively. In all cases, these values can be considered as evidence of low statistical power.

© 2016 British Academy of Management.
Single- and Multiple-Informant Research Designs

Table 4. Standardized parameter estimates and robust standard errors (in parentheses) for the multiple-key-informant design

<table>
<thead>
<tr>
<th>Item factor loadings</th>
<th>Multiple key informant (N = 73)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP</td>
<td></td>
</tr>
<tr>
<td>( \lambda_1 ) (improvements in customer satisfaction)</td>
<td>0.666 (–)</td>
</tr>
<tr>
<td>( \lambda_2 ) (customer retention)</td>
<td>0.612 (0.289)*</td>
</tr>
<tr>
<td>( \lambda_3 ) (improvements in communication with customers)</td>
<td>0.869 (0.401)*</td>
</tr>
<tr>
<td>( \lambda_4 ) (reduction of complaints from customers)</td>
<td>0.647 (0.256)*</td>
</tr>
<tr>
<td>( \lambda_5 ) (adjusting to customers’ behaviour patterns)</td>
<td>0.624 (0.357)*</td>
</tr>
<tr>
<td>( \lambda_6 ) (approaching customers quickly)</td>
<td>0.415 (0.203)*</td>
</tr>
<tr>
<td>HPWS</td>
<td></td>
</tr>
<tr>
<td>( \lambda_7 ) (selective staffing)</td>
<td>0.672 (0.195)*</td>
</tr>
<tr>
<td>( \lambda_8 ) (comprehensive training)</td>
<td>0.935 (–)</td>
</tr>
<tr>
<td>( \lambda_9 ) (developmental performance appraisal)</td>
<td>0.549 (0.144)*</td>
</tr>
<tr>
<td>( \lambda_{10} ) (equitable reward systems)</td>
<td>0.298 (0.211)</td>
</tr>
<tr>
<td>HRF</td>
<td></td>
</tr>
<tr>
<td>( \lambda_{11} ) (functional flexibility)</td>
<td>0.777 (–)</td>
</tr>
<tr>
<td>( \lambda_{12} ) (skill malleability)</td>
<td>0.882 (0.180)*</td>
</tr>
<tr>
<td>( \lambda_{13} ) (behaviour flexibility)</td>
<td>0.646 (0.175)*</td>
</tr>
<tr>
<td>Direct effects</td>
<td></td>
</tr>
<tr>
<td>HPWS ( \rightarrow ) HRF</td>
<td>0.130 (0.164)</td>
</tr>
<tr>
<td>HPWS ( \rightarrow ) FP</td>
<td>−0.020 (0.135)</td>
</tr>
<tr>
<td>HRF ( \rightarrow ) FP</td>
<td>0.252 (0.139)</td>
</tr>
<tr>
<td>Indirect effect</td>
<td></td>
</tr>
<tr>
<td>HPWS ( \rightarrow ) FP</td>
<td>0.033 (0.038)</td>
</tr>
<tr>
<td>Total effect</td>
<td>0.013 (0.126)</td>
</tr>
</tbody>
</table>

FP, firm performance; HPWS, high performance work systems; HRF, human resource flexibility. *p < 0.05.

Supplementary analyses. With the aim of examining whether the statistical power of the model would explain the differences between the results using the single- and multiple-key-informant research designs, an additional analysis was carried out to test the existence of possible significant relationships between HPWS, HRF and FP using a multiple-informant design with an extended sample size. In this analysis we merge data used in the multiple-informant analysis with the ‘target’ information obtained using the single-informant research design. Here, we refer to ‘target’ information as the responses reported by the selected key (or well-informed) informant. In our illustration this corresponds to using information on the HRF and FP items reported by sales managers and using the information on the HPWS items reported by HR managers. Using this additional information we extend the sample by including 49 additional cases with information on the HRF and FP items reported by sales managers, and 108 cases with information on the HPWS items reported by HR managers. With the addition of these 157 cases (49 + 180), the extended sample size for the supplementary analysis is 230 cases.

As expected, increasing the sample size does not improve the fit of the model. The goodness-of-fit indices indicate that the model still does not provide an adequate fit to the data (\( \chi^2 = 115.184, \text{df} = 104, p < 0.05 \)). The factor item loadings are all statistically significant with high values (except for \( \lambda_{10} \)) and slightly smaller standard errors than in the previous analyses. The increased sample size allows us to find a significant relationship between HRF and FP, but no influence of HPWS on HRF or FP. Again, these results are different from those obtained in the single-informant design, providing additional evidence for our research question 2 that there are differences in the proposed relationships between HRM and performance when a multiple-key-informant design is adopted, in comparison with a single-informant design.

Discussion and conclusions

The purpose of this paper is to extend the discussion about the importance of the research design adopted in the HR—performance literature. By evidencing problems associated with single-informant designs, we propose using an alternative approach based on the use of multiple designs.

© 2016 British Academy of Management.
key informants. We addressed these issues in the context of the HPWS and FP relationship. In what follows we summarize the main theoretical and practical implications of our study.

**Theoretical implications**

In our illustration, results for the single-informant design show significant measurement differences (i.e. differences in factor structure and factor loadings) and differences in the significance of the relationships between HPWS, HRF and FP depending on whether the respondent was an HR manager or a sales manager, which suggests that the assumption of 'parallel raters' should be questioned (i.e. respondent matters). There are several explanations for these differences. Based on social cognition theory (Hastie and Park, 1986), Homburg et al. (2012) demonstrated that the informants' position in the organizational hierarchy significantly influences the reliability of their responses. The quantity and quality of the information possessed by HR managers and sales managers may vary, as may their degree of involvement in the questions analysed in our study. For instance, HR managers may be more knowledgeable about the organization's HR practices than sales managers, while sales managers can provide more accurate information about the flexibility of employees in the commercial department. In addition, implicit theories about the relationship between HRM and performance may be more prevalent when the HR manager is the informant, given the knowledge that he or she may have about the benefits of certain HR practices for FP, but less relevant when the sales manager evaluates the consequences of the HR practices. When systematic differences between raters are present, pooling responses into a single dataset could lead to misinterpretation of the results.

A related issue is how differences between sales and HR managers should be interpreted. Previous studies in the HRM field (e.g. Wright et al., 2001) considered differences across raters (mainly in assessing HR practices) as measurement error attributed to the lack of rater reliability. They implicitly adopt the 'normative accuracy perspective' (Vandenberg, Lance and Taylor, 2005) in which raters are assumed to be parallel informants. However, this assumption is often particularly restrictive (Murphy and De Shon, 2000) because raters are not equally knowledgeable and may observe different aspects of the construct domain (Lance et al., 2010). In our illustration, we believe that differences between sales and HR managers should not be interpreted as measurement error but measures provided from different perspectives. For instance, recent research lines in HR management recognize the existence of different perspectives in measuring HR practices when a distinction is proposed between exposed and actual practices (Nishii and Wright, 2008).

**Practical implications**

Given the methodological nature of this paper, our practical implications are mostly addressed to researchers in the HRM field. In our illustration, discrepancies between sales and HR managers are observed in the significance of the relationships between HPWS, HRF and FP. A hypothetical researcher who ignores these differences and assumes that HR managers and sales managers are parallel raters (i.e. alternative forms of measurement) could consider it appropriate to pool responses from these two groups of informants and estimate the research model based on the pooled dataset. By doing so, the researcher would observe a significant relationship among all the constructs in the model, as indicated in the first column of Table 2, and would therefore draw conclusions based on these relationships. We suggest caution is due in the decision to pool data provided by informants occupying diverse positions in the firm. In this regard, Rungtusanatham et al. (2008) suggest analysing the measurement equivalence across different groups prior to performing the statistical analyses. If there are differences between the information provided by the different types of raters, then the statistical results should be discussed separately for each rater group. These authors also claim that the worst thing to do is to ignore the issue of measurement equivalence and provide a discussion of the results for the pooled data that fails to appreciate how the lack of measurement equivalence affects the validity of the results.

The second implication of our illustration refers to the appropriateness of collecting data from multiple key informants in order to increase the validity of the measures. Some shortcomings can arise in using the single-informant approach when both criterion and dependent variables are assessed by the same informant in the same questionnaire, creating spurious correlations among the variables.
of interest and the ubiquitous problem of CMV. In order to avoid the CMV problem, some authors suggest alternative data collection strategies, based on splitting the questionnaire into different sections and addressing each section to the best-informed respondent (Huselid and Becker, 2000). In our illustration we found differences in the results obtained using the single-informant and the multiple-key-informant research designs in terms of the relationships among the constructs of the model. In particular, results from the multiple-key-informant design show no significant relationships between HPWS, HRF and FP but some of these relationships were statistically significant when only one informant provided all the information. We cannot claim that CMV explains the significant relationships in the single-informant design, as we are aware that the small sample size and the low statistical power for detecting significant relationships may affect the results, but nonetheless we recommend researchers collect data from multiple key informants.

Past research on the relationship between HRM and firm performance has shown interest in this question. The work of Gerhart, Wright and McMahan (2000), Huselid and Becker (2000) and Gerhart et al. (2000), published in Personnel Psychology, deserves particular mention. These papers opened up a debate about the most appropriate data collection strategy. Gerhart, Wright and McMahan (2000), and also Wright et al. (2001), implicitly adopting the parallel rater assumption, recommended using several respondents in a firm to assess the HR variables as a way to increase the reliability of the measures and to reduce the ‘error’ due to raters. However, Huselid and Becker (2000) argue that the validity of the measures does not increase by adding multiple respondents who have insufficient information. Rather, these authors recommend collecting data based on the opinions of respondents that are really key (or well-informed) informants. Based on their own experience, Huselid and Becker (2000) suggest that it is common practice among managers when responding to surveys to assign different sections of the questionnaire to people in the organization who are more knowledgeable about each area. On this issue we agree with Spector and Brannick (2010, p. 403) when they suggest that ‘researchers who conduct studies using such methods [self-report surveys answered by a single person] often find themselves subject to editor and/or reviewer complaints that results are suspect due to the likelihood of common method variance affecting results’. We believe that multi-source data collection studies are more easily accepted in top-tier journals because this methodological design avoids the risk of CMV.

Given our recommendation to collect data from multiple key informants, we agree with Rungtusanatham et al. (2008) that researchers’ efforts should focus on selecting the key informant before data collection begins. In this endeavour it is also important to select the raters depending on the consistency between the raters’ perspectives or frames of reference and the substantive research objectives. Following Wright et al.’s (2001) recommendation, in our illustration we chose the best informed raters based on their position in the firm, and also on prior empirical work in the HRM field. However, some improvements could be made in choosing the key informants. A plausible strategy to empirically evaluate the appropriateness of the selected informants is to collect information from several ‘a priori key informants’ and then use measures of intrarater agreement, such as James, Demaree and Wolf’s (1984, 1993) rWG or other alternative measures of agreement (Le Breton and Senter, 2007). Lack of agreement might indicate that those respondents are not as informed as originally thought. If this is the case, some additional steps can be followed to select the key informant. For instance, Kumar, Stern and Anderson (1993) suggest two approaches to evaluate informants’ competency, namely, using overall evaluations of informant competency (e.g. respondent’s tenure in the firm) and/or including specific measures in the survey that assess the level of an informant’s knowledge on the issues included in the questionnaire. Similarly, Wright et al. (2001) suggest asking the respondent for his or her confidence in rating items. Once the key informant has been identified, we agree with Wright et al. (2001), who state that researchers should provide clear instructions indicating who should complete each part of the questionnaire to avoid naïve or ill-informed respondents.

Despite its advantages, analysing data from multiple key informants brings additional difficulties for researchers such as missing data that arise

6The authors are grateful to an anonymous reviewer for suggesting this point.
when only one of the key informants in the organization provides responses. In the illustration, we used the full information ML approach to missing data. Other approaches such as multiple-group SEM (Allison, 1987) or multiple imputation (Little and Rubin, 1987), currently available in several commercial software packages, could also be applied to deal with this issue. In particular, a Monte Carlo simulation study would assess which method works best in estimating the population values for different sample sizes and with different amounts of missing data. Future research could usefully explore these issues as a way to provide more convincing empirical evidence on the use of the multiple-key-informant approach.

Finally, the multiple-key-informant design can also be applied when there are many possible key informants for a specific construct. This is the case, for instance, when employees' job satisfaction is the construct of interest and a large number of employees can be considered as key informants. In this case, the approach can be extended to a multilevel analysis in which multiple informants from the same organization are surveyed. Furthermore, multilevel analysis could also be a more efficient way of analyzing data coming from multiple-key-informant research designs. In the multilevel analysis the type of informant (e.g. HR director or sales manager) can be taken into account as an individual-level variable, addressing the possible bias associated with particular groups of informants. In addition, using additional informants would allow their interrater agreement to be assessed.

### Limitations and future research

As with all empirical studies our results are not free of limitations. One is the small sample size used in the analysis. The statistical power to detect significant relationships is low, and this reduces the likelihood of detecting significant relationships between HPWS, HRF and FP. We have addressed this point by conducting supplementary analyses using an extended sample size. With a large sample size a significant relationship between HRF and FP is found. This suggests that more research is needed to thoughtfully explore these relationships.

A second limitation is that the analysis does not control for the influence of other variables that may explain the differences in the responses provided by the HR and sales managers. Although we have provided arguments to justify why differences are expected to be found between informants who hold different positions (amount of knowledge or information about the constructs, implicit theories, social desirability etc.), in the literature it is well known that other contextual variables and firm characteristics may also explain these differences. For example, firm characteristics such as size and complexity, level of communication and information throughout the organization, degree of centralization, geographical distance, among others, could potentially confound the results and make them less reliable.

Specifically, in our illustration we use information provided by HR and sales managers belonging to different organizations, and therefore results may be caused or explained by our use of different types of respondents from different organizations and by our inability to explicitly control for other alternative explanations. Although in the research design the organizations were randomly assigned to each type of informant (and it is expected that randomization controls for possible systematic differences in the characteristics of the firms assigned to each type of informant) it would be interesting to explicitly test the effect on the data and results of using different types of informant when information is collected using the same questionnaire administered to key informants in the same organization. Comparing responses from multiple key informants in the same firm would allow unmeasured firm characteristics to be controlled for, ruling out the possible influence of such potential confounding factors. Unfortunately, this information is not available in our study.

In sum, based on the theoretical review in the paper, we suggest that the respondent matters in studies that collect data through surveys. Thus, we recommend researchers interested in analysing the HRM–performance relationship place more emphasis on the careful selection of the key informants in each organization prior to collecting data, and on collecting data through multiple informants to avoid CMV when the research questions require more than one informant per firm.

---

7The authors are grateful to an anonymous reviewer for suggesting this point.

8The authors are grateful to an anonymous reviewer for suggesting this point.
## Appendix 1: Survey research approaches in the HRM—performance literature

<table>
<thead>
<tr>
<th>Research approach</th>
<th>Description of the approach</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach 3</strong></td>
<td>Multiple informants: all the informants in each unit of analysis provide answers to all questions within the same questionnaire to assess measurement reliability</td>
<td>Datta, Guthrie and Wright (2005), Snell and Youndt (1995), Subramony et al. (2008), Veld, Paauwe and Boselie (2010), Youndt and Snell (2004) (N = 5)</td>
</tr>
</tbody>
</table>

**Notes:** This review was based on empirical studies that use surveys to collect data included in Jiang et al.’s (2012) meta-analysis, except for doctoral dissertations, papers presented at academic meetings and working papers. Obviously, in papers where the dependent variable (i.e. organizational performance) is measured by archival data, the common source bias is not a significant issue. Our approaches 1 and 2 correspond to Rungtusanatham et al.’s (2008) approaches C and A, respectively. Approaches 3 and 4 correspond to Rungtusanatham et al.’s (2008) approach B that involves the use of multiple informants.

*These papers partially gather some characteristics of approach 3 since they account for measurement reliability by providing multiple informants to some questions within the same questionnaire. For instance, they analyse employee perceptions through multilevel analysis.

© 2016 British Academy of Management.
Appendix 2: Questionnaire measurement scales

Firm performance

(Likert type seven-point scale: 1, worse than our competitors; 4, equal to our competitors; 7, much better than our competitors)

- \( V_1 \) Improvements in customer satisfaction
- \( V_2 \) Customer retention
- \( V_3 \) Improvements in the communication with customers
- \( V_4 \) Reduction in the number of complaints or claims from customers
- \( V_5 \) Being able to adjust to customers’ behaviour patterns
- \( V_6 \) Approaching customers quickly

High performance work systems

- \( V_7 \) Selective staffing
  - How extensive is the employee selection process for a job in this department? (1, not extensive; 4, i.e. use of few staffing techniques; 7, very extensive; i.e. use of many different techniques)
  - How long does it take to select someone for a position in this department once the job becomes open? (1, short, less than one week; 4, medium, about six months; 7, long, more than one year)
  - How many people are involved in the selection decision? (1, one person; 7, seven persons)
  - How much money is generally spent in selecting people for a job, as a percentage of the firm’s profits? (1, very little; 4, moderate amount; 7, great deal)
  - How many applicants are screened for each person hired for a job (1, one; 7, 20 or more)

- \( V_8 \) Comprehensive training
  - How formal or structured are the training processes in this department? (1, very unstructured; 7, very structured)
  - What percentage of employees in your department have received training this past year? (1, 0%; 4, 50%; 7, 100%)
  - On average, how many hours of formal training does a typical member of your department receive per year? (1, 0 h; 7, >60 h)

- \( V_9 \) Developmental performance appraisal
  - What percentage of employees in your department are covered by performance appraisal systems? (1, 10% of employees or less; 4, 40% of employees; 7, 70% of employees or more)
  - How closely are pay raises, promotions etc. tied to performance appraisal in this department? (1, not closely; 4, occasionally; 7, daily)
  - How often is performance discussed with employees? (1, rarely; 4, occasionally; 7, daily)
  - How would you describe the approach used to discuss performance? (1, tell/sell; 4, tell/listen; 7, problem solving)
  - How many people provide input to the performance evaluation of each employee in the department? (1, one; 7, seven or more)

- \( V_{10} \) Equitable reward systems
  - How would you rate pay levels in this unit relative to other firms? (1, low; 4, same; 7, high)
  - How would you rate the pay levels in this unit relative to past years? (1, lower than past; 4, same; 7, higher than past)
  - How wide is the range in pay across members in this department? (1, narrow; 4, moderate; 7, very wide)
  - How closely is pay tied to individual performance, in terms of percentage of the salary? (1, <10%; 2, 10%–20%; 3, 20%–30%; 4, 30%–40%; 5, 40%–50%; 6, 50%–60%; 7, >60%)

© 2016 British Academy of Management.
Single- and Multiple-Informant Research Designs

30%−40%; 5, 40%−50%; 6, 50%−60%; 7, >60%)

Human resource flexibility

(Likert type seven-point scale; 1, applies to very few employees; 7, applies to most of the employees)

$V_{11}$ Functional flexibility

- Our employees can switch to new jobs with similar responsibilities to their current jobs within a short time
- Our employees can switch to more qualified jobs within a short time
- Employees in our department have multiple skills that can be applied to tasks corresponding to other jobs

$V_{12}$ Skill malleability

- Employees in this department try to constantly update their skills and abilities
- Employees in this department learn new procedures and processes introduced in their jobs within a short period
- Employees in this department anticipate future skill requirements that may be needed to perform their jobs
- When employees in this department are not able to perform a specific task, they quickly learn how to do it

$V_{13}$ Behaviour flexibility

- When employees detect problems in performing their jobs, they voluntarily try to identify the causes of these problems
- Most of the changes that have taken place in this department were introduced by employees
- Employees in this department act efficiently when a problem emerges, even in cases in which they do not have full information about the problem
- Employees in this department act efficiently under uncertain and ambiguous circumstances

References

Single- and Multiple-Informant Research Designs


Hastie, R. and B. Park (1986). ‘The relationship between memory and judgment depends on whether the judgement is memory-based or online’, *Psychological Review*, 93, pp. 258–268.


Single- and Multiple-Informant Research Designs


human resource management. Some of the results of her researches have been presented in national and international conferences and published in journals such as *Journal of Management, Human Resource Management, International Journal of Human Resource Management* and *Personnel Review*.

Vicente Roca-Puig is an associate professor at the Department of Business Management and Marketing at Universitat Jaume I (Castellón, Spain). He teaches courses at the MBA and PhD level on human resource management. He is currently working on high-commitment management and contingent workers in small firms and its interrelationship to various organizational strategies. He has published in *Journal of Management, Journal of Operations Management, Personnel Review, International Journal of Human Resource Management* and *Management Research*.

Ana B. Escrig-Tena is an associate professor at the Department of Business Management and Marketing at Universitat Jaume I (Castellón, Spain). She obtained her PhD in business management from the same university. Her research interests include the contribution of human resource management to the implementation of quality management. She has published in such journals as *Journal of Operations Management, Journal of Management, Decision Sciences* and others.

© 2016 British Academy of Management.
Author Queries

Dear Author

Please respond to the following queries by marking up your proofs with the necessary changes/additions. Please write your answers clearly on the query sheet if there is insufficient space on the page proofs.

Q1: Author: Please confirm that given names (red) and surnames/family names (green) have been identified correctly.
Q2: Author: Please confirm the change to Ghee (from Chee) so that this reference matches the references.
Q3: Author: Please confirm the short title or provide another one, 50 characters maximum.
Q4: Author: Please confirm SEM in full.
Q5: Author: Bonnet is spelt Bonett in the references. Which is correct please?
Q6: Author: Please check the wording of the first column.
Q7: Author: Please confirm SRMR in full.
Q8: Author: Please confirm the change to Ghee (from Chee) so that this reference matches the references.
Q9: Author: Bentler 2006 isn’t cited in the text. Please insert a mention or delete the reference.
Q10: Author: Borucki and Burke isn’t cited in the text. Please insert a mention or delete the reference.
Q11: Author: Dempster et al. isn’t cited in the text. Please insert a mention or delete the reference.
Q12: Author: Should the initials for Fey be G. F. or C. F.?
Q13: Author: Huselid et al. isn’t cited in the text. Please insert a mention or delete the reference.
Q14: Author: Katz et al. isn’t mentioned in the text or in the appendices. Please insert a mention or delete the reference.
Q15: Author: Please correct the page range.
Q16: Author: Muthen and Muthen isn’t cited in the text. Please insert a mention or delete the reference.
Q17: Author: Please give the location of the publishers.
Q18: Author: Please give the volume number.
Q19: Author: Wright and Snell isn’t cited in the text. Please insert a mention or delete the reference.