

CHRISTIAN HOMBURG, MARTIN KLARMANN, MARTIN REIMANN, and
OLIVER SCHILKE*

In an effort to establish and enhance the accuracy of key informant data, organizational survey studies are increasingly relying on triangulation techniques by including supplemental data sources that complement information acquired from key informants. Despite the growing popularity of triangulation, little guidance exists as to when and how it should be conducted. Addressing this gap, the authors develop hypotheses linking a comprehensive set of study characteristics at the construct, informant, organizational, and industry levels to key informant accuracy. Two studies test these hypotheses. The first study is a meta-analysis of triangulation applications. Using data from 127 studies published in six major marketing and management journals, the authors identify antecedents to key informant reliability. The second study, using eight multi-informant data sets, analyzes antecedents to key informant validity. The results from these studies inform survey researchers as to which conditions particularly call for the use of triangulation. The authors conclude by offering guidelines on when and how to employ triangulation techniques.

Keywords: organizational survey research, triangulation, reliability, validity, key informant design, multiple data sources

What Drives Key Informant Accuracy?

Key informant surveys are an essential data source in marketing and management research. However, given the complex and subjective task key informants need to perform, some investigators have questioned the accuracy of their responses (e.g., March and Sutton 1997). Therefore, to ensure the accuracy (i.e., reliability and validity) of key informant data, researchers are increasingly advocating collecting supplementary data, such as information from archival sources and additional respondents (e.g., Van Bruggen, Lilien, and Kacker 2002). This strategy has been variously labeled “convergent methodology” (Campbell and Fiske 1959), “convergent validation” (e.g., Slovic 1962),

and “triangulation” (Webb et al. 1966) and can be understood as “the combination of methodologies in the study of the same phenomenon” (Denzin 1978, p. 291). We primarily use Webb et al.’s (1966) term “triangulation,” borrowed from navigation, in which it denotes the “use [of] multiple reference points to locate an object’s exact position” (Jick 1979, p. 602).

Although triangulation is increasingly common, it is costly and not always easy to implement (Rindfleisch et al. 2008). For example, given the issue of “oversurveying” (Rogelberg and Stanton 2007), managers are reluctant to participate in surveys, especially if a colleague has already taken part. At the same time, archival data are often available for only a small subsample of the firms surveyed and cause problems of matching data sources and establishing a common metric. Making things worse, triangulation does not add substantial value if the responses of the first key informant are largely accurate.

Therefore, researchers could benefit from being able to predict under what circumstances they most need triangulation of key informant responses. Such predictions presuppose knowledge of the factors systematically affecting the accuracy of key informants, and our article addresses this issue. In a first study, we conduct a meta-analysis of previous triangulation efforts to establish whether and, in particu-

*Christian Homburg is Professor of Business Administration and Marketing and Chair of the Marketing Department, University of Mannheim, and Professorial Fellow, Department of Management and Marketing, University of Melbourne (e-mail: homburg@bwl.uni-mannheim.de). Martin Klarmann is Professor of Marketing, Department of Economics and Business Engineering, Karlsruhe Institute of Technology (e-mail: martin.klarmann@kit.edu). Martin Reimann is a doctoral candidate, Department of Psychology, University of Southern California (e-mail: mreimann@usc.edu). Oliver Schilke is a doctoral candidate, Department of Sociology, University of California, Los Angeles (e-mail: schilke@ucla.edu). The authors contributed equally to the research and are listed in alphabetical order. Christine Moorman served as associate editor of this article.

lar, when key informant responses are reliable. To do so, we use data from 127 studies (corresponding to a total number of $n = 11,874$ observations) that report key informant reliability when compared with data from other sources and that were published between 1998 and 2010 in six major marketing and management journals. In a second study, we analyze eight multi-informant data sets from different marketing research contexts to determine whether and particularly when key informant responses are valid.

Our research makes at least three important contributions. First, whereas previous studies have focused mostly on either reliability or validity, we adopt an integrative perspective by investigating both aspects of accuracy. Second, previous research on method biases in survey research has typically investigated people's self-reports. We focus on key informant reports that involve aggregations over people, tasks, organizational units, or events (Seidler 1974), which may introduce other biases and problems. Third, and most important, previous research has predominantly focused on describing overall levels of response accuracy and only conducted some exploratory analyses to identify antecedents. Our aim is to develop and test hypotheses regarding a broad set of potential drivers of key informant accuracy. We analyze the impact of construct characteristics on reliability and validity. Moreover, we investigate the impact of informant, organization, industry, and method characteristics on reliability. Knowledge of the effects of these antecedents will allow researchers to make an informed choice about when triangulation is most needed.

GOALS AND METHODS OF TRIANGULATION

The ultimate goal of triangulation is to ensure the accuracy of key informant data. Two types of error appear that commonly affect key informant response accuracy (e.g., Churchill 1979): systematic measurement error (also referred to as "method error") and random measurement error. Therefore, a key informant response x can be expressed as the sum of the true value t_x , a systematic measurement error s_x , and a random measurement error e_x :

$$(1) \quad x = t_x + s_x + e_x.$$

If random errors are uncorrelated with the true values and systematic errors and if true values and systematic errors are uncorrelated, the correlation between x and a second variable y (measured through the same informant) can be written as follows:

$$(2) \quad r_{x,y} = \frac{\text{Cov}(t_x, t_y) + \text{Cov}(s_x, s_y)}{\sqrt{\text{Var}(t_x) + \text{Var}(s_x) + \text{Var}(e_x)} \sqrt{\text{Var}(t_y) + \text{Var}(s_y) + \text{Var}(e_y)}}.$$

From this equation, the differential consequences of the two types of error become apparent (Cote and Buckley 1988). Random measurement error attenuates the correlation. A covariance between the systematic error components s_x and s_y biases results in a way that depends on the direction and magnitude of this covariance. Only if systematic error terms are uncorrelated do they affect results the same way as random measurement error. This is only the case if x is influenced by a different method factor than y and if these two

method factors are uncorrelated. As prior research shows (Cote and Buckley 1987), this occurrence is unlikely.

The distinction between random and systematic measurement error is directly linked to the distinction between reliability and validity (Churchill 1979). Using triangulation to counter the threats of random measurement error is intended to ensure the reliability, and using it to counteract the threat of systematic measurement error is intended to ensure the validity of key informant responses.

In addressing issues of reliability and validity with triangulation, researchers may pursue two distinct strategies. First, they may aim to establish the accuracy of key informant responses by using additional data to estimate the degree of random and systematic measurement error present in the responses. Second, they may aim to enhance the accuracy of key informant data by combining those data with information from different data sources. Accordingly, as Figure 1 shows, triangulation can have four goals, each associated with a different set of methods.

First, to establish the reliability of key informants, researchers can use correlational approaches. If data sources are not interchangeable (i.e., the informants are distinguishable according to a systematic criterion), Pearson's r is the suitable choice. If data sources are interchangeable, the intraclass correlation coefficient (ICC)—more specifically the ICC(1) (Bliese 2000)—is appropriate. In addition, although it does not measure reliability in a strict sense (Kozlowski and Hat-trup 1992), the $rwg_{(j)}$ agreement index (James, Demaree, and Wolf 1984) offers a heuristic approach to assessing random error (James, Demaree, and Wolf 1993).

Figure 1
GOALS OF TRIANGULATION

Facet of Key Informant Response Accuracy	Validity	
	Establish Validity	Enhance Validity
Reliability	Available triangulation methods: • Multitrait-multimethod (MTMM) matrix analysis • CFA with M method factors • CFA with $M - 1$ method factors • Correlated Uniqueness Model • Direct Product Model	Available triangulation method: • SEM with method factors
	Establish Reliability Available triangulation methods: • Correlational approaches (e.g., Pearson's r , intraclass correlation) • Agreement approaches (e.g., $rwg_{(j)}$, AD_M)	Enhance Reliability Available triangulation method: • Aggregation across multiple data sources
Establish Key Informant Response Accuracy		Enhance Key Informant Response Accuracy
Desired Triangulation Result		

Second, if researchers are interested in enhancing the reliability of key informant data, they can aggregate the information from different data sources (Van Bruggen, Lilien, and Kacker 2002) and assess the reliability of the composite using the ICC(2) (Bliese 2000).

Third, as Equation 2 suggests, researchers wanting to establish the validity of key informant data must assess the covariance between systematic errors for at least two constructs. To do so, previous research has proposed numerous approaches that combine the analysis of multitrait multi-method (MTMM) correlation matrices (Campbell and Fiske 1959) with confirmatory factor analysis (CFA) (Bagozzi and Yi 1991; Podsakoff et al. 2003). They include the CFA with M and M – 1 method factors (Eid 2000; Kumar and Dillon 1990), the correlated uniqueness model (Marsh, Byrne, and Craven 1992), and the direct product model (Kumar and Dillon 1992).

Fourth, the logic of the MTMM approaches described previously can also be adapted to enhance the validity of key informant data. By integrating factors representing the different data sources in a structural equation model, researchers can estimate trait interrelationships while controlling for systematic error (Cote and Buckley 1987; Doty and Glick 1998).

This article's objective is to make generalizations about situational drivers of key informant accuracy. For this purpose, we conducted a first study that investigates drivers of key informant reliability and a second study that analyzes antecedents to key informant validity.

STUDY 1: ANTECEDENTS TO KEY INFORMANT RELIABILITY

Hypotheses

In deriving specific study characteristics that drive key informant reliability, we rely on social cognition theory (e.g., Hastie and Park 1986). This theory is useful for analyzing the processes underlying survey response formation and the emergence of random response errors (Harrison, McLaughlin, and Coalter 1996; Tourangeau, Rips, and Rasinski 2000).

Two concepts from social cognition theory are central for our research: (1) availability of relevant information and (2) difficulty of retrieval. First, according to social cognition theory, only some of the issues assessed in survey research are already present in an informant's memory. On a continuum of information availability, at the more available end, the respondent has relevant information that is easily reproduced, whereas at the less available end, the respondent's memory does not contain relevant information (Tourangeau, Rips, and Rasinski 2000). In the latter case, informants must form a response based on related cues, introducing the possibility of making imprecise inferences that adversely affect the reliability of the response.

Second, human long-term memory contains significantly more information than a single mental task requires (Baddeley 1986). Retrieval denotes the process of bringing information held in long-term memory to an active state, in which it enters the short-term memory to be used (Jobe, Tourangeau, and Smith 1993). The retrieval process can be subject to problems, such as retrieving wrong or generalized information (Fazio et al. 1986), which also reduces reliability.

We propose that availability of information and difficulty of retrieval, and thus informant reliability, differ markedly as a function of four sets of factors: (1) construct characteristics, (2) informant characteristics, (3) organizational characteristics, and (4) industry characteristics. Because we also investigate formal considerations, we consider method characteristics as a fifth set.

Construct characteristics. We suggest that five construct characteristics (temporal reference, objectivity, salience, specificity, and local scope) affect key informant reliability. We expect the construct's temporal reference (defined as its allusion to the present vs. the past) to be related to key informant reliability. First, temporal reference is substantially linked to the availability of information. The greater the difference between the time of the interview and the time to which the construct refers, the more likely the respondent will have forgotten relevant details. Second, passage of time negatively affects the precision of retrieval owing to interference effects (Gillund and Shiffrin 1984). It increases the likelihood that the informant faces other issues and also stores them in memory. Then more information search is required, decreasing the chance of precise recall. Thus:

H₁: The reliability of key informant responses is greater for constructs pertaining to the present (vs. the past).

The complexity of the information retrieval process is greater if the nature of the construct requires key informants to engage in subjective assessments (low objectivity) rather than to report on factual phenomena that are clearly linked to objective referents (high objectivity). In general, cognitive processes involved in making subjective assessments on ambiguous phenomena are more difficult to perform (Golden 1992; Starbuck and Mezias 1996), whereas reporting on more objectively verifiable data is less demanding, decreasing the risk of random response error (Feldman and Lynch 1988). Thus:

H₂: The reliability of key informant responses is greater for more objective constructs (vs. more subjective constructs).

Next, we put forward the notion that key informant reliability depends on the investigated construct's salience, defined as the degree to which its subject matter relates to a conspicuous issue versus a routine process. Research in psychology has shown that people can easily recall information on landmark events (e.g., Pillemer et al. 1988). At the same time, the occurrence of multiple similar routine events reduces the likelihood of precisely retrieving any one of them (Huber and Power 1985). People are less prone to forget emotionally arousing events because they are more likely to notice them initially and to reflect on them afterward. Given that salient events are more elaborately coded and cause greater emotional impact, both availability and retrievability of relevant information are greater. Thus:

H₃: The reliability of key informant responses is greater for constructs pertaining to more salient events (vs. more routine events).

Whether a construct pertains to characteristics of specific people (e.g., the key informant's superior) or to nonpersonal phenomena (e.g., team-level or organizational-level occurrences) has implications for the complexity of information

retrieval. When informants make inferences about nonpersonal phenomena or perform aggregations over people, retrieval complexity increases, leading to unreliable responses. Silk and Kalwani's (1982) study on the reliability of organizational purchasing measures finds a lower reliability when these constructs were more global (i.e., involved numerous interactions) than when the constructs that were more specific (i.e., referred to a particular instance). In a similar vein, we hypothesize the following:

- H₄: The reliability of key informant responses is greater for constructs pertaining to characteristics of specific people (vs. nonpersonal entities).

Key informants can more easily provide consistent information on constructs related to phenomena internal to the firm than on phenomena related to the firm environment. The idea that informants remember organizational internal issues more precisely than environmental issues emerges from the notion that people store information that pertains to a unit to which they belong and that they experience firsthand in a more detailed form than events they merely hear or read about (Sudman et al. 1994). Therefore, information availability should be greater and information retrieval easier for constructs pertaining to the organization than for constructs pertaining to its environment. Thus:

- H₅: The reliability of key informant responses is greater for constructs pertaining to phenomena internal to the firm (vs. phenomena related to the firm environment).

Informant characteristics. We consider the informant characteristics of hierarchical position, functional position, and tenure as antecedents to reliability. Again building on the two mechanisms derived from social cognition theory, we suggest that informants will be more reliable if they have better access to information and if they often deal with the survey issues, which facilitates retrieval.

First, employees at different hierarchical levels vary in the quantity and quality of information sources available to them; top managers are typically informed about organizational activities and outcomes continuously and virtually instantly. Thus, informants in higher hierarchical positions tend to have superior access to information. In addition, high-level employees mostly participate firsthand in organizational decisions related to many topical areas key informant surveys investigate (Hambrick 1981). Even if lower-level managers are formally informed about organizational decisions, their involvement tends to be lower, thus exacerbating retrieval from memory when they act as key informant. We hypothesize the following:

- H₆: The reliability of key informant responses is greater for informants in higher hierarchical positions (vs. lower hierarchical positions).

Second, role theory indicates that people occupying a specialized role within a social structure provide more precise information on role-related aspects of that structure than people occupying general positions (Houston and Sudman 1975). Imprecise responses are more likely with respondents whose roles are not closely linked with the phenomenon under study. It follows that employees with a focus on a specific organizational function are able to pro-

vide more reliable responses than those with a generalist position.

- H₇: The reliability of key informant responses is greater for informants with a focus on a specific organizational function (versus generalists).

Third, we propose that informant tenure—the duration the respondent has worked in the unit on which he or she is reporting—is associated with response reliability. People with long tenure have a greater awareness of many organizational issues because of their prolonged opportunity to experience organizational life (Golden 1992; Starbuck and Mezias 1996). Thus:

- H₈: The reliability of key informant responses is greater for informants with longer tenure (vs. shorter tenure).

Organizational characteristics. At the organizational level, we view size and age as being related to information availability and retrievability and thus key informant reliability. We expect organizational size to negatively affect informant reliability. Imagine a key informant being asked about the marketing strategy of a large multidivisional firm such as General Electric. Given the diversity of operations and the great number of employees, the information gathering and computational effort required to provide precise firm-level responses is enormous. In contrast, a key informant reporting on a one-employee firm should have less difficulty overseeing the firm's operations (Seidler 1974) and providing reliable responses. Thus:

- H₉: The reliability of key informant responses is greater for smaller organizations (vs. larger organizations).

We propose that the age of the organization positively affects response reliability. Older organizations tend to have established information processing and dissemination routines as well as more resources for management information systems than younger organizations. Thus:

- H₁₀: The reliability of key informant responses is greater for older organizations (vs. younger organizations).

Industry characteristics. We also consider industry characteristics' effect on the reliability of key informant responses—in particular, industry concentration, research-and-development (R&D) intensity, and dynamism. In concentrated industries, competition tends to be more intense and competitors behave more aggressively because they vie for the same pool of customers (Barnett 1997). As a consequence, fiercely competing firms typically do not share information openly, making it more difficult for managers to obtain market information (Slater and Narver 1994). Furthermore, concentrated, highly competitive industries are typically characterized by greater pressure on prices and high levels of advertising (Porter 1980), which is why managers in these industries lack the time to carry out comprehensive analyses. Instead, they must often rely at least partly on their intuitive assessment of the circumstances (Homburg, Grozdanovic, and Klarmann 2007). Both the lack of available information and constant time pressure preventing detailed analyses reduces managers' ability to provide precise survey responses. Thus:

- H₁₁: The reliability of key informant responses is greater for less concentrated industries (vs. more concentrated industries).

Organizations in industries with high R&D spending usually compete on innovations, which requires them to constantly gather new ideas and insights both from the environment and from inside the organization (Cohen and Levinthal 1990). Therefore, information collection is highly routinized, affording key informants efficient access to information. Thus:

- H₁₂: The reliability of key informant responses is greater for industries with higher R&D intensity (vs. lower R&D intensity).

Industry dynamism refers to the extent and frequency of unforeseen changes in the firm environment (e.g., Jap 1999). In highly dynamic industries, firms' resource configurations and strategies quickly become obsolete. Frequent changes hamper key informants' ability to provide precise information about their organization at a particular point in time. Thus:

- H₁₃: The reliability of key informant responses is greater for more stable industries (vs. more dynamic industries).

Method characteristics. We consider two characteristics of the triangulation method: type of data and informant similarity. We reflect on these antecedents last because we expect them to affect the estimated but not the actual reliability of key informants. First, we expect that using archival data will yield higher reliability estimates than using survey data. Survey items introduce random measurement error, whereas archival data is usually more precise (Starbuck and Mezias 1996). Because random error attenuates correlations (Nunnally 1978), comparing two surveys will lead to smaller reliability estimates than comparing survey data with archival data. Thus:

- H₁₄: Key informant reliability estimates will be greater if triangulation is conducted using archival data (vs. survey data).

Second, when additional survey data are used for triangulation, reliability estimates will be higher if the type of informant used for triangulation is similar to that of the first key informant. Because researchers are advised to choose the first informant on the basis of presumed access to information (Kumar, Stern, and Anderson 1993), responses from other (potentially less well informed) informant types are likely to contain more random error. Thus, reliability estimates will be less attenuated if responses from similar informants are used for triangulation. Thus:

- H₁₅: Key informant reliability estimates will be greater if triangulation is conducted using key informants of a similar type (vs. dissimilar key informants).

Methodology

Collection of studies. To test these hypotheses, we conducted a meta-analysis of key informant triangulation. We performed a census of six major marketing and management journals (*Academy of Management Journal*, *International Journal of Research in Marketing*, *Journal of Marketing*, *Journal of Marketing Research*, *Journal of the Academy of Marketing Science*, and *Strategic Management Journal*) between 1998 and 2010. For inclusion in our meta-analysis, a study had to have used key informants to analyze organizations or organizational subunits and, for at least one of the

constructs, had to have collected additional data from a second source. The 182 studies meeting our selection criteria constitute the overall sample for our meta-analysis. Figure 2 presents information on sample characteristics. (A complete list of articles is available on request.)

As Figure 2 shows, these studies employed several triangulation methods. Their inclusion in the meta-analysis required a numerical summary measure of key informant reliability. Here, we considered correlations (particularly Pearson's r and the ICC(1)) and agreement (particularly the $rwg_{(j)}$ index that James, Demaree, and Wolf [1984] propose). Other than correlations, which assess the degree to which ratings are "proportional, when expressed as deviations from their means" (Tinsley and Weiss 1975, p. 359), agreement measures "the degree to which ratings ... are interchangeable" (Bliese 2000, p. 351). Of the studies, 127 provided information on one or both of these types of measures. Correlational reliability information was available for 320 constructs from 105 key informant samples (9814 observations) and agreement information for 219 constructs from 60 samples (4982 observations). We obtained both correlations and agreement for 108 constructs from 38 samples. In total, our meta-analysis is based on data for 431 constructs from 127 studies (total number of observations = 11,874).

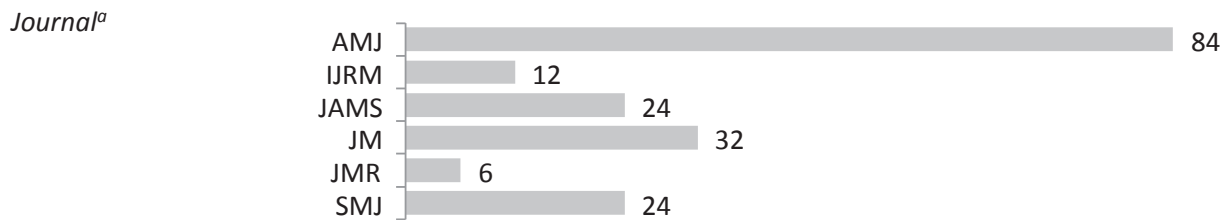
Meta-analytic procedures. A meta-analysis of key informant triangulation faces several specific challenges. First, as outlined previously, studies use different types of effect size measures (correlation and agreement) that are not comparable across a common metric (Bliese 2000). Therefore, in this study, we conduct separate analyses for correlation and agreement.

Second, studies use different correlation coefficients to establish key informant reliability (Pearson's r and ICC). Although these coefficients are derived under different assumptions (McGraw and Wong 1996), they share a common substantive meaning. Therefore, we do not distinguish between ICC and Pearson correlations when aggregating reliability results (Conway, Jako, and Goodman 1995). However, for the purpose of hypothesis testing, we transform both types of correlations to Fisher's (1925) z -scale using appropriate transformations (McGraw and Wong 1996). Moreover, when testing antecedents to reliability, we control for the type of index.

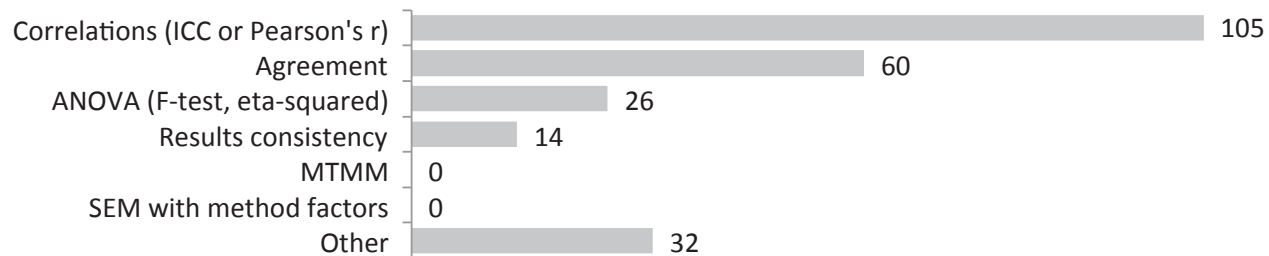
Third, triangulation results also depend on methodological artifacts resulting from specific characteristics of a study's design—namely, the random measurement error attributable to the questionnaire items (Borenstein et al. 2009) and the degree of similarity of the phenomena measured across data sources. To account for these factors, we adapt psychometric meta-analysis (Hunter and Schmidt 2004). Before aggregation, we corrected correlations individually for measurement error and construct similarity. We compute corrected correlations as $r^{\text{corr}} = r^{\text{obs}}/(\alpha \times \text{similarity})$, where r^{obs} is the observed correlation and α is the composite reliability of the measurement instrument. (If no α was reported, we used $\alpha = .84$, which is the mean reliability of all constructs in our sample.) We coded the degree of similarity between the construct measured through the focal key informant and the construct measured through additional data on a scale from .2 ("some association") to 1 ("identical"). In most cases, identical constructs were used, result-

Figure 2
META-ANALYSIS SAMPLE CHARACTERISTICS

Overall Sample (All Studies Using Triangulation)

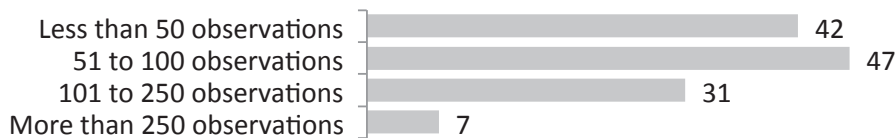


Triangulation Method Employed^b

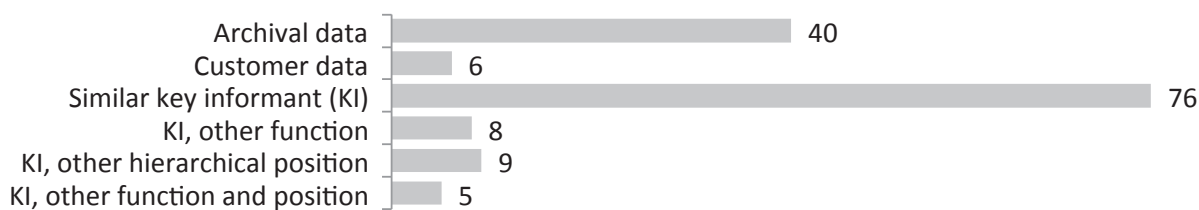


Sample Used in Meta-Analysis (All Studies Reporting Numerical Values for Correlations and/or Agreement)

Triangulation Sample Sizes (Number of Observations with Data from at Least Two Sources)



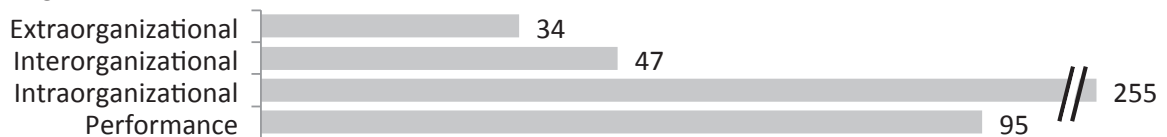
Type of Second Data Source^b



Number of Triangulated Constructs



Type of Triangulated Constructs^c



^aAMJ = *Academy of Management Journal*, IJRM = *International Journal of Research in Marketing*, JAMS = *Journal of the Academy of Marketing Science*, JM = *Journal of Marketing*, JMR = *Journal of Marketing Research*, and SMJ = *Strategic Management Journal*.

^bStudies may fall into multiple categories.

^cNumber of constructs.

ing in a mean similarity of .942 (median and mode are 1). Because of their ad hoc nature (Dunlap, Burke, and Smith-Crowe 2003), agreement measures were not corrected.

Coding. To measure the construct characteristics evoked by our hypotheses, we relied on judgments from four marketing researchers. Using seven-point semantic differentials, they separately provided scores for construct characteristics. We then established consensus among the experts when scores differed on average by 2 or more (Kumar, Stern, and Anderson 1993). The resulting measures show high reliability, with ICC(2) ratings ranging from .71 (temporal reference) to .84 (local scope). Web Appendix A (www.marketingpower.com/jmr_webappendix) provides details on the coding process.

With regard to informant characteristics, to study the impact of an informant's hierarchical position, we coded studies on a five-point scale (1 = "frontline employee," and 5 = "chief executive officer"). To study the effect of an informant's functional position, we coded studies depending on whether the key informants held a specialist or a generalist position. Finally, we measured average key informant tenure as the natural logarithm of the average number of years that the key informants had worked in the unit on which they were reporting.

With regard to organizational characteristics, we measured average organizational size by the natural logarithm of the average number of employees of the firms in a study. In addition, we studied the influence of average organizational age, or the natural logarithm of the average number of years the organizations in a given study have been in existence.

Pertaining to industry characteristics, we captured industry concentration using the sum of the sales market shares of the four largest firms in the industry (Biggadike 1979). Furthermore, we measured R&D intensity as the ratio of industry R&D expenditures to total sales (e.g., Lee and Mahmood 2009) and industry dynamism through industry sales volatility (Sutcliffe 1994). Web Appendix B (www.marketingpower.com/jmr_webappendix) provides more information on the coding of industry characteristics.

Regarding method characteristics, to measure the type of data we used a dummy variable indicating whether triangulation was conducted using archival data or whether customer data was used. We coded informant similarity using three dummy variables indicating whether the second key informant held the same hierarchical and/or functional position. Finally, we control for the type of correlation index used, using a dummy for ICC(1).

Results

Descriptive analysis. Before we tested our hypotheses, we were interested in the average key informant reliability

in our sample. To compute a mean across the 105 studies reporting correlational information, we first aggregated correlations across constructs to arrive at a study-level estimate. Then, we computed a weighted mean across studies using sample sizes as weights (Hunter and Schmidt 2004). This procedure yielded a mean correlation of .612 ($p < .01$). In a similar manner, we computed the mean agreement across the 60 studies reporting agreement and found a mean key informant agreement of .876. In addition, we computed weighted means for different construct domains using four categories: performance, intraorganizational, interorganizational, and extraorganizational constructs. Table 1 shows these descriptive results.

Investigation of publication bias. To address the potential issue of a publication bias in reported triangulation results, we first computed Olkin's fail-safe N (Lipsey and Wilson 2001). Regarding correlations, we found that 24 unpublished studies with a correlation of zero would reduce the weighted average from .612 to .500. Reducing mean agreement to .700, indicating medium agreement, would require 15 studies with an agreement of zero (Brown and Hauenstein 2005). Compared with the number of studies reporting reliability, these fail-safe Ns are quite large.

Second, we employed the "trim and fill" method (Duval and Tweedie 2000), which assumes that publication bias becomes apparent through an asymmetric distribution of study results. Therefore, the sample is first "trimmed"—that is, the most extreme studies are removed until a symmetric distribution is achieved. Then, an unbiased mean effect is computed. In a next step, the sample is "filled"—that is, the algorithm adds the removed studies back to the data while imputing a mirror result below the unbiased effect. Applying this procedure to correlational measures of reliability, we found that no studies below the mean seemed to be missing. With regard to agreement, the results indicate that 11 studies are likely missing. However, after correcting for publication bias, the weighted mean agreement is .841, suggesting only a minor bias.

Hypothesis testing. The level of the data available for hypothesis testing varies. For construct and method characteristics, we have access to complete information at the construct level. However, for the remaining antecedents, data are only available at the study level and are typically incomplete. Therefore, we employed different approaches to test these antecedents.

Because construct characteristics and method characteristics are nested in studies, we tested hypotheses using multi-level regression and weighted individual observations (Geyskens et al. 2009). With correlations as the dependent variable, we used the inverse of the standard errors of the

Table 1
AVERAGE CORRELATIONS AND AGREEMENT VALUES FOR DIFFERENT CONSTRUCT CATEGORIES

	Correlations				Agreement			
	Studies	Constructs	Observations	Mean Correlation	Studies	Constructs	Observations	Mean Agreement
Overall	105	320	9814	.612	60	219	4982	.872
Performance	50	84	5229	.764	14	19	1034	.924
Intraorganizational	65	174	5826	.502	56	176	4879	.874
Interorganizational	15	39	1302	.649	7	12	416	.875
Extraorganizational	16	23	1310	.612	6	12	1048	.865

study effect as weights (Thompson and Sharp 1999). For agreement, we used the square root of the sample size as weights because standard errors are not available (Dunlap, Burke, and Smith-Crowe 2003). Table 2 reports the results; for exact model specifications, see Web Appendix C (www.marketingpower.com/jmr_webappendix).

Our data support H_1 , which predicts that key informant reliability is greater for constructs pertaining to the present rather than the past ($b^{Cor} = .094, p < .01$; $b^{Ag} = .021, p < .01$). Likewise, H_2 is supported: Constructs referring to objective information are associated with significantly higher correlation and agreement values ($b^{Cor} = .072, p < .01$; $b^{Ag} = .007, p < .01$). In addition, as H_3 predicted, correlations and agreement are significantly higher when constructs refer to salient events ($b^{Cor} = .129, p < .01$; $b^{Ag} = .013, p < .01$). However, we find no effect of a construct's specificity ($b^{Cor} = -.017, p > .05$; $b^{Ag} = .002, p > .05$). Thus, the data do not support H_4 . Finally, agreement is higher for constructs referring to the organization versus the environment ($b^{Ag} = .013, p < .01$). However, the effect is not significant with correlations as dependent variable ($b^{Cor} = .015, p > .05$). Thus, H_5 is only partially supported.

Moreover, our data support H_{14} , which predicts that reliability estimates are higher if archival data are used for triangulation ($b^{Cor} = .450, p < .01$). The results provide little support for H_{15} . Although correlational reliability estimates are higher if key informants from the same function are used for triangulation ($b^{Cor} = .101, p < .05$), informant similarity does not have any effect otherwise.

To test the remaining hypotheses, we computed the mean reliability for each study for use as dependent variable. We applied weighted regression, using the weights described previously multiplied by the square root of the number of constructs in the study to account for the precision of the aggregates. To maximize usable observations, we analyzed each antecedent separately (Franke and Park 2006), with the exception of industry characteristics for which usable observations

are identical. Table 3 reports the results; for exact model specifications, see Web Appendix C (www.marketingpower.com/jmr_webappendix).

Regarding informant characteristics, H_6 is supported. Correlations and agreement are higher if the key informant holds a higher position ($b^{Cor} = .319, p < .01$; $b^{Ag} = .014, p < .05$). H_7 is not supported; an informant's functional position has no effect ($b^{Cor} = -.273, p > .05$; $b^{Ag} = .005, p > .05$). Regarding average informant tenure, we find an effect on agreement but not on correlations ($b^{Cor} = .057, p > .05$; $b^{Ag} = .021, p < .05$). Thus, H_8 is only partially supported.

Regarding organizational characteristics, we find that, consistent with H_9 , correlations are smaller if organization size is high, but we observe no such effect for agreement ($b^{Cor} = -.131, p < .05$; $b^{Ag} = -.012, p > .05$). Thus, H_9 is partially supported. H_{10} , which predicts greater reliability for older organizations, is not supported ($b^{Cor} = .011, p > .05$; $b^{Ag} = .012, p > .05$).

Among the industry characteristics, we find a significant, negative effect of industry concentration on correlations and agreement ($b^{Cor} = -1.368, p < .05$; $b^{Ag} = -.204, p < .01$), consistent with H_{11} . In support of H_{12} , R&D intensity significantly affects the reliability of responses ($b^{Cor} = 5.126, p < .05$; $b^{Ag} = .393, p < .05$). Industry dynamism is not related to key informant reliability ($b^{Cor} = -.194, p > .05$; $b^{Ag} = .712, p > .05$). Thus, H_{13} is not supported.

STUDY 2: ANTECEDENTS TO KEY INFORMANT VALIDITY

Hypotheses

Next, we present hypotheses on antecedents to key informant validity. We focus on construct characteristics because the results of Study 1 suggest that they play a central role in driving response behavior. In addition, the data collection for the meta-analysis revealed practically no applied studies reporting validity, limiting our access to large-scale data on study level antecedents.

Table 2
THE IMPACT OF CONSTRUCT AND METHOD CHARACTERISTICS: RESULTS OF MULTILEVEL REGRESSION

Hypothesis	Independent Variable	Dependent Variable	
		Correlations	Agreement
Construct Characteristics			
H ₁	Present-focused	.094** (.014)	.021** (.003)
H ₂	Objective information	.072** (.017)	.007** (.002)
H ₃	Salient events	.129** (.015)	.013** (.002)
H ₄	Characteristics of specific people	−.010 ^{n.s.} (.009)	.002 ^{n.s.} (.001)
H ₅	Internal to the firm	.015 ^{n.s.} (.011)	.013** (.001)
Method Characteristics: Second Data Source			
H ₁₄	Archival data	.450** (.055)	—
H ₁₄	Survey data from customers	.152 ^{n.s.} (.083)	—
H ₁₅	Survey data from key informants with same hierarchical background	.041 ^{n.s.} (.048)	.015 ^{n.s.} (.012)
H ₁₅	Survey data from key informants with same functional background	.101* (.051)	−.005 ^{n.s.} (.011)
Control Variable			
	Reliability measured through ICC	−.443** (.047)	—
	N (studies)	105	60
	N (constructs)	320	219
	N (observations)	9814	4982

* $p \leq .05$.

** $p \leq .01$.

^{n.s.} $p > .05$ (not significant).

Notes: Unstandardized parameters are shown, and standard errors appear in parentheses.

Table 3
THE IMPACT OF INFORMANT CHARACTERISTICS, ORGANIZATIONAL CHARACTERISTICS, AND INDUSTRY CHARACTERISTICS: RESULTS OF WEIGHTED LEAST SQUARES REGRESSION

Hypothesis	Independent Variable	Dependent Variable: Correlations						Dependent Variable: Agreement					
		Model C1	Model C2	Model C3	Model C4	Model C5	Model C6	Model A1	Model A2	Model A3	Model A4	Model A5	Model A6
H ₆	<i>Informant Characteristics</i>												
	Informant hierarchical position	.319** (.055)						.014* (.007)					
H ₇	Informant specific organizational function	-.273 ^{n.s.} (.161)							.005 ^{n.s.} (.017)				
H ₈	Informant tenure		.057 ^{n.s.} (.112)							.021* (.008)			
H ₉	<i>Organizational Characteristics</i>												
	Organization size			-.131* (.067)							-.012 ^{n.s.} (.011)		
H ₁₀	Organization age				.011 ^{n.s.} (.100)							.012 ^{n.s.} (.008)	
H ₁₁	<i>Industry Characteristics</i>												
	Industry concentration					-1.368* (.661)							-.204** (.066)
H ₁₂	Industry R&D intensity					5.126* (2.291)							.393* (.188)
H ₁₃	Industry dynamism					-.194 ^{n.s.} (10.99)							.712 ^{n.s.} (1.063)
	R-square	.247	.027	.049	.092	.000	.152	.061	.002	.125	.092	.000	.261
	N (studies)	103	105	63	39	28	62	56	60	31	14	11	38

* $p \leq .05$.

** $p \leq .01$.

^{n.s.} $p > .10$ (not significant).

Notes: Unstandardized parameters are shown, and standard errors appear in parentheses.

While social cognition theory suggests that availability of relevant information and difficulty of retrieval introduce random error, scholars note that these same issues can also introduce the risk of systematic error (Podsakoff et al. 2003). That is, for constructs for which informants lack pertinent knowledge and/or struggle to retrieve such knowledge from their long-term memory, these informants may provide a random best guess (thus threatening reliability) or may use some heuristic to guide their response (thus threatening validity). Systematic response tendencies that may come into play when information relevant to a construct is unavailable or difficult to retrieve include implicit theories, affectivity, central tendency, and leniency biases, among others (Podsakoff et al. 2003). Thus, the five construct characteristics discussed in the section on reliability may also influence the extent of systematic error.

Temporal reference. If constructs refer to the distant past, locating and retrieving adequate information from memory is difficult. As a consequence, informants may report what they believe might have plausibly happened rather than what they specifically remember has happened in that past time period, increasing the potential of systematic tendencies to affect the response process. For example, Huber and Power (1985) argue that retrospective accounts are subject to attributional bias (in which informants believe that their past decisions were more rational than they actually were), which can introduce systematic error when respondents reconstruct the past.

H₁₆: The validity of key informant responses is greater for constructs pertaining to the present (vs. the past).

Objectivity. Reporting objective data is less demanding, and thus the risk of distorting influences is lower than if key informants must engage in subjective assessments (Davis, Douglas, and Silk 1981). In contrast, constructs requiring informants to perform subjective assessments open up the possibility for informants to (knowingly or unknowingly) report systematically erroneous information.

H₁₇: The validity of key informant responses is greater for more objective constructs (vs. more subjective constructs).

Salience. As we argued previously, because salient events are more elaborately coded and have greater emotional impact, information availability and retrievability tend to be greater for salient events than for routine processes. Consequently, we expect systematic response errors to play a lesser role when salience is high.

H₁₈: The validity of key informant responses is greater for constructs pertaining to more salient events (vs. more routine events).

Specificity. Not only might the task of aggregating information over multiple people when providing responses on nonpersonal entities such as organizations cause random computation mistakes, but its high cognitive demand may also lead uncertain respondents to adopt systematic response tendencies. For example, high cognitive demand placed on respondents has been associated with the response style of acquiescence (Paulhus 1991), the tendency to agree with items regardless of content. Thus:

H₁₉: The validity of key informant responses is greater for constructs pertaining to characteristics of specific people (vs. nonpersonal entities).

Local scope. The lack of firsthand experience when reporting on environmental (as opposed to firm-internal) issues introduces the risk of systematic error. When relevant firsthand experience is unavailable, respondents' imagination may fill in information gaps (Huber and Power 1985), which may lead them to systematically over- or underreport.

H₂₀: The validity of key informant responses is greater for constructs pertaining to phenomena internal to the firm (vs. phenomena related to the firm environment).

Methodology

Data. To employ the MTMM techniques necessary to assess key informant validity, we need intertrait, intratrait, intermethod, and intramethod correlations from multiple informants on multiple traits. Our search for applications of triangulation described in the previous study (Figure 2) indicates that between 1998 and 2010, no study in six major journals reports this information. Moreover, with the exception of Phillips (1981), studies on applications of MTMM techniques focus mostly on self-reports, with method factors representing measurement techniques (Cote and Buckley 1987; Doty and Glick 1998; Lance et al. 2010).

Thus, we needed to turn to a different data source to test antecedents to key informant validity. We relied on eight own data sets from the field of marketing strategy, which used at least two key informants from each organization. These data sets are from cross-sectional, cross-industry surveys, thus reflecting typical conditions of key informant studies in marketing research (Rindfleisch et al. 2008). From these studies, we used 127 constructs. Table 4 provides an overview of the key informant samples and lists exemplary constructs.

Assessment of key informant validity. From the various methods available to analyze MTMM matrices using CFA (Bagozzi, Yi, and Phillips 1991), we relied on the correlated traits correlated methods (CTCM) approach (Lance, Noble, and Scullen 2002) to separate true score variance, systematic measurement error, and random measurement error. In this approach, a latent variable is specified for each trait (i.e., construct) and each method (i.e., informant), and each indicator loads on one trait factor and one method factor. Moreover, traits are allowed to correlate freely with other traits, and methods are allowed to correlate freely with other methods. Web Appendix D (www.marketingpower.com/jmr_webappendix) provides a formal and visual representation of this specification.

More specifically, to analyze key informant validity, we specified a series of CTCM models. To deal with the relatively common issue of nonconvergence, particularly in smaller samples (Rindfleisch et al. 2008), we relied on three approaches. First, instead of including all traits from one data set in a complex single model, the models we specified always included a set of four traits. To generalize across different trait combinations, we randomly assigned traits to these models and considered each trait in multiple models (20 on average). All models included two method factors because two types of key informants were used in all our samples. Second, consistent with our hypotheses that refer

Table 4
MULTI INFORMANT SAMPLES USED IN STUDY 2

<i>Sample Size (Number of Firms and/or Business Units)</i>	<i>Position of First Key Informant</i>	<i>Position of Second Key Informant</i>	<i>Number of Constructs</i>	<i>Selected Constructs</i>
126	Head of marketing	Head of sales	31	Customer orientation (culture), competitive intensity, responsiveness to customers and competitors
98	Marketing team leader	Marketing team members ^a	20	Competitive intensity, market dynamics, task-related competence, social skills
124	Key account team leader	Key account team member	13	Task-related competence, social skills, team performance, customer share of revenues
113	Head of sales	Head of planning	13	Centralization, market complexity, market performance, decision complexity, return on sales
215	Head of marketing or head of sales	Head of management accounting	12	Competitive intensity, market performance, cost leadership strategy, differentiation strategy
133	Head of marketing home country	Head of subsidiary foreign country	13	Market complexity, market performance, market growth, information flow, decision-making autonomy
136	Sales representative	Customers ^b	7	Customer loyalty, customer orientation, flexibility, relationship orientation
112	Head of marketing or sales	Head of accounting	18	Formalization, centralization, market complexity, market performance, organizational responsiveness

^aFor most team leaders, multiple responses from team members were available. We averaged their responses to create a team-leader team-member dyad.

^bFor all sales representatives, multiple responses from customers were available. We averaged their responses to create a sales representative customer dyad.

to validity at the construct level, we constrained loadings on each method factor to be equal for all items measuring the same construct using the same key informant. While reducing the number of improper solutions, these constraints do not substantially affect our results. Third, we excluded all models from further analysis for which standardized loadings exceeded 1 (Lance et al. 2010), which is a sign of nonconvergence.

For each resulting model, we computed the variance attributable to the type of key informant used by squaring the loadings on the method factor for each indicator (Cote and Buckley 1987). We then averaged the method variance obtained this way across all items of the construct. Next, we averaged this estimate across all models including the respective construct to arrive at one estimate of key informant bias per construct. It represents the amount of variance of the construct's indicators that is due to the types of key informants and serves as the dependent variable for testing the hypotheses of Study 2.

Coding. To measure the construct characteristics evoked by our hypotheses, we used the same expert rating approach as in the first study. Interrater reliability of the four judges was again high, with ICC(2, k) values ranging from .71 (objectivity) to .87 (local scope).

Results

Descriptive analysis. Before testing our hypotheses, we analyzed the percentage of variance in our data due to the key informant types. Consistent with previous studies (e.g., Cote and Buckley 1987), we computed the mean share of key informant variance at the item level. To account for the fact that random measurement error averages out if a construct is measured through multiple items, we also computed the exact mean share of key informant variance at the construct level using the formula described in Web Appendix E (www.marketingpower.com/jmr_webappendix). In Table 5,

Table 5
AVERAGE SHARE OF KEY INFORMANT VARIANCE FOR
DIFFERENT CONSTRUCT CATEGORIES

	<i>Studies</i>	<i>Constructs</i>	<i>Observations</i>	<i>Mean Share of Key Informant Variance</i>	
				<i>Item Level</i>	<i>Construct Level</i>
Overall	8	127	1057	.190	.264
Performance	8	23	1057	.086	.123
Intraorganizational	8	66	1057	.240	.337
Interorganizational	7	12	931	.230	.332
Extraorganizational	8	26	1057	.137	.173

we report descriptive results with regard to the construct categories previously employed in Study 1.

Overall, we find that 19% (26.4%) of indicator (construct) variance in our samples is due to the informant type used. Regarding the construct categories, we find that systematic error is particularly low for performance-related constructs (8.6%; 12.3%). Systematic measurement error is also below average for constructs referring to the firm environment (13.7%; 17.3%). In contrast, systematic measurement error is more problematic with respect to intraorganizational (24.0%; 33.7%) and interorganizational constructs (23.0%; 33.2%).

Hypothesis testing. Because our data are hierarchically structured (constructs are nested in studies), we use multi-level regression. Web Appendix F (www.marketingpower.com/jmr_webappendix) reports the exact model specifications. Table 6 summarizes the results. The share of key informant variance at the item level and the share of key informant variance at the construct level serve as the dependent variables. Therefore, low levels of the dependent variable represent high levels of key informant validity. Consistent with H₁₆, key informant variance is lower when

Table 6
THE IMPACT OF CONSTRUCT CHARACTERISTICS ON THE
EXTENT OF KEY INFORMANT VARIANCE: RESULTS OF
MULTILEVEL REGRESSION

		<i>Dependent Variable: Share of Key Informant Variance</i>	
<i>Hypothesis</i>	<i>Independent Variable</i>	<i>Item Level</i>	<i>Construct Level</i>
	<i>Construct Characteristics</i>		
H ₁₆	Present-focused	-.028 (.011)**	-.012 (.016)n.s.
H ₁₇	Objective information	-.066 (.008)**	-.076 (.012)**
H ₁₈	Salient events	-.034 (.012)**	-.040 (.017)*
H ₁₉	Characteristics of specific people	.001 (.01)n.s.	.000 (.001)n.s.
H ₂₀	Internal to the firm	.009 (.005)n.s.	.002 (.007)n.s.
	N (studies)	8	
	N (constructs)	127	
	N (observations)	1057	

* $p \leq .05$.

** $p \leq .01$.

n.s. $p > .05$ (not significant).

Notes: Unstandardized parameters are shown, and standard errors appear in parentheses.

the construct refers to the present rather than the past. This effect is significant at the item level ($b^I = -.028, p < .05$) but not at the construct level ($b^C = -.012, p > .05$). Thus, support for H₁₆ is mixed. Furthermore, key informant variance is significantly lower for constructs referring to objective information ($b^I = -.066, p < .05$; $b^C = -.076, p < .05$) and salient events ($b^I = -.034, p < .05$; $b^C = -.040, p < .05$), in support of H₁₇ and H₁₈. Key informant variance barely differs depending on whether constructs refer to characteristics of specific people versus social entities ($b^I = .001, p > .05$; $b^C = .000, p > .05$), which leads us to reject H₁₉. Finally, there is no support for H₂₀, predicting that key informant variance is lower when the construct refers to phenomena internal to the firm ($b^I = .009, p > .05$; $b^C = .002, p > .05$).

DISCUSSION

Implications for Empirical Research Using Triangulation

Our findings have implications for marketing and management researchers who consider using a key informant survey as their primary data source. On the basis of our empirical analyses, we offer insights on seven important questions they are facing.

Are key informant reports accurate? Our results allow us to empirically determine the overall accuracy of key informants. To evaluate these results, an important first step is to identify specific levels of key informant reliability and key informant validity that are deemed acceptable.

For correlational approaches to reliability, researchers might consider the .70 reliability cutoff for the sum of a construct's indicators (Nunnally 1978) and the .60 threshold for aggregates across organizational informants (Glick 1985). However, reported r and ICC(1) traditionally measure reliability of individual informants instead of aggregates.

The literature on CFA offers insights into appropriate levels of individual indicators that seem more relevant. Bagozzi and Yi (1988) require the reliability of individual indicators on average to be no smaller than .50. However, they make this recommendation under the assumption that

multiple indicators are present to measure a construct. Thus, instead of relying on a universal .50 threshold to assess the reliability of individual key informants, the research context needs to be taken into consideration. Particularly, individual key informant reliability should be relatively greater if only one key informant is used. Moreover, reliability requirements should also depend on the typical effect sizes in a research domain. We aimed to analytically determine the level of key informant reliability needed to identify true small, medium, and large effects in different sample sizes (for details, see Web Appendix G at www.marketingpower.com/jrm_webappendix). Our analyses reveal that to identify true small effects ($f^2 = .02, r = .14$) in a sample of 500 firms, individual key informant reliability must be .62 or higher. To find medium effects ($f^2 = .15, r = .36$) in a sample of 100 firms, a key informant reliability of .54 is sufficient (with smaller reliabilities needed in larger samples). Given the mean reliability of .612 that we find in our meta-analysis, key informant data may be problematic if researchers expect small effects, but they are adequate for identifying medium and large effects. Finally, when rather distant archival proxies are used for triangulation, small reliability estimates may not be surprising. Therefore, when interpreting reliability estimates, researchers should also take the conceptual closeness of the data points into account, as we do in our meta-analysis. For the $rwg_{(j)}$ agreement measure, Brown and Hauenstein (2005) suggest a .80 threshold as being indicative of strong agreement. The average key informant agreement of .872 that we find in our first study compares favorably with this cutoff value.

To the best of our knowledge, no established thresholds exist for evaluating systematic measurement error. Thus, to interpret the results from our second study, we describe additional analyses in Web Appendix G (www.marketingpower.com/jrm_webappendix). First, we assessed the risk of systematic error creating an artificial effect where no true effect is present. In samples of 500 firms, an 8% share of systematic error could suffice to produce such an artificial effect, whereas in samples of 100 firms, a 19% share is required. Thus, the results reported in Table 5 indicate that there is a risk of finding artificial effects, especially in large samples. Second, we explored the case of systematic error masking a true effect. If the covariance between systematic error components has the opposite sign of the true effect, even a 5% share of systematic error could mask true small effects in large samples. At the same time, a 20% (28%) share of systematic error is needed to mask a medium (large) effect under these conditions. Thus, the average degree of bias in our data sets suggests that there is a risk that true small and medium effects are not correctly identified using key informants.

Our finding of a mean item-level key informant variance of 19% is consistent with the results of Cote and Buckley (1987), who report for item variance in self-reports a share of method influences of 15.8% in marketing and 23.8% in other business disciplines. At the same time, it is worth stressing that our results do not imply that research using single key informants is necessarily biased. We report the mean share of systematic error across key informants, but consistent with Phillips (1981), there is variability between key informants in our data. For example, in the last data set listed in Table 4, we find a mean share of item-level key

informant variance of 13.1%. However, if we consider only the more accurate key informant for each construct, this value drops to 7.4%. Thus, key informant responses can be accurate.

Our results support Spector's (2006) claim that it is an urban legend that everything measured with the same method automatically shares common method variance. Indeed, key informant accuracy can be linked to a diverse set of conditions, which we discuss in the next paragraphs. Thus, we suggest that relying on key informants can be useful if the circumstances are right.

For which phenomena are key informant responses particularly accurate? We find key informant accuracy to vary systematically depending on the construct being measured. Key informants are significantly more reliable for constructs that refer to the present, that point to objectively verifiable referents, and that address salient events. In addition, there is some evidence that they are less reliable for constructs referring to the firm environment. Comparing standardized regression coefficients shows that the routine versus salience factor has the strongest effect for correlations ($\beta_{\text{Cor}} = -.156$), followed by present versus past ($\beta_{\text{Cor}} = .105$) and subjective versus objective ($\beta_{\text{Cor}} = -.104$). Regarding agreement, the strongest effects occurred for present versus past ($\beta_{\text{Ag}} = .193$), environment versus organization ($\beta_{\text{Ag}} = -.191$), routine versus salience ($\beta_{\text{Ag}} = -.161$), and subjective versus objective ($\beta_{\text{Ag}} = -.083$). In addition, we find key informant validity to be greater for constructs pointing to objective referents and addressing salient issues.

Scholars can use these findings to assess the suitability of key informants for their particular study. For example, key informant responses on performance outcomes (objectively verifiable and salient) are likely to be relatively accurate. However, for measures of organizational culture (subjective and routine processes), key informants are less likely to be highly accurate.

Which key informants provide particularly reliable responses? Our results suggest that reliability is linked to the informant's hierarchical position and tenure. Informants in high hierarchical positions and with longer tenure are more reliable. As Kumar, Stern, and Anderson (1993) note, response reliability is thus linked to the experience of the particular informant.

For which organizations are key informant responses particularly reliable? Overall, our results indicate that organizational characteristics do not substantively affect key informant reliability. Thus, key informants appear to be a suitable data source for various kinds of organizations. One exception seems noteworthy: Reliability (as measured through correlational approaches) is significantly lower for large organizations, suggesting that key informant methodology is particularly useful for research on small and medium-sized enterprises.

In which industries are key informant responses particularly reliable? Key informants appear to be less reliable when industry concentration is high (e.g., the four largest firms have a market share greater than 40%). In addition, key informants are more reliable in R&D-intensive industries.

When should researchers employ key informant triangulation? The results of this research indicate that scholars should employ triangulation if at least one of two conditions is met: (1) Key informant response accuracy is expected to be

low (relevant conditions are discussed previously), and/or (2) little empirical evidence on key informant response reliability exists in the particular research domain. As Figure 2 illustrates, this is especially the case for constructs referring to interorganizational relationships and constructs referring to the organizational environment.

How should researchers implement triangulation? First, with respect to the goals of triangulation, we suggest that researchers address both reliability and validity issues. Moreover, given the degree of systematic error that we find, an advisable step is to pursue the goal of enhancing validity using the methods referred to in Figure 1. Second, regarding the number of triangulated constructs, a prudent approach is to collect data from multiple sources for at least two (preferably three) constructs, which is a requirement for applying CFA-based MTMM approaches. Third, our analyses suggest cautious use of absolute threshold values to assess reliability. As our elaboration of thresholds reflects, the implications of reliability and validity for hypotheses testing strongly depend on effect size, sample size, and the type of data used. (Lower reliability values can be expected if survey data is used for triangulation instead of archival data.) Thus, we recommend assessing triangulation coefficients in light of these study-specific parameters. Fourth, our search for studies using triangulation revealed the need for improved reporting of triangulation. Only 127 studies (of 182) provide numerical reliability information. In addition to information on key informant validity, this should always be made available.

Implications for Further Methodological Research on Triangulation

Our results also have some implications for further methodological research. First, researchers could attempt to uncover the exact cognitive mechanisms behind our findings. For example, we find that key informants respond more accurately if constructs refer to objectively verifiable referents. Our social cognition explanation is that in these situations, information availability/retrievability is greater. However, if key informants are strategically inclined to misrepresent their perceptions, except when they know their responses can ultimately be verified, our findings would not differ.

Second, further research could investigate how key informant reliability and validity are linked to other aspects of response behavior, such as response times. Fast response times could indicate high information availability and thus higher response accuracy. Third, whereas most methodological studies related to key informant research rely on MTMM-based methods to analyze validity, applied empirical studies do not employ these methods, perhaps owing to estimation problems common to many MTMM model specifications. These problems have led to recommendations regarding sample size ($n \geq 250$), number of informants ($k \geq 3$), and number of triangulated constructs ($t \geq 3$) (Lance, Noble, and Scullen 2002) that most studies in our meta-analysis do not meet. Thus, on the basis of insights from Figure 1 regarding typical triangulation designs in applied research, researchers should develop approaches to detecting systematic error with more realistic requirements.

REFERENCES

- Baddeley, Alan (1986), *Working Memory*. Oxford: Clarendon.
- Bagozzi, Richard P. and Youjae Yi (1988), "On the Evaluation of Structural Equation Models," *Journal of the Academy of Marketing Science*, 16 (1), 74–94.
- and ——— (1991), "Multitrait-Multimethod Matrices in Consumer Research," *Journal of Consumer Research*, 17 (4), 426–39.
- , ———, and Lynn W. Phillips (1991), "Assessing Construct Validity in Organizational Research," *Administrative Science Quarterly*, 36 (3), 421–58.
- Barnett, William P. (1997), "The Dynamics of Competitive Intensity," *Administrative Science Quarterly*, 42 (1), 128–60.
- Biggadike, E. Ralph (1979), *Corporate Diversification: Entry, Strategy, and Performance*. Boston: Harvard University Press.
- Bliese, Paul D. (2000), "Within-Group Agreement, Non-independence, and Reliability: Implications for Data Aggregation and Analysis," in *Multilevel Theory, Research, and Methods in Organizations*, Katherine J. Klein and Steve W. J. Kozlowski, eds. San Francisco: Jossey-Bass.
- Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein (2009), *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons.
- Brown, Reagan D. and Neil M.A. Hauenstein (2005), "Interrater Agreement Reconsidered: An Alternative to the r_{WG} Indices," *Organizational Research Methods*, 8 (2), 165–84.
- Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56 (2), 81–105.
- Churchill, Gilbert A. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64–73.
- Cohen, Wesley M. and Daniel A. Levinthal (1990), "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, 35 (1), 128–52.
- Conway, James M., Robert A. Jako, and Deborah E. Goodman (1995), "A Meta-Analysis of Interrater and Internal Consistency Reliability of Selection Interviews," *Journal of Applied Psychology*, 80 (5), 565–79.
- Cote, Joseph A. and M. Ronald Buckley (1987), "Estimating Trait, Method, and Error Variance: Generalizing Across 70 Construct Validation Studies," *Journal of Marketing Research*, 24 (August), 315–18.
- and ——— (1988), "Measurement Error and Theory Testing in Consumer Research: An Illustration of the Importance of Construct Validation," *Journal of Consumer Research*, 14 (4), 579–82.
- Davis, Harry L., Susan P. Douglas, and Alvin J. Silk (1981), "Measure Unreliability: A Hidden Threat to Cross-National Marketing Research?" *Journal of Marketing*, 45 (Spring), 98–109.
- Denzin, Norman K. (1978), *The Research Act*. New York: McGraw-Hill.
- Doty, D. Harold and William H. Glick (1998), "Common Methods Bias: Does Common Methods Variance Really Bias Results?" *Organizational Research Methods*, 1 (4), 374–406.
- Dunlap, William P., Michael J. Burke, and Kristin Smith-Crowe (2003), "Accurate Tests of Statistical Significance for r_{WG} and Average Deviation Interrater Agreement Indexes," *Journal of Applied Psychology*, 88 (2), 356–62.
- Duval, Sue and Richard Tweedie (2000), "A Nonparametric 'Trim And Fill' Method of Accounting for Publication Bias in Meta-Analysis," *Journal of the American Statistical Association*, 95 (449), 89–98.
- Eid, Michael (2000), "A Multitrait-Multimethod Model with Minimal Assumptions," *Psychometrika*, 65 (2), 241–61.
- Fazio, Russell H., David M. Sanbonmatsu, Martha C. Powell, and Frank R. Kardes (1986), "On the Automatic Activation of Attitudes," *Journal of Personality and Social Psychology*, 50 (2), 229–38.
- Feldman, Jack M. and John G. Lynch (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73 (3), 421–35.
- Fisher, R.A. (1925), "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society*, 22 (1), 700–725.
- Franke, Georg R. and Jeong-Eun Park (2006), "Salesperson Adaptive Selling Behavior and Customer Orientation: A Meta-Analysis," *Journal of Marketing Research*, 43 (November), 693–702.
- Geyskens, Inge, Rekha Krishnan, Jan-Benedict E.M. Steenkamp, and Paulo V. Cunha (2009), "A Review and Evaluation of Meta-Analysis Practices in Management Research," *Journal of Management*, 35 (2), 393–419.
- Gillund, Gary and Richard M. Shiffrin (1984), "A Retrieval Model for Both Recognition and Recall," *Psychological Review*, 91 (1), 1–67.
- Glick, William H. (1985), "Conceptualizing and Measuring Organizational and Psychological Climate: Pitfalls in Multilevel Research," *Academy of Management Review*, 10 (3), 601–616.
- Golden, Brian R. (1992), "The Past Is the Past—Or Is It? The Use of Retrospective Accounts as Indicators of Past Strategy," *Academy of Management Journal*, 35 (4), 848–60.
- Hambrick, Donald C. (1981), "Strategic Awareness Within Top Management Teams," *Strategic Management Journal*, 24 (3), 263–79.
- Harrison, David A., Mary E. McLaughlin, and Terry M. Coalter (1996), "Context, Cognition, and Common Method Variance: Psychometric and Verbal Protocol Evidence," *Organizational Behavior and Human Decision Processes*, 68 (3), 246–61.
- Hastie, Reid and Bernadette Park (1986), "The Relationship Between Memory and Judgment Depends on Whether the Judgment Is Memory-Based or Online," *Psychological Review*, 93 (3), 258–68.
- Homburg, Christian, Marko Grozdanovic, and Martin Klarmann (2007), "Responsiveness to Customers and Competitors: The Role of Cognitive and Affective Organizational Systems," *Journal of Marketing*, 71 (July), 18–38.
- Houston, Michael J. and Seymour Sudman (1975), "A Methodological Assessment of the Use of Key Informants," *Social Science Research*, 4 (2), 151–64.
- Huber, George P. and Daniel J. Power (1985), "Retrospective Reports of Strategic-Level Managers: Guidelines for Increasing Their Accuracy," *Strategic Management Journal*, 6 (2), 171–80.
- Hunter, John E. and Frank L. Schmidt (2004), *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 2d ed. Thousand Oaks, CA: Sage Publications.
- James, Lawrence R., Robert G. Demaree, and Gerrit Wolf (1984), "Estimating Within-Group Interrater Reliability with and Without Response Bias," *Journal of Applied Psychology*, 69 (1), 85–98.
- , ———, and ——— (1993), " r_{WG} : An Assessment of Within-Group Interrater Agreement," *Journal of Applied Psychology*, 78 (2), 306–309.
- Jap, Sandy D. (1999), "Pie-Expansion Efforts: Collaboration Processes in Buyer-Supplier Relationships," *Journal of Marketing Research*, 36 (November), 461–75.
- Jick, Todd D. (1979), "Mixing Qualitative and Quantitative Methods: Triangulation in Action," *Administrative Science Quarterly*, 24 (4), 602–611.
- Jobe, Jared B., Roger Tourangeau, and Albert F. Smith (1993), "Contributions of Survey Research to the Understanding of Memory," *Applied Cognitive Psychology*, 7 (7), 567–84.
- Kozlowski, Steve W. and Keith Hattrup (1992), "A Disagreement About Within-Group Agreement: Disentangling Issues of Consistency Versus Consensus," *Journal of Applied Psychology*, 77 (2), 161–67.

- Kumar, Ajith and William R. Dillon (1990), "On the Use of Confirmatory Measurement Models in the Analysis of Multiple-Informant Reports," *Journal of Marketing Research*, 27 (February), 102–111.
- and — (1992), "An Integrative Look at the Use of Additive and Multiplicative Covariance Structure Models in the Analysis of MTMM Data," *Journal of Marketing Research*, 29 (February), 51–64.
- Kumar, Nirmalya, Louis W. Stern, and James C. Anderson (1993), "Conducting Interorganizational Research Using Key Informants," *Academy of Management Journal*, 36 (6), 1633–51.
- Lance, Charles E., Bryan Dawson, David Birkelbach, and Brian J. Hoffman (2010), "Method Effects, Measurement Error, and Substantive Conclusions," *Organizational Research Methods*, 13 (3), 435–55.
- , Carrie L. Noble, and Steven E. Scullen (2002), "A Critique of the Correlated Trait-Correlated Method and Correlated Uniqueness Models for Multitrait-Multimethod Data," *Psychological Methods*, 7 (2), 228–44.
- Lee, Chang-Yang and Ishtiaq P. Mahmood (2009), "Inter-Industry Differences in Profitability: The Legacy of the Structure-Efficiency Debate Revisited," *Industrial and Corporate Change*, 18 (3), 351–80.
- Lipsey, Mark W. and David B. Wilson (2001), *Practical Meta-Analysis*. Thousand Oaks, CA: Sage Publications.
- March, James G. and Robert I. Sutton (1997), "Organizational Performance as a Dependent Variable," *Organization Science*, 8 (6), 698–706.
- Marsh, Herbert W., Barbara M. Byrne, and Rhonda Craven (1992), "Overcoming Problems in Confirmatory Factor Analyses of MTMM Data: The Correlated Uniqueness Model and Factorial Invariance," *Multivariate Behavioral Research*, 27 (4), 489–507.
- McGraw, Kenneth O. and S.P. Wong (1996), "Forming Inferences About Some Intraclass Correlation Coefficients," *Psychological Methods*, 1 (1), 30–46.
- Nunnally, Jum C. (1978), *Psychometric Theory*. New York: McGraw-Hill.
- Paulhus, Delroy L. (1991), "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes*, John P. Robinson, Phillip R. Shaver, and Lawrence S. Wright, eds. San Diego: Academic Press.
- Phillips, Lynne W. (1981), "Assessing Measurement Error in Key Informant Reports: A Methodological Note on Organizational Analysis in Marketing," *Journal of Marketing Research*, 18 (November), 395–415.
- Pillemer, David B., Lynn R. Goldsmith, Abigail T. Panter, and Sheldon H. White (1988), "Very Long-Term Memories of the First Year in College," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (4), 709–715.
- Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003), "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, 88 (5), 879–903.
- Porter, Michael E. (1980), *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: The Free Press.
- Rindfleisch, Aric, Alan J. Malter, Shankar Ganesan, and Christine Moorman (2008), "Cross-Sectional Versus Longitudinal Survey Research: Concepts, Findings, and Guidelines," *Journal of Marketing Research*, 45 (August), 261–79.
- Rogelberg, Steven G. and Jeffrey M. Stanton (2007), "Introduction Understanding and Dealing with Organizational Survey Nonresponse," *Organizational Research Methods*, 10 (2), 195–209.
- Seidler, John (1974), "On Using Informants: A Technique for Collecting Quantitative Data and Controlling for Measurement Error in Organizational Analysis," *American Sociological Review*, 39 (12), 816–31.
- Silk, Alvin J. and Manohar U. Kalwani (1982), "Measuring Influence in Organizational Purchase Decisions," *Journal of Marketing Research*, 19 (May), 165–81.
- Slater, Stanley F. and John C. Narver (1994), "Does Competitive Environment Moderate the Market Orientation-Performance Relationship?" *Journal of Marketing*, 58 (January), 46–55.
- Slovic, Paul (1962), "Convergent Validation of Risk Taking Measures," *Journal of Abnormal and Social Psychology*, 65 (1), 68–71.
- Spector, Paul E. (2006), "Method Variance in Organizational Research: Truth or Urban Legend?" *Organizational Research Methods*, 9 (2), 221–32.
- Starbuck, William H. and John M. Mezias (1996), "Opening Pandora's Box: Studying the Accuracy of Managers' Perceptions," *Journal of Organizational Behavior*, 17 (2), 99–117.
- Sudman, Seymour, Barbara Bickart, Johnny Blair, and Geeta Menon (1994), "The Effects of Level of Participation on Reports of Behavior and Attitudes by Proxy Reporters," in *Autobiographical Memory and the Validity of Retrospective Reports*, Norbert Schwarz and Seymour Sudman, eds. New York: Springer.
- Sutcliffe, Kathleen M. (1994), "What Executives Notice: Accurate Perceptions in Top Management Teams," *Academy of Management Journal*, 37 (5), 1360–78.
- Thompson, Simon G. and Stephen J. Sharp (1999), "Explaining Heterogeneity in Meta-Analysis: A Comparison of Methods," *Statistics in Medicine*, 18 (20), 2693–2708.
- Tinsley, Howard E.A. and David J. Weiss (1975), "Interrater Reliability and Agreement of Subjective Judgements," *Journal of Counseling Psychology*, 22 (4), 358–74.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski (2000), *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.
- Van Bruggen, Gerrit H., Gary L. Lilien, and Manish Kacker (2002), "Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses," *Journal of Marketing Research*, 39 (November), 469–78.
- Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest (1966), *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.

What Drives Key Informant Accuracy?

Christian Homburg, Martin Klarmann, Martin Reimann, and Oliver Schilke

WEB APPENDIXES

WEB APPENDIX A:

DETAILS ON THE CODING OF CONSTRUCT CHARACTERISTICS IN STUDY 1

To measure the *construct characteristics* evoked by the first study's hypotheses, we relied on expert judgments from four veteran marketing researchers (all with considerable publication experience in leading journals of the field), who provided scores for five construct characteristics. Two of them are co-authors of the paper. To provide the experts with necessary background information, an assessment sheet was prepared for each construct considered in the meta-analysis. Information included the construct name, the study the construct was included in, the construct definition, the corresponding paragraph from the measurement section, and the questionnaire items. A sample assessment sheet is reproduced below.

The experts were asked to assess the construct characteristics using five seven-point semantic differentials (also reproduced below). When wording these semantic differentials, we took care to ensure that they would only refer to characteristics of the constructs, without requiring any assumptions about the study's key informants, organizations, or industries. We also instructed the expert judges to solely assess the constructs based on the assessment sheet. To ensure that the experts had all necessary information to provide valid judgments, we conducted a pretest of the assessment sheets using a subsample of twenty randomly selected constructs. After the pretest, the experts all agreed that they had sufficient information and were not forced to make assumptions about informants, organizations, or industries.

After assessing the constructs separately, we followed Kumar, Stern, and Anderson (1993) by establishing consensus among the experts in cases where scores differed on average by 2 or more. The resulting measures of construct characteristics showed high inter-rater reliability, with ICC(2,k) ratings ranging from .71 (temporal reference) to .86 (local scope).

Sample Assessment Sheet

Group innovativeness

Innovation, Market Orientation, and Organizational Learning: An Integration and Empirical Examination. By: Hurley, Robert F.; Hult, G. Tomas M.. Journal of Marketing, Jul98, Vol. 62 Issue 3, p42-54,

Definition:

Innovativeness is the notion of openness to new ideas as an aspect of a firm's culture. Innovativeness of the culture is a measure of the organization's orientation toward innovation.

Measurement:

All group culture measures (innovativeness, power sharing, participative decision making, learning and development, and support and collaboration) represent an aggregation of the individual scores of the group members to arrive at mean scores for the group. This follows from theory that posits culture as collective perceptions maintained at the group level but carried at the individual level (Van Maanen and Barley 1985).

Culture was operationalized by using an instrument developed by Burke (1989) to measure people's perceptions of group culture. All culture measures used a five-point scale ranging from "not descriptive" = 1 to "very descriptive" = 5. The respondents were asked to indicate the extent to which different items described their group.

Items (R = item is reverse scored):

Innovativeness

1. Technical innovation, based on research results, is readily accepted.
2. Management actively seeks innovative ideas.
3. Innovation is readily accepted in program/project management.
4. People are penalized for new ideas that don't work (R).
5. Innovation in XYZ is perceived as too risky and is resisted (R).

	Past 1 2 3 4 5 6 7 Present						
To what extent does this construct refer to the past vs. the present?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Objectively verifiable referents				Subjective states and attitudes		
	1	2	3	4	5	6	7
To what extent does this construct refer to objectively verifiable referents vs. subjective states and attitudes?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Salient Events				Routine Processes		
	1	2	3	4	5	6	7
To what extent does this construct refer to salient events vs. routine processes?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Specific Individuals				Abstract Social Entity		
	1	2	3	4	5	6	7
To what extent does this construct refer to specific individuals vs. abstract social entities?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Characteristics of the Organization				Characteristics of the Environment		
	1	2	3	4	5	6	7
To what extent does this construct refer to characteristics of the organization vs. to characteristics of the organizational environment?.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

WEB APPENDIX B:

DETAILS ON THE CODING OF INDUSTRY CHARACTERISTICS IN STUDY 1

In our meta-analysis, we consider three *industry characteristics*. First, we study the effect of *industry concentration*. We measure it using the sum of the sales market shares of the industry's four largest firms (Biggadike 1979). The required industry composition information was available for 70 studies in our sample. Also, we required information on the year in which the survey had taken place. Here, exact information was provided in 27 studies, with an average time-span of six years between data collection and publication. For the remaining 43 studies we approximated the year of the survey by using this average. We then collected information from COMPUSTAT on industry concentration for the five years prior to data collection. For each study, industry concentration was calculated as a weighted average across the five data points, with weights reflecting the industry composition of the study's sample.

Second, we analyze the effect of *industry R&D intensity*. While collecting data in a similar way as described above, R&D intensity was measured as the ratio of industry R&D expenditures to total industry sales (e.g., Lee and Mahmood 2009) across the five years preceding a study's data collection.

Third, we consider sales volatility, which is indicative of *industry dynamism* (Sutcliffe 1994). To compute this index, we regressed industry sales in the five years prior to data collection on a variable representing the time period. The standard error of this regression was then divided by the mean of the dependent variable, yielding the sales volatility index (Sutcliffe 1994).

WEB APPENDIX C:

EXACT MODEL SPECIFICATIONS FOR HYPOTHESIS TESTING IN STUDY 1

In Tables 2 and 3 in the paper, we report the effects of hypothesized antecedents to informant reliability. Here, we reproduce the exact specification of the models in equation form.

Models estimating the impact of construct characteristics and method characteristics

Data on construct characteristics and method characteristics are available at the construct level. Because oftentimes reliability data for multiple constructs come from a single study, the data is structured hierarchically. Hence, we use multilevel regression with a random intercept (capturing study-wide influences) and fixed effects of the antecedents to informant reliability.

Specification of the model that considers correlational reliability as dependent variable:

$$\text{Cor}_{ij} = \alpha_j + \beta_1 \text{PRESENT}_{ij} + \beta_2 \text{OBJECTIVE}_{ij} + \beta_3 \text{SALIENCE}_{ij} + \beta_4 \text{SPECIFIC}_{ij} + \beta_5 \text{INSIDE}_{ij} \\ + \beta_6 \text{OBJ}_{ij} + \beta_7 \text{CUST}_{ij} + \beta_8 \text{HIERDIF}_{ij} + \beta_9 \text{FUNDIF}_{ij} + \beta_{10} \text{INDEX}_{ij}$$

Specification of the model that considers key informant agreement as dependent variable:

$$\text{Ag}_{ij} = \alpha_j + \beta_1 \text{PRESENT}_{ij} + \beta_2 \text{OBJECTIVE}_{ij} + \beta_3 \text{SALIENCE}_{ij} + \beta_4 \text{SPECIFIC}_{ij} + \beta_5 \text{INSIDE}_{ij} \\ + \beta_6 \text{HIERDIF}_{ij} + \beta_7 \text{FUNDIF}_{ij}$$

where

Cor_{ij} : Reliability of construct i from study j measured through Pearson's r or ICC(1)

Ag_{ij} : Informant agreement for construct i from study j measured through rwg(J) index

PRESENT_{ij} : Degree to which construct i from study j refers to the present (versus the past)

OBJECTIVE_{ij} : Degree to which construct i from study j refers to objectively verifiable referents (versus subjective states and attitudes)

SALIENCE_{ij}: Degree to which construct i from study j refers to salient events (versus routine processes)

SPECIFIC_{ij}: Degree to which construct i from study j refers to specific individuals (versus specific individuals)

INSIDE_{ij}: Degree to which construct i from study j refers to characteristics of the organization (versus characteristics of the environment)

OBJ_{ij}: Dummy variable indicating whether objective data was used to assess key informant reliability for construct i in study j

CUST_{ij}: Dummy variable indicating whether customer data was used to assess key informant reliability for construct i in study j

HIERDIF_{ij}: Dummy variable indicating whether data from key informants with a different hierarchical background was used to assess key informant reliability for construct i in study j

FUNDIF_{ij}: Dummy variable indicating whether data from key informants with a different functional background was used to assess key informant reliability for construct i in study j

INDEX_{ij}: Dummy variable indicating whether Cor_{ij} was measured through ICC(1).

As described in the paper, individual observations were weighted for the purpose of model estimation using the inverse of the standard errors of Cor_{ij} as weights with correlations as DV and the square root of the size of the sample used to compute Ag_{ij} as weights with agreement as DV.

Models estimating the impact of informant characteristics, organizational characteristics, and industry characteristics

Data on informant characteristics, organizational characteristics, and industry characteristics were available at the study level. Hence, as described in the paper, we aggregated construct level reliabilities to a study estimate and used weighted regression analysis to estimate the impact of the different antecedents. In the following we reproduce the exact model

specifications of models using average correlations in study j ($\overline{\text{Cor}}_j$) and average agreement in study j ($\overline{\text{Ag}}_j$) as dependent variables.

Specification of the models with informant hierarchical position as independent variable:

$$\overline{\text{Cor}}_j = \alpha + \beta \text{POSITION}_j \quad \overline{\text{Ag}}_j = \alpha + \beta \text{POSITION}_j$$

where POSITION_j is the hierarchical position of the first key informant, coded as reported.

Specification of the models with informant specialization as independent variable:

$$\overline{\text{Cor}}_j = \alpha + \beta \text{SPECIAL}_j \quad \overline{\text{Ag}}_j = \alpha + \beta \text{SPECIAL}_j$$

where SPECIAL_j is a dummy variable, indicating whether the first key informants holds a specialist position (versus a generalist position).

Specification of the models with average informant tenure as independent variable:

$$\overline{\text{Cor}}_j = \alpha + \beta \text{TENURE}_j \quad \overline{\text{Ag}}_j = \alpha + \beta \text{TENURE}_j$$

where TENURE_j is the natural logarithm of the average tenure of the first key informant in the unit of analysis.

Specification of the models with average organization size as independent variable:

$$\overline{\text{Cor}}_j = \alpha + \beta \text{ORGSIZE}_j \quad \overline{\text{Ag}}_j = \alpha + \beta \text{ORGSIZE}_j$$

where ORGSIZE_j is the natural logarithm of the average number of employees in the first key informant's organization.

Specification of the models with average organization age as independent variable:

$$\overline{Cor}_j = \alpha + \beta \text{ORGAGE}_j \quad \overline{Ag}_j = \alpha + \beta \text{ORGAGE}_j$$

where ORGAGE_j is the natural logarithm of the average number years the first key informant's organization is in existence.

Specification of the models with industry characteristics as independent variables:

$$\overline{Cor}_j = \alpha + \beta_1 \text{CONCEN}_j + \beta_2 \text{RDINT}_j + \beta_3 \text{DYN}_j \quad \overline{Ag}_j = \alpha + \beta_1 \text{CONCEN}_j + \beta_2 \text{RDINT}_j + \beta_3 \text{DYN}_j$$

where

CONCEN_j is the sum of the sales market shares of the four largest firms in the industry of the key informants' firms during the five years before data collection

RDINT_j is ratio of R&D expenditures to total sales in the industry of the key informants' firms during the five years before data collection

DYN_j is the industry sales volatility in the industry of the key informants' firms during the five years before data collection.

As described in the paper, individual observations were weighted during model estimation. For models with correlations as dependent variable we used the inverse of the standard errors of \overline{Cor}_j multiplied with the square root of the number of constructs that was averaged as weights. With agreement as DV, we used the square root of the size of the average sample sized used to compute \overline{Ag}_j multiplied with the square root of the number of constructs that was averaged as weights.

WEB APPENDIX D:

MTMM-MODEL SPECIFICATIONS USED IN STUDY 2

To measure the degree of systematic error variance in the key informant data sets used in Study 2, we relied on an approach that combines Multi-Trait Multi-Method (MTMM) matrices with confirmatory factor analysis (CFA). In this context, a number of different approaches have been developed (Bagozzi et al. 1991, Podsakoff et al. 2003). In this section of the Web Appendix, we justify our specific choice and then reproduce the exact model specification both in equation form and visually.

To estimate the amount of variance that is due to the specific type of key informant used, we relied on the Correlated Traits Correlated Methods (CTCM) approach (Lance, Noble, and Scullen 2002). In this specification, each indicator loads on one trait factor (i.e., construct) and one method factor (i.e., informant). Moreover, traits are allowed to correlate freely with other traits, and methods are allowed to correlate freely with other methods.

We chose the CTCM approach for three reasons. (1) “[I]t is theoretically the most faithful to the design of the MTMM matrix” (Lance et al. 2010, p. 443). (2) It is consistently used in related studies (e.g., Cote and Buckley 1987; Doty and Glick 1998; Lance et al. 2010). (3) It avoids the upward bias in trait loadings of the correlated uniqueness (CU) specification that is often considered the best alternative to the CTCM approach (Lance, Noble, and Scullen 2002). It is worth noting that by using the CTCM approach, we assume that traits and methods are independent of each other (Cote and Buckley 1987). However, the direct product (DP) approach that relaxes this assumption (Bagozzi et al. 1991) does not allow the separation of trait and method variance (Podsakoff et al. 2003) that is central to our analysis.

The key characteristics of the CTCM models we estimated are as follows:

- *Two methods*: In all CTCM models two method factors were specified (representing the two types of informants used in our datasets)
- *Four traits*: In all CTCM models four trait factors were specified (representing four randomly selected constructs from a single dataset).
- *Varying number of indicators*: The number of indicators per trait typically varied but they were always equally distributed across informant types.
- *Constrained loadings on method factors*: Our theoretical considerations imply that all items of a construct that are measured through the same key informant contain the same amount of systematic error variance. Therefore (and to reduce the risk of model non-convergence), we constrained the loadings on the method factor to be equal across all items measuring the same construct using the same key informant.
- *No correlations between traits and methods*: In our models traits were allowed to correlate freely with other traits and methods were allowed to correlate freely with other methods. However, no correlations between traits and methods were specified.

Hence, for each model, we estimated the following latent variable covariance matrix:

$$\begin{pmatrix} 1 & \Phi_{21} & \Phi_{31} & \Phi_{41} & 0 & 0 \\ \Phi_{21} & 1 & \Phi_{42} & \Phi_{42} & 0 & 0 \\ \Phi_{31} & \Phi_{32} & 1 & \Phi_{43} & 0 & 0 \\ \Phi_{41} & \Phi_{42} & \Phi_{43} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Phi_{65} \\ 0 & 0 & 0 & 0 & \Phi_{65} & 1 \end{pmatrix}$$

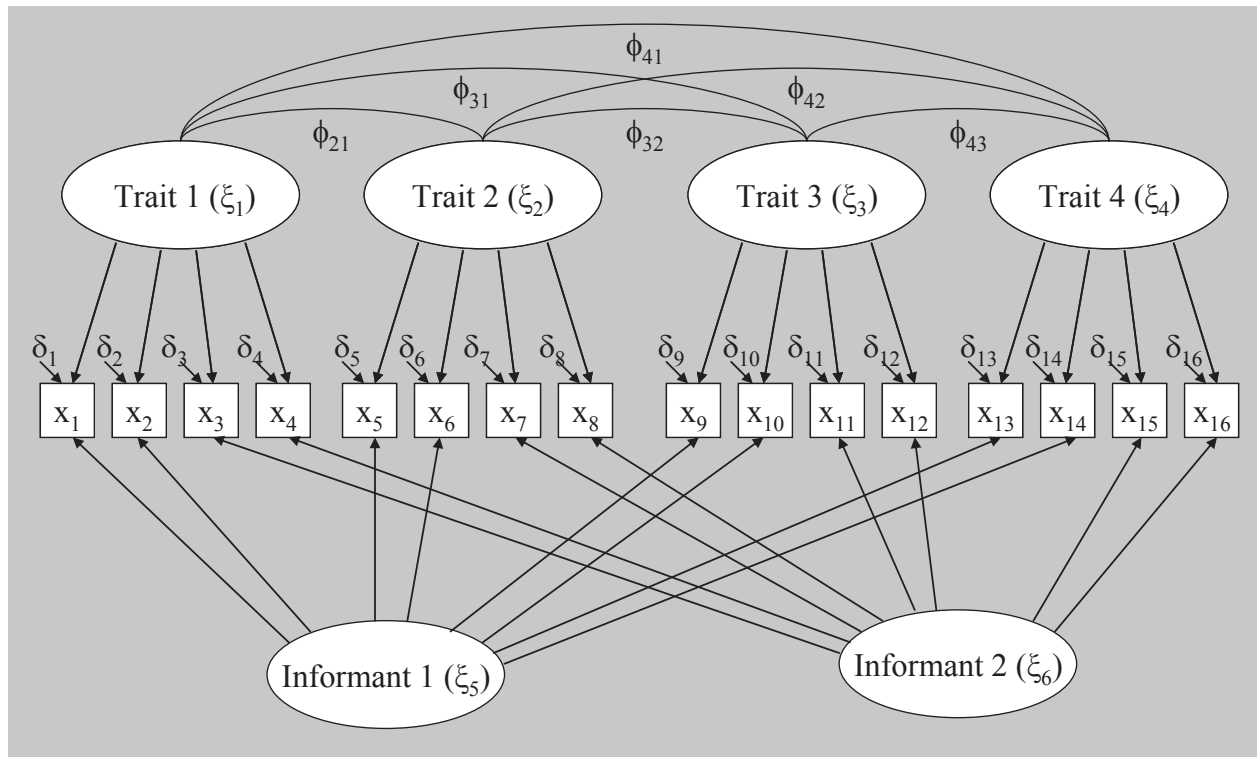
where Φ_{ij} represents the covariance of latent variable ξ_i with ξ_j . Note that we established a scale for the latent variables by setting their variance to 1 (Brown 2006).

The corresponding measurement model is shown below. Here, we assume that all traits are measured using two items per informant. In our data this number varied.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \\ x_{15} \\ x_{16} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 & 0 & 0 & \lambda_{51} & 0 \\ \lambda_{12} & 0 & 0 & 0 & \lambda_{51} & 0 \\ \lambda_{13} & 0 & 0 & 0 & 0 & \lambda_{63} \\ \lambda_{14} & 0 & 0 & 0 & 0 & \lambda_{63} \\ 0 & \lambda_{25} & 0 & 0 & \lambda_{55} & 0 \\ 0 & \lambda_{26} & 0 & 0 & \lambda_{55} & 0 \\ 0 & \lambda_{27} & 0 & 0 & 0 & \lambda_{67} \\ 0 & \lambda_{28} & 0 & 0 & 0 & \lambda_{67} \\ 0 & 0 & \lambda_{39} & 0 & \lambda_{59} & 0 \\ 0 & 0 & \lambda_{310} & 0 & \lambda_{59} & 0 \\ 0 & 0 & \lambda_{311} & 0 & 0 & \lambda_{611} \\ 0 & 0 & \lambda_{312} & 0 & 0 & \lambda_{611} \\ 0 & 0 & 0 & \lambda_{413} & \lambda_{513} & 0 \\ 0 & 0 & 0 & \lambda_{414} & \lambda_{513} & 0 \\ 0 & 0 & 0 & \lambda_{415} & 0 & \lambda_{615} \\ 0 & 0 & 0 & \lambda_{416} & 0 & \lambda_{615} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \\ \delta_{14} \\ \delta_{15} \\ \delta_{16} \end{pmatrix}$$

where x_k is the k^{th} observed variable, λ_{ik} is the loading of observed variable k on latent variable ξ_i and δ_k is the random measurement error of the k^{th} observed variable. Note that equality constraints for factor loadings are represented using identical indices for two loadings.

An illustrative visualization of the CTCM model specification we used appears below
(again, we assume that each trait is measured through two indicators per informant):



WEB APPENDIX E:

FORMULAS USED TO DERIVE SHARE OF SYSTEMATIC MEASUREMENT ERROR

Table 5 in the paper reports the average share of key informant variance across the constructs in our study. We report the share of systematic error at the item level and the share of systematic error at the construct level. In the following, we reproduce the formulas used to arrive at these measures for individual constructs (i.e., before averaging across all constructs in our sample). To measure the average item level key informant variance ISys for a given construct we used the following formula:

$$\text{ISys} = (\lambda_{11}^2 + \lambda_{12}^2) / 2 \quad (\text{E.1})$$

where

λ_{11} is the standardized factor loading of the items measured through informant I1 on the factor representing I1. In this context, it is worth mentioning again that we specified loadings on the method factor to be equal for all indicators measuring a construct through the same informant. In all models we only considered two informants.

To account for the fact that random measurement error averages out if multiple indicators of a construct are used to form a composite, we also computed key informant variance at the construct level. To do so, we adapted the classic formula for composite reliability to indicate the share of systematic error in the sum of a constructs items measured through the same informant. We reproduce the formula used for the first informant below.

$$\text{CSys}_{\text{I1}} = \frac{\left(\sum_{k=1}^K \lambda_{11k} \right)^2}{\left(\sum_{k=1}^K \lambda_{1k} \right)^2 + \left(\sum_{k=1}^K \lambda_{11k} \right)^2 + \left(\sum_{k=1}^K \theta_{\delta k} \right)} \quad (\text{E.2})$$

where

λ_{I1k} is the standardized factor loading on the factor representing informant I1 of the k^{th} item of trait T measured through informant I1.

λ_{Tk} is the standardized factor loading on the trait factor of the k^{th} item of trait T measured through informant I1.

$\theta_{\delta k}$ is the variance of the random error of the k^{th} item of trait T measured through informant I1.

The term in the numerator captures the variance specific to informant 1. The left term in the denominator captures the variance due to the trait T, the right term in the denominator captures the variance due to the random error. It can be seen that the impact of random measurement error becomes relatively smaller if the number of indicators increases. Hence, for constructs measured through a large number of items, the formula simply indicates the ratio of systematic error variance to the sum of systematic error variance and trait variance.

For our analyses, we again considered the average of systematic error across both informants asked to assess the construct. Hence, results reported in the article are based on the average share of systematic measurement CSys across both informants. For a given construct, this is computed according to the following formula:

$$\text{CSys} = (\text{CSys_I1} + \text{CSys_I2}) / 2 \text{ (E.3).}$$

WEB APPENDIX F:

EXACT MODEL SPECIFICATION FOR HYPOTHESIS TESTING IN STUDY 2

Table 6 in the paper reports the effects of hypothesized antecedents to key informant validity. Again, the data is hierarchical in nature, with constructs being nested in data sets. Therefore, we use multilevel regression with a random intercept (capturing data set-wide influences) and fixed effects of construct characteristics as antecedents to key informant validity. We specified two models, one with the degree of systematic error measured at the item level, the other with the degree of systematic error measured at the construct level. We reproduce the exact model specifications below:

$$ISys_{ij} = \alpha_j + \beta_1 PRESENT_{ij} + \beta_2 OBJECTIVE_{ij} + \beta_3 SALIENCE_{ij} + \beta_4 SPECIFIC_{ij} + \beta_5 INSIDE_{ij}$$

$$CSys_{ij} = \alpha_j + \beta_1 PRESENT_{ij} + \beta_2 OBJECTIVE_{ij} + \beta_3 SALIENCE_{ij} + \beta_4 SPECIFIC_{ij} + \beta_5 INSIDE_{ij}$$

where

$ISys_{ij}$: Average share of systematic error variance in the indicators used to measure construct i from dataset j

$CSys_{ij}$: Share of systematic error variance in the composite of the items used to measure construct i from dataset j

$PRESENT_{ij}$: Degree to which construct i from dataset j refers to the present (versus the past)

$OBJECTIVE_{ij}$: Degree to which construct i from dataset j refers to objectively verifiable referents (versus subjective states and attitudes)

$SALIENCE_{ij}$: Degree to which construct i from dataset j refers to salient events (versus routine processes)

SPECIFIC_{ij}: Degree to which construct i from dataset j refers to specific individuals (versus abstract social entities)

INSIDE_{ij}: Degree to which construct i from dataset j refers to characteristics of the organization (versus characteristics of the environment of the organization)

CR_{ij}: Construct reliability of construct i from dataset j based on a trait-only confirmatory factor analysis considering only construct and its indicators measured through the first informant.

WEB APPENDIX G:

ANALYTICAL APPROACH TO DETERMINING ACCEPTABLE LEVELS OF RANDOM AND SYSTEMATIC MEASUREMENT ERROR IN KEY INFORMANT RESPONSES

A key question in interpreting the results of our studies is whether they indicate key informant responses on average to be reliable and/or valid. However, with the exception of key informant agreement indices, the literature offers little guidance and/or thresholds for what can be considered acceptable levels of key informant reliability and systematic key informant measurement error. Therefore, we conducted some additional analyses reported below.

Our approach is based on the idea that measurement error becomes problematic if it seriously affects the interpretation of a study's substantive results (Doty and Glick 1998), in particular those pertaining to hypothesis testing. We focused on the effect of measurement error on the test of bivariate relationships between constructs using correlation coefficients and the corresponding t-test (Cohen et al. 2003).

We outline our considerations first with regard to key informant reliability and then with regard to key informant validity.

Key informant reliability and its role in significance testing of correlations

The t-test for correlation coefficients is our starting point. If two constructs are assessed using one key informant per firm, the resulting correlation coefficient r_{xy} can be tested using the following test statistic (Cohen et al. 2003):

$$t = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \sqrt{n - 2} \quad (G.1),$$

where n is the number of firms in the sample and t is t-distributed with $(n-2)$ df if H_0 (no relationship) holds.

Based on this test statistic it is possible to determine the lowest correlation coefficient r^* that is still significant given a specific sample size and a critical value C of the t-distribution based on the Type I error probability α .

$$r_{xy}^* = \frac{C}{\sqrt{(n-2) + C^2}} \quad (G.2),$$

We consider lack of key informant reliability to pose a serious problem if it is so small that a true correlation r between two variables is attenuated by random measurement error in a way that it becomes lower than r^* (i.e., if random error prevents us from finding true effects).

Without losing generalizability, we can assume that the true score variance is 1. We again consider the situation where one key informant per firm with key informant reliability Rel is used to assess x and y . Then, the attenuation of a true correlation r_{xy} through random measurement error can be expressed using the following equation.

$$r_{xy}^{obs} = Rel \times r_{xy} \quad (G.3),$$

where r_{xy}^{obs} is the observed correlation.

By equating r_{xy}^{obs} from equation (G.3) and r_{xy}^* from equation (G.2), one can derive the amount of reliability required to find a true effect given sample size n , α , and the size of the true effect. With regard to sample sizes, we consider three values for n (100, 250, 500) that correspond to relatively small, medium, and large samples one can expect to find in typical key informant research. We follow common practice and specified $\alpha = .05$. With regard to the magnitude of the true effects, we refer to Cohen's (1988) f-squared statistic to distinguish between small, medium, and large effects. Thus, a true correlation of .14 corresponds to a small effect, a correlation of .36 corresponds to a medium effect, and a correlation of .51 corresponds to a large effect.

Using these values for n , α , and r_{xy} , we can derive the level of reliability needed to find true effects at various effect size/sample size combinations. They are reported in the table below.

	Key informant reliability required to correctly test for a		
	... true small effect ($r=.14$)	... true medium effect ($r=.36$)	... true large effect ($r=.51$)
Small sample ($n=100$)	-- ^{a)}	.54	.38
Medium sample ($n=250$)	.88	.34	.24
Large sample ($n=500$)	.62	.25	.17

^{a)} At a sample size of $n=100$ it is impossible to correctly identify true small effects, even if reliability is 1.00.

Relating these tabulated values to the average key informant reliability of .612 found in our meta-analysis suggests that it is possible to identify true medium and large effects even in small samples. At the same time, key informant research needs to employ large samples ($n=500$) in order to identify true small effects.

Key informant validity and its role in significance testing of correlations

We employed a similar approach to identify the consequences of systematic measurement error in key informant responses. Our starting point here is equation (2) from the paper that we reproduce here for convenience:

$$r_{xy}^{obs} = \frac{Cov(t_x, t_y) + Cov(s_x, s_y)}{\sqrt{Var(t_x) + Var(s_x) + Var(e_x)} \cdot \sqrt{Var(t_y) + Var(s_y) + Var(e_y)}} \quad (G.4)$$

where t represents the true score component, s the systematic error component and e the random error component of a key informant's responses on x and y .

Based on equation (G.4), two problems of systematic measurement error for significance testing of correlations can be distinguished. (1) If $Cov(t_x, t_y) = 0$, then a systematic covariance between the error components can create artificial effects. (2) If $Cov(t_x, t_y) \neq 0$, then $Cov(s_x, s_y)$

can mask a true effect, if it is opposite in sign to $\text{Cov}(t_x, t_y)$. Below, we analyze at what levels of systematic measurement error these two problems may arise.

Before we turn to these analyses, it is worth noting that we focus on situations where $\text{Cov}(s_x, s_y) = \text{Var}(s)$. This is an assumption in MTMM models using CFA techniques. In these models, all items based on responses by one type of informant load on the same method factor. In other words, the model implies a perfect correlation between systematic error components on all items measured through the same method (i.e., informant). Hence, other than it might appear, $\text{Cov}(s_x, s_y)$ does not refer to correlations between methods but to correlations between systematic errors within one method.

First, we look at the risk of artificial effects caused by systematic measurement error. To simplify the analysis, we assume that $\text{Var}(t_x) = \text{Var}(t_y) = \text{Var}(t)$ and $\text{Var}(e_x) = \text{Var}(e_y) = \text{Var}(e)$. Referring to the paragraph above, it is also apparent that $\text{Var}(s_y) = \text{Var}(s_x) = \text{Var}(s)$ in CTCM models. Based on these assumptions, we can restate equation (G.4) as follows:

$$r_{xy}^{\text{obs}} = \frac{\text{Cov}(t_x, t_y)}{\text{Var}(t) + \text{Var}(s) + \text{Var}(e)} + \frac{\text{Cov}(s_x, s_y)}{\text{Var}(t) + \text{Var}(s) + \text{Var}(e)} \quad (\text{G.5})$$

As $\text{Cov}(t_x, t_y) = 0$ (because we assume no true effect) and $\text{Cov}(s_x, s_y) = \text{Var}(s)$, equation (G.5) can also be expressed as:

$$r_{xy}^{\text{obs}} = 0 + \frac{\text{Var}(s)}{\text{Var}(t) + \text{Var}(s) + \text{Var}(e)} \quad (\text{G.6})$$

Hence, if no true relationship between two variables exists, the observed correlation is equal to the share of systematic error variance Sys inherent in the key informant's responses. Therefore, for the purpose of assessing the percentage of systematic error variance necessary to create an artificial effect, it is simply necessary to solve:

$$r_{xy}^* = \text{Sys} \quad (\text{G.7})$$

Again, the degree of problematic error variance depends on the sample size n and α . We focus on $\alpha = .05$. As a result, in samples with $n=100$, a systematic error share of 19% is sufficient to create artificial correlations that are statistically significant. In samples with $n=250$, a systematic error share of 12% is sufficient, in samples with $n=500$, a share of 8%. Because the average share of systematic error in our datasets is 19%, there is obviously a substantial risk that relying on single key informants may create artificial effects where no true effects are present.

In a second step, we looked at the risk that systematic error masks true effects. This can happen when $\text{Cov}(s_x, s_y)$ has the opposite sign of $\text{Cov}(t_x, t_y)$. To simplify our analysis, we assume that there is no random measurement error. This assumption is particularly appropriate for situations where methods that account for attenuation through random measurement error are used, such as CFA. In this case, equation (G.5) can be restated as:

$$r_{xy}^{\text{obs}} = \frac{\text{Cov}(t_x, t_y)}{\text{Var}(t) + \text{Var}(s)} - \frac{\text{Cov}(s_x, s_y)}{\text{Var}(t) + \text{Var}(s)} \quad (\text{G.8})$$

Again the right fraction is analogous to Sys, i.e., the share of systematic measurement error. The left fraction can also be restated in terms of the share of systematic measurement error. To do so, we additionally assume without loss of generalizability that the true score variance is 1. Then (G.8) becomes:

$$r_{xy}^{\text{obs}} = r_{xy} - (1 + r_{xy})\text{Sys} \quad (\text{G.9})$$

Equating (G.9) with (G.2) allows us to solve for the share of systematic error (opposite to the true effect) necessary to mask an existing true effect. Relevant parameters again include n , α , and the magnitude of the true effect. Based on $\alpha = .05$, we list the resulting combinations of true effect magnitudes and sample sizes in the table below.

	Share of systematic measurement error (opposite to the true effect) needed to mask a....		
	... true small effect ($r=.14$)	... true medium effect ($r=.36$)	... true large effect ($r=.51$)
Small sample ($n=100$)	-- ^{a)}	.12	.21
Medium sample ($n=250$)	.01	.18	.26
Large sample ($n=500$)	.05	.20	.28

^{a)} At a sample size of $n=100$ it is impossible to correctly identify true small effects, even without systematic error.

These figures show that low key informant validity can create problems in finding true small effects if the systematic error is opposite in sign to the true effect. Moreover, the average of 19% (26.4%) systematic measurement error share at the item (construct) level in our data points to an elevated risk that medium-sized effects are masked.

In our analyses pertaining to the effects of systematic measurement error, we assume that one key assumption of the CTCM specification holds, namely that there is no correlation between traits and methods. Below, we briefly consider what happens if this assumption is relaxed. Then, (G.4) needs to be restated as:

$$r_{xy}^{obs} = \frac{\text{Cov}(t_x, t_y) + \text{Cov}(s_x, s_y) + \text{Cov}(t_x, s_x) + \text{Cov}(t_y, s_y)}{\sqrt{\text{Var}(t_x) + \text{Var}(s_x) + 2\text{Cov}(t_x, s_x)} \cdot \sqrt{\text{Var}(t_y) + \text{Var}(s_y) + 2\text{Cov}(t_y, s_y)}} \quad (\text{G.10})$$

Hence, understanding the consequences of systematic measurement error becomes considerably more complex. Three situations can be distinguished:

- $\text{Cov}(t_x, s_x)$ and $\text{Cov}(t_y, s_y)$ have the same sign, which also corresponds to the sign of the true effect. Then, the risk that the systematic error masks a true effect may be reduced.
- $\text{Cov}(t_x, s_x)$ and $\text{Cov}(t_y, s_y)$ have the same sign, which corresponds to the sign of $\text{Cov}(s_x, s_y)$. Then, both risks resulting from systematic measurement error described may be elevated.

- $\text{Cov}(t_x, s_y)$ and $\text{Cov}(t_y, s_y)$ have opposing signs. Then, their effects are likely to cancel each other out, particularly if both covariances have about the same magnitude.

REFERENCES USED IN WEB APPENDICES

- Bagozzi, Richard P., Youjae Yi and Lynn W. Phillips (1991), "Assessing construct validity in organizational research," *Administrative Science Quarterly*, 36 (3), 421-58.
- Biggadike, E. Ralph (1979), *Corporate diversification: entry, strategy, and performance*. Boston, MA: Harvard University.
- Brown, Timothy A. (2006), *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Cohen, Jacob (1988), *Statistical power analysis for the behavioral science*. Hillsdale, NJ: Erlbaum Associates.
- Cohen, Jacob, Cohen, Patricia, West, Stephen G., Aiken, Leona S. (2003), *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum Associates.
- Doty, D. Harold and William H. Glick (1998), "Common methods bias: does common methods variance really bias results," *Organizational Research Methods*, 1 (4), 374-406.
- Cote, Joseph A. and M. Ronald Buckley (1987), "Estimating trait, method, and error variance: generalizing across 70 construct validation studies," *Journal of Marketing Research*, 24 (3), 315-18.
- Kumar, Nirmalya, Louis W. Stern, and James C. Anderson (1993), "Conducting interorganizational research using key informants," *Academy of Management Journal*, 36 (6), 1633-51.
- Lance, Charles E., Carrie L. Noble, and Steven E. Scullen (2002), "A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data," *Psychological Methods*, 7 (2), 228-44.
- , Bryan Dawson, David Birkelbach, and Brian J. Hoffman (2010), "Method effects, measurement error, and substantive conclusions," *Organizational Research Methods*, 13 (3), 435-55.
- Lee, Chang-Yang and Ishtiaq P. Mahmood (2009), "Inter-industry differences in profitability: the legacy of the structure-efficiency debate revisited," *Industrial and Corporate Change*, 18 (3), 351-80.
- Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003), "Common method biases in behavioral research: a critical review of the literature and recommended remedies," *Journal of Applied Psychology*, 88 (5), 879-903.
- Sutcliffe, Kathleen M. (1994), "What executives notice: accurate perceptions in top management teams," *Academy of Management Journal*, 37 (5), 1360-78.

Copyright of Journal of Marketing Research (JMR) is the property of American Marketing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.