Cranet and Comparative Research

Conceptual and Methodological Issues in Comparative HRM

Research: The Cranet Project as an Example

DRAFT: Please cite as

Abstract

The “Cranfield Project on International Strategic Human Resource Management” (Cranet) is an international research network which regularly conducts a survey of human resource management (HRM) issues in more than 40 countries. As a result, Cranet provides rich data of the human resource department’s institutionalization, HRM practices, and organizational characteristics. Whereas scholars in other research fields (e.g., organizational behavior, marketing) have intensively discussed scientific standards for cross-national and cross-cultural studies, there is a need for more comprehensive discussion in the HRM field. The present article discusses four methodological challenges of comparative research (i.e., survey administration, sampling strategies and nonresponse bias, equivalence of constructs and measurement, and analytical techniques) and illustrates these issues based on the design and methodology of the Cranet project.
Conceptual and Methodological Issues in Comparative HRM Research: The Cranet Project

International human resource management (HRM) focuses on the strategies, systems, and practices that companies apply to leverage their employees' potential worldwide (Lazarova, Morley, & Tyson, 2008). In this context, comparative research on international HRM covers a broad range of cross-cultural and cross-national topics such as global staffing practices in multinational enterprises (e.g., Collings, Morley, & Gunnigle, 2008; Collings, Scullion, & Morley, 2007), social policy practices and employee participation across countries (Brandl, Mayrhofer, & Reichel, 2008; Poutsma, Kalmi, & Pendleton, 2006), the cultural dependence of vocational training (e.g., Beck, Kabst, & Walgenbach, 2009), foreign subsidiary autonomy in setting HRM practices (Fenton-O’Creevy, Goodeham, & Nordhaug, 2008) or HRM consequences of different employment systems (e.g., Gunnigle, Collings, & Morley, 2007; Heraty, Morley, & Cleveland, 2008). Over the years, the interest in comparative research on international HRM has increased steadily (for reviews, see Lazarova et al., 2008; Schuler, Budhwar, & Florkowski, 2002).

Whereas methodological issues salient in comparative research have been widely discussed in fields like organizational behavior (e.g., Schaffer & Riordan, 2003; Tsui, Nifadkar, & Yi, 2007) or marketing (Baumgartner & Steenkamp, 2001; He, Merz, & Alden, 2008), such a discussion is still missing in the field of HRM. However, with an increasing number of comparative studies in HRM research, it is important to elaborate which methodological standards are crucial for comparative HRM studies and to discuss how these methodological standards can be implemented into practice in a well-established research project on international strategic HRM.
In order to contribute to the academic literature on comparative HRM research, the aim of our paper is twofold. First, we summarize and discuss main methodological issues salient in comparative research in general, chronologically along the process of comparative research. For instance, researchers must secure that samples represent their national population and do not deviate in other matters or use equivalent measures of the addressed construct or measurement characteristics (e.g., interpretation of the scale units). Comparative studies impose their own difficulties such as the administration of the survey which has to be coordinated and carried out in multiple countries. In particular, the present paper discusses 1) the requirement for an integrated and coordinated survey design, 2) standardized survey execution, 3) the diagnosis and treatment of nonresponse bias, and 4) the analytical techniques suited for comparative research (i.e., moderated regression, multilevel analysis, and structural equation modeling).

Second, whereas previous discussions on comparative research relied on a review of selected articles from various fields, we examine how methodological standards are applied in a large research project in international HRM (drawing on the example of the Cranfield Project on International Strategic Human Resource Management; briefly Cranet). Cranet is a large scale network of scholars from universities across the world, representing more than 40 countries. Since 1989, Cranet regularly performs a survey inquiring into HRM policies and practices in its member countries. Up to now, this comprehensive survey of major HRM areas is unique in comparative HRM research with regard to its size, age, and composition of network members from all over the world. Focusing on this project enables discussing how and if standards are treated in a large integrated network of researchers and, thus, can inform readers about the typical struggles in comparative HRM research. In particular, a large network like Cranet faces the dilemma that, on one hand, the value of comparative research increases with the number of
participating countries but, on the other hand, meeting the standard and conducting the survey in an integrated, standardized, and methodologically adequate way is getting more difficult at a certain size. In this regard, we found issues where Cranet meets standards and others where Cranet still could improve. Hence, the present paper provides practical insights for researchers conducting comparative studies.

The Cranfield Project on International Strategic Human Resource Management

Cranet is a network of universities and business schools from over 40 countries worldwide. Since 1989, the Cranet network collects data on HRM practices and policies via a standardized survey approximately every four to five years (Brewster & Hegewisch, 1994; Brewster, Mayrhofer, & Morley, 2000; Kabst, Giardini, & Wehner, 2009). The survey covers major areas of HRM policies and practices like recruitment and staffing, training and development, compensation and benefits, and employee relations and communication.

Whereas early Cranet surveys included only European countries, the network has expanded continuously and today also includes non-European countries like Australia, Canada, Japan, South Korea, South Africa, or the USA. The growth of the network was associated with a shift in the themes under study. In the early years, the prevalent research question was if HRM practices converge in the European Union or whether countries keep their own HRM policies and practices (e.g., Brewster, 2004; Brewster & Hegewisch, 1994; Gooderham, Nordhaug, & Ringdal, 1999). More recently, the focus shifted to issues with a broader interest, for example, the relationship between HRM practices and performance (Apospori, Nikandrou, Brewster, & Papalexandris, 2008; Nikandrou, Apospori, Panayotopoulou, Stavrou, & Papalexandris, 2008) or the influence of culture on training and development (Beck et al., 2009; Nikandrou, Apospori, & Papalexandris, 2008). In the field of comparative HRM research, Cranet is unique regarding its
size, its growth, as well as its continuity in performing the survey. However, the growth of the network over years challenges Cranet with respect to methodological issues. These challenges are not unique to the Cranet network but concern comparative research in general. In the following, we discuss methodological issues which are of major concern for comparative HRM research and illustrate how Cranet deals with these issues in practice.

**Survey Design**

Scholars (i.e., Schaffer & Riordan, 2003) stress that preparation and execution of comparative studies necessitate particular efforts which single country studies do not require. For example, whereas single country studies concern a rather homogenous group of researchers coming from one country, comparative studies concern a multitude of researchers coming from different countries, with their own culturally based research interests, working styles, different competencies, and sets of facilitators (e.g., funding) or restrictions. The most basic task in a comparative research project is, hence, to transcend the cultural differences of the researchers and converge to a standardized procedure of how the study has to be conducted. As most salient issues, we identify the a) necessity to apply a common research design, b) the development of equivalent measures, and c) sound and comparable sampling.

First, the study has to apply a *common research design*. If data are collected by questionnaire in one country and by a personal interview in another, it is difficult to compare the countries. In the Cranet context, all countries collect data with a standardized questionnaire which is addressed to the most senior HR manager in the organization. Some of the HRM topics are surveyed in each wave to investigate changes over time. For example, questions referring to staffing procedures and training reveal changes over time as well as differences across countries. In addition, the questionnaire is continuously adapted to include new questions in order to
analyze if novel HR practices become legitimized HR practices over time. For instance, questions about downsizing and outsourcing were added to reflect the increasing prevalence of these practices. Cranet reexamines and adapts the questionnaire before each survey round (Lazarova et al., 2008). In addition, Cranet arranges regular meetings (approximately every six months) to discuss potential research topics, measures, and finally, the adaptation of the questionnaire. All these efforts undertaken by Cranet are valuable and in line with comparative research standards.

Second, participating countries should jointly develop equivalent measures, that is, are understood by respondents from different countries in the same way (Hui & Triandis, 1985). Beyond translating existing measures, participating countries should discuss culturally sensitive measures in order to develop measures which are not culturally biased. General survey research (Groves, 1996; Krosnick, 1999) offers a wide array of probing techniques that can be applied to assure that respondents understand questionnaire items as intended. In the context of comparative HRM research, probing should aim at an equal understanding of the practices. HRM researchers may feel that this is obvious but, for instance, the literature on part-time employment (Barling & Gallagher, 1996) has shown that the part-time concept differs across countries. Hence, simply asking for the percentage of applied part-time jobs may yield substantial differences that are only due to the unspecific question formulation.

To achieve these standards, Cranet uses translation/back-translation techniques (Brislin, 1970; Hui & Triandis, 1985) to establish the semantic or linguistic equivalence of the measures. In addition, a steering group consisting of selected members of the network is appointed before each new survey round starts. The task for the steering group is to develop the new (master) questionnaire based on the previous questionnaire, however, taking in account methodological
improvements as well as advancements in HR practice and policy. Measures that may have
different meanings across countries or that may not be applicable to a specific institutional
setting are addressed. The final questionnaire is discussed with all network members during the
regular meetings (which takes place approximately every six months).

Third, Schaffer and Riordan (2003) note that the sampling strategy employed in the
countries should also be as similar as possible in order to draw equivalent samples. Because
sampling was intensively discussed in cross-cultural survey research (Armstrong & Overton,
1977; Brehm, 1993; Groves et al., 2004; Häder & Gabler, 2003) and in the Cranet network itself
(e.g., Brewster, Tregaskis, Hegewisch, & Mayne, 2000; Groves et al., 2004; Häder & Gabler,
2003), we will treat this issue more in-depth.

Appropriate sampling of respondents from a population is crucial for any study as it
serves as the foundation for the estimation of population parameters (i.e., means, percentages,
correlations, and regression coefficients) and significance tests. In cross-national or cross-
cultural comparisons, proper sampling is necessary to enable correctly attribution of differences
in the sample statistics to differences between the compared populations. However, practice is
often far from ideal standards due to restrictions in time, money, or human resources. In this
regard, applied research has to rely on convenience samples which represent the intended
population to an unknown degree.

Statistical theory states that significance tests require random sampling. Without random
sampling, the spread of the sampling distribution is unknown and estimates of standard errors
and, thus, p-values may be incorrect. Nonrandom sampling may be the result of either coverage
bias (e.g., some units of the population are not considered) or selection bias (i.e., some units
have a low or zero probability to take part in the study). If, in addition, the sample systematically
differs from the population in some characteristic, the estimate of the target parameter (e.g., percentage of a HRM practice) may be biased. For this reason, Häder and Gabler (2003) recommend creating a sampling frame (i.e., a list of population members) that is as comprehensive as possible to enable that each member of the population has a nonzero probability to be selected. The sample should be a random draw from the sampling frame. One presumption of a comprehensive sampling frame is the development of a clear and detailed definition of the population which should be shared by all participating countries. To avoid selection bias, researchers have to secure that members, once selected, have an equal chance to participate.

In the Cranet network, each participating country is in charge of developing a list of organizational addresses and personal contacts. Available international databases in Europe, however, are often incomplete and lack important information (e.g., public sector organizations, contact persons, and comprehensive organizational addresses). Moreover, an international database for all organizations in all countries is missing. Therefore, the Cranet partners themselves decide which database should be used for the Cranet survey within the country (Häder & Gabler, 2003; Tregaskis, Mahoney, & Atterbury, 2004). For example, some countries use commercial databases (e.g., France, Germany, and United Kingdom), other countries use databases from professional associations (e.g., Greece and Israel), governmental databases (e.g., Sweden and Finland), or a combination of all sources (e.g., Portugal and Turkey).

Cranet defined the population for all participating countries as private and public organizations with more than 200 employees. With regard to organizational size, Brewster et al. (2000) argue that formal HRM functions are more likely in larger organizations. For example, an examination of differences in expenditures in training and development across nations may lead
to biased results when comparing countries with many large companies (e.g., 200 or more employees on average) to countries with many small companies (e.g., 200 or less employees on average). Although in some participating countries a large number of organizations are smaller than 200 employees (e.g., Netherlands or Lithuania), Cranet sets this threshold for comparative HRM studies using Cranet data (Tregaskis et al., 2004). In addition to size, the Cranet network agrees on a population which contains private as well as public organizations (Brewster, Hegewisch, Mayne, & Tregaskis, 1994). Cranet members are responsible for including a representative proportion of public sector organizations to their list of organizational addresses. Thus, some countries had to integrate other lists to get a more representative sample (e.g., Belgium, Brewster et al., 1994; Tregaskis et al., 2004).

Although these efforts can be appreciated, Cranet's sampling frame may be further optimized in order to meet ideal sampling standards so that every organization fitting the population criterion has the same chance to participate in the study. As it may be impossible to construct a sampling frame of organizations, the easier alternative may be a sampling frame of macro-units (postal districts, counties, area mapping) from which a number of macro-units can be randomly drawn (Häder & Gabler, 2003). In this regard, multistage sampling may be a possibility: In the first stage, macro-units would be sorted into strata (i.e., categories defined on basis of some important population characteristics). In the second stage, organizations would be randomly sampled from these strata. Another option is to apply cluster sampling, that is, to randomly sample a number of macro-units and try to administer the survey to all organizations in the unit. All efforts to increase the comprehensiveness of the sampling frame result in a higher approximation of the sample to an ideal random sample.
To sum up, Cranet undertakes comprehensive efforts, especially with regard to develop a common research design (i.e., standardized questionnaire) and/or to develop equivalent measures (i.e., by translation/backtranslation techniques) in each participating country. However, as it is the case for any empirical research project, there is room for improvement in order to further approach ideal sampling standards. Although we admit that it may not be possible to create a complete sampling frame, researchers should not deter from increasing its comprehensiveness. Cranet only defines its target population but is lacking a consistent sampling frame. In addition, we recommend that pretesting of the questionnaire should be conducted with higher intensity in order to achieve valid and comparable measurement of HRM practices.

Survey Execution

After the study has been designed and the questionnaire has been developed, participating countries should execute the study in an integrated and standardized way. Substantial differences in the procedures can otherwise lead to different nonresponse rates (see the section below) or differences in the results of statistical analyses which are erroneously attributed to differences between the countries or cultures.

First, the manner in which the survey is executed (i.e., "procedural equivalence", Schaffer & Riordan, 2003) needs to be as similar as possible. For example, all participating countries should use the same means of distributing the questionnaire, develop a similar cover letter and instructions for completing the questionnaire, and contact the participants in the same way to remind nonrespondents. It may, however, be sometimes necessary to adapt the cover letter, instructions, and text of the reminder to cultural specifics. In these instances, rigid standardization may be more harmful than beneficial but the differences in the administration should be jointly discussed, documented, and coded¹.
With regard to the distribution of the questionnaire, the common procedure in Cranet is to send it out via postal mail (Brewster et al., 1994; Brewster, Tregaskis et al., 2000; Tregaskis et al., 2004). Cranet, however, has refrained from standardizing the cover letter or text of the reminder for all countries. The local partner schools in each country are responsible for the study execution (Brewster et al., 1994). Therefore, each research partner in Cranet plays a vital role for the execution of the survey in terms of subject, cultural, and academic expert (Brewster et al., 1994; Tregaskis et al., 2004). After each survey round, the participating countries document their method of conducting the survey and send the documentation to the research team of the Cranfield School of Management. Since the first survey round in 1989, the Cranfield team has been the central coordinator and has also been in charge of the expansion of the project. This central role of Cranfield ensures that the study has always been executed as similar as possible.

Second, it is necessary to execute the survey in a similar time frame. For instance, a change of environmental conditions during the data collection process may contaminate the quality of data across different national settings (Brewster et al., 1994; Schaffer & Riordan, 2003). For example, studying employee retention management across nations may be seriously biased when data collection has taken place before the financial crisis in country A, during the financial crisis in country B, and after the financial crisis in country C. Differences across nations may simply be due to inconsistencies in terms of the timing of data collection. Hence, when influential economic or institutional conditions change within the survey period, differences between the countries in the prevalence of some HRM cannot be attributed to national or cultural differences. Therefore, administering the survey in a consistent time frame serves to hold influential background variables constant. If differences in the time frame occur,
each country should code the time frame. In this regard, correlations between the time frame and the average value of some HRM practice could indicate the amount of bias.

The members of the Cranet network agree on a period of approximately 18 months in which all of the participating countries undertake the survey. Whereas in the early years the time frame for undertaking the survey in all countries was only 12 months, the continuous growth of the network (up to 41 member countries today) made the extension of the time frame necessary. The survey period of 18 months takes account for different institutional settings as well as different funding opportunities of the network members. A longer time frame increases feasibility of undertaking a survey in a large network, however, may also cause limitations or at least a greater need for methodological documentation. The example of the financial and economic crisis during the survey period of the 7th round of the Cranet (2008-2009) illustrates that a documentation of the precise survey time of each country is necessary. Research questions like extent and measures of downsizing, training and development, or recruitment may need to be interpreted in respect to the timing of the survey in each country.

In summary, while Cranet generally meets methodological standards, the growth of the network brings about challenges with the similar study execution. Here, Cranet faces a dilemma. On one hand, increasing the number of participating countries is valuable as it broadens the empirical basis for its comparative studies. On the other hand, the increase of participating countries makes it difficult to meet the methodological standards. For instance, each participating country is responsible for the funding of the study (e.g., printing and sending out the questionnaire). Therefore, the size of the network makes it difficult to ensure that the funding is equally available at the same time. As a result, each Cranet member defines independently the point of time (within the given time frame of 18 months) when the study is executed. Moreover,
due to problems with funding some Cranet members switch to less expensive methods of collecting the data than via postal mail, such as online survey. We recommend that exceptions in survey execution should be documented in order to thoughtfully examine and interpret the data (e.g., select the countries for data analysis or discuss potential consequences of deviations from methodological standards).

**Diagnosis and Treatment of Nonresponse Bias**

Survey nonresponse is a substantial threat for any study and different nonresponse rates across countries make differences in the results ambiguous. Nonresponse means that parts of the addressed sample rejects participating in the study, is not able to participate (e.g., lack of time), or forgets to participate. Nonresponse bias occurs when nonrespondents differ systematically from respondents in important characteristics. Survey research shows that respondents differ from nonrespondents in personality traits as well as organizational attitudes. For instance, Rogelberg et al. (2003) found that nonrespondents were less conscientious and agreeable than respondents. In addition, individuals lacking the intention to participate in a study were less satisfied with their job, less committed to their company, and had a higher intention to quit their job (Rogelberg, Luong, Sederburg, & Cristol, 2000). Hence, it has to be assumed that samples in organizational and HRM research differ systematically from the population with regard to some characteristics. Newman (2009) noted that “nonresponse is a behavioral indicator of one or more latent constructs, and these constructs can be substantive forces in empirical models” (p. 22).

Whether deviations of the sample from the population regarding some characteristic lead to biased estimates depends on the role of the characteristic in the research model. In this regard, Groves et al. (2004) distinguished between descriptive aims of the study (e.g., describing the percentage of a HRM practice) and analytical aims (e.g., estimating regression coefficients).
Whereas estimating relationships between variables are usually robust against nonresponse bias (Brehm, 1993; Groves et al., 2004; Krosnick, 1999), descriptive statistics are often biased. That means that results from regression analyses or structural equation modeling should less likely be affected by nonresponse bias (see, for an opposite view, Short, Ketchen, & Palmer, 2002) but differences in percentages or means may be biased. The issue becomes especially crucial in comparative HRM research when differences in the percentage of an applied HRM practice are more likely to result from differences between nonresponse rates than from differences between national characteristics. Generally, we can specify the circumstances under which parameter bias occurs as follows.

Biases of descriptive parameters (e.g., means, percentages) are likely if the sample differs from the population in the mean of a characteristic which correlates with the target variable for which descriptive statistics are calculated. For example, if assessment centers are less often applied in organizations with high work load in the HRM department and work load decreases participation in the study, then the percentage of applied assessment centers would be higher in the sample than in the overall population. Biases of relationship parameters (e.g., correlation or regression coefficients) mainly occur when the characteristic functions as a moderator of the relationship in question. For instance, if work load moderates the effect of assessment centers on performance because the preparation or execution of the assessment center has a lower quality and, again, the sample consists of organizations with lower work load, then the relationship between use of assessment centers and performance would be overestimated. In addition, relationship parameters become biased when the sample differs from the population in other aspects that affect the relationship (e.g., lower reliability of the measures, variance restriction). Newman (2009) emphasizes that a low response rate per se is not critical as the
sample can be nevertheless representative for the population regarding the estimated parameter. However, his study showed that the combination of nonresponse bias and a low response rate has substantial effects on the results of statistical analyses.

To reduce nonresponse bias, Cranet applies a number of established procedures (Kessler, Little, & Groves, 1995; 2009). For instance, participating countries develop a short cover letter (of around 12 lines, see Kessler et al., 1995) that introduces Cranet, the research topic, and presents contact data. These procedures intend to increase interest and motivation to participate in the study. In addition, nonrespondents are contacted by telephone. To assess the potential extent of nonresponse bias, Cranet compares early and late respondents (Armstrong & Overton, 1977). If late and early respondents differ significantly in respect to constructs that may function as influential variables, nonresponse may jeopardize the data.

While these procedures are widely accepted in the literature, we recommend additional procedures to assess the extent of nonresponse bias. For instance, investigating the response behavior itself should become a research topic for the Cranet network. Research has differentiated between active and passive nonrespondents (Rogelberg et al., 2003). Active nonrespondents reject participating in the study, whereas passive nonrespondents intend to participate but lack time or resources or simply forget to participate. Because it is the active nonrespondents who differ substantially from respondents (Rogelberg et al., 2003), nonrespondents could be categorized in passive and active based on expressions when being reminded to fill out the questionnaire. For example, the person’s explanations about his/her reasons not to participate could be used to assess if systematic causes prevented participation. As explained above, such systematic causes may be related to the research model in question. In summary, the proposed strategies could enable estimating the extent of active versus passive
nonrespondents, deliver further information about possible causes of nonresponse and, most importantly, enable comparing the extent of nonresponse bias across countries.

As a final remark, we emphasize that all efforts to increase the response rate have their limitations (Kessler et al., 1995). As Newman (2009) notes, data gathered from respondents who would not have responded under normal conditions, may even bias the sample more due to differential willingness to follow the reminder. In addition, persons who respond only after being reminded may give less reliable and valid answers. Therefore, research has to find an optimal trade-off to achieve a trustworthy sample.

Analytical Techniques Suited for Comparative Research

As a fourth issue, we discuss methodological approaches that could be fruitful for comparative HRM research in general and for Cranet in particular (i.e., moderated regression, multilevel analysis, and structural equation modeling). Whereas moderated regression and multilevel analyses have already been applied in the Cranet network (e.g., Nikandrou, Apospori, Panayotopoulou et al., 2008; Stavrou, 2005), some have not been applied so far (e.g., structural equation modeling). While moderated regression is a straightforward approach to analyze differences across countries, multilevel analysis and structural equation modeling, offer advantages over traditional methods. Regarding multilevel analysis, Cranet is an appropriate candidate because of the large number of countries being considered. In a similar vein, structural equation modeling (SEM) may be perceived as inappropriate for HRM research because of the used measures (i.e., single measures of HRM practices). We argue that SEM can be used to express the variety of the researcher’s theoretical views about HRM constructs. However, there are disadvantages of using single measures because the often recommended tests of measurement equivalence (e.g., Steenkamp & Baumgartner, 1998) cannot be conducted.
Moderated regression. Reviews on cross-cultural research (Schaffer & Riordan, 2003; Tsui et al., 2007) stress the importance of moderator analyses. In a moderator analysis, the researcher analyzes if a (moderator) variable alters the relationship between two variables (Sharma, Durand, & Gur-Arie, 1981). In cross-national research, the country is assumed to function as a moderator. In such contexts, beyond descriptively comparing correlations or regression coefficients across countries, moderated regression analysis should be used. The analysis consists of using dummy variables representing the countries together with the independent variable of interest and the product between the dummies and the independent variable as predictors (Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2003; Frazier, Barron, & Tix, 2004; West, Aiken, & Krull, 1996). As a consequence, the regression coefficient of the independent variable (i.e., its “first-order effect”) quantifies its relationship with the dependent variable in the reference country with the dummy-code of zero. The regression coefficients of the dummy variables denote their mean difference in the dependent variable of the reference country. Finally, the regression coefficients of the product terms signify the differences between the country in question and the reference country in the regression effect of the independent variable. An alternative to a regression model with dummy variables is to apply a multiple-group analysis in which regression coefficients are compared across countries (Marsh, Wen, & Hau, 2006; Sharma et al., 1981).

Multilevel analyses. One problem with the application of moderator analysis in cross-cultural designs is that countries are equated with cultures. Countries differ in more than the cultural characteristic which could have led to the observed difference in the regression coefficient. Multilevel analyses (or hierarchical linear modeling, random coefficient modeling) seems more suitable, however, they require a large number of countries. Multilevel analyses are a complex form of linear regression and are especially suited for comparative studies involving
many nations or cultures. The term “multilevel” refers to the nested data structure that is implied by comparative research designs, meaning that individuals (lower level) are nested in collectives (higher level). Multilevel analyses incorporate constructs on both levels—that is, individual characteristics on the lower level, and national or cultural characteristics on the higher level. In the HRM context, lower level constructs may be, for instance, HRM practices, and typical higher level constructs are labor legislation or cultural dimensions (e.g., power distance, Hofstede, 1980; House, Hanges, Javidan, Dorfman, & Gupta, 2004). Multilevel analyses are particularly useful for comparative research as higher level constructs are often proposed to influence lower-level constructs or effects of these constructs on other lower level-constructs.

Multilevel random coefficient models, in particular, estimate variance in the dependent variable (or intercept, respectively) as well as the slope of lower-level regression models and attempt to explain these variances with higher-level constructs. For instance, a lower-level regression model could specify the effect of an HRM practice on effectiveness. Consequently, the intercept of this model as well as the effect (i.e., regression coefficient) would be treated as outcomes of a higher level construct (e.g., culture). Hence, the variation in the effects of the HRM practice across countries could be attributed to variations of the cultural or national characteristic.

Structural equation modeling. As a third methodological approach, structural equation modeling (SEM) allows specifying hypotheses on the structure of relationships among observed and latent variables. SEM is especially useful for comparative research as it allows comparing groups by specifying a model for each group and testing equality constraints imposed on parameters of interest (e.g., regression effects). In addition, new developments allow incorporating measures with different metrics (e.g., rating scales, count data, proportional data,
or binary data) and to specify continuous as well as categorical latent variables (Muthén, 2002; Skrondal & Rabe-Hesketh, 2005). Finally, SEM allows assessing measurement equivalence which is regarded as a prerequisite for any further analyses in comparative research (Steenkamp & Baumgartner, 1998).

However, considering HRM practices in SEM imposes some difficulties. First, it is unclear how HRM practices can be incorporated in SEM with latent variables. In this regard, we will propose different hypothetical structures that could reflect a researcher’s thinking about an HRM construct in question. Second, the use of single items to measure HRM practices prevents testing measurement equivalence in most cases. This, again, highlights the importance of pretesting procedures that we discussed earlier to ensure that respondents from different countries understand the measure in an identical way. We note, however, that the use of single indicators does not prevent the use of SEM. We present an option to use information about the measurement error (e.g., from research with two or more respondents per organization) to specify a latent variable model.

In order to apply SEM to HRM research, the specification of a latent variable model affords that the HRM construct in question is theoretically connected to latent variables. Hypothetical constructs are often vague terms with regard to their dimensionality and their ontological status (i.e., does the construct refer to an existing entity or is it a simply an umbrella term for a set of features, cf. Borsboom, Mellenbergh, & van Heerden, 2003). Latent variables, in contrast, represent singular dimensions which are related to observed and/or other latent variables. Figure 1 illustrates five examples of how measured HRM practices (e.g., amount of flexible work, part-time, and child-care support) can be related to latent variables.
In the case the researcher focuses on a unidimensional construct, there are two possibilities how this construct can be transferred to a latent variable model. In case the focus is on a single HRM practice, this single practice may be conceived as a latent variable in order to differentiate the true practice and its measurement (see Figure 1a). When, for instance, a researcher’s theory proposes that flexible working time decreases the turnover rate, s/he would specify a model in which the amount of flexible working schedules is the latent variable of interest. We will discuss the problem of specifying latent variables with single indicators later. Alternatively, the researcher may focus on different practices as mere reflections of a single underlying latent variable (e.g., the organization’s concern for work-life issues) which is of main theoretical interest. Such a conceptualization leads to a standard common factor model with multiple HRM practices involving work-life policies specified as reflective indicators of a single latent variable (Figure 1b). Figure 1c contains the same structure but again highlights that the true practices underlie the common factor and not the ratings.

In case of a multidimensional HRM construct involving a bundle of practices (i.e., practices A, B, and C in Figure 1b – e), however, the researcher has to clarify the relationship between the dimensions (i.e., HRM practices) and the overall construct and what the researcher’s causal theory says about relationships between the construct and external (e.g., dependent variables). As an example for a multidimensional HRM construct, the researcher may conceive the practices as several facets of the overall construct. For instance, flexible working time may be just one facet of the overall construct “work-family practices”. This multidimensionality renders the construct ambiguous regarding the latent structure it imposes.

As one possibility, the higher-order common factor model (Figure 1b and 1c) was discussed earlier. Another possibility may be that the researcher conceptualizes work-family
practices in terms of a *formative structure* (Bollen & Lennox, 1991; Howell, Breivik, & Wilcox, 2007; MacKenzie, Podsakoff, & Jarvis, 2005; Wilcox, Howell, & Breivik, 2008). In this regard, a latent variable (i.e., functioning of the work-family practices) is a consequence of the set of practices (Figure 1d). In such a model, the practices have effects on the dependent variable(s) (e.g., turnover) channeled by that single latent variable. It should be noted that such a model has rigid demands to fulfill to be meaningful (e.g., proportional effects of the formative indicators and the effects are mediated by the single latent variable; see Franke, Preacher, & Rigdon, 2008). A different view on the work-family construct may be that the construct is simply an umbrella term which serves as a common label for the set of practices (Figure 1e). In this case, the model in question could impose causal effects of each of the practices on turnover or a diverse set of unique effects on several dependent variables. Analyzing the relationship between turnover and a composite score of practices is feasible but renders the ontological status of the score and the meaning of the relationship ambiguous (Borsboom, 2006).

As one application of SEM in comparative research, *measurement equivalence* has been intensively discussed as a prerequisite of any quantitative comparisons (He et al., 2008; Schaffer & Riordan, 2003; Steenkamp & Baumgartner, 1998). Post-hoc tests of measurement equivalence, however, require that a latent variable is measured by multiple indicators. In this regard, the most important tests of equivalence concern the questions of whether a) the measurement structure is comparable across countries (*configural equivalence*), b) the factor loadings that link the latent variable to its measures are equal across countries (*metric equivalence*), and c) the indicator intercepts which are part of the regression equation, that links the latent variable to each of its measures, are equal (*scalar equivalence*). Configural and metric equivalence refer to the notion of an equal meaning of the construct; metric equivalence, in
particular, refers to an equal usage of the metric, and scalar equivalence refers to the requirement of equal systematic biases in the investigated countries.

Because tests of measurement equivalence require multiple indicators (i.e., the common factor model, Figure 1b), Cranet researchers may be inclined to use different practices as multiple measures of a latent variable. As discussed, however, such a model implies a specific meaning of the latent variable--namely, as a common cause of the different practices--which the researcher possibly does not theorize. If the researcher intends to measure some kind of an HRM strategy as a latent common cause of a set of practices, the common factor model may be appropriate and can be compared across countries. If, however, the latent target variable is theorized as a single HRM practice (i.e., Figure 1a), solutions may be a) to use multiple indicators referring to the same HRM practice or b) to use more than one rater. Whereas the first option enables correcting for random errors in responding to the questions, the second option enables analyzing errors in the perception of the true HRM practice (i.e., lack of validity). This may be advantageous as research shows that different HR managers estimating the prevalence of HRM practices show low agreement (Gerhardt, Wright, McMahan, & Snell, 2000). The rest of the aforementioned model structures (Figure 1d and 1e)--that is, the construct as a formative measurement structure or umbrella term--does only enable limited possibilities to formally test for construct, metric, and scalar equivalence. Whereas the model structure can be compared across countries, metric and scalar equivalence cannot be tested.

One may regard the use of single items to measure latent variables as a drawback against the usual multiple indicator tradition in organizational and social sciences. In this regard, we would like to point to the ongoing debate on the use of single-indicator-latent variables (see, for instance, the special issue in Structural Equation Modeling, 2000, vol. 7). For instance, Hayduk
proposes to fix the measurement error of the indicator to a theoretically reasonable value in order to incorporate the researcher’s measurement theory into the model. As an alternative, empirical information (e.g., correlations between different raters of the same practice) may be used to fix the measurement error of the practice measure in order to specify the latent practice properly. For example, some research may have shown that the correlation between two ratings of the prevalence of an HRM practice is \( r = .50 \). Under the assumption that both subjective ratings of the (objective) HRM practice equally reflect the practice, such a correlation would result from an HRM practice variable influencing both ratings with .71 (i.e., square root of .50), thus, leaving 50 percent of each rater’s variance due to idiosyncratic random and systematic error. In the target study, consequently, the researcher could fix the error variance of the single rater response to \((1 - r)\) times the indicator’s variance and the factor loading to a latent practice variable to 1—thus, leading to a specification of the latent variable as the supposed objective HRM practice (and not just the respondent’s perception of the practice).

Although this strategy makes some assumptions (e.g., all raters perceive the objective practice with an equal perception error and the observed correlation used as a baseline focuses on the same population as the target study does), it may lead to more valid results than relying on practice ratings that imply high error. Using multiple indicators for the same HRM practice, in contrast, would enable correcting for random error but would still leave unclear exactly how far the perception of the respondent matches the objective HRM practice of interest.

**Summary and Conclusions**

This paper has addressed four topics which are important for comparative HRM research and discussed these topics using the Cranet project as a practical example. First, we relied on reviews on cross-cultural methods (e.g., Schaffer & Riordan, 2003) and emphasized that
comparative research demands specific efforts to plan, coordinate, and standardize the research process. For instance, the network of researchers has to develop equivalent measures, create a questionnaire with identical form, administer the questionnaires in the same way, and execute the process of reminding nonrespondents in a manner as similar as possible. We illustrated how Cranet performs these tasks but also recommended to increase efforts with regard to pretesting the questions and standardizing the survey process.

Second, we discussed sampling of respondents in greater detail as other comparative reviews do. In this regard, we argued that statistical tests rely on random sampling, as otherwise the standard errors may be biased. Moreover, when the sample systematically differs from the population, parameters of interest (e.g., percentage of coverage of some HRM practice, or regression coefficient) may be biased. We argued further that the probability of this bias depends on the very characteristic on which the sample deviates from the population and discuss how this characteristic is related to the research model (i.e., predictor of target variables or moderator of relationships). The goal of this section, however, was not to implement unrealistic standards but prevent the formation of myths about the unimportance of sampling (Highhouse & Gillespie, 2009). As applied researchers, we know that random samples are a fiction. However, we think that improving current practice is a valuable goal and acknowledging the possible consequences of nonrandom sampling prevents pitfalls when interpreting differences across cultural or national populations. In this regard, we acknowledge the efforts already undertaken by Cranet, however, discuss the potential for improvement regarding the comparativeness and representativeness of Cranet's samples. To achieve better comparability, Cranet may further consider how differences in response rates and related issues can be managed in order to further approach the ideal standards of sample selection and nonresponse bias.
Third, we discussed moderated regression, multilevel analyses, and structural equation modeling as fruitful analytical techniques from which Cranet research could benefit. To date, Cranet primarily utilizes methods like linear regression analysis or logistic regression analysis with only main effects. The potential value of moderated regression analyses, multilevel analyses, and SEM have not yet been systematically developed in Cranet publications. Although moderator analyses have a long tradition in comparative research, Cranet’s consideration of moderator variables is scarce. Cranet publications may gain from a more intensive consideration of dummy variables as moderators. It must be noted, however, that during the past years Cranet has increasingly utilized multiple-group analysis in which variables under consideration are compared across countries.

Multilevel analysis, also not yet systematically applied by Cranet researchers, seems especially suitable and promising for future research. Cranet, as one of the very few comparative HRM research networks, offers a sizable database on the organizational and country level, which allows for multilevel analysis. Cranet delivers a suitable database for multilevel analyses as the number of participating countries is above recommended sample sizes on the higher level thus providing sufficient power and efficient parameter estimates.

SEM has also not yet found systematic application in Cranet publications. Without doubt, Cranet’s focus on HRM practices and along with this, its use of single items seems to make the application of SEM not an obvious choice for researchers. However, as discussed the measurement error of the indicator may be fixed to a reasonable value in order to incorporate the researcher’s measurement theory into the model. We also discussed measurement invariance in the Cranet context. In this regard we argued that the standard model discussed in the literature on
equivalence (i.e., the common factor model) may be inappropriate in many cases, and alternative model structures may be more appropriate.
References


methodological myths and urban legends - Doctrine, verity and fable in the
organizational and social sciences (pp. 7-36). New York: Routledge.


*Journal of Business Research, 61*, 1219-1228.
Footnotes

1 Our recommendation to code national specifics in conducting the study is similar to the use of codes in meta-analyses to quantify differences between features of primary studies (Lipsey & Wilson, 2001)

2 See the website of the European Social Survey [www.europeansocialsurvey.org](http://www.europeansocialsurvey.org)

3 Whether the model in Figure 1b and 1c is regarded as expressions of a unidimensional or multidimensional HRM construct depends on the focus on the researcher: It is unidimensional if the researcher focuses on the single latent variable; it is multidimensional if the researcher focuses on the bundle of practices
Figure Captions

*Figure 1.* Options to transfer unidimensional (Figure 1 a) and multidimensional HRM constructs (Figures 1b – 1e) into latent variable models.
a) An HRM practice as a latent variable (Note: As there is only one measure of the HRM practice, the measurement error cannot be estimated)

b) A common factor model (i.e., the latent variable causes the ratings of the practices)

c) A common factor model (i.e., the latent variable explains/causes the true practices)

d) A formative measurement model (i.e., the latent variable refers to the causal mechanism by which the practices influence the dependent variables)

e) The HRM practices as a multivariate set (i.e., the practices have unique effects on the causal variables)