Journal of Marketing Theory and Practice

ISSN: 1069-6679 (Print) 1944-7175 (Online) Journal homepage: http://www.tandfonline.com/loi/mmtp20

Multilevel Analysis in Marketing Research: Differentiating Analytical Outcomes

Jan Wieseke , Nick Lee , Amanda J. Broderick , Jeremy F. Dawson & Rolf Van Dick

MULTILEVEL ANALYSES IN MARKETING RESEARCH: DIFFERENTIATING ANALYTICAL OUTCOMES

Jan Wieseke, Nick Lee, Amanda J. Broderick, Jeremy F. Dawson, and Rolf Van Dick

Marketing scholars are increasingly recognizing the importance of investigating phenomena at multiple levels. However, the analyses methods that are currently dominant within marketing may not be appropriate to dealing with multilevel or nested data structures. We identify the state of contemporary multilevel marketing research, finding that typical empirical approaches within marketing research may be less effective at explicitly taking account of multilevel data structures than those in other organizational disciplines. A Monte Carlo simulation, based on results from a previously published marketing study, demonstrates that different approaches to analysis of the same data can result in very different results (both in terms of power and effect size). The implication is that marketing scholars should be cautious when analyzing multilevel or other grouped data, and we provide a discussion and introduction to the use of hierarchical linear modeling for this purpose.

Issues concerning levels of analysis permeate organizational and, by association, marketing research (Klein, Dansereau, and Hall 1994). In theoretical terms, the level of analysis in a given study refers to the object of interest. For example, is one trying to explain individual, group, or organizational performance? Levels issues also extend to *how* one is trying to explain that object of interest. For example, is individual-level performance explained by an organizational-level variable such as corporate culture, or are there any effects of group membership on individual performance? Consideration of levels issues is also vital in data analysis terms, and organizational research has recently seen a profusion of studies incorporating multilevel analysis (also called hierarchical linear modeling) techniques designed to take account of such issues (e.g., Bliese 2000; Bliese, Chan, and Ployhart 2007; Kozlowski and Klein 2000).

However, even though multilevel approaches are well known and accepted in disciplines such as organizational psychology, it will be shown subsequently that they do not currently occupy a similar position in marketing research, despite some isolated uses of the techniques (e.g., Jong, de Ruyter, and Lemmink 2004; Pieters and Wedel 2004). In particular, the discipline lacks an integrative overview and introduction to multilevel analysis that also illuminates key reasons to use multilevel analysis, the likes of which appeared in the marketing literature in the 1990s regarding structural equation modeling (SEM) (e.g., Baumgartner and Homburg 1996; Steenkamp and van Trijp 1991). As a result, similar to the situation of SEM prior to the 1990s, despite the appearance of a number of articles in the specialist, methodological, and nonmarketing literatures, many marketing researchers may be unfamiliar with the use of key multilevel analysis methods and their potential benefits. Nevertheless, even though multilevel analysis is uncommon within mainstream marketing scholarship, it is clear that central marketing questions directly involve theories that concern different levels of analysis. For example, in organizational marketing research, the level of theoretical interest may be at the employee, unit/team, or firm level. In consumer marketing research, individual consumer, product

Jan Wieseke (Ph.D., Philipps University Marburg, Germany), Assistant Professor of Marketing, Department of Business Administration and Marketing I, University of Mannheim, Germany, jwieseke@bwl.uni-mannheim.de.

Nick Lee (Ph.D., Aston Business School), Senior Lecturer and Marketing Research Group Convenor, Aston Business School, Aston University, Birmingham, UK, n.j.lee@aston.ac.uk.

Amanda J. Broderick (Ph.D., DeMontford University), Professor of Marketing and Head of Marketing and Advertising Applied Research Group, Coventry University Business School, Coventry, UK, amanda.broderick@coventry.ac.uk.

Jeremy F. Dawson (M.Sc., Sheffield University), RCUK Research Fellow, Aston Business School, Aston University, Birmingham, UK, j.f.dawson@aston.ac.uk.

Rolf Van Dick (Ph.D., Philipps University Marburg, Germany), Professor of Social Psychology, Institute of Psychology, Goethe University Frankfurt, Germany, and Chair of Social Psychology and Organizational Behavior (part time), Aston Business School, Aston University, Birmingham, UK, van.dick@psych.uni-frankfurt.de.

Journal of Marketing Theory and Practice, vol. 16, no. 4 (fall 2008), pp. 321–339.

ISSN 1069-6679 / 2008 $9.50 + 0.00.
DOI 10.2753/MTP1069-6679160405

category, intra- or international sociobehavioral groups are commonly considered.

Most empirical studies in marketing have to date focused exclusively on analyzing data collected at a single level of analysis, often due to multiple levels of data being unavailable or inaccessible. However, although it is the case that collecting empirical data with regard to levels issues may be more demanding, it is also clear that using only a single level of analysis may (although not always) inadequately account for many marketing research issues (Liao and Chuang 2004; van Bruggen, Lilien, and Kacker 2002). In response, scholars have recently recognized the importance of investigating and drawing conclusions regarding the influence of phenomena at different levels of analysis, and increasing numbers of studies are focusing on two or more levels (e.g., Homburg and Stock 2004; Jong, de Ruyter, and Lemmink 2004).

To analyze multiple levels of data simultaneously, or otherwise take levels issues into account analytically, one needs to consider that in most cases, data sets are of a hierarchical, or nested, nature. For example, employees are nested in work groups or under managers, those groups in turn are nested in functions, which are nested in organizations, and so on. In such samples, the data points of multiple individuals are usually not independent. For example, employees in the same organizational unit with the same manager tend to be more similar to each other than they are to employees in different units, because of factors such as selection processes, the leadership style of the manager, and the common history they share. Thus, the intraclass correlation between variables measured among employees from the same organizational unit or team will be higher than the average correlation between variables measured on employees from different organizational units (Hox 1995; Raudenbush and Bryk 2002). This may cause serious difficulties in data analysis, as standard statistical tests assume that the observations are independent. "If this assumption is violated (and in multilevel data this is usually the case) the estimates of the standard errors of conventional statistical tests are much too small, and this results in many spuriously 'significant' results" (Hox 2002, p. 5).

Furthermore, issues concerning the hierarchical nature of organizational data are relevant even if the researcher has no interest in modeling multilevel relationships. More specifically, group membership is likely to have an impact on many key criterion variables (see Bliese and Hanges 2004). For example, variables such as performance, satisfaction, commitment, role conflict, and many others are likely to depend in some way on the particular work group or organization (etc.) that subjects are nested in. In such situations, where group membership may affect the constructs of interest, ignoring the nonindependence of data points when multiple respondents are collected from a given unit (say, collecting multiple members of a single organization or team) can potentially lead to misleading results in the same way as described above for the multilevel context.

Although a limited amount of empirical research in marketing has appeared that takes such multilevel and hierarchical data issues into account (Pieters and Wedel 2004; van Dolen et al. 2002), this is not yet common. Nevertheless, applying standard statistical approaches to the analysis of multilevel, or naturally grouped, data risks misinterpretations due to the inherent nature of multilevel data sets detailed above (Osborne 2000). Furthermore, with increasing calls for the aggregation of multiple responses from groups or organizations (e.g., van Bruggen, Lilien, and Kacker 2002), it is important to address the potential nonindependence problems that may result from the use of multiple responses from groups or organizations.

In this paper, we address how studies involving multiple levels have been approached in marketing as compared with more general organizational research, identify appropriate methods for the analysis of cross-level effects, and provide some information regarding the differential outcomes of competing analytical approaches. However, although we focus on multilevel research, our findings and assertions are just as relevant to researchers who examine single-level theories where respondents' group membership is likely to influence the dependent variables—that is, where the nonindependence of multiple respondents from the same group is an issue (see Bliese and Hanges 2004). We employ a Monte Carlo simulation, inspired by a classic marketing study, to compare the outcomes of utilizing three alternative analytical approaches to nested or hierarchical/multilevel data. Finally, we introduce alternative multilevel approaches and outline key issues and applications of these analytical techniques.

PREVAILING EMPIRICAL APPROACHES TO MULTILEVEL THEORY TESTING IN MARKETING

To analyze relations between constructs of different aggregation levels, there has been an increased use of multilevel research settings within social science, especially in the disciplines of education and medicine (e.g., Goldstein, Browne, and Rashbash 2002; Leeuw and Kreft 1986). This trend is also reflected in the management and, to some extent, marketing science literature. In order to gain an appreciation of the different methods that have been employed in marketing research to analyze multilevel research questions, we performed an analysis of marketing publications in the recent past. Following meta-analytic procedures, a census of all empirical articles between 1998 and 2006 in the *Journal of Marketing Research, Journal of Marketing, International Journal of Research in Marketing, Journal of Retailing,* and *Academy of Management Journal* that addressed topics of a multilevel nature was undertaken. The selection of these journals was based primarily on the marketing journal rankings of Baumgartner and Pieters (2003) and Hult, Neese, and Bashaw (1997). Furthermore, a key criterion was to include journals that were primarily general in their coverage, and specifically not solely focused on consumer research. Even though consumer research can be conducted within a hierarchical class structure (e.g., Broderick, Greenley, and Mueller 2007; Macintosh and Lockshin 1997), and consumers can often be considered to be nested within large-scale entities such as regions and countries (MacKenzie 2001), the predominant research focus in, for example, the *Journal of Consumer Research,* is not to predict or explain micro-level issues with macro-level variables. In view of this, we considered it necessary to choose a feasible number of journals that would take in the broadest selection of relevant marketing issues, as well as capture the research that was generally rated as being of the highest quality by academics. Although it is certainly the case that other journals could have been selected,[1] it is considered that the present set strikes a good balance.

The *Academy of Management Journal* was also used to compare the multilevel research of marketing science with that typical of management science. The *Academy of Management Journal* was considered to be the most appropriate single journal to compare with the marketing literature for a number of reasons. In particular, the *Academy of Management Journal* is an outlet focusing on theory testing, which does not allow purely conceptual work to be submitted. This is not the case in other top management journals such as *Administrative Science Quarterly* or the *Academy of Management Review.* It would not have been an illustrative comparison to have included instead journals such as the latter, where the editorial policy was considerably different to the marketing journals that were selected.

In total, 190 empirical studies with multilevel research questions that were tested empirically were identified, with 119 appearing in the four marketing journals, and 71 appearing in the *Academy of Management Journal.* Figure 1, summarizing the distribution of articles by journals, clearly indicates the increasing importance of multilevel studies in the marketing and management science literature since 1998.

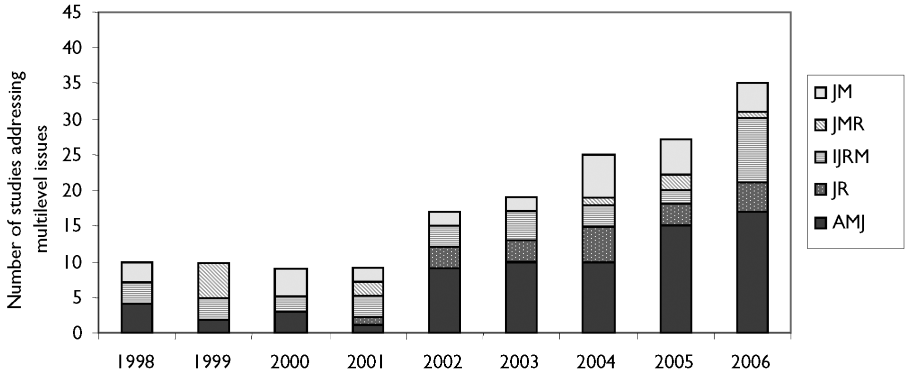
Multilevel studies in marketing and management science can be divided into two dominant groups. One group deals with intra-organizational levels and one focuses on the interaction between intra- and extra-organizational levels. In the first group, levels of employees and managers are investigated most often (e.g., Bettencourt 2004; Piercy, Cravens, and Lane 2003). Investigations of the influence of superordinate managers' variables on subordinate employees' variables are dominant here. In the second group, business-to-business dyads (e.g., Ping 2003; Wathne and Heide 2004) are analyzed predominantly, followed by employee–customer relationships (e.g., Tax, Brown, and Chandrashekaran 1998; van Dolen et al. 2002). To date, three-level studies, such as focusing on the influence of managers on employees as well as on the employees' impact on customers simultaneously, are very rare (for exceptions, see Bell and Menguc 2002; Liao and Chuang 2004).

These multilevel approaches usually reflect a hierarchical structure of nested entities. That is, for example, when superordinates and their employees are of interest, data are typically organized as detailed in Figure 2. A number of entities on level n are related to a single entity on level $n + 1$. Therefore, persons A_1–A_4 are all exposed to context A and B_1–B_4 are all exposed to context B (Bryk and Raudenbush 1992; Hofmann 1997).

In general, six main methodological approaches to multilevel theory issues can be found in the literature. The most commonly used methodological approach in multilevel research endeavors is to survey members of one level about their perceptions of variables that are relevant at two or more levels (e.g., Klein and Kim 1998; Lankau and Scandura 2002; Liu and Leach 2001; Piercy, Cravens, and Lane 2003; Ragins, Cotton, and Miller 2000; Ramus and Steger 2000; Smidts, Pruyn, and van Riel 2001). For example, employees are asked to indicate their own work satisfaction and organizational commitment as well as the level of support provided by their superordinate managers (Speier and Venkatesh 2002). This approach incorporates a risk of common method bias (Bell and Menguc 2002; Netemeyer et al. 1997). It also raises the issue of informant bias, which could result from the specific hierarchical position of the surveyed subjects. To control these difficulties, separate data collection on each level of interest is commonly recommended (Bryk and Raudenbush 1992; Kidwell, Mossholder, and Bennett 1997).

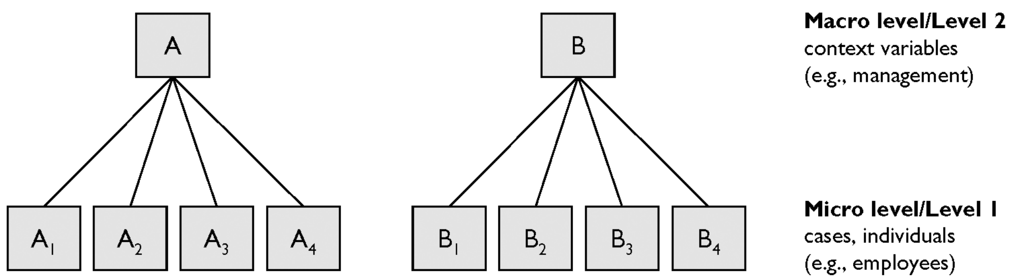
If data of two or more levels have been collected, one option is to *aggregate the data sets* (usually on the micro

Figure 1
Empirical Multilevel Studies in Marketing and Management Journals



Notes: JM = Journal of Marketing; JMR = Journal of Marketing Research; IJRM = International Journal of Research in Marketing; JR = Journal of Retailing; AMJ = Academy of Management Journal.

Figure 2
A Hierarchical Structure of Nested Entities



Note: Hierarchical linked data sets in a two-level design.

level; e.g., Hartline, Maxham, and McKee 2000; Sarin and Mahajan 2001), so that, for example, each superordinate is linked with an average score of his or her subordinates. The disadvantage of this approach lies in the loss of information, because possible meaningful variances at the micro level are ignored (Hofmann 1997). If these variances are substantive, misleading results can be returned. In the studies reviewed above, only a small minority of studies tested the assumption of a sufficient correspondence among the ratings within the micro level, by, for example, using intraclass correlation coefficients (ICCs) or within- and between-group analysis (WABA), before aggregating the data (e.g., Lam, Chen, and Schaubroeck 2002). Another disadvantage of data aggregation can be seen in the often-drastic decline of sample size and, therefore, statistical power.

Also commonly found is the "key informant" approach, where dyads with one subject on each level are taken into account (e.g., Bettencourt 2004; DeCarlo, Rody, and DeCarlo 1999; Jap 1999; Lam, Chen, and Schaubroeck 2002; Madjar, Oldham, and Pratt 2002; Moorman, Blakely, and Niehoff 1998; Siguaw, Simpson, and Baker 1998; Tepper and Taylor 2003). This approach does not capture information on the other subjects at the micro level and can be criticized because the risk of increased correlations between systematic measurement errors is enhanced. Therefore, obtaining data from multiple informants has been recommended as superior to such an approach (Liu and Leach 2001; van Bruggen, Lilien, and Kacker 2002).

A fourth option to match data in multilevel designs is the disaggregation of the data sets (usually on the macro level). In this approach, each unit on the lower level (e.g., employee) is allocated to a score at the macro level (e.g., business unit) within which it is nested. Statistical analyses in this case are based on the sample of the lower level (e.g., Dellande, Gilly, and Graham 2004; McAllister and Bigley 2002; Saparito, Chen, and Sapienza, 2004). Problematic in this approach is the violation of the assumption of independent observation, which is a central assumption of most classical statistical procedures (Bryk and Raudenbush 1992). Only in the unlikely case of total independence of the individual answering tendencies to the characteristic of the higher level would this approach be reasonable. Otherwise a distortion of standard error estimates occurs and results in an increase of Type I error (Kidwell, Mossholder, and Bennett 1997). A further problem of disaggregation lies in the fact that variables concerning the higher level are analyzed on the basis of the larger lower-level sample size. Again, this can affect the estimation of standard errors and the statistical conclusions (Hofmann 1997). In light of these risks, it is unsurprising that few of the observed studies use the disaggregation procedure.

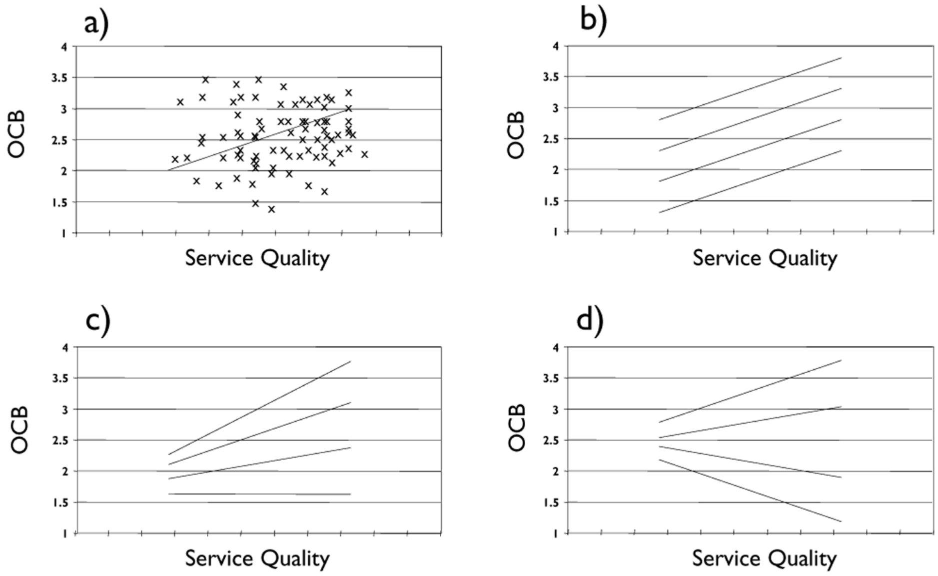
The fifth approach to handle multilevel data is the use of *hierarchical linear models* (HLM), also called *random coefficient* models (Leeuw and Kreft 1986; Longford 1993). This method was developed to overcome the aforementioned difficulties in the analysis of multilevel data (Hox 1995). HLM explicitly take into account the nesting of micro- and macro-level phenomena (Kozlowski and Klein 2000). They explicitly recognize that individuals *within* a particular group may be more similar to one another than to individuals in other groups and, therefore, may not provide independent observations (Hofmann 1997). They also account for macro-level effects that occur through the interactions with micro-level elements (Kozlowski and Klein 2000). The major advantage of the HLM is the possibility to link multiple levels simultaneously in a single regression equation (Goldstein 1995). Nevertheless, it is important to note that HLM is not always the method of choice for testing multilevel theoretical models, and most of the methods above can be appropriate in certain specific situations.

In the surveyed journals, it can be seen that there was a significantly higher proportion of the *Academy of Management Journal* studies that made use of this method than the few that employed it in the marketing journals (Jong, de Ruyter, and Lemmink 2004; Pieters and Wedel 2004; van Dolen et al. 2002). In fact, over the eight-year period, only seven studies in the four marketing journals surveyed employed HLM methods (6 percent), whereas in the *Academy of Management Journal,* this number is more than double (14, or 20 percent). Indeed, in 2006 alone, there are six articles in the *Academy of Management Journal* that use HLM methods—a proportion of 35 percent (six from 17 articles), suggesting a significant increasing trend over time, which is not evident in the four marketing journals. However, this is not intended to imply that management science is somehow "superior" to marketing science—or that management researchers always analyze their data in the most appropriate manner—but merely to observe that HLM methods appear to be more common in management than marketing research at the present time.

In addition to methodological issues, different conceptual patterns of relationships may be evident in multilevel research designs that may be unable to be captured without a hierarchical linear modeling approach. Consider the following example detailed in Figure 3. Figure 3 illustrates a hypothetical example of the relationship between organizational citizenship behavior (OCB) (i.e., showing extra effort or going the "extra mile," cf. Organ 1997) of salespeople (y-axis) and service quality as perceived by customers across retail stores. Assume that 150 sales employees have responded to an OCB scale and, from each of them, 10 customer surveys on their perceived service quality are available. The regression line in Figure 3a describes this relationship for the whole sample, whereas Figures 3b–d provide separate regression lines for different retail stores. Individual employees' data points are not included for simplification purposes. Figure 3a shows a positive linear relationship between employees' OCB and perceived service quality, indicating that the more an employee engages in OCB, the better the perceived service quality by the customers. Around the regression line, the variation of data points of the observed individuals can be seen. Figure 3b shows the same relationship between salespersons' OCB and customers' service quality evaluations for different stores. Differences occur with respect to the intercepts across stores. In Figure 3c, an interaction between stores and the relation between OCB and perceived service quality can be seen. Here the relationship is the strongest in stores in which employees express high OCB. This fanlike effect is also expressed by the significant differences between stores when high service quality scores are focused. Another example for a cross-level interaction is illustrated in Figure 3d. Here, in stores with high OCB scores, there is a positive relationship between OCB and service quality, whereas in stores with low OCB scores, a negative correlation can be found. By analyzing the sample as a whole in Figure 3, one might draw differing conclusions from those drawn from a multilevel approach. For example, in the Figure 3d regression slope, the whole sample is zero, but the single stores can be either positive or negative.

Although we have shown that analytical approaches designed to address multilevel data sets are comparatively more common in management science, the marketing literature has preferred alternative approaches. MacKenzie summarizes that in marketing, "researchers have tended to emphasize either a micro- or macro-level perspective without recogniz-

Figure 3
Possible Relations of Micro-Level Variables when Macro-Level Variance Is Considered



ing the interaction between the two" (2001, p. 164). This focus is most likely to be due to unfamiliarity with either (1) the hierarchical linear modeling approach, or (2) the differences in results that may result from varying analytical techniques. To this end, the following section details a simulation study designed to provide a direct comparison of the differential outcomes of competing analytical approaches, in a marketing-relevant context.

A COMPARISON OF MULTILEVEL ANALYTICAL APPROACHES

Method

Our study was designed to demonstrate the potential differences between alternative modeling approaches under varying conditions of nonindependence, within a set of contexts likely to be both commonly faced, and familiar to, marketing researchers. In order to do so, we based our simulated data on a selection of the results reported in Atuahene-Gima and Li (2002).[2] In order to select the Atuahene-Gima and Li paper, we explored existing literature in top marketing journals, which analyzed research questions where the answers clearly depended on assumptions regarding the nonindependence of cases at the micro level. Such questions should concern respondents who were nested into groups, and where concepts in the research were at multiple levels. A final theoretical criterion was that it should be conceptually sound to expect that the concepts in the study are likely to be influenced somehow by between-group variation associated with macro-level factors. Methodologically speaking, as regression-based analysis methods are by far the most commonly published in the marketing literature, we also looked to choose a paper that utilized such a methodology.

The Atuahene-Gima and Li (2002) paper uses two samples of data, one from China and one from the United States. For our purposes, we focus on the Chinese data set. Due to the inherent difficulty in collecting data in China, Atuahene-Gima and Li were forced to use multiple respondents from organizations. Specifically, three members of each of 150 selected firms participated in the study. Two hundred fifteen completed questionnaires were returned. However, there is no information given on how many of the 150 firms were represented, or the average number of employees per firm who returned questionnaires. Thus, the data are intrinsically structured on at least two levels—employees are naturally grouped into firms. After accounting for missing data, the sample was reduced to 157. Although no specific information is given, it must be assumed that Atuahene-Gima and Li treated multiple respondents from the same firm as independent observations. To test their hypotheses, Atuahene-Gima and Li chose to use a regression approach across each sample.

For the purposes of our analysis, we selected four individual hypotheses that together exemplify the main types of multilevel hypotheses likely to be encountered by marketing researchers. First, we chose to replicate H1a, which argued that increased output control used by the manager should be negatively related to supervisee trust, and H2a, which stated that increased process control should be related positively to supervisee trust. Both of these hypotheses link manager-level variables to employee-level outcomes, but are tested using solely employee-level data. Group-level effects are likely to be inherent to such a methodology, and therefore individuals from the same group should not necessarily be treated as independent. Second, we examined H5a, that role ambiguity is negatively related to supervisee trust, and H6a, that supervisee trust is related positively to sales performance. Both of these latter hypotheses relate two employee-level variables to each other. However, there are a number of group-level factors likely to affect the variance of the constructs (e.g., manager or firm factors), and thus again it is not clear that cases from the same firm should be treated as independent. These four hypotheses encompass different types of multilevel hypotheses that may be commonly considered by marketing researchers. In testing these hypotheses, Atuahene-Gima and Li uncovered a range of effect sizes. We simulated our data using the correlations[3] between the measures reported by Atuahene-Gima and Li (but regression coefficients are also reported, it is not possible to simulate our data using these). Table 1 presents the relevant correlations for our four hypotheses.

An important characteristic of multilevel research is the level of nonindependence of variables between higher-level units. In this case, it is not known how similar responses of individuals from the same organization are to each other, compared with responses of individuals from other organizations. If there is little systematic difference in responses between employees from different units (low nonindependence), then it may be expected that individual-level analysis (as Atuahene-Gima and Li appear to have conducted) would be acceptable. If there are high levels of agreement in responses from employees from the same organization (high nonindependence), then aggregating the data to the unit level before analysis may be a suitable method (although this would reduce sample size and power). However, if there are moderate levels of nonindependence—some systematic differences between units, but still substantial variation within organizations—neither of these methods is appropriate, and hierarchical linear modeling may be the best method. The amount of members in each group is also likely to have an impact on the results. Specifically, both to illustrate possible marketing research situations and to show the influence of group size in general, we systematically varied the group size from three (Atuahene-Gima and Li's stated number of sampled cases per organization) to 10 and 20.

We simulated data with dependent variables in each of these $3 \times 3 \times 4$ conditions. Based on the figures quoted by Bliese (2000), we used for low nonindependence an intraclass correlation coefficient (ICC(1)) of 0.01, moderate nonindependence of 0.12, and high nonindependence of 0.30. For each of these conditions, we used S-Plus (www.insightful.com) to generate data sets with the correlations in Table 1, a sample size of 80 groups (which corresponds to an estimate of how many firms may have been represented in Atuahene-Gima and Li's sample assuming a moderate to high response rate from organizations), and with the varying group sizes. We then tested the four hypotheses on each data set using three alternative methods—ordinary regression analysis at the individual level, regression analysis on data aggregated to the organizational level, and hierarchical linear modeling.

Table 1
Intercorrelations of Variables Quoted by Atuahene-Gima and Li (2002)

	1	2	3	4
1. Output Control				
2. Process Control	0.40			
3. Supervisee Trust	0.29[1]	0.48[2]		
4. Role Ambiguity	–0.11	0.08	–0.01[3]	
5. Sales Performance	0.22	0.47	0.37[4]	–0.02

Notes: [1] The association examined in H1a. [2] The association examined in H2a. [3] The association examined in H5a. [4] The association examined in H6a.

Results and Discussion

For each hypothesis, each level of nonindependence, each group size, and each method of analysis, the proportion of times the hypothesis was found to be supported (with $p < 0.05$) was recorded (the observed power), along with the standardized regression coefficient (beta). Summaries of these for each condition are shown in Table 2.

The first factor that is evident from the results in Table 2 is that there are clear differences between the results returned using aggregate-level regression and the two other analysis methods. In every condition (team size and ICC) the aggregate-level effect sizes were systematically larger than those obtained using individual-level regression or HLM. What is particularly interesting is that where it differed at all, the observed power for these large effect sizes was, in general, lower than that reported by the other two methods. Even though it is not surprising that aggregate-level results should be of lower power (because the sample size is, by definition, smaller), it is interesting that it should be lower even though the effect sizes were so much larger

Table 2
Results of Simulation Studies

	Individual-Level Regression		Aggregate-Level Regression		Hierarchical Linear Modeling	
	Percent Significant	Average Effect Size (Standard Deviation)	Percent Significant	Average Effect Size (Standard Deviation)	Percent Significant	Average Effect Size (Standard Deviation)
Team Size = 3						
Effect Size = –0.01						
ICC(1) = 0.01	5.7	–0.015 (0.065)	5.9	–0.010 (0.114)	5.2	–0.012 (0.066)
ICC(1) = 0.12	5.0	–0.011 (0.065)	4.9	–0.009 (0.124)	5.5	–0.012 (0.066)
ICC(1) = 0.30	5.0	–0.007 (0.065)	5.6	–0.011 (0.149)	6.0	–0.009 (0.061)
Effect Size = 0.37						
ICC(1) = 0.01	100.0	0.369 (0.058)	93.3	0.372 (0.103)	100.0	0.370 (0.057)
ICC(1) = 0.12	100.0	0.368 (0.057)	96.4	0.440 (0.109)	100.0	0.363 (0.054)
ICC(1) = 0.30	100.0	0.367 (0.054)	98.5	0.552 (0.115)	100.0	0.328 (0.056)
Effect Size = 0.29						
ICC(1) = 0.01	99.5	0.291 (0.060)	75.2	0.289 (0.107)	99.5	0.288 (0.059)
ICC(1) = 0.12	99.7	0.290 (0.058)	83.3	0.355 (0.113)	99.4	0.283 (0.059)
ICC(1) = 0.30	99.5	0.290 (0.059)	92.8	0.446 (0.120)	99.4	0.255 (0.057)
Effect Size = 0.48						
ICC(1) = 0.01	100.0	0.481 (0.051)	99.8	0.489 (0.095)	100.0	0.479 (0.050)
ICC(1) = 0.12	100.0	0.480 (0.050)	100.0	0.564 (0.096)	100.0	0.468 (0.052)
ICC(1) = 0.30	100.0	0.478 (0.050)	100.0	0.675 (0.098)	100.0	0.435 (0.052)
Team Size = 10						
Effect Size = –0.01						
ICC(1) = 0.01	7.1	–0.011 (0.037)	5.1	–0.011 (0.121)	6.6	–0.009 (0.036)
ICC(1) = 0.12	6.1	–0.010 (0.035)	5.0	–0.020 (0.163)	6.7	–0.009 (0.035)
ICC(1) = 0.30	5.8	–0.009 (0.036)	5.7	–0.036 (0.219)	5.7	–0.010 (0.030)
Effect Size = 0.37						
ICC(1) = 0.01	100.0	0.371 (0.029)	94.7	0.397 (0.106)	100.0	0.370 (0.030)
ICC(1) = 0.12	100.0	0.369 (0.030)	99.9	0.671 (0.128)	100.0	0.349 (0.032)
ICC(1) = 0.30	100.0	0.368 (0.031)	100.0	1.005 (0.132)	100.0	0.293 (0.030)

Effect Size = 0.29									
ICC(1) = 0.01	100.0	0.289	(0.033)	78.1	0.314	(0.111)	100.0	0.289	(0.033)
ICC(1) = 0.12	100.0	0.289	(0.033)	96.3	0.554	(0.142)	100.0	0.273	(0.031)
ICC(1) = 0.30	100.0	0.291	(0.033)	99.9	0.883	(0.155)	100.0	0.227	(0.029)
Effect Size = 0.48									
ICC(1) = 0.01	100.0	0.480	(0.027)	99.9	0.514	(0.097)	100.0	0.480	(0.028)
ICC(1) = 0.12	100.0	0.479	(0.028)	100.0	0.806	(0.109)	100.0	0.457	(0.027)
ICC(1) = 0.30	100.0	0.477	(0.027)	100.0	1.095	(0.099)	100.0	0.394	(0.029)
Team Size = 20									
Effect Size = –0.01									
ICC(1) = 0.01	6.0	–0.010	(0.025)	3.5	–0.011	(0.120)	6.0	–0.008	(0.025)
ICC(1) = 0.12	7.3	–0.010	(0.025)	4.8	–0.046	(0.207)	6.7	–0.010	(0.024)
ICC(1) = 0.30	8.0	–0.010	(0.026)	5.3	–0.060	(0.291)	5.7	–0.008	(0.021)
Effect Size = 0.37									
ICC(1) = 0.01	100.0	0.370	(0.022)	96.2	0.425	(0.111)	100.0	0.369	(0.021)
ICC(1) = 0.12	100.0	0.370	(0.022)	100.0	0.926	(0.140)	100.0	0.343	(0.021)
ICC(1) = 0.30	100.0	0.369	(0.022)	100.0	1.391	(0.130)	100.0	0.284	(0.021)
Effect Size = 0.29									
ICC(1) = 0.01	100.0	0.290	(0.023)	81.9	0.339	(0.112)	100.0	0.291	(0.023)
ICC(1) = 0.12	100.0	0.291	(0.024)	99.9	0.793	(0.154)	100.0	0.267	(0.022)
ICC(1) = 0.30	100.0	0.290	(0.024)	100.0	1.322	(0.175)	100.0	0.218	(0.022)
Effect Size = 0.48									
ICC(1) = 0.01	100.0	0.479	(0.020)	99.8	0.547	(0.105)	100.0	0.480	(0.018)
ICC(1) = 0.12	100.0	0.480	(0.020)	100.0	1.030	(0.111)	100.0	0.450	(0.020)
ICC(1) = 0.30	100.0	0.480	(0.021)	100.0	1.390	(0.096)	100.0	0.378	(0.020)

in most cases. Also, in a number of conditions (especially large team sizes and higher nonindependence), the effect sizes returned for aggregate-level regression were greater than one, and thus improper.

Comparing individual-level regression with HLM, one can see that the effect sizes for the HLM analyses were systematically lower in all conditions except where low nonindependence was specified (i.e., ICC(1) of 0.01). Furthermore, this effect consistently increased as the level of nonindependence increased. This is, of course, not surprising, as Bliese and Hanges (2004) show that this is likely to be exactly the case. As team size increased, the differences also appeared to generally get more substantial.

Note that we are unable to say which is the optimal approach simply on the basis of these results. In specific terms, we can suggest that the results reported by Atuanhene-Gima and Li (2002) would have almost certainly been different depending on the analysis method that was used. That said, it is impossible to say which was the "correct" method to use in that situation, because we have no information on the nonindependence of the observations, or the number of cases sampled per organization. Nevertheless, the prevalence of what could be called improper effect sizes (greater than one) for the aggregate-level results do suggest that aggregation is a potentially risky strategy for researchers, and may result in inflated effect sizes even while power is reduced. What is undeniable is that the results found—both in terms of effect sizes and significance—can depend on the type of analysis used for different levels of nonindependence. Therefore, it is essential that researchers consider what the most appropriate method of analysis for their data is, otherwise misleading results can occur. The information given in other sections of this paper should provide considerable help to researchers in making the correct decision. In particular, the superior effect sizes reported in the individual-level results should be balanced against theory, which suggests that nonindependence of observations may, in fact, have inflated these (e.g., Bliese and Hanges 2004), implying that HLM may be a superior alternative in most situations where nested data are employed.

USING HIERARCHICAL LINEAR MODELS TO ANALYZE CROSS-LEVEL EFFECTS

The results of the simulation above have demonstrated the differential outcomes of individual-level analysis, data aggregation, and hierarchical linear modeling on a naturally nested data set. However, this data set contained data at a single level, and hierarchical linear modeling was used to show how the nonindependence of nested data points can influence analysis results. A further advantage of the multilevel method of HLM is the possibility of linking several levels in one regression equation, which takes into account that β_0 and β_1 may vary in different contexts or subgroups. The membership of individuals in a certain group or context is indicated by the index *j*. In general, multilevel regression models assume a hierarchical linkage of the data, where a measured criterion variable on the micro-level is explained by predictor variables on both micro- and macro-level Y_{ij} (Hox 1998). Therefore, the following basic equation for multilevel regression models with two levels is (Goldstein 1995; Hox 2002):

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}. \tag{1}$$

Different regression equations for each group or context *j* of level 2 (macro level) are estimated. The β_j weights are modeled through the predictor variables Z_j of the macro level:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + \mu_{0j} \tag{2}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + \mu_{1j}. \tag{3}$$

Here μ indicates the residual values on level 2. By replacing Equation (1) with Equations (2) and (3), the following equation results:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + \mu_{1j} X_{ij} + \mu_{0j} + e_{ij}. \tag{4}$$

Here the equation part $\gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij}$ contains the fixed coefficient and is therefore called the *fixed part.* The remaining part, $\mu_{1j} X_{ij} + \mu_{0j} + e_{ij}$, is labeled as *randomized part.* The term $Z_j X_{ij}$ indicates interactions between predictor variables, which lead to variations of the regression coefficients β_{1j} of the predictor variables X_{ij} (Hox 1995). Most studies include more than one predictor variable on each level. With *P* predictor variables on the micro level and *Q* predictor variables on the macro level, the following general equation results:

$$y_{ij} = \gamma_{00} + \gamma_{p0} X_{pij} + \gamma_{0q} Z_{qj} + \gamma_{pq} Z_{qj} X_{pij} + \mu_{pj} X_{pij} + \mu_{0j} + e_{ij}. \tag{5}$$

Steps in Multilevel Modeling

The classical procedure to examine multilevel models consists of five steps (Hox 1995). These are analogous to a hierarchical linear regression, which examines the effects of macro-level variables when controlled for micro-level predictors. First, an intercept-only model is calculated, which consists solely of a constant, but no predictor variables. The constant is allowed to vary across both levels, so that the variance of each level can be estimated. Second, predictors of the micro level are included in the regression equation. The fit for this model is compared with the fit of the intercept-only model. This results in a χ^2 value for the fit difference, which can be tested for significance using the number of added parameters as degrees of freedom. In case of a significant improvement of model fit, each of the added parameters are examined for significance. Third, it is evaluated whether any of the slopes of the explanatory variables has a significant variance component between the groups. Fourth, explanatory variables of the macro level are added to the regression equation. Again, tests for the significance of the increase in fit for this model, as well as significance tests for the added parameters, are undertaken. The last step is to include cross-level interactions between explanatory group-level variables and micro-level explanatory variables that have a significant slope variation, and to conduct analogous significance tests.

Estimation of Parameters and Significance Tests

To estimate parameters in HLM, iterative estimation approaches are used that search for a converging model meeting the maximum-likelihood criterion (Goldstein 1995; Hox 1995, 1998, 2002). The starting point for an iterative procedure is the estimation of fixed parameters. On this basis, the randomized values are calculated followed by the fixed parameters, and so on, until the procedure converges (Goldstein 1995; Hox 1995). Tests for significance are done for both single parameters as well as complex models. The tests divide parameter estimates by their standard error, resulting in a *z*-value equivalent quotient, which is normally distributed (Hox 2002). Fit of a predicted model is derived from the difference in likelihood values between this model and the basic model (Goldstein 1995; Rasbash et al. 2000). These differences follow a χ^2 distribution. The degrees of freedom for significance testing are drawn from the number of added parameters (Duncan, Jones, and Moon 1995).

Prerequisites for Hierarchical Linear Models

Prerequisites and assumptions for HLM are generally similar to those of ordinary least squares (OLS) regression analysis (Hox 1998). Thus, a linear relationship between predictor and criterion variables is supposed. Residual variances e_{ij} on the micro level are assumed to be normally distributed, having a mean of zero and a common variance in all groups. Residuals on the macro level, μ_{0j} and μ_{pj}, should also follow a normal distribution having a mean of zero. In addition, they should be unrelated to micro-level errors. Even though a normal distribution for dependent variable values in HLM is required, this is of minor importance for independent variables.

However, the assumption that may be of most interest to researchers is that of sample size. In fact, one reason for the low take-up of multilevel analysis methods in marketing could be a perception that sample size demands are difficult to meet. With regard to the required sample size, requirements at the macro level take primacy (Hox 1998). By nature, conditions for testing intralevel relationships are enhanced with a larger number of individuals (e.g., team members), whereas cross-level analysis needs a large number of aggregation units (e.g., teams). But an increase in aggregation units while the number of individuals is reduced has a more positive effect than vice versa. Hox and Maas (2002) demonstrated in a simulation study that small sample sizes ($n < 50$) at the macro level lead to distorted estimates of standard errors on this level. On the other hand, the sample size at the micro level had no influence on the accuracy of the parameter estimation. Accordingly, authors such as Lee (2003) and Mossholder, Bennett, and Martin (1998) have worked with sample sizes of $n \geq 3$ on the micro level. The more restrictive requirements on the macro level are underlined by the following example from Snijders and Bosker:

> A relevant general remark is that the sample size at the highest level is usually the most restrictive element in the design. For example, a two-level design with 10 groups, i.e. a macro-level sample of 10, is at least as uncomfortable as a single-level design with a sample size of 10. Requirements on the sample size at the highest level, for a hierarchical linear model with q explanatory variables at this level, are at least as stringent as requirements on the sample size in a single level design with q explanatory variables. (1999, p. 140)

Thus, there would at first glance appear to be a trade-off here between, on one hand, the difficulty of generating a multilevel data set of sufficient size, and on the other hand, using a single-level approach, such as multiple regression or SEM. The major factors in such a decision are primarily conceptual, and concern the key constructs in any given theory that is being explored. In particular, if there are key independent variables that vary at the higher level, then one must collect enough data points at this level to satisfy the assumptions of *any* multivariate technique, whichever is to be used. As stated by Kozlowski and Klein: "levels and units should be consistent with the nature of the phenomenon of interest. Principle: Unit specification (formal versus informal) should be driven by the theory of the phenomena in question" (2000, pp. 19–20). These sample size assumptions are broadly similar whether one is using HLM, multiple regression, or SEM.

Specifically, in theoretical terms, one should be clear about the highest level at which their key independent variables vary, and ensure that they collect enough data points at this level. What this means is that, for example, if something like "team culture" is a key predictor of individual performance, one should collect representatives of enough separate teams to satisfy standard sample size assumptions for any multivariate method, rather than collect multiple members of only a few teams and treat them as independent data points because—as was shown above—this approach may cause misleading results. If multiple members of teams are collected, they should be treated in an appropriate way (whether this be multilevel analysis, aggregation, or key informant) rather than used as what would be a misleadingly "large" sample.

As such, the question of whether it is preferable to use a multilevel approach with a prohibitively small higher-level sample size, or use another technique that does not take into account levels issues, or nonindependence within groups, is somewhat rhetorical. At the theory-development stage, the researcher must be clear about what level he or she needs to collect data points at, *whichever* analysis technique is being used. Erroneously assuming that multiple lower-level members of higher-level groups (e.g., teams) can be treated as independent, when key theoretical variables are manifestly influenced by group membership, will lead to misrepresentation of a greater or lesser degree, and thus sample size estimations should always use the highest level of theoretical variation as their base.

That said, it is unarguable that some decision criteria for researchers working with theories where levels issues are likely to play a role would help in the process of determining whether the additional difficulty of collecting data from multiple group members is outweighed by the increased robustness of the results from multilevel analysis. In other words, how should the researcher trade off sample size demands against methodological rigor? The results reported here provide some detailed indications, and it can be seen that the higher the intraclass correlation coefficient within the lower level group, the less appropriate it is to treat individuals within a group as independent data points. However, determining the intraclass correlation requires data to be available in the first place. When planning an empirical study, it is vital to consider the impact that higher-level constructs are likely to have on lower-level variables (whether this be as hypothesized or control relationships). The stronger the effect that higher-level variables are likely to have, the greater the benefit of taking the higher level as the unit of analysis to determine the relevant sample size. For example, if one expects a group-level variable such as leadership style to have a substantive impact on individual performance, one should treat the individual employees in each team as nonindependent and then determine which analysis method to employ—whether HLM, or multiple regression using single key informants, or aggregates of multiple team members. The stronger the influence of the higher-level variable, then the greater the benefit to assuming nonindependence and designing a sample as appropriate.

Extensions of, and Alternatives to, Hierarchical Linear Modeling

As mentioned above, a standard assumption of HLM, and indeed, most regression-based techniques, is that the relationship between predictor and criterion variable is linear. However, just as generalized linear modeling allows the extension of OLS regression to analyze data parametrically from nonnormal distributions, HLM has been extended to cope with nonnormal individual-level outcomes (e.g., Goldstein 1991). Most common among these is probably binary response data—the equivalent of binary logistic regression at a single level—but commonly available software allows analysis of more complex distributions, including Poisson and negative binomial distributions (Goldstein et al., 1998).

Hierarchical linear modeling is also increasingly used to model longitudinal (repeated measures) data. In this scenario, individuals (or cases) are taken as the higher level, and individual observations at a given time period are taken as the lower level. This allows longitudinal data analysis with far fewer observations than would be necessary for traditional time-series analysis, but without the constraints on numbers of observations placed by repeated measures analysis via the general linear model. (For further information, see Bliese and Ployhart, 2002.)

Although HLM has become the most common technique for analyzing multilevel data, there are other options available. Three articles in a special issue of *Leadership Quarterly* (Bliese and Halverson 2002; Gavin and Hofmann 2002; Markham and Halverson 2002) analyzed a single multilevel data set in three ways. As well as HLM, other methods used were within- and between-entity analysis (WABA) (Dansereau, Alutto, and Yammarino 1984), and random group resampling (RGR). WABA is a method of "assessing the importance of entity membership when examining the relationship between constructs at multiple levels of analysis" (Markham and Halverson 2002, p. 35). On the other hand, RGR "provides a tool for statistically determining whether group-level relationships are the result of true group phenomena (group effects) or the result of aggregating individual level to the group level (grouping effects)" (Bliese and Halverson 2002, p. 53). As these descriptions suggest, each method has its own advantages, but each also has limitations. Castro (2002) provided a useful comparison of these three methods in the same volume. Drawing on Castro's comparison, and the strengths and limitations of each of the three methods as acknowledged by the authors of the original three papers, we present a summary of the advantages and disadvantages of each method in Table 3.

Statistical Software for Hierarchical Linear Modeling

As analytic approaches using HLM have become more widely used by researchers, more software has been produced to allow researchers to analyze data in this way. We will briefly discuss two specialist HLM software packages—HLM6 and MLwiN—and identify other packages that have incorporated HLM modules.

HLM6 (Raudenbush, Cheong, and Congdon 2004) was specifically designed to analyze HLM. It is based on a user-friendly format, allowing the user to specify models in a step-by-step basis, first stating the number of levels in the model (two or three), then choosing the data files, and going on to specify which of the variables in these files form the model. It reads both level 1 and level 2 files simultaneously, allowing easy specification of cross-level interactions. Recent versions have incorporated the ability to model nonnormal (e.g., binomial) dependent variables. Although HLM6 does not allow manipulation of data, it does allow a wide range of formats for data input, including SPSS, STATA, Excel, SAS, S-PLUS, and ASCII. Another software package designed especially for the computation of multilevel models is MLwiN (Goldstein et al. 1998). MLwiN is less user-friendly than HLM6 in the sense that it is not as directive and requires greater user knowledge, but allows greater flexibility and a much wider range of possible models, including multivariate response models, Bayesian modeling, and bootstrap estimation. It is a Windows-updated version of a syntax-based program, MLn, which can still be used in the new software, allowing manipulation of data within the file.

Comprehensive statistical software to have included HLM functions include S-PLUS (*lme*) and SAS (PROC MIXED). Bliese (2002) gives a useful introduction to modeling multilevel data in these packages. The obvious advantage of using comprehensive statistical software is that the same data can be analyzed in different ways without the need for the transfer of data between packages. The disadvantages to the use of S-PLUS and SAS are that, unlike HLM6, knowledge of the underlying programming language is necessary, and the range of models available is less wide than in MLwiN. Models estimated in S-PLUS can also be estimated using the open-source language R (Ihaka and Gentleman 1996). Mplus (Muthén and Muthén 2006) is an SEM package that allows for multilevel structures. Thus, it allows estimation of confirmatory factor analyses, path models, and other structural equation models where the data are hierarchical in nature. This means that a much wider set of linear models can be tested.

IMPLICATIONS, LIMITATIONS, AND CONCLUSIONS

In this paper, we presented the state-of-the-art of multilevel research in marketing science, and explored key issues regarding the use and appropriateness of multilevel analysis within marketing research. Our study has clear and substantive implications for marketing scholars and practitioners. First, our analysis of existing marketing literature shows that there is a need for empirical and analytical approaches within marketing research to catch up to multilevel marketing theory development in taking account of the nonindependence of nested data structures, and incorporating multiple levels of analysis. In particular, in light of recent recommendations in the literature that marketing scholars should make efforts to collect multiple responses per work group or organization and aggregate them (e.g., van Bruggen, Lilien, and Kacker 2002), further information on the implications of such a strategy—which we provide here—is surely welcome. As well as this, the growing importance of considering varying levels of analysis, and conceptualizing relationships across multiple levels of the firm and firm–customer interface, mean that a marketing-specific consideration of multilevel analysis issues is overdue.

Our findings provide clear evidence of issues concerning nested data that, together with the emergence of an increasing number of marketing-related studies dealing either empirically or conceptually with multiple levels of analysis, suggests the importance of a consideration of multilevel research approaches. Implications of the alternate methodological approaches to multilevel research issues have been identified, which may include misinterpretations of the examined data where there is nonindependence between individual data points. As such, the findings reported here have important implications for marketing practice in

Table 3
A Comparison of Hierarchical Linear Models, WABA, and Random Group Resampling (Evidence Drawn from Bliese and Halverson 2002; Bryk and Raudenbush 1992; Castro 2002; Gavin and Hofmann 2002; Markham and Halverson 2002)

	Hierarchical Linear Models	WABA	Random Group Resampling
Basic Assumptions	• The error term of each level 1 unit should have a mean of zero and the residuals should be normally distributed. • Level 1 predictors are independent of the level 1 error term. That is, the covariance between the level 1 predictors and the error term should equal zero. • Level 2 error terms have a mean of zero and adhere to a multivariate normal distribution. • Level 2 predictors are independent of all level 2 error terms. • The level 1 error terms are independent of level 2 error terms.	• The data should be multivariate normal, with constant error variance and independent error terms, except for nonindependence defined by the level structure.	• Random sampling is necessary. • Same assumptions as those of underlying methods (OLS regression and WABA).
Units of Theory	• HLM is designed to test cross-level and multilevel models or relationships among variables at different levels of analysis.	• WABA is a nontheoretical approach. It can be used to help determine units of theory in an exploratory manner, but does not require prior theoretical units. • WABA I partitions variance of individual variables into within-group and between-group variance—similar to analysis of variance. • WABA II partitions covariance (and therefore correlations) into the within- and between-group levels, and determines whether the correlation at each level is significant.	• RGR group-mean analysis looks at unit-level variables (the group). • Analyses conducted at the group level, using group averages. • Tests whether group-level relationships between variables and group-level moderator are due to group effects and not to the aggregation process (grouping effects). • Analysis based on the assumption that a certain hypothesis proposed at the individual level would have applicability at the group level.

Strengths	• The method is well suited to test cross-level moderator effects models. • HLM allows researchers to identify and partition different sources of variance in outcome variables. • Enables analysis of longitudinal relationships. • Software to perform HLM is plentiful.	• No prior theory is necessary—very useful for exploratory analysis and theory development. • Relatively straightforward to understand (correlation coefficients are directly analogous to individual-level versions). • Enables assessment of both statistical and practical significance.	• RGR group-mean procedure allows evaluation of whether group results (or other units of interest) are based on group or grouping effects. • Creates many pseudo groups (as many as actual groups) and uses statistics other than only group means to compare among them. • RGR procedure can easily be extended to other areas as well (ex. WABA II correlation, WABA I etas).
Restrictions/ Limitations	• HLM restricts the dependent variable to be operationalized at the lowest level of analysis and therefore cannot be used to test hypotheses that have a dependent variable at a lower level of analysis. • HLM does not provide tests of appropriateness of aggregation or nonaggregation. • Sample size requirements are a limitation when one needs to consider groups as units of analysis. • When using HLM for longitudinal analysis, the assumption that level 1 residuals are assumed to be independent may be violated.	• Less useful for hypothesis testing, unless all variables and relationships are supposed to be random at both levels. • Incapable of examining cross-level interactions. • Relies on eta values, which have been shown to be related to group size. • Can only examine moderation by means of subgroup analysis. • Inconclusive results can lead to ambiguity in interpretation because there is no theoretical underpinning.	• For RGR group-mean analyses, variables must be operationalized at the group level using group averages. This prevents the evaluation of other potential levels of analysis at which effects could be operating. • The variability within each group is assumed to be error. By not evaluating the variance within groups, important information could be overlooked (such as the variable operating at both within- and between-groups levels).

an indirect manner. Specifically, the more robust results achieved by the use of the multilevel modeling approach posited herein should lead directly to more useful and beneficial practical implications from future marketing research. In other words, marketing practitioners should be able to have more confidence in, and thus gain more advantage from, the results of research conducted with adequate attention paid to the nested nature of multilevel and grouped data where this is relevant. Furthermore, the use of multilevel analysis approaches in research can provide practitioners with a more accurate picture of what level of the organization they should directly target for any performance improvement efforts. By contrast, research that does not take into account multilevel structures where they are relevant is only able to give part of the picture to practitioners, forcing them to make assumptions about the direct role of other levels.

Results from the Monte Carlo simulation suggested that where medium to high levels of nonindependence exist, treating nested data as independent using standard OLS approaches may inflate effect sizes, and that therefore multilevel methods may be considered more appropriate. Of course, our simulation is not without limitations. In particular, our aim to clearly relate the simulation to a marketing-relevant context meant that we were bound by the characteristics of the particular study we chose to emulate. In the present case, this meant that our simulation does not include any "intermediate" effect sizes (in the range of 0.1–0.25), which may have shown more marked differences in observed power across the study conditions. In addition, we were unable to simulate data with "known" population parameters, which did not allow us to determine which of the methods was most accurate at returning these parameters under different conditions. Although the latter design is common to most Monte Carlo studies, this was not our aim presently. Our aim was to show how parameter estimates and power can change with different analysis methods and conditions, in a highly marketing-relevant context. Even though we were able to show that the choice of analysis method was a key influence on the results, we were unable to objectively show which was the "best" method. Nevertheless, existing methodological research does shed light on this issue (e.g., Bleise and Hanges 2004), and can be consulted by the interested reader. Finally, we were, of course, unable to include either (1) all possible combinations of sample size, group size, and effect size as experimental conditions, or (2) model a number of other factors that may influence the choice of the most appropriate methodology (such as measure reliability or population distribution). Despite the interesting and substantive differences we found, future research should look to investigate more potential factors that can influence the results of the analysis of multilevel and nested data.

As has become the norm in other disciplines of organizational and behavioral science research, we expect HLM to become a standard procedure in marketing research, eventually becoming a method as commonly used as SEM. Nevertheless, as we have made clear, HLM is not a panacea, nor is it appropriate to all situations, and our twofold aim was to expand understanding of the HLM approach, as well as to clarify the most appropriate situations in which it should be used. Having said this, it is undoubtedly the case that increased familiarity with multilevel issues and procedures will be of use to most marketing scholars. Second, there may be huge potential in the reanalysis of existing data, to look for new answers to old questions by utilizing the power of multilevel tools. Third, although multilevel modeling is likely to be of significant use to the marketing researcher, we would issue a word of caution about overenthusiastic application of multilevel techniques to new and existing marketing problems. As is the case with all analytic techniques, multilevel modeling should be considered an addition to, not a replacement for, our existing methodological repertoire.

We described here the basics of multilevel modeling and introduced some of the techniques that have become more and more common in the organizational fields in the recent past. Of course, this does not mean that methods will not progress, and we would like to conclude with an outlook on what the future may bring. The first of these developments certainly will be the combination of multilevel modeling and SEM. For example, the most recent versions of both Mplus (Muthén and Muthén 2006) and EQS (version 6.1) (Bentler 2005) have started to build multilevel modeling functionality into their SEM systems and these will certainly influence the way we think about more complex models in the near future. The second development may enhance our ability to test multilevel models (which have heretofore solely been concerned with micro-level criterion variables; Snijders and Bosker 1999), by incorporating techniques to predict macro-level criterion variables from variables measured at the individual level (e.g., Croon and van Veldhoven 2007).

NOTES

1. *Marketing Science* was considered for inclusion due to its high quality and quantitative emphasis, but it was decided not to in-

clude this journal as its content was not as general or mainstream as the other included journals.

2. Note that we are not implying that the methods used by Atuahene-Gima and Li (2002) were incorrect, nor are we reanalyzing their data. Rather, we are using the structure of their data, and individual-level correlations, to represent those that might be found by marketing researchers.

3. It would also have been possible to use covariances to simulate the data, and no differences would have been observed in effect sizes or significance/power levels.

REFERENCES

Atuahene-Gima, Kwaku, and Haiyang Li (2002), "When Does Trust Matter? Antecedents and Contingent Effects of Supervisee Trust on Performance in Selling New Products in China and the United States," *Journal of Marketing,* 66 (3), 61–82.

Baumgartner, Hans, and Christian Homburg (1996), "Applications of Structural Equation Modeling in Marketing and Consumer Research: A Review," *International Journal of Research in Marketing,* 13 (2), 139–161.

——, and Rik Pieters (2003), "The Structural Influence of Marketing Journals: A Citation Analysis of the Discipline and Its Subareas Over Time," *Journal of Marketing,* 67 (4), 123–139.

Bell, Simon J., and Bulent Menguc (2002), "The Employee–Organization Relationship, Organizational Citizenship Behaviors, and Superior Service Quality," *Journal of Retailing,* 78 (2), 131–146.

Bentler, Peter M. (2005), *EQS Structural Equation: Program Manual,* Encino, CA: Multivariate Software.

Bettencourt, Lance A. (2004), "Change-Orientated Organizational Citizenship Behaviors: The Direct Effect and Moderating Influence of Goal Orientation," *Journal of Retailing,* 80 (3), 165–180.

Bliese, Paul D. (2000), "An Introduction to Multilevel Modeling Techniques," *Personnel Psychology,* 53 (4), 1062–1065.

—— (2002), "Multilevel Random Coefficient Modeling in Organizational Research: Examples Using SAS and S-PLUS," in *Modeling in Organizational Research: Measuring and Analyzing Behavior in Organizations Advances in Measurement and Data Analysis,* Fritz Drasgow and Neal Schmitt, eds., San Francisco: Jossey-Bass, 401–445.

——, and Ronald R. Halverson (2002), "Using Random Group Resampling in Multilevel Research: An Example of the Buffering Effects of Leadership Climate," *Leadership Quarterly,* 13 (1), 53–68.

——, and Paul J. Hanges (2004), "Being Both Too Liberal and Too Conservative: The Perils of Treating Grouped as Though They Were Independent," *Organizational Research Methods,* 7 (4), 400–417.

——, and Robert E. Ployhart (2002), "Growth Modeling Using Random Coefficient Models: Model Building, Testing, and Illustrations," *Organizational Research Methods,* 5 (4), 362–387.

——, David Chan, and Robert E. Ployhart (2007), "Multilevel Methods. Future Directions in Measurement, Longitudinal Analyses, and Nonnormal Outcomes," *Organizational Research Methods,* 10 (4), 551–563.

Broderick, Amanda J., Gordon E. Greenley, and Rene D. Mueller (2007), "The Behavioral Homogeneity Evaluation Framework: Using Consumer Involvement in International Segmentation," *Journal of International Business Studies,* 38 (5), 1–25.

Bryk, Anthony, and Stephen W. Raudenbush (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods,* Newbury Park, CA: Sage.

Castro, Stephanie L. (2002), "Data Analytic Methods for the Analysis of Multilevel Questions: A Comparison of Interclass Correlation Coefficients, $r_{wg(j)}$, Hierarchical Linear Modeling, Within- and Between-Analysis, and Random Group Resampling," *Leadership Quarterly,* 13 (1), 69–93.

Croon, Marcel A., and Marc J.P.M. van Veldhoven (2007), "Predicting Group Level Variables from Variables Measured at the Individual Level: A Latent Variable Multilevel Model," Working Paper, Faculty of Social Sciences, Tilburg University, Tilburg, the Netherlands.

Dansereau, Fred., Joseph A. Alutto, and Francis J. Yammarino (1984), *Theory Testing in Organizational Behavior: The Varient Approach,* Englewood Cliffs, NJ: Prentice Hall.

DeCarlo, Thomas E., Raymond C. Rody, and James E. DeCarlo (1999), "A Cross National Example of Supervisory Management Practices in the Sales Force," *Journal of Personal Selling & Sales Management,* 19, 1 (Winter), 1–14.

Dellande, Stephanie, Mary C. Gilly, and John L. Graham (2004), "Gaining Compliance and Losing Weight: The Role of the Service Provider in Health Care Services," *Journal of Marketing,* 68 (3), 78–91.

Duncan, Craig, Kelwyn Jones, and Graham Moon (1995), "Blood Pressure, Age and Gender," in *A Guide to MLn for New Users,* Geoff Woodhouse, ed., London: Institute of Education, 59–85.

Gavin, Mark B., and David A. Hofmann (2002), "Using Hierarchical Linear Modeling to Investigate the Moderating Influence of Leadership Climate," *Leadership Quarterly,* 13 (1), 15–33.

Goldstein, Harvey (1991), "Nonlinear Multilevel Models, with an Application to Discrete Response Data," *Biometrika,* 78 (1), 45–51.

—— (1995), *Multilevel Statistical Models,* 2d ed., London: Hodder Arnold.

——, William Browne, and Jon Rashbash (2002), "Tutorial in Biostatistics: Multilevel Modeling of Medical Data," *Statistics in Medicine,* 21, 3291–3315.

——, Jon Rasbash, Ian Plewis, David Draper, William Browne, Min Yang, Geoff Woodhouse, and Michael Healy (1998), *A User's Guide to MLwiN,* London: Institute of Education.

Hartline, Michael D., James G. Maxham, III, and Daryl O. McKee (2000), "Corridors of Influence in the Dissemination of Customer-Oriented Strategy to Customer Contact Service Employees," *Journal of Marketing,* 64 (2), 35–50.

Hofmann, David A. (1997), "An Overview of the Logic and Rationale of Hierarchical Linear Models," *Journal of Management,* 23 (6), 723–744.

Homburg, Christian, and Ruth M. Stock (2004), "The Link Between Salespeople's Job Satisfaction and Customer Orientation in a Business-to-Business Context: A Dyadic Analysis," *Journal of the Academy of Marketing Science,* 32 (2), 144–158.

Hox, Joop (1995), *Applied Multilevel Analysis,* Amsterdam: TT-Publikaties.

—— (1998), "Multilevel Modeling: When and Why," in *Classification, Data Analysis, and Data Highways,* Ingo Balderjahn, Rudolf Mathar, and Martin Schader, eds., New York: Springer Verlag, 147–154.

—— (2002), *Multilevel Analysis: Techniques and Applications,* Mahwah, NJ: Lawrence Erlbaum.

——, and Cora J.M. Maas (2002), "Sample Sizes for Multilevel Modeling," in *Social Science Methodology in the Millennium: Proceedings of the Fifth International Conference on Logic and Methodology,* 2d expanded edition, Jörg Blasius, Joop Hox, Edith de Leeuw, and Peter Schmidt, eds., Opladen: Leske & Budrich Verlag.

Hult, Tomas G.M., William T. Neese, and Edward R. Bashaw (1997), "Faculty Perceptions of Marketing Journals," *Journal of Marketing Education,* 19 (1), 37–52.

Ihaka, Ross, and Robert Gentleman (1996), "A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics,* 5 (3), 299–314.

Jap, Sandy D. (1999), "Pie-Expansion Efforts: Collaboration Processes in Buyer–Supplier Relationships," *Journal of Marketing Research,* 36 (4), 461–475.

Jong, Ad. de, Ko de Ruyter, and Jos Lemmink (2004), "Antecedents and Consequences of the Service Climate in Boundary-Spanning Self-Managing Service Teams," *Journal of Marketing,* 68 (2), 18–35.

Kidwell, Ronald E., Jr., Kevin W. Mossholder, and Nathan Bennett (1997), "Cohesiveness and Organizational Citizenship Behavior: A Multilevel Analysis Using Work Groups and Individuals," *Journal of Management,* 23 (6), 775–793.

Klein, Howard J., and Jay S. Kim (1998), "A Field of the Influence of Situational Constraints Leadership–Member Exchange, and Goal Commitment on Performance," *Academy of Management Journal,* 41 (1), 88–95.

Klein, Katherine J., Fred Dansereau, and Rosalie J. Hall (1994), "Levels Issues in Theory Development, Data Collection, and Analysis," *Academy of Management Review,* 19 (2), 195–229.

Kozlowski, Steve W.J., and Katherine J. Klein (2000), "A Multilevel Approach to Theory and Research, in Organizations: Contextual, Temporal, and Emergent Processes," in *Multilevel Theory, Research and Methods in Organizations, Foundations Extensions, and New Directions,* Katherine J. Klein and Steve W.J. Kozlowski, eds., San Francisco: Jossey-Bass, 3–90.

Lam, Simon S.K., Xiao-Ping Chen, and John Schaubroeck (2002), "Participative Decision Making and Employee Performance in Different Cultures: The Moderating Effects of Allocentrism/Idiocentrism and Efficacy," *Academy of Management Journal,* 45 (5), 905–914.

Lankau, Melenie J., and Terri A. Scandura (2002), "An Investigation of Personal Learning in Mentoring Relationships: Content, Antecedents, and Consequences," *Academy of Management Journal,* 45 (4), 779–790.

Lee, Byung Hee (2003), "Using Hierarchical Linear Modeling to Illustrate Industry and Group Effects on Organizational Commitment in a Sales Context," *Journal of Managerial Issues,* 15 (3), 353–368.

Leeuw, Jan de, and Ita G.G. Kreft (1986), "Random Coefficient Models," *Journal of Educational Statistics,* 11 (1), 55–85.

Liao, Hui, and Aichia Chuang (2004), "A Multilevel Investigation of Factors Influencing Employee Service Performance and Customer Outcomes," *Academy of Management Journal,* 47 (1), 41–58.

Liu, Annie H., and Mark P. Leach (2001), "Developing Loyal Customers with a Value-Adding Sales Force: Examining Customer Satisfaction and the Perceived Credibility of Consultative Salespeople," *Journal of Personal Selling & Sales Management,* 21, 2 (Spring), 147–156.

Longford, Nicholas T. (1993), *Random Coefficient Models,* New York: Oxford University Press.

Macintosh, Gerard, and Lawrence S. Lockshin (1997), "Retail Relationships and Store Loyalty: A Multi-Level Perspective," *International Journal of Research in Marketing,* 14 (5), 487–497.

MacKenzie, Scott B. (2001), "Opportunities for Improving Consumer Research Through Latent Variable Structural Equation Modeling," *Journal of Consumer Research,* 28 (1), 159–166.

Madjar, Nora, Greg R. Oldham, and Michael G. Pratt (2002), "There's No Place Like Home? The Contributions of Work and Nonwork Creativity Support to Employees' Creative Performance," *Academy of Management Journal,* 45 (4), 757–767.

Markham, Steven E., and Ronald R. Halverson (2002), "Within- and Between-Entity Analysis in Multilevel Research: A Leadership Example Using Single Level Analyses and Boundary Conditions," *Leadership Quarterly,* 13 (1), 35–52.

McAllister, Daniel J., and Gregory A. Bigley (2002), "Work Context and the Definition of Self: How Organizational Care Influences Organization-Based Self Esteem," *Academy of Management Journal,* 45 (5), 894–904.

Moorman, Robert H., Gerald H. Blakely, and Brian P. Niehoff (1998), "Does Perceived Organizational Support Mediate the Relationship Between Procedural Justice and Organizational Citizenship Behavior?" *Academy of Management Journal,* 41 (3), 351–357.

Mossholder, Kevin W., Nathan Bennett, and Christopher L. Martin (1998), "A Multilevel Analysis of Procedural Justice Context," *Journal of Organizational Behavior,* 19 (2), 131–141.

Muthén, Linda K., and Bengt O. Muthén (2006), *Mplus User's Guide,* 4th ed., Los Angeles: Muthén and Muthén (available at www.statmodel.com/download/usersguide/Mplus%20Users%20Guide%20v41.pdf).

Netemeyer, Richard G., James S. Boles., Daryl O. McKee, and Robert McMurrian (1997), "An Investigation into the Antecedents of Organizational Citizenship Behaviors in a Personal Selling Context," *Journal of Marketing,* 61 (3), 85–98.

Organ, Dennis W. (1997), "Organizational Citizenship Behavior: It's Construct Clean-Up Time," *Human Performance,* 10 (2), 85–97.

Osborne, Jason W. (2000), "Advantages of Hierarchical Linear Modeling: Practical Assessment," *Research and Evaluation,* 7 (1), 1–6.

Piercy, Nigel F., David W. Cravens, and Nikala Lane (2003), "Sales Manager Behavior Control Strategy and Its Consequences: The Impact of Manager Gender Differences," *Journal of Personal Selling & Sales Management,* 23, 3 (Summer), 221–237.

Pieters, Rik, and Michel Wedel (2004), "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing,* 68 (2), 36–50.

Ping, Robert A., Jr. (2003), "Antecedents of Satisfaction in a Marketing Channel," *Journal of Retailing,* 79 (4), 237–248.

Ragins, Belle Rose, John L. Cotton, and Janice S. Miller (2000), "Marginal Mentoring: The Effects of Type of Mentor, Quality of Relationship, and Program Design on Work and Career Attitudes," *Academy of Management Journal,* 43 (6), 1177–1194.

Ramus, Catherine A., and Ulrich Steger (2000), "The Roles of Supervisory Support Behaviors and Environmental Policy in Employee 'Ecoinitiatives' at Leading-Edge European Companies," *Academy of Management Journal,* 43 (4), 605–626.

Rasbash, Jon, William Browne, Harvey Goldstein, Min Yang, Ian Plewis, Michael Healy, Geoff Woodhouse, David Draper, Ian Langford, and Tobby Lewis (2000), *A User's Guide to MLwiN,* London: Centre for Multilevel Modeling Institute of Education, University of London.

Raudenbush, Stephen W., and Anthony S. Bryk (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods,* 2d ed., Thousand Oaks, CA: Sage.

——, ——, Yuk Fai Cheong, and Richard T. Congdon (2004), *HLM 6: Hierarchical Linear and Nonlinear Modeling,* Lincolnwood, IL: Scientific Software International.

Saparito, Patrick A., Chao C. Chen, and Harry J. Sapienza (2004), "The Role of Relationship Trust in Bank–Small Firm Relationships," *Academy of Management Journal,* 47 (3), 400–410.

Sarin, Shikhar, and Vijay Mahajan (2001), "The Effect of Reward Structures on the Performance of Cross-Functional Product Development Teams," *Journal of Marketing,* 65 (2), 35–53.

Siguaw, Judy A., Penny M. Simpson, and Thomas L. Baker (1998), "Effects of Supplier Market Orientation on Distributor Market Orientation and the Channel Relationship: The Distributor Perspective," *Journal of Marketing,* 62 (3), 99–111.

Smidts, Ale., Ad Th.H. Pruyn, and Cees B.M. van Riel (2001), "The Impact of Employee Communication and Perceived External Prestige on Organizational Identification," *Academy of Management Journal,* 49 (5), 1051–1062.

Snijders, Tom A.B., and Roel J. Bosker (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling,* London: Sage.

Speier, Cheri, and Viswanath Venkatesh (2002), "The Hidden Minefields in the Adoption of Sales Force Automation Technologies," *Journal of Marketing,* 66 (3), 95–111.

Steenkamp, Jan-Benedict E.M., and Hans C.M. van Trijp (1991), "The Use of LISREL in Validating Marketing Constructs," *International Journal of Research in Marketing,* 8 (4), 283–300.

Tax, Stephen S., Stephen W. Brown, and Murali Chandrashekaran (1998), "Customer Evaluations of Service Complaint Experiences: Implications for Relationship Marketing," *Journal of Marketing,* 62 (2), 60–76.

Tepper, Bennett J., and Edward C. Taylor (2003), "Relationships Among Supervisors' and Subordinates' Procedural Justice Perceptions and Organizational Citizenship Behaviors," *Academy of Management Journal,* 46 (1), 97–105.

van Bruggen, Gerrit H., Gary Lilien, and Manish Kacker (2002), "Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses," *Journal of Marketing Research,* 39 (4), 469–478.

van Dolen, Willemijn, Jos Lemmink, Ko de Ruyter, and Ad de Jong (2002), "Customer–Sales Employee Encounters: A Dyadic Perspective," *Journal of Retailing,* 78 (4), 265–279.

Wathne, Kenneth, and Jan B. Heide (2004), "Relationship Governance in a Supply Chain Network," *Journal of Marketing,* 68 (1), 73–89.