

**MANAGING KNOWLEDGE IN THE DARK:
AN EMPIRICAL STUDY OF THE RELIABILITY OF
COMPETENCE EVALUATIONS[#]**

**Niklas Arvidsson^{*}
Jerker Denrell^{**}
Udo Zander^{***}**

[#] We would like to thank the companies involved in this study as well as Askus AB. We are grateful for the comments of Julian Birkinshaw, Peter Hagström, and Lars-Gunnar Mattsson. Please direct all correspondence to Udo Zander, Institute of International Business, Stockholm School of Economics, Box 6501, 113 83, Stockholm, Sweden.

^{*} SMG Consulting, 114 88 Stockholm, Sweden. Phone: +46-8-50743226, Email: Niklas.Arvidsson@smg.se

^{**} Institute of International Business, Stockholm School of Economics, Box 6501, 113 83, Stockholm, Sweden. Phone: +46-8-7369504, Fax: +46-8-31 99 27, Email: Jerker.Denrell@hhs.se

^{***} Institute of International Business, Stockholm School of Economics, Box 6501, 113 83, Stockholm, Sweden. Phone: +46-8-7369513, Fax: +46-8-31 99 27, Email: Udo.Zander@hhs.se

MANAGING KNOWLEDGE IN THE DARK: AN EMPIRICAL STUDY OF THE RELIABILITY OF COMPETENCE EVALUATIONS

ABSTRACT

An implicit assumption in the literature on knowledge management and transfer is that participants in organizations know where capabilities reside. This paper reports an in-depth empirical study of capabilities on the agenda of knowledge management efforts in large leading multinational companies. The results show that evaluation of these capabilities is a complex task. The average inter-rater correlation for capabilities designated as strategic by top management is only 0.25. The results of our empirical study have important implications for creating, retaining, and transferring knowledge in organizations.

Keywords: Knowledge Management, Capabilities, Inter-rater Reliability.

Introduction

The nature and quality of knowledge is at the heart of the division of tasks within and between firms and the subsequent communication and coordination problems. The appointment of Chief Knowledge Officers (CKOs) in firms manifests the appreciation of the importance of these issues in the business community. The function of CKOs is not only limited to addressing issues of information technology and its use, but to oversee the accumulation, use, duplication, transfer, maintenance, and further development of the firm's knowledge base. Increasingly, efforts are also made to identify, measure and report the value of the firm's "intellectual capital" in annual reports.

The focus on knowledge mirrors recent developments in academia, where thoughts and ideas on firm capabilities, knowledge and its management have flourished since the late 1980s¹. The "capabilities perspective" has arisen as a candidate for a "strategic theory of the firm", trying to explain the existence, boundaries, organization, and competitive advantage of the firm. In the capabilities view, the answers to the questions what firms are and why they exist have been grounded in attempts to understand what firms know how to do particularly well.

The capabilities view begins with ideas such as heterogeneous, firm-specific, path-dependent, and hard to imitate production and organization knowledge. The firm is described as a repository of capabilities, which are carried by the community of firm employees and structured by so-called organizing principles. Relative to the market, the firm is viewed as a superior institutional structure for knowledge creation and transfer (Kogut and Zander, 1992; Ghoshal and Moran, 1996). The transfer and management of knowledge become central to explanations of the existence and the competitive advantage of firms.

This perspective implies that identifying capabilities and expertise, knowing "who in the organization is good at what" (Argote, 1999), is a major component in knowledge management and transfer, and at the heart of what makes multinational companies successful. If knowledge is to be "managed" and transferred, it is essential that participants in organizations know (and agree on) where capabilities reside. The purpose of this paper is to examine this issue empirically.

A large body of research examines whether productivity gains acquired in one organization (or part of an organization) transfer to another (Argote, Beckman and Epple 1990, Darr, Argote and Epple 1995, Zander and Kogut 1995, Epple, Argote and Murphy 1996, Haunshild and Miner 1997, Baum and Ingram 1998). Underlying much of the interest is to study productivity increases through transfer of so called 'best practices'. Multinational companies (MNCs) have been described as a prime candidate for capitalizing on differences in capabilities, since they have access to units around the world embedded in differently attractive local environments (Vernon, 1979; Hedlund, 1986; Bartlett and Ghoshal, 1989). Often they appoint subsidiaries so called "centers of excellence" or give them "global product mandates" to indicate where in the international network of subsidiaries "best practice" resides (Holm and Pedersen, 2000; Moore and Birkinshaw, 1998). The different units vary in levels of capabilities due to differences in learning and forgetting rates, and the proficiency of individuals, including managers, engineers and direct production workers. Other factors are differences in the units' technology, tooling and layout; their structure, routines and methods of coordination; and their understanding of who in the organization is good at what (Argote, 1993).

Several empirical studies have examined obstacles to successful transfer of knowledge (Kogut and Zander, 1993; Zander and Kogut, 1995; Kostova, 1996; Szulanski, 1996; Argote, 1999, Gupta and Govindarajan, 2000). Although diverse, the most common obstacles to successful transfer of knowledge identified in this literature are related to the characteristics of knowledge to be transferred, the absorptive capacity of the recipient, and the character of the relationship between source and recipient. Although important, it goes without saying that that another fundamental requirement for successful transfer and management of knowledge is the ability to identify expertise and capable organizational units. Moreover, evaluating capabilities involves a series of complex judgments, with numerous possibilities for bias and error. Thus, there are no reasons to suspect that evaluation of capabilities is flawless. Only rarely, however, have the difficulties of evaluating capabilities in complex, multi-unit organizations, like

multinational companies, been addressed in empirical detail. As a result, it is difficult to know whether the problem of identifying expertise and evaluating capabilities presents a significant obstacle to knowledge management in some of the world's largest firms.

In this paper we empirically examine the difficulties of capability evaluation in large multinational firms. Based on a large data set of evaluations of subsidiary marketing capabilities performed by top management in headquarters and by subsidiary management, we demonstrate that inter-rater reliability is generally low. This holds even for competencies that are claimed by management to be of strategic importance. The results suggest that there is a significant probability that capable or incapable units or individuals will be erroneously classified. We present several implications for knowledge transfer and knowledge management within firms.

The Accuracy of Capability Evaluation in Large Firms

In a background study to this paper one of the present authors examined the efforts of an US-based multinational firm to leverage its geographically dispersed knowledge. This firm had designed a center of excellence structure for sales and marketing activities. The study explored the background to the initiative, its purpose and the process in which the initiative had been born. Unexpectedly, it was found that the firm had severe problems related to how the skills of the center of excellence managers were assessed. There was widespread doubt among managers in this firm whether the centers were strong in their presumed area of expertise.

We will argue in this section that there are reasons to believe that the problem in the above example is common. More precisely, consider a situation in which the capabilities of individuals or organizational units are ranked by management or peers. Reliability can then be defined as inter-rater correlation and (criterion) validity by the correlation of ratings with objective measures of competence. We will argue that there are reasons to believe that the reliability and validity of ratings of competence is often poor. Our argumentation will proceed in two stages. First, we will consider the capability evaluation process and the sequence of

judgments involved in this process. Using the literature in cognitive and social psychology, we will examine the potential for bias in this process. Second, we will review empirical studies in personnel psychology that examine the reliability and validity of performance appraisal and identification of expertise. We will show that reliability and validity in these contexts are frequently poor.

The Capability Evaluation Process

Evaluating the capability of individuals or organizational units can be considered as a process involving a number of phases (Feldman, 1981; Ilgen and Feldman, 1983). Very stylized, this process may be divided into the three phases. First, information about relevant behaviors and outcomes is observed, categorized, and stored. Second, at the time of evaluation information is recalled. Third, the available information is interpreted and combined into an aggregate evaluation of the capability of the individual or the organizational unit.

As noted by Feldman (1981), the literature in social and cognitive psychology have demonstrated that each of these phases may be subject to bias. First, the information observed is usually not a random sample but a highly selective sample. The selection of the sample may be systematically influenced by psychological biases such as availability and salience (Taylor, 1982) and by a tendency to seek out observations in order to confirm prior beliefs (Snyder and Swann, 1978). Within the selected sample, self-fulfilling prophecies (Rosenthal and Jacobsen, 1968; Snyder, 1984), prior beliefs (Lord, Ross, and Lepper, 1979), and “halo-effects” (Nathan and Lord, 1983) may bias the perception of individual behaviors and outcomes. Second, research on memory has established that recall of information is subject to numerous biases (Murphy and Balzer, 1986; Fiske and Taylor, 1991). For example, individuals are most likely to remember information congruent with their general impression or prototype of an individual (Feldman, 1981; Nathan and Lord, 1983; Fiske and Taylor, 1991). Indeed, individuals may even recall events, which are consistent with their prototype of the individual, but never took place (Philips,

1984). Third, combining information in order to form an aggregate capability evaluation requires causal inferences, usually from data subject to noise. Several biases of causal attribution and statistical intuitions have been documented in the literature (Nisbett and Ross, 1980; Kahneman, Slovic, Tversky, 1982). In addition, several studies have demonstrated that individuals are poor at combining various pieces of information to form a global judgment (Dawes, Faust and Meehl, 1989). Informal integration of information is typically inconsistent and unreliable.

Examined in this way it is clear that the process of capability evaluation contain numerous possibilities for error and bias. Such error and bias have been documented in experiments as well as in field studies of performance appraisal. For example, when department heads rated lower-level supervisors in one field study, the ratings correlated only 0.2 with an objective record of their performance. However, the ratings correlated 0.6 with how long the department head had known the foreman, and 0.65 with her liking for the foreman (Stockford and Bissell, 1949). In another study from the Korean War, certain bombing teams were consistently identified as being more accurate than others. In reality, accuracy was inconsistent from day to day and performance was entirely random (Hemphill and Sechrest, 1952). Several laboratory experiments of the ability to identify expertise, using auditors and loan officers as well as various survival simulation exercises, also found that the ability to recognize expertise was limited or at chance levels (Yetton and Bottger, 1982; Trotman, Yetton, and Zimmer, 1983; Miner, 1984; Libby, Trotman, and Zimmer, 1987; Littlepage et. al., 1995; Littlepage, Robinson, and Reddington, 1997).

Consider the implications of these findings for the process of evaluating capabilities in a large complex multinational company. The exposure to differently dynamic institutional environments also lead to varying learning (and forgetting) rates among units, which makes the need for knowledge transfer essential. However, subsidiaries are active selling and producing very different products and services under very different market conditions around the world, which makes it hard to compare performance. Moreover, managers are of different nationalities and subsidiaries in the international network are geographically and culturally distant from each other

and headquarters. As a result, evaluators from headquarters can usually only get a select sample of observations of subsidiaries. These observations are probably not random but driven by cultural and geographic distance as well as by the need to control subsidiaries who are perceived as less competent. The interpretations of behaviors are also likely to be influenced by differences in national and functional culture. In sum, there seem to be good reasons to believe that the capability evaluation process in multinational companies is not free from error and bias.

Empirical studies of reliability and validity

Although suggestive, examples of how the process of capability evaluation may be biased do not demonstrate that evaluations generally are of poor validity and reliability. As the case of visual illusions make clear, the possibility of bias does not imply that bias is the rule rather than the exception. Several pieces of evidence, however, from field studies in firms, suggest that poor reliability and validity is common in situations where ability and performance are evaluated informally.

First, numerous studies in personnel psychology have examined inter-rater reliability of job performance appraisal. In these studies, peers on various dimensions of ability and job duties rate the job performance of individuals. In a review of this literature Viswesvaran et al. (1996) concluded that the average inter-rater correlation was 0.52. King, Hunter and Schmidt (1980), Schmidt et al. (1985), and Hunter and Hirsh (1987) have reported similar figures. In an interesting study Rothstein (1990) examined how the inter-rater reliability depended on the experience of the rater and the length of the period of observation. Based on parallel ratings of 9975 supervisors from 79 different organizations Rothstein (1990) demonstrated that even raters with 20 years of experience and typically long periods of observation of the same individual only achieved an average inter-rater correlation of 0.6. Such low correlation coefficients obviously place an upper bound on the validity of rankings. Commenting on this, Rothstein noted that “To the extent that organizational rewards are allocated on the basis of supervisory ratings of performance, the

maximum reliability of 0.6 suggests that organizationally administered outcomes can never be more than loosely coupled to actual performance.” (Rothstein, 1990, p. 326).

Second, several studies have examined the correlation between self-evaluations and supervisor evaluations of job performance and ability. In a meta-analysis of this literature, based on 36 studies, Harris and Schaubroeck (1988) reported an average correlation of 0.35. In addition, they noted that the correlation tend to be particularly low for managerial and professional workers. Some studies have also examined inter-rater reliability of evaluations of organizational units from respondents from different functions and hierarchical levels. For example, a study in hospitals demonstrated that there were large differences in how administrative and operative staff evaluated the market orientation of operative units (Wenn, La Tour, Bobby, 1994). The correlation between operative and administrative managers’ evaluations ranged between 0.03 and 0.18. Similarly, a study of knowledge flows in multinational firms conducted by Gupta and Govindarajan (2000), found correlation coefficients of 0.23 and 0.38 between headquarters and subsidiary respondents' answers to questions about the extent to which a given subsidiary was an important recipient or source of knowledge. Although this does not provide direct evidence for differences in evaluations of capabilities, such low correlations between evaluations of important recipients and sources of transfer suggest that respondents had different perceptions of the identity of competent subsidiaries. One may note that such differences in perceptions between individuals at different hierarchical levels are consistent with studies that have documented systematic differences in perception and cognition within firms (Simon and Dearborn, 1958; Stevenson, 1976; Walker, 1985; Ireland et al., 1987). Individuals from different departments, functions, and hierarchical levels typically perceive the organization, the environment, and the strength and weaknesses of the firms differently.

Conclusion and Hypotheses

The above argumentation can be summarized as follows. First, capability evaluations involve a series of judgments. There is considerable evidence that these kinds of judgments are subject to bias and error. Presumably, capability evaluations in large complex organization are subject to these biases and errors. Second, there is empirical evidence suggesting that the reliability and validity of capability evaluations is generally low. The most important systematic evidence comes from studies of the reliability of peer and self-supervisor job performance appraisals. Although based on ratings of individuals, and typically of individuals in less complex jobs, the evaluation task examined in these studies are in many respects similar to the task of evaluating capabilities for the purpose of managing the process of knowledge transfer.

These different pieces of evidence suggest the general proposition that the reliability and validity of competence ratings in firms will be poor. The purpose of this paper is to empirically examine this proposition. Since objective measures of competence are very difficult to obtain we will, for the most part, concentrate on the reliability of ratings. However, it should be noted that validity, defined as the correlation between ratings and objective measures, requires reliable ratings. Thus, estimates of reliability provide information about validity.

For statistical analyses, the general proposition that reliability is “poor” obviously needs to be specified. To do so it is useful to consider what an acceptable correlation coefficient would be. Typically, an average inter-item correlation coefficient of at least 0.7 would be required to classify scales as reliable (Nunnally, 1978). Using this level, we may formulate the main hypothesis and the null hypothesis as follows:

Hypothesis (H1): The correlation between parallel rankings of capabilities in a multi-unit firm is below 0.7.

Null Hypothesis (H0): The correlation between parallel rankings of capabilities in a multi-unit firm is 0.7.

Since we are interested in examining the hypothesis that reliability is low, the appropriate null-hypothesis is not zero but some positive number close to 1. The choice of 0.7 is arbitrary but can be defended as the lowest level at which ratings can be argued to be reliable. If we are able to reject this null hypothesis, we are able to reject the possibility that the data comes from a population in which ratings are reliable.

Calculating the p-value for a one-sided test of the null hypothesis that the correlation is 0.7 is considerably more complex than calculating the p-value of a test of the null hypothesis that the correlation is zero. To obtain the p-value we make use of computer simulations. To explain how this works, consider first what the p-value means for a one-sided test of a correlation coefficient of 0.3 using a null hypothesis of zero. This value is the probability that a sample from a population with a true correlation of zero will, due to sampling variance, be observed to have a correlation of 0.3 or above. The p-value is usually calculated using a transformation. However, it could also be obtained using computer simulations. To do so, we generate a random sample of observations from a population where the true correlation is zero. The sample size should be of the same size as in the observed sample. We then notice whether the correlation is at or above 0.3 and repeat this 10,000 times. The proportion of times the correlation coefficient is at or above 0.3 is the p-value. To obtain p-values for a one sided test of a null hypothesis of 0.7 we use a similar approach. 10,000 samples of parallel ratings are generated from a population in which the average correlation coefficient is 0.7. The p-value is the proportion of samples in which the correlation is at or *below* the correlation in the data (for more detail, see the appendix). In addition to using computer simulations, we also calculate approximate one-sided 95% confidence intervals for the correlation coefficients and compare the upper bound of the confidence interval with 0.7.²

Data and Measures

The Sample

The data for this study is based on a survey of evaluations of marketing capabilities of subsidiaries performed by corporate management as well as by subsidiary management in seven large multinational firms. The quantitative data was complemented with field studies including extensive interviews with corporate and subsidiary management in all firms. As a part of a larger research project on knowledge management we approached multinational firms with knowledge management and transfers of best practices on their agenda. This resulted in a sample of seven multinational firms. The sample consists of Ericsson, Pharmacia & Upjohn, Volvo, Sandvik Steel, Sandvik Coromant, Skandia and DeLaval³. At the time of our study, the firms' main global competitors were Nokia (Finland) and Motorola (USA); Novartis (Switzerland) and Pfizer (USA); AUDI (Germany) and BMW (Germany); Sumitomo Metals (Japan); Kennametal (USA) and Mitsubishi (Japan); and Westfalia (Germany) and Bowmatic (USA) respectively. For the sample, median global revenues were 3,379 million USD and the median number of employees was 7,918 in 1998 (See Table 1). Our data was collected in 1997 and 1998 and covers each firm's operations at that time. Our sample includes four firms from the Fortune Global 500 list, which is unique given the nature and focus of our study. In addition, the other three firms are global leaders in their respective industries.

Insert Table 1 around here

A questionnaire with questions about the marketing capabilities of subsidiaries was distributed to corporate management and subsidiary managers within each firm. In each firm, corporate managers helped us identify corporate and subsidiary respondents to our survey. The criteria for selecting respondents at the corporate level were that they had to have a corporate responsibility for marketing activities and an objective to leverage marketing skills globally. The

subsidiary respondents were to have a formal responsibility for local marketing activities. The questionnaire was mailed to subsidiary managers in 209 subsidiaries covering 50 countries as well as to 25 corporate and/or division managers in the seven firms. The survey was constructed in English since the corporate managers informed us this was appropriate given that each of the firms used English as their official corporate language. Each questionnaire included a personalized cover letter which explained the purpose of the study, provided assurance regarding the confidentiality of the collected responses and signaled the support our study had from each firm's top management. Each survey also had a pre-addressed envelope to encourage each respondent to answer and reply promptly. We called each respondent who had not answered in due time to explain the study and to increase the response rate. The collected responses cover 223 managers in 176 subsidiaries as well as 22 corporate managers (see Table 2). The overall response rate is an outstanding 84 percent. Furthermore, the lowest response rate for any company is 70%. The data includes multiple subsidiary responses from 36 subsidiaries and multiple corporate responses regarding 108 subsidiaries. In the units with multiple respondents, we assigned the manager with the greatest formal authority over decision-making to be the lead respondent for that particular subsidiary. The other respondents became non-lead respondents.

Insert Table 2 around here

To achieve comparability among firms, the study is focused on capabilities in marketing. The decision to focus on marketing is, in a sense, arbitrary. However, several firms in the sample were technology-intensive multinational firms, marketing similar products in different countries. Thus, there was an interest in leveraging marketing capabilities across counties and regions. It might be argued that marketing capabilities are more difficult to evaluate than other capabilities where the output can be more easily measured and compared. Although this might be a concern for multinational firms with large variation in the products marketed in different countries, it is

less of a concern for the firms in this sample. In addition, there are several other functions, such as R&D, where the outputs and capabilities are difficult to measure.

Measures

Empirically examining the reliability of capability evaluations in firms requires careful consideration of the measures and data employed. Unreliable measures and unclear questions can easily give a false impression of unreliability (Starbuck and Mezias, 1996). At the same time questions should be representative of capability evaluations occurring within firms rather than deliberately designed to reduce measurement error. To deal with these problems, the items used in the questionnaire were developed in conjunction with corporate management with the following goals in mind. First, the capabilities studied should be of high strategic value to the firms in our sample. Second, it was important that the respondents in each firm – corporate as well as subsidiary managers – had a clear and shared definition of the actual capabilities. Third, the capabilities studied had to be on the agenda of each firm’s knowledge management efforts. Ideally, the capabilities should be regularly measured and evaluated with the aim of increasing the firm’s overall leverage and use of the specific capability. To accomplish these objectives and to minimize response error due to unclear and unfamiliar language, we asked a team of corporate managers to provide the definitions. This team always included several corporate managers and typically included the CEO and/or the corporate and regional marketing managers. The corporate team helped us tailor the questionnaire to each firm in terms of capability dimension definitions, and organizational terminology.

As a result of this exercise four to six firm-specific marketing capabilities were defined for each firm (see Table 3). In each firm, respondents in subsidiaries were given a list of all firm specific capability dimensions and the instruction: “For each category, circle a number reflecting your local company’s ability in relation to other <firm name> local companies.” A 7-point Likert scale, ranging from “*Much worse*” to “*Much better*” with “*About Average*” as the midpoint, was used.

Similarly, corporate respondents were given a list of firm specific capabilities and were instructed: “Rank each subsidiary relative to other <firm name> subsidiaries on a scale were 1 = much below average, 4 = average, and 7 = much above average.”

Insert Table 3 around here

The questionnaire also asked respondents in subsidiaries to identify the firm’s most capable subsidiary in each capability dimension. Two other response alternatives were provided: that all units were equally capable or that the manager did not know. This data provided information about how subsidiary managers evaluated other subsidiaries. Ideally, we wanted data on how each subsidiary evaluated all other subsidiaries. However, given the number of subsidiaries we decided that this would not be feasible.

Finally, we asked both corporate and subsidiary respondents to assess the performance of subsidiaries. Using a 7-point Likert scale, ranging from “*Much below average*” to “*Much above average*” with “*Average*” as the midpoint, respondents at subsidiaries were asked to evaluate “Overall sales revenue”, “Sales Revenue Growth”, and “Overall Market Share”. Using the same scale, corporate respondents were asked to evaluate each subsidiary on these dimensions.

Interviews

Interviews were made at three different points in time based on three different objectives. First, interviews were made with corporate managers – CEOs, Global Marketing Managers and/or Vice Presidents – with the objective to tailor-made the study to each firm’s organizational structure and processes in addition to develop the quantitative survey. The criteria for selecting interviewees at the corporate level were that they had to have a corporate responsibility for marketing activities and an objective to leverage marketing skills globally. Second, interviews were made with corporate and subsidiary managers to corroborate preliminary findings and to

elaborate and understand causes to and consequences of the findings. The subsidiary interviewees were to have a formal responsibility for local marketing activities. Finally, interviews were made primarily with corporate managers – and in some cases with subsidiary managers – to present and discuss the final conclusions. All in all, interviews covering 80 hours were made with 56 corporate and subsidiary managers in seven multinational firms (see Table 2).

Results

Description of the Data

The survey generated data on rankings of relative ability performed by subsidiary managers and corporate managers in seven firms and for 176 subsidiaries and three to six firm specific capability dimensions.⁴ Overall, subsidiary and corporate managers agreed about the relative ability of subsidiaries 28% of the time. However, in 72 % of all cases subsidiary and corporate managers disagreed. Figure 1 shows a frequency distribution of the difference between the evaluation of subsidiary managers and the evaluation of corporate managers. As Figure 1 shows, absolute differences of 2 or more were not uncommon. In fact, in 35 % of all cases, the absolute difference was 2 or more. In line with the literature on job performance appraisal (Harris and Schaubroeck, 1988), there was also a tendency for the evaluations of subsidiary managers to be higher than the evaluations of corporate managers. Indeed, the mean difference between the evaluations of subsidiary managers and the evaluations of corporate managers was 0.42 which is significantly different from zero ($t = 6.81$, $n = 704$). Moreover, the mean subsidiary evaluation was 4.52 ($n = 818$), while the mean corporate evaluation was 4.14 ($n = 729$). Notwithstanding the tendency for subsidiary evaluations to be high, it should be noted that in 27% of all cases the corporate evaluations were higher than subsidiary evaluations.

Insert Figure 1 around here

Analysis

As a result of large differences, both positive and negative, between corporate and subsidiary rankings, the correlation between rankings for specific capability dimensions tended to be low. As an illustration, Figure 2 shows the frequency distribution for the difference between subsidiary and corporate evaluations for a specific capability dimension in one firm. Although evaluations not infrequently agree there are several cases in which they disagree substantially. As a result, evaluations tend to be weakly correlated. In this case, the (Pearson) correlation between subsidiary and corporate evaluations was 0.21.

Insert Figure 2 around here

To examine inter-rater reliability between subsidiary and corporate managers more systematically we computed Pearson correlation coefficients for evaluations performed by corporate and subsidiary managers. The 32 correlation coefficients, for the different capability dimensions in all firms, are listed in Tables 4.1 – 4.7. The average correlation coefficient is 0.25 and the standard deviation is 0.27. The median correlation coefficient is 0.25, the maximum is 0.78 and minimum is –0.39. Of all correlation coefficients, 59 % are lower than 0.3, 9 % are between 0.3 and 0.4, 16 % are between 0.4 and 0.5, and 16 % are higher 0.5. Figure 3 shows a histogram of the 32 correlation coefficients.

Insert Table 4 around here

Insert Figure 3 around here

The fourth column of Tables 4.1-4.7 presents the p-values for a one-sided test of the null hypothesis that the correlation is 0.7. The null hypothesis can be rejected at significance level of 5 % for 26 correlation coefficients (81%) and could not be rejected for six (19%) correlation

coefficients. Similar results were obtained when comparing the upper bound of a one-sided 95 % (approximate) confidence interval with a correlation coefficient of 0.7. The upper bound of the confidence interval was at least 0.7 for seven (22%) of the 32 correlation coefficients.

The fact that we are able to reject the null hypothesis that the correlation is 0.7 for most correlation coefficients suggests that there are significant differences between the subsidiary and corporate evaluations. Since the ratings of subsidiary managers are generally higher than the ratings of corporate managers, however, one might argue that this result could mainly be explained by overly positive, and inaccurate, self-evaluations by subsidiaries. Although this might explain part of the result, there are two reasons for why it is not the whole explanation. First, if self-evaluations were overly positive and corporate evaluations were accurate, it does not immediately follow that the correlation between evaluations would be low. If the evaluations of subsidiary managers were always one unit higher than the evaluations of corporate managers, the correlation would in fact be +1. Only if the ratings of subsidiary managers were overly positive and also highly variable would a low correlation follow. Second, the data demonstrates that in 27% of all cases, the difference between subsidiary and corporate managers is negative rather than positive. Such negative differences imply that mechanisms other than overly positive self-evaluations are responsible for the results.

Further information about differences in evaluations can be obtained by using the answers to the question of the most capable subsidiary. Each respondent in subsidiaries was asked to identify the most capable subsidiary. Adding the number of times a subsidiary was designated most capable by respondents from different subsidiaries, we obtained a number, an index of “peer evaluations” for each subsidiary and capability dimension. We computed the rank correlation between the list of these numbers and the evaluations provided by subsidiary managers and corporate managers for the specific capability dimensions. The average rank correlation between corporate evaluations and “peer evaluations” was 0.27 and the average rank correlation between subsidiary evaluations and “peer evaluations” was 0.17. The general trend

that corporate evaluations were more highly correlated with “peer evaluations” was present in five out of the seven firms. Interestingly, this pattern has also been observed in numerous studies of job performance rankings (Harris and Schaubroeck, 1988). It suggests that corporate evaluations might be more accurate than self-evaluations. In the present case, however, the magnitude of this effect is small.

Finally, it is interesting to compare the above figures with the correlation between corporate and subsidiary managers’ evaluations of “Overall sales revenue”, “Sales Revenue Growth”, and “Overall Market Share”. The average correlation, aggregating across firms and the three dimensions, was 0.52. Although not close to 1, it is substantially higher than the correlations between evaluations of capabilities.

Insights From Field Studies

In the final round of interviews, we presented and discussed the results in managerial meetings in each firm as well as with managers from all firms. This provided yet another opportunity to test the validity and reliability of the findings. These presentations resulted in heated discussions. There were often strong opinions on how capability evaluation processes were characterized. Some strongly distrusted our results and some trusted them. In general, managers tended to question the results in the beginning but – sometimes reluctantly – accept them in the end of the discussions.

As for the reasons given by managers for discrepancies in evaluation of capabilities, only a few comments had to do with the difficulties of observing, categorizing and storing evaluation data. Since a set of agreed upon measures of strategic capabilities existed in all seven multinationals, the availability and salience of evaluation information was not seen as a problem. Neither did managers complain about evaluators seeking out observations to confirm prior beliefs, nor did they emphasize problems related to selectively recalling performance in different units. An interesting finding in the interviews was that headquarter evaluators seemed to spend a

disproportionate amount of time on what were seen as very capable and very incapable units. A typical quote from a corporate manager was: “In our company headquarter managers use an 80-20 rule. This means that 80% of the attention is spent on 20% of the subsidiaries – often the least capable ones”. What was especially mentioned by a handful of subsidiary managers confronted with non-satisfactory evaluations of their units’ capabilities were problems to find the theoretically most accurate and articulable measures for proper evaluation. Headquarters managers in general agreed that measures could in theory be improved. However, they were like most subsidiary managers reasonably happy with the measures currently used.

By far, most comments during the interviews circled around the evaluators’ interpretation and combination of information into an aggregate opinion. Managers in firms where inter-rater reliability was *high*, in their comments emphasized high levels of social interaction as an explanation. Managers specified interaction as dependent on personal friendship, long tenure, and self-selection into a firm with a strong and homogenous organization culture. Consequently, large differences in evaluators’ assessments were explained by a lack of face-to-face meetings between evaluators and the evaluated. Incentive systems that did not encourage cross-border cooperation and open exchange of ideas and experience were another explanation for low inter-rater reliability. Also infrequent job-rotation between subsidiaries and between subsidiaries and corporate levels was brought up frequently as a reason for not agreeing on the evaluation of capabilities.

Interestingly, discussion of evaluations with managers often ended in humorous apologetic remarks related to national cultures. Typical explanations why certain units would never be accepted as capable by certain evaluators were: “the Mexicans will never listen to the Americans...” and “the French always ignore the Germans...”. The degree to which evaluators liked the individual managers of evaluated units was another commonly mentioned factor hypothesized to influence the evaluations of capabilities. In an extremely strong case, a Swiss subsidiary manager expressed the following: “My unit is seen as incapable by headquarters since

my high demands on corporate support makes me problematic for the headquarters guys”. When confronted with this statement, the evaluator explained the situation as follows: “This guy is a proud gentleman who has a very high opinion of himself. He has poor self-assessment and poor judgement”.

All in all, there are a variety of possible reasons for differences between how corporate and subsidiary managers evaluate the capabilities of specific units. Interviews at a firm with few differences indicated the main reasons to be that managers stayed with the firm for a long time and that they met face-to-face quite often. In our opinion, this is associated with developing a shared interpretation of information rather than just access to information. The interviews indicate the importance of shared interpretations of information as explaining why there would not be low levels of inter-rater reliability regarding the capabilities of marketing units. However, it is still difficult to know whether deep social interaction and prolonged personal relationships lead to collective groupthink only or enlightened shared insight.

Discussion and Conclusion

Our results show that evaluations of capability in multi-unit firms are seldom highly reliable. The study thus contributes in an important way to the literature on knowledge transfer and management. Obviously it is only a first step towards understanding the identification of capable units. Empirically examining evaluations of capabilities is methodologically challenging. In our case, some readers might argue that the inter-rater differences are the result of poor measures. However, this study made use of measures designed and used on a daily basis by top management. To the extent that the measures are poor, the use of them will still have important effects.

The results of our empirical study have important implications for creating, retaining, and transferring knowledge in organizations. If inter-rater reliability is low when capabilities are evaluated in multi-unit firms, the creation of new knowledge can become problematic. Firms

often select staff for innovation projects from what is seen as the most capable units. Our results show that these perceptions may not be valid. If so, false expectations regarding competencies of members in the innovation teams may create problems in the innovation process. In multinational companies, the potential problems may be exacerbated by differing functional and national cultures.

As for retention of knowledge, it goes without saying that uncertainty and disagreement around what to retain complicates knowledge management systems. Regarding transfer, there is a risk of transferring knowledge from less capable units. This can be especially problematic if certain units are formally assigned the role of “centers of excellence”. Recently, centers of excellence have been suggested as a central part of knowledge management systems. “Centers of excellence represent the best practice of managing knowledge. Formally charged with the responsibility of leveraging and/or making knowledge available throughout the firm... the center of excellence provides a focal point for knowledge development and dissemination and replaces an old-fashioned reliance of informal, word-of-mouth mechanisms.” (Moore & Birkinshaw 1998: p. 81). Our simple question based on the empirical findings is what happens if the evaluation of the center of excellence is not uniform and uncontested?

An obvious implication for empirical researchers in knowledge management is the danger of relying on single respondents, for example one level of managers in an organization. Agreement cannot be assumed when evaluating strategically important capabilities in leading firms. One can add that differences in evaluations of capabilities obviously also have implications for motivation, incentives, as well as for reputation and careers of individuals as well as sub-units in organizations. For example, to the extent that agreement on capabilities and performance is lacking, processes of social influence will account for more of the variance in outcomes than if there is agreement (Pfeffer et al 1984).

The results also have implications for the discussion about meta-knowledge and the development of cognitive structures (Larson & Christiansen 1993). The term meta-knowledge is

used to indicate that group members can develop an understanding of the knowledge that is possessed by other group members. Our interview data suggests that the concept of meta-knowledge in the minds of multinational company managers be tied to social integration. If meta-knowledge has indeed been established in an organization through frequent face-to-face communication, personal friendship, job rotation, and supportive incentive systems, agreement as to which units are capable will be promoted. It should, however be remembered that there is still no guarantee that the admired units are necessarily the most capable. In March et al's 1991 terminology, agreement produced in these social processes may lead to reliability but not necessarily to high validity. Furthermore, low inter-rater reliability in the evaluation of capabilities is not necessarily a bad thing. Disagreement between managers and employees in a firm's different levels and units may stimulate productive discussion and may have beneficial motivational effects.

References

- Argote, L., "Group and Organizational Learning Curves: Individual System, and Environmental Components," *British Journal of Social Psychology*, 32, 31-51, 1993.
- Argote, L. *Organizational Learning – Creating, Retaining and Transferring Knowledge*, Boston: Kluwer Academic, 1999.
- Argote, L., Beckman, S. L., and D. Epple, "The Persistence and Transfer of Learning in Industrial Settings," *Management Science*, 36, 140-154, 1990.
- Bartlett, C. A., and S. Ghoshal, *Managing Across Borders – The Transnational Solution*, Cambridge, MA: Harvard Business School Press, 1989.
- Baum, J. A. C., and P. Ingram, "Survival-Enhancing Learning in the Manhattan Hotel Industry, 1898-1980," *Management Science*, 44, 996-1016, 1998.
- Conner, K. R., "A Historical Comparison of Resource-Based Theory and Five Schools," *Journal of Management*, 17:1, 121-154, 1991.
- Conner, K. R., and C. K. Prahalad, "A Resource-Based Theory of the Firm: Knowledge Versus Opportunism," *Organization Science*, 7:5, 1996.
- Darr, E., Argote, L., and D. Epple, "The Acquisition, Transfer and Depreciation of Knowledge in Service Organizations: Productivity in Franchises," *Management Science*, 41, 1750-1762, 1995.
- Dawes, R. M., D. Faust, and P. E. Meehl, "Clinical Versus Actuarial Judgment," *Science*, 243, 1668-1674, 1989.
- Epple, D., Argote, L., and K. Murphy, "An Empirical Investigation of the Micro Structure of Knowledge Acquisition and Transfer Through Learning by Doing," *Operations Research*, 44, 77-86, 1996.
- Feldman, J. M., "Beyond Attribution Theory: Cognitive Processes in Performance Appraisal," *Journal of Applied Psychology*, 66, 127-148, 1981.
- Fiske, S. T., and S. E. Taylor, *Social Cognition*, 2nd Ed. McGraw Hill Inc, New York, 1991.

- Foss, N. J., "Theories of the Firm: Conceptual and Competence Perspectives," *Journal of Evolutionary Economics*, 1993.
- Ghoshal S., and P. Moran, "Bad for Practice: A Critique of the Transaction Cost Theory," *Academy of Management Review*, 1996.
- Grant, R. M., "Toward a Knowledge-Based Theory of the Firm," *Strategic Management Journal*, 17, 109-122, 1996.
- Gupta, A. and V. Govindarajan, "Knowledge flows within multinational corporations," *Strategic Management Journal*, 21, 473-496, 2000.
- Harris, M. H. and J. Schaubroeck, "A Meta-Analysis of Self-Supervisor, Self-Peer, and Peer-Supervisor Ratings," *Personnel Psychology*, 41, 43-62, 1988.
- Haunshild, P. R., and A. S. Miner, "Modes of Interorganizational Imitation: The Effects of Outcome Saliency and Uncertainty," *Administrative Science Quarterly*, 42, 472-500, 1997.
- Hedlund, G., "The Hypermodern MNC – A Heterarchy?" *Human Resource Management*, 25:1, 1986.
- Hemphill, J. K. and L. B. Sechrest, "A comparison of Three Criteria of Air Crew Effectiveness in Combat over Korea," *Journal of Applied Psychology*, 36, 323-327, 1952.
- Holm, U., and T. Pedersen, editors, *The Emergence and Impact of MNC Centres of Excellence – A Subsidiary Perspective*. MacMillan Press, 2000.
- Hunter, J. E. and H. R. Hirsh, "Applications of Meta-Analysis." In C. L. Cooper and I. T. Robertson (eds.), *International Review of Industrial and Organizational Psychology*, 66, 127-148, 1987.
- Ilgen, D. R. and J. M. Feldman, "Performance Appraisal: A process Focus." In B. M. Staw and L. L. Cummings (eds.), *Research in Organizational Behavior*, 5, 141-197, Greenwich, Connecticut, JAI Press Inc, 1983.
- Ireland, R. D., M. A. Hitt, R. A. Bettis, and D. A. De Porras, "Strategy Formulation Processes: Differences in Perceptions of Strengths and Weaknesses and Environmental Uncertainty by Managerial Level," *Strategic Management Journal*, 8, 469-485, 1987.

- Kahneman, D., P. Slovic, and A. Tversky (eds.) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, 1982.
- King, L. M., J. E. Hunter, and F. L. Schmidt, "Halo in a Multidimensional, Forced-Choice Performance Evaluation Scale," *Journal of Applied Psychology*, 65, 507-516, 1980.
- Kogut, B. and U. Zander "What firms do? Coordination, identity, and learning," *Organization Science*, 7, 502-518, 1996.
- Kogut, B. and U. Zander, "Knowledge of the firm and the evolutionary theory of the multinational corporation," *Journal of International Business Studies*, 625-645, 1993.
- Kogut, B. and U. Zander, "Knowledge of the Firm, Combinative Capabilities and the Replication of Technology," *Organization Science*, 3, 383-397, 1992.
- Kostova, T. *Success of the transnational transfer of organizational practices within multinational companies*. Doctoral Dissertation, University of Minnesota, USA. UMI, 1996.
- Larson, J. R. and Christensen, "Groups as Problem-solving Units: Toward a New Meaning of Social Cognition," *British Journal of Social Psychology*, 32, 5-30, 1993.
- Libby, R., K. T. Trotman, and I. Zimmer, "Member Variation, Recognition of Expertise, and Group Performance," *Journal of Applied Psychology*, 72, 81-87, 1987.
- Littlepage, G. E., W. Robinson, and K. Reddington, "Effects of Task Experience and Group Experience on Group Performance, Member Ability, and Recognition of Expertise," *Organizational Behavior and Human Decision Processes*, 69, 133-147, 1997.
- Littlepage, G. E., G. W. Schmidt, E. W. Whisler, and A. G. Frost, "An Input-Process-Output Model of Influence and Performance in Problem-Solving Groups," *Journal of Personality and Social Psychology*, 69, 877-889, 1995.
- Lord, C. G., L. Ross, and M. R. Lepper, "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence," *Journal of Personality and Social Psychology*, 37, 2098-2109, 1979.

- Madhok, A., "The Organization of Economic Activity: Transaction Costs, Firm Capabilities, and the Nature of Governance," *Organization Science*, 7:5, 577-590, 1996.
- March, J. G., L. S. Sproull, and M. Tamuz, "Learning from Samples of One or Fewer," *Organization Science*, 2, 1-13, 1991.
- Miner, C. F., "Group versus Individual Decision Making: An Investigation of Performance Measures, Decision Strategies, and Process Losses/Gains," *Organizational Behavior and Human Performance*, 33, 112-124, 1984.
- Moore, K., and J. Birkinshaw, "Managing Knowledge in Global Service Firms: Centers of Excellence," *Academy of Management Executive*, 12:4, 1998.
- Murphy, K. R. and W. K. Balzer, "Systematic Distortions in Memory-Based Performance Evaluations: Consequences for Rating Accuracy," *Journal of Applied Psychology*, 71, 39-44, 1986.
- Nathan, B. R. and R. G. Lord, "Cognitive Categorization and Dimensional Schemata: A Process Approach to the Study of Halo in Performance Ratings," *Journal of Applied Psychology*, 68, 102-114, 1983.
- Nisbett, R. E. and L. Ross. *Human Inference: Strategies and Shortcomings of Social Judgment*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
- Nunnally, J. C. *Psychometric Theory*. McGraw-Hill, New York, 1978.
- Pfeffer, J., Salancik, G. R. and H. Leblebici, "Uncertainty and Social Influence in Organizational Decision Making," In: M. W. Meyer, J. H. Freeman, M. T. Hannan, J. W. Meyer, W. G. Ouchi, J. Pfeffer, and W. R. Scott, *Environments and Organizations*, San Francisco: Jossey-Bass, 306-393, 1984.
- Philips, J. S., "The Accuracy of Leadership Ratings: A Cognitive Categorization Perspective," *Organizational Behavior and Human Performance*, 28, 143-163, 1984.
- Rosenthal, R. and L. Jacobsen. *Pygmalion in the Classroom: Teacher expectations and children's intellectual development*. Holt, Rinehart & Winston, New York, 1968.

- Rothstein, H. R., "Interrater Reliability of Job Performance Ratings: Growth to Asymptote Level With Increasing Opportunity to Observe," *Journal of Applied Psychology*, 75, 322-327, 1990.
- Schmidt, F. L., J. E. Hunter, K. Pearlman, and H. R. Hirsh, "Forty questions about Validity Generalization and Meta-Analysis," *Personnel Psychology*, 38, 697-798, 1985.
- Simon, H. A. and D. C. Dearborn, "Selective Perception: The Identifications of Executives," *Sociometry*, 21, 140-144, 1958.
- Snyder, M., "When Beliefs Creates Reality," In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*: 18, 247-305. Academic Press, New York, 1984.
- Snyder, M. and W. B. Swann, "Hypothesis-testing Processes in Social Interaction," *Journal of Personality and Social Psychology*, 36, 1202-1212, 1978.
- Spender, J. C., "Competitive Advantage from Tacit Knowledge? Unpacking the Concept and Its Strategic Implications," In: B. Moingeon, and A. Edmondson, eds., *Organizational Learning and Competitive Advantage*, London: Sage Publications, 1996.
- Starbuck, W. H. and J. M. Mezias, "Opening Pandora's box: Studying the accuracy of manager's perceptions," *Journal of Organizational Behavior*, 17, 99-117, 1996.
- Stevenson, H., "Defining Corporate Strengths and Weaknesses," *Sloan Management Review*, Spring, 51-68, 1976.
- Stockford, L. and H. W. Bissell, "Factors Involved in Establishing a Merit Rating Scale," *Personnel*, 26, 94-116, 1949.
- Szulanski, G., "Exploring Internal Stickiness," *Strategic Management Journal*, 17, 27-44, 1996.
- Taylor, S. E., "The Availability Bias in Social Perception and Interaction." In D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases*: 190-200. Cambridge University Press, Cambridge, 1982.
- Trotman, K. T, P. W. Yetton, and I. Zimmer, "Individual and Group Judgments of Internal Control Systems," *Journal of Accounting Research*, 21, 286-292, 1983.

- Vernon, R., "The Product Cycle Hypothesis in a New International Environment," *Oxford Bulletin of Economics and Statistics*, 41, 255-267, 1979.
- Viswesvaran, C., D. S. Ones, and F. L. Schmidt, "Comparative Analysis of the Reliability of Job Performance Ratings," *Journal of Applied Psychology*, 81, 557-574, 1996.
- Walker, G. "Network Position and Cognition in a Computer Software Firm," *Administrative Science Quarterly*, 30, 103-130, 1985.
- Wenn, B., S. A. LaTour, and J. Bobby, "Differences in perceptions of hospital marketing orientation between administrators and marketing officers," *Hospital and Health Services Administration*, 39, 341-355, 1994.
- Winter, S., "Knowledge and Competence as Strategic Assets," In: D. Teece, editor, *The Competitive Challenge - Strategies for Industrial Innovation and Renewal*, Cambridge, MA: Ballinger, 159-184, 1987
- Yetton, P. W. and P. C. Bottger, "Individual versus Group Problem Solving: An Empirical Test of Best Member Strategy," *Organizational Behavior and Human Performance*, 29, 307-321, 1982.
- Zander, U., *Exploiting a Technological Edge – Voluntary and Involuntary Dissemination of Technology*, Stockholm: IIB, 1991.
- Zander, U., and B. Kogut, "Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: An Empirical Test," *Organization Science*, 6:1, 76-92, 1995.

Appendix: Monte Carlo simulation of p-values for a one-sided test of the null hypothesis that the correlation is 0.7

Consider a sample with n pair-wise observations with an observed correlation of r . We assume that the underlying distribution of the observations is two-dimensional normal. To calculate the p-value for the null-hypothesis that the correlation is 0.7, we first generate a list, $X = \{x_1, \dots, x_n\}$, of n normally distributed random variables with mean zero and variance 1. We then generate a list, $Y = \{y_1, \dots, y_n\}$, of n random variables where y_n is a normally distributed variable with mean rx_n and variance $(1-r^2)$. By construction the average correlation between X and Y is 0.7 (note that the distribution of the correlation coefficient is independent of the variance of X and Y , thus the variance and mean of X is arbitrary). We compute the correlation between X and Y in the random sample and note whether it is at or below 0.7. This procedure is repeated 10,000 times. The proportion of times the correlation coefficient is at or below 0.7 is the p-value.

Tables and Figures

Table 1: Overview of Firms in the Sample (Sales and Employees refer to 1998)

Name	Sales (MUSD)	Employees	Units in our study
Skandia	95	196	5
Sandvik Coromant	1042	7918	31
Sandvik Steel	1339	7660	39
Ericsson	3379	93949	46
Volvo Car Corporation	11074	29564	29
Pharmacia & Upjohn	7176	31700	26
Alfa Laval Agri	10275	4100	33

Table 2: Overview of response rates and interviews by firm.

Firm	Subsidiary (lead resp.)	Corporate	Interviews	Duration (hrs)
Skandia	5	3	15	18
Sandvik Coromant	31	4	6	9
Sandvik Steel	32	3	8	16
Ericsson	32	3	9	12
Volvo Car Corporation	25	2	7	9
Pharmacia & Upjohn	19	2	6	6
Alfa Laval Agri	32	5	5	5
Total	176	22	56	75

Table 3: Firm-specific capability dimensions.

Firm	Firm-specific activities
Firm 1	The ability to attract new customers; Underwriting; Actuarial work; Claims; and Technical service
Firm 2	New product introduction; New customer prospecting; Customer segmentation and targeting; Efficiency in inquiry and order-handling; Sourcing of products outside <Firm>; and Sales staff training and development
Firm 3	New product introduction; New customer prospecting; Distribution sales; Implementing corporate campaigns; Servicing special tools customers
Firm 4	Adding value to the "Time-To-Customer" process; Adding value to the "Time-To-Market" process; Forecasting future demand for products; The ability to integrate sales levels with manufacturing; and The ability to integrate activities with R&D units
Firm 5	Product planning; Logistics and distribution; Managing dealer relationships; Communicating the <Firm> Brand; and After-care of customers (after-sales services)
Firm 6	Strategic marketing; Sales and support to general practitioners; Sales and support to hospitals; and Promotion, advertising and public relations
Firm 7	Sales per cow; Capital-goods sales management; Service levels in after-market services; Total number of items sold; and Growth of customer base

Table 4.1: Correlation between evaluations of units' capabilities by subsidiary and corporate managers for Firm 1⁵. Boldface is used to p-values > 0.05 and upper bounds >= 0.7.

Dimension	N	Correlation	p-value	Upper bound of confidence interval
The ability to attract new customers	5	-0.071	0.053	0.580
Claims handling	5	-0.375	0.0193	0.327
Technical service	5	-0.253	0.0296	0.442

Table 4.2: Correlation between evaluations of units' capabilities by subsidiary and corporate managers for Firm 2. Boldface is used to p-values > 0.05 and upper bounds >= 0.7.

Dimension	N	Correlation	p-value	Upper bound of confidence interval
New-product introduction	25	+0.280	0.0034	0.548
New-customer prospecting	25	+0.266	0.0024	0.537
Distribution sales	24	+0.780	0.7767	0.881
Implementing corporate campaigns	24	+0.292	0.0044	0.562
Servicing special tools customers	24	+0.300	0.005	0.568

Table 4.3: Correlation between evaluations of units' capabilities by subsidiary and corporate managers for Firm 3. Boldface is used to p-values > 0.05 and upper bounds >= 0.7.

Dimension	N	Correlation	p-value	Upper bound of confidence interval
New-product introduction	32	+0.207	0.0006	0.462
New-customer prospecting	32	-0.041	0	0.244
Customer segmentation and targeting	32	-0.064	0	0.222
Efficiency in inquiry and order handling processes	32	+0.412	0.0088	0.622
Sourcing of products outside <Firm>	32	+0.230	0.0002	0.481
Sales staff training and development	32	+0.045	0	0.323

Table 4.4: Correlation between evaluations of units' capabilities by subsidiary and corporate managers for Firm 4⁶. Boldface is used to p-values > 0.05 and upper bounds >= 0.7.

Dimension	N	Correlation	p-value	Upper bound of confidence interval
Adding value to the "Time to Market" process	11	+0.571	0.2228	0.816
Adding value to the "Time to Customer" process	11	+0.308	0.0426	0.671
Forecasting future demand for products	20	+0.219	0.0033	0.529
Ability to integrate sales levels with manufacturing	11	+0.233	0.0271	0.624

Table 4.5: Correlation between evaluations of units' capabilities by subsidiary and corporate managers for Firm 5. Boldface is used to p-values > 0.05 and upper bounds >= 0.7.

Dimension	N	Correlation	p-value	Upper bound of confidence interval
Product planning	23	+0.219	0.0019	0.511
Logistics and distribution	23	+0.213	0.0013	0.507
Managing dealer relationships	24	+0.457	0.0383	0.680
Communicating the <Firm> brand	24	+0.214	0.0018	0.502
After-care of customers (after-sales service)	24	+0.488	0.0505	0.700

Table 4.6: Correlation between evaluations of units' capabilities by subsidiary and corporate managers for Firm 6. Boldface is used to p-values > 0.05 and upper bounds >= 0.7.

Dimension	N	Correlation	p-value	Upper bound of confidence interval
Strategic Marketing	10	-0.138	0.0035	0.362
Sales and support to General Practitioners	15	+0.333	0.0279	0.647
Sales and support to Hospitals	16	+0.572	0.1882	0.786
Promotion, advertising and public relations	16	-0.022	0.001	0.370

Table 4.7: Correlation between evaluations of units' capabilities by subsidiary and corporate managers for Firm 7. Boldface is used to p-values > 0.05 and upper bounds >= 0.7.

Dimension	N	Correlation	p-value	Upper bound of confidence interval
Sales per cow	31	+0.486	0.0305	0.678
Capital-goods sales management	30	+0.425	0.014	0.637
Service levels in after-market services	28	+0.564	0.1039	0.739
Total number of items sold	28	+0.634	0.2444	0.785
Growth of customer base	30	+0.059	0	0.344

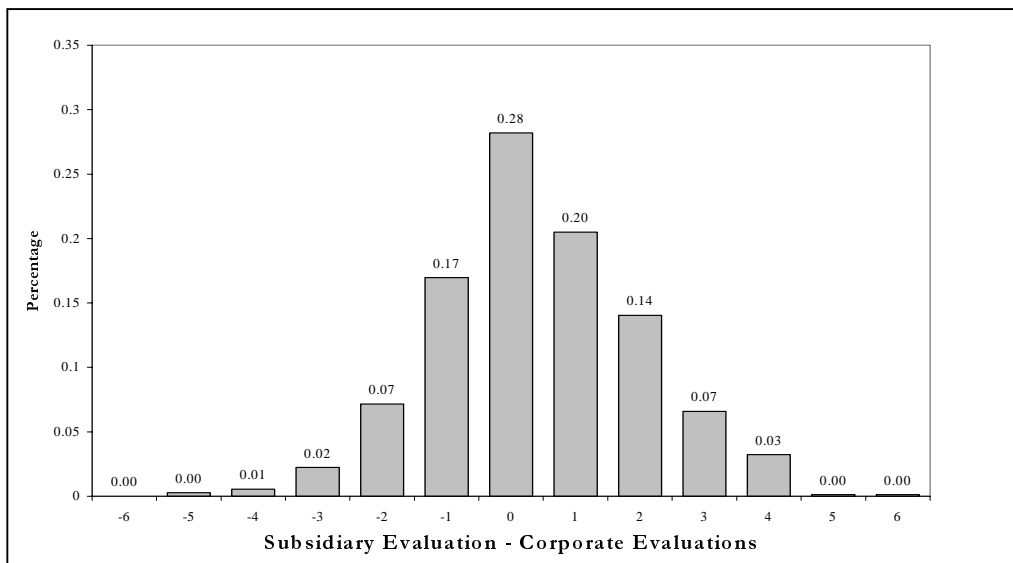


Figure 1: Frequency distribution of the difference between subsidiary and corporate evaluations. N = 704.

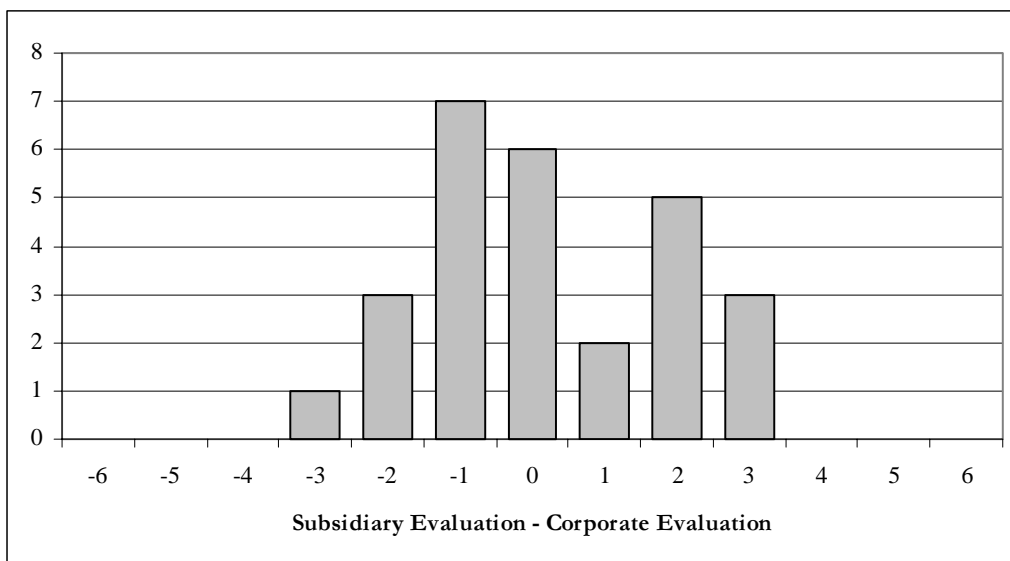


Figure 2: Frequency distribution of the difference between subsidiary and corporate evaluations for one capability dimension in one firm.

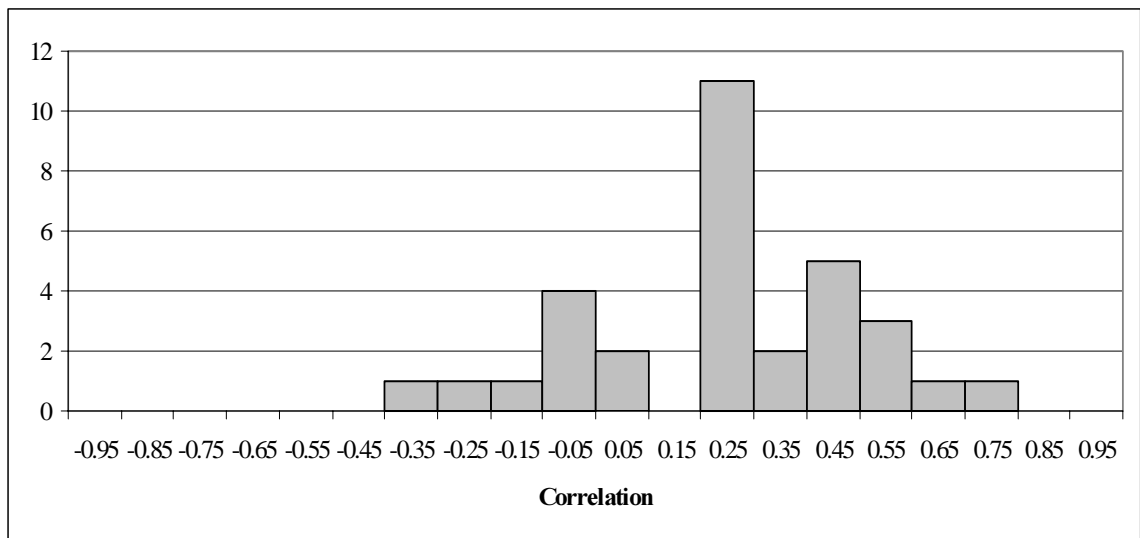


Figure 3: Histogram of the (Pearson) correlation between subsidiary and corporate evaluations. $N = 32$. The x-axis shows the midpoint of an interval with width 0.1. Intervals include their upper boundary but not their lower boundary.

Endnotes:

¹ Winter (1987), Zander (1991), Conner (1991), Kogut and Zander (1992), Kogut and Zander (1993), Foss (1993), Zander and Kogut (1995), Spender (1996), Conner and Prahalad (1996), Grant (1996), Madhok (1996), and Kogut and Zander (1996).

² Approximate confidence intervals for correlation coefficients can be obtained by using the fact that $(1/2)\ln((1+r)/(1-r))$, where r is the correlation coefficient, is approximately normally distributed.

³ Ericsson is a Swedish-based leading global actor in telecommunications including mobile telephone operations. Pharmacia & Upjohn is a major US-Swedish pharmaceutical company. Volvo Cars has a Swedish background but was recently acquired by GM. Steel and Coromant are independent divisions within the Swedish leading steel and tooling firm Sandvik. Skandia is a leading Swedish insurance company. DeLaval, which was called Alfa Laval Agri at the time of our study, is a global leader in the milking industry and is a division of the Tetra-Laval Group.

⁴ Since there were no responses regarding three capability dimensions (two in one firm and one in another firm) they were excluded from the analysis below.

⁵ Since there were no responses regarding the capability dimensions "Underwriting" and "Actuarial work", these are excluded from the analysis.

⁶ Since there were no responses regarding the capability dimension "Ability to integrate activities with R&D units", it is excluded from the analysis.