# **Evaluation and Leadership**

An Explorative Study of Differences in Evaluative Style

Jan Noeverman and Bas A.S. Koene

#### 1 Introduction

Many large organizations have staff departments where people are being paid to develop, implement and maintain all kinds of systems that are intended to influence and motivate participants to behave in the best interest of the organization. High investments are made in sophisticated management control systems, performance evaluation systems, accounting information systems, et cetera. But do these systems really affect people in the way that they are intended to?

Some early studies (Argyris, 1953; Hopwood, 1972) found that using accounting information for evaluating the performance of subordinate managers could enhance feelings of tension and evoke dysfunctional behaviours such as manipulating accounting information and inferior relationships with peers and superiors. These studies also suggest that participants' behaviour is not affected by the adequacy of the system only, but as much by how managers use these systems. In the past few decades, many studies have tried to relate the way in which managers evaluate the performance of their subordinates to behavioural and attitudinal outcomes at the subordinate level. Especially the reliance on accounting performance measures (RAPM) in evaluating performance has been a key focus of research. This so-called RAPM literature, building on studies by Hopwood (1972) and Otley (1978), is problematic in several ways. The main problem is the measurement of one of the key variables in this line of research, i.e. performance evaluation style. Performance evaluation style or evaluative style refers to the manner in which superiors evaluate the performance of their subordinates.

This paper reports some preliminary findings of a research project exploring differences in how leaders evaluate the performance of subordinates in a Dutch food-processing organization, and the causes and consequences of these differences. The research project aims at identifying and categorizing different patterns of managers' evaluative style qualitatively, through interviews with both subordinate managers and their superiors, and quantitatively, administering a paper and pencil survey to all the subordinates in the units of the managers under study. Since the interview data were not available yet for analysis in this paper, the analysis here is limited to a quantitative analysis of how perceived and preferred evaluative style, as indicated by subordinates, relate to each other and to the (perceived) leadership style of their managers (their direct superiors). The data for this analysis were gathered through a questionnaire that was completed by the subordinates working in the units that were also included in the qualitative part of this study.

Section two of this paper reviews the literature and provides the theoretical background of the research project. The research method for the project as a whole, and for the quantitative analysis in this paper will be reported in section three. Section four reports our findings on the relationship between perceived and ideal evaluative style of subordinates and their manager's leadership style. Finally, in section five we will discuss findings and shortcomings of this study, and how some of the shortcomings of the present study will be addressed in the qualitative part of this study, and we will point out directions for further research.

Drs. J. Noeverman is a university teacher at the section Financial and Management Accounting of the Faculty of Economics, Erasmus University Rotterdam. Dr. B.A.S. Koene is a university teacher at the section Organization of the Faculty of Economics, Erasmus University Rotterdam.

# 2 Evaluative Style in Context: Theoretical Background

# 2.1 Towards a New Measure of Performance Evaluation Style

Much research has been undertaken into the behavioural consequences of using accounting performance measures for performance evaluation purposes. Many of these studies tried to reconcile the contradictory findings from two leading studies in this field: a study by Hopwood (1973) and the replication of this study in a slightly different organisation by Otley (1978). In spite of the large number of studies in this field, a precise understanding of the processes underlying some of the findings is still lacking. This is not surprising, since there are numerous problems with the existing literature. Hartmann and Moers (1999), for example, have argued that the statistical testing in many studies in this field is inappropriate, sometimes even incorrect, making interpretation and comparison of the findings difficult. As important as this conclusion may be, better statistical testing in the future is no guarantee for (more) meaningful findings in this field in the future. A more important issue is the measurement of performance evaluation style in the literature. Many studies have used the instrument developed by Hopwood (1973), some (e.g. Ross, 1994; 1995) in an uncritical manner, others (e.g. Brownell (1985), Brownell and Hirst (1986)) in confusing ways by making such modifications minor and major to the instrument that it becomes uncertain to what extent this modified instrument measures the same or different constructs as those measured by the instrument in earlier studies. Unfortunately, this has led to contradicting and confusing findings in literature, which are hard to interpret, understand and compare. In a recent paper, Otley and Fakiolas (2000) review the measurement of performance evaluation style in the literature. They divide RAPM studies into four groups, based on the similarity of the method used to measure evaluative style to that used by Hopwood (1973). Since they discuss both Hopwood's instrument, and other instruments in literature in great detail, for our purpose it will suffice to limit our discussion to the conclusions from their review.

Firstly, they conclude that the different instruments in the four groups differ with respect to the items they use. Furthermore, even studies using the same instrument differ in whether they use ranking scores, absolute scores, contrasts or an aggregate score in further analysis.

Secondly, different measures also represent

different underlying concepts.

Finally, Hopwood's original instrument was carefully developed to measure different styles within a specific organization. The instrument matched the language in the organization, and the items he used to differentiate between the different styles had distinctive meanings for the participants in the organization. Hopwood used it to effectively categorize complex phenomena that he observed in the organization. The simple instrument succeeded in distinguishing between different styles of evaluation, but these styles were not as simple as the instrument to measure them. It is questionable whether the uncritical application of Hopwood's measure in cross-sectional questionnaires in later studies is meaningful without prior investigation that any single item on the scale means the same thing in each organization and that the terminology captures the same phenomena.

These conclusions make clear that future studies on the behavioural consequences of evaluative style cannot rely on existing scales to measure evaluative style, unless researchers are able to account for the reliability, relevance and meaning of the concepts captured by the instrument. In future research, careful consideration should be given to the concept and measurement of evaluative style, and future categorizations should be based on distinctions that are meaningful in the particular organization studied.

Yet, this does not imply that existing scales to measure performance evaluation style and distinctions made between styles are necessarily useless for future studies. For, as Otley and Fakiolas (2000) state, 'a number of distinct dimensions to evaluative style are becoming apparent, and effort should be devoted to constructing better measurement instruments to detect these.' According to Otley and Fakiolas, these dimensions are:

- (a) Hopwood's rigid use of budgets versus a more flexible use:
- (b) Hopwood's short-run versus long-run emphasis (which may be a subdimension of (a) above);
- (c) As above, but referring to any quantitative targets, not just financial budgets;
- (d) An emphasis on absolute (quantitative) measures of performance rather than comparisons with pre-set targets.

In our opinion, the extension of evaluative style to other quantitative targets, not just financial budgets (see (c) above) could be important. There is growing evidence that the role of financial budgeting and budgetary targets for motivational and

appraisal purposes is limited in contemporary organizations (see for example, Marginson, 1999). Performance measurement frameworks like the Balanced Scorecard (Kaplan and Norton, 1992; 1993, 1996) and the Performance Pyramid (Lynch and Cross, 1992) are only a few examples of the development in the accounting field towards the inclusion of more non-financial measures in performance measurement systems. It is likely to assume that organisations that have included nonfinancial items in their measurement system will use some of these measures for performance evaluation purposes too. And even when superiors do not use such measures in performance evaluation, the mere existence of these measures may affect the perceived style of evaluation. Otley (1978) showed a similar phenomenon for financial performance measures. Although his findings show that the managers' intended evaluative style corresponded reasonably with the style as perceived by his subordinates, but that these same subordinates significantly overrated the importance of quantitative financial measures in the manager's evaluative style, as soon as these measures were available. Future studies of evaluative style should therefore be aware of the possibility that non-financial quantitative performance measures play an important role in perceptions of performance evaluation in contemporary organizations.

Although the dimensions identified by Otley and Fakiolas (2000) could serve as a starting point towards a new measure of evaluative style, unfortunately, they do not discuss how these different dimensions relate to different patterns of evaluative style. For example, does 'a rigid use of budgets' imply an evaluation characterised by comparisons against pre-set targets, and does 'a flexible use of budgets' imply an emphasis on absolute (quantitative) measures of performance? Is 'a rigid use of quantitative targets' short-term oriented and 'a flexible use' long-term oriented? But most important of all, how can we make sure that these, and not other dimensions, are important when superiors evaluate the performance of their subordinates?

To explore relevant characteristics and dimensions of evaluative style, and the relationships between different dimensions, we studied 12 superiors and their subordinates in twelve units of a Dutch foodprocessing company (for details on the sample, see section 3). Evaluative styles were charted using interviews with managers and subordinates, and with a questionnaire survey amongst the subordinates.

The interviews will be used to develop a new

categorization of evaluation styles trying to take some of the shortcomings of the research to date into account.

The questionnaire, amongst other things, measured aspects of perceived performance evaluation style that seemed theoretically relevant when trying to identify different patterns in how superiors evaluate their subordinates. Furthermore, it measured leadership characteristics of the 12 superiors. This paper reports results from the survey, exploring three issues regarding perceived performance evaluation style of superiors. First, the difference between perceived (real) and preferred (ideal) evaluative style. Second, the issue whether perceived evaluative style is an individual or group level phenomenon. Third and finally, the relationship between leadership characteristics and aspects of performance evaluation style.

## 2.2 Preferred (Ideal) and Perceived (Real) Evaluative Style

What seems to us has been a very important observation by Otley (1978) and neglected to a great extent in later research, is his repeated conclusion that the prevalent norms and values of the organization studied exert a significant effect on the appraisal of a manager's evaluative style (p.131, 133, and 145). Otley suggests this is the most important reason for the assumption that the effects found by Hopwood are conditional on the organisational context. The importance of norms and values is shown in Otley's study in two of his conclusions:

- 1 'There thus appears to be a norm of equal influence on budget setting which, when violated in either direction (either by the imposition of a budget or by leaving it mainly for the unit manager to set himself), causes increases in felt tension.' (1978: 131)
- 2 'The hypothesis that job-related tension is associated with the extent to which a manager is in agreement with the way in which his performance is evaluated receives further support from comparing the criterion of evaluation a manager considers ought to be most important in his evaluation with the criterion he has ranked as being the three most important in practice... As a manager increasingly disagrees with the appropriateness of the major criterion used to evaluate his performance, so his feelings of job-related tension increase.' (1978:132)

In summary, 'there is some evidence that such tension is associated with the degree of agreement a manager has with the criteria being used to evaluate his performance and the extent to which he considers he ought to participate in setting his own budget.' (1978: 132)

The discussion above suggests that it is disagreement with evaluation criteria that could help us explain behavioural and attitudinal consequences of performance evaluation. Therefore, we asked respondents to indicate both the importance that their superior attaches to different aspects of the performance evaluation process (real evaluative style), and how much emphasis they think he should attach to those items (ideal evaluative style). This enables us to calculate a comparison score for different aspects of evaluation, which may be used as a measure of the extent in which an individual agrees with how he is evaluated. In this paper, comparisons are made between mean preferred and mean perceived importance scores on different aspects of evaluative style at organisational and group-level. We expect that on average some aspects of performance evaluation receive more emphasis than participants feel those aspects should receive, while other aspects on average receive less emphasis than participants prefer, and still other aspects receive the same emphasis as prefered. However, it is impossible to predict for which aspects the preferred emphasis will be higher than. lower than, or equal to the perceived emphasis.

Furthermore, since the mean preferred importance score on each aspect of evaluation is assumed to reflect the norms and values held by participants in this organisation regarding performance evaluation, we expect that there is more agreement among subordinates on the preferred importance scores on aspects of evaluative style than on the perceived importance scores on those aspects. In short, we expect:

- 1 There are significant differences between the mean perceived and preferred importance scores of some but not all aspects of performance evaluation; for some aspects these differences are positive, for others negative.
- *II* There is more agreement among subordinates on the preferred importance scores of aspects of evaluative style than on the perceived importance scores on those aspects.

#### 2.3 Individual or Group Level Phenomenon

Another issue that has not received much attention in the literature on evaluative style is whether evaluative style should be understood as a phenomenon occurring at the level of individuals, or at the level of groups. Is there significant agreement among subordinates on the evaluative style of their superior, or are there large differences? Based on the limited evidence available (Hopwood (1973) and Otley (1978)), with respect to the reported perceived importance attached to aspects of evaluative style, we expect that there will be considerable within- and between-group differences. Therefore, in this paper we explore whether the perceived emphasis on different aspects of evaluative style differs across leaders, and whether there is (considerable) agreement among the scores of subordinates working under the same superior. However, since we argued before that preferred evaluative style is supposed to reflect norms and values, we do not expect the scores on the items measuring preferred evaluative style to vary across leaders. Therefore, in this paper the following propositions are explored:

- *III There are significant differences in means between leaders on the items in the perceived evaluative style.*
- *IV* There is considerable agreement in the importance scores on the items in the perceived evaluative style within groups of subordinates reporting to the same superior.
- *V* There are no significant differences in means between leaders on the items in the preferred evaluative style.

#### 2.4 Leadership Style and Evaluative Style

Evaluating subordinates is just one of many things expected by leaders in an organisation. How a superior evaluates his subordinates, i.e. his evaluative style, is most likely part of the more general way in which a superior approaches his job as a leader. Therefore, we expect that differences in evaluative style are part of more general differences in leadership style. Many different definitions of leadership exist, but in this study we adopted Den Hartog's (1997, 2) definition of leadership as 'the ability of an individual to influence, motivate, and enable others to contribute toward the effectiveness and success of the organisations of which they are members.' Although leadership style has been largely ignored in the management control and accounting literature (Otley and Pierce, 1995), two studies in the accounting literature on evaluative style address the relationship between evaluative style and leadership. Using the two dimensions of leadership style taken from the Ohio State University leadership studies, i.e. 'Consideration' and 'Initiating Structure', Hopwood (1973) was able to show that respondents gave both budgetconstrained and profit conscious supervisors high ratings on the 'Initiating Structure' dimension,

whereas non-accounting supervisors were given significantly lower ratings on this dimension. Furthermore, profit-conscious and non-accounting supervisors were given significantly higher ratings on the 'Consideration' dimension than budgetconstrained supervisors were. Using the same two dimensions in a survey under all audit seniors in three Big Six audit firms, Otley and Pierce (1995) concluded that a leadership style characterised by high structure and low consideration was associated with the highest level of dysfunctional behaviour (under-reporting of time, and audit quality reduction behaviour). In contrast, a leadership style characterised by low structure and high consideration was associated with the lowest level of dysfunctional behaviour.

When studying the relationship between leadership and aspects of evaluative style, however, some consideration should be given to the development in the leadership field. Research on leadership has changed from an emphasis on exchange, reward and control towards an emphasis on vision, value transformation, symbolic behaviour and management of meaning. Recent leadership theories attempt to explain how leaders are able to change people rather than respond to them, and leaders are seen as effective when they succeed in changing and creating the environment, rather than complementing the environment. These new types of leadership are often referred to as 'transformational' and 'charismatic' leadership, as opposed to the traditional 'transactional' or 'instrumental' leadership styles. Recent-

#### Table 1: Leadership Subscales in the ILO-Questionnaire

#### **Inspirational scales**

- 1 Vision (9 items)
- 2 Charisma (8 items)
- 3 Intellectual stimulation (7 items)
- 4 Individualized consideration (5 items)
- 5 Motivating and performance expectations (13 items)
- 6 Role modelling (3 items)
- 7 Demonstrating trust in subordinates (3 items)
- 8 Teambuilding (5 items)

#### Transactional scales

1 Contingent reward ( 4 items)

2 Active management-by-exception (8 items)

#### Passive leadership (laissez faire)

1 Passive (4 items)

#### Other leadership scales

1 Integrity (7 items) 2 Participative leadership (6 items)

Source: Den Hartog, 1997

ly, a new Dutch questionnaire, the Inspirational Leadership in Organisations (ILO) questionnaire was developed to measure thirteen a-priori dimensions underlying these types of leadership (see Den Hartog, 1997 for a detailed discussion on the development of this questionnaire, and the theoretical details behind it). Table 1 provides an overview of the 13 a-priori scales, and the type of leadership they refer to.

Since inspirational leadership indicates a leader who is motivating, stimulating, supporting, changing his subordinates, inspirational leadership emphasises interpersonal aspects of leadership. Therefore, we expect that inspirational leaders will tend to attach high importance to interpersonal, qualitative aspects of performance evaluation, such as subordinates' explanations of their performance, possibilities to improve/develop their performance, their plans for the years to come, et cetera. On the other hand, transactional leaders point out to subordinates 'what is expected of them and providing them with feedback when their behaviour is not up to standard' (Den Hartog, 1997: 50). Therefore, we expect transactional leaders to attach high importance to quantitative measures of performance and deviations from targets.

Since we have already argued that if preferred evaluative style reflects norms and values held within the organization, preferred evaluative style does not vary across leaders (section 2.4), we expect that preferred evaluative style is not affected by leadership.

Finally, regarding the level of analysis, leadership too could be studied at the individual and group level. Therefore, in this study we explored both the agreement among subordinates regarding leadership style of their superiors, and differences in mean scores on the different subscales of leadership among leaders. To summarize, with regard to leadership, we expect that:

- *VI* The more supervisors are seen as being inspirational leaders, the more importance will be attached to interpersonal, qualitative aspects of performance evaluation.
- VII The more supervisors are seen as being transactional, the more importance will be attached to quantitative measures of performance and deviations from targets.
- *VIII There is no significant relationship between leadership and preferred evaluative style.*
- *IX* There are significant differences in means of the leadership subscales between leaders.
- *X* There is considerable agreement on the

leadership subscale scores within groups of subordinates reporting to the same superior.

# **3 Research Method**

In this section we describe our research method. The first two paragraphs of this section discuss research site and sample. The last two sections variable measurement and data analysis.

# 3.1 Research Site

As a first criterion for the contemporary organization in which we wanted to do our research on different patterns of performance evaluation style, the research site should have similar units, and preferably be a branch-type organization. The limitation to one single organisation and the similarity of units were deemed necessary to make sure that differences that were measured could not be attributed to differences in the performance evaluation system, but would be traceable to differences in use of the system.

Based on this criterion, we were able to get access to a large Dutch food-processing organization. The specific organisation for this study was selected for two reasons:

- 1 we expected that non-financial measures played an important role in this organization;
- 2 it is a large organization, with several production sites, which makes it relatively easy to interview managers in the same functions, at the same level in the organisation who report to and are evaluated by superiors in the same function, and all work under the same performance evaluation system.

Tal	ole 2:	Number	of	Subordinates	per	Superior
-----	--------	--------	----	--------------	-----	----------

Superior	Function*	Sample	Response
(n = 12)		(n = 75)	(n = 56)
1	М	2	2
2	М	4	4
3	Р	8	6
4	Р	16	10
5	Р	6	5
6	Р	10	10
7	М	3	3
8	Р	6	4
9	М	3	2
10	Р	7	3
11	Р	8	5
12	М	2	2
* M = marke	ting; P producti	on	

# NOVEMBER 2000

#### 3.2 Research Method and Sample

As was clarified before, the research in this paper is part of a larger study exploring different patterns of evaluative style, and it's causes and consequences. An extensive field study was conducted, using information from interviews, organizational documents and questionnaires. In total, twelve superiors and their subordinates were asked to participate in this research project. Seven superiors were Works Managers, i.e. heads of production sites: the other five superiors were Marketing Managers, i.e. heads of marketing departments. These twelve superiors and their units were selected by the company's HR manager, based on our request that we wanted around ten different units showing as much similarity as possible, but at the same time with expected differences in evaluative style. In each unit, we attempted to interview two or three subordinates reporting to the superior, who were working in a management function. This criterion was applied to warrant the comparability of responses. It was important since at the production sites not all subordinates working under the Works Manager held a management function: production heads also reported to the Works Manager, but they were not considered managers. Therefore, at the production sites only subordinates in a staff function were interviewed.

#### Questionnaire Survey

A questionnaire in Dutch was used, measuring concepts such as job related tension, upward mobility aspiration, trust, leadership style subscales, and evaluative style. The questionnaire helps to assess the significance of the relationships between different variables explored in this project. Furthermore, the questionnaire enabled us to include subordinates in our research project that could not be interviewed. The questionnaire was left with the respondent directly after the interview with the request to complete it sometime during the next week and send it back. A return envelope was provided. After interviewing subordinates that reported to one superior, a questionnaire was sent to all other subordinates reporting to the same superior, being a manager or not. In total, 75 questionnaires were sent, and 60 questionnaires returned. Four questionnaires were omitted from the sample due to incomplete answers on the key variables, i.e. real and ideal evaluative style. The criterion was that for each of these two variables containing 18 items at least 14 items should have been filled out. If less than 14 items were filled out, the questionnaire was omitted from the sample. This applied to four questionnaires, which all contained no answers to all

18 items. Of the remaining 56 questionnaires, 6 lacked an answer on the item 'performance delivered in the past few years (in de afgelopen jaren geleverde prestaties)'. Further analysis of these six questionnaires revealed that those six respondents had either been working in their present function less than a year or they had been working under their current supervisor for less than a year, which made this item inapplicable. This however applied to 21 other individuals who did answer this question. But because it is doubtful whether these answers have real meaning, we decided to drop this item in the further analysis. Of the 56 questionnaires returned, 13 questionnaires came from marketing managers, 18 from managers at a production site, and 25 from production heads. The number of questionnaires sent and returned per superior are reported in Table 2.

## 3.3 Variable Measurement

*Perceived Performance Evaluation Style* To measure perceived performance evaluation style, respondents were asked the following question:

When evaluating your performance as a manager (actual situation), how much emphasis does your superior put on (Als uw baas uw prestaties als manager beoordeelt (huidige situatie), hoeveel nadruk leg hij/zij dan op):

- 1 Short-term goals (korte termijndoelen).
- 2 Long-term goals (lange termijndoelen).
- 3 Financial information (financiële informatie).
- 4 Non-financial information (niet-financiële informatie).
- 5 Deviations from agreed performance (afwijkingen van afgesproken prestaties).
- 6 Your explanation of your performance (uw uitleg van uw prestaties).
- 7 Objectively measurable performance (objectief meetbare prestaties).
- 8 His/her own intuition (zijn/haar eigen intuïtie).
- Performance delivered in the past year (in het afgelopen jaar geleverde prestaties).
- 10 Your plans for the next year (uw plannen voor het komende jaar).
- 11 Positive aspects of your performance (positieve aspecten van uw presteren).
- 12 Negative aspects of your performance (negative aspecten van uw presteren).
- 13 Performances of the unit that you are responsible for ((prestaties van de eenheid waarvoor u verantwoordelijk bent.)
- 14 Personal, individual performance (persoonlijke, individuele prestaties).
- 15 Causes of failings in performance (oorzaken

voor het achterblijven van prestaties).

- 16 Possibilities to improve/develop your performance (mogelijkheden tot verbetering/ontwikkeling van uw prestaties).
- 17 Information he/she gets from the conversation with you (informatie die hij/zij ontvangt uit het gesprek met u.)

Respondents were asked to answer on a 5-point Likert-scale with 1 = no, 2 = low, 3 = some, 4 = high, 5 = very high.

## Preferred Performance Evaluation Style

To measure preferred performance evaluation style, respondents were asked: 'When evaluating your performance, how much emphasis should your boss in your opinion in the ideal situation put on:' (Hoeveel nadruk zou uw baas bij het beoordelen van uw prestaties naar uw mening in de ideale situatie moeten leggen op:), followed by the same 17 items as described above.

# Leadership Style

To measure leadership style, four subscales (21 items) from the ILO-questionnaire developed by Den Hartog (1997) were used<sup>1</sup>. The a-priori subscales used in our study to measure leadership were active management-by-exception, demonstrating trust and confidence in subordinates, individual consideration, and team-building. Of these scales, active management-by-exception measures an aspect of transactional leadership, while the other three subscales all measure aspects of inspirational leadership. Sample items from each scale can be found in Box 1. These subscales were selected because we expected them to measure aspects of leadership that were seen as important in the particular organization (team-building), or were most likely related to evaluative style.

Box 1 Sample items from the ILO-subscales used in this study

_	Active management-by-exception
	'points it out to me when my work is not up to par'
_	Demonstrating confidence and trust in
	subordinates
	shows confidence in my ability to contribute
	to the goals of this unit'
_	Individualized consideration
	'is genuinely concerned about the growth
	and development of subordinates'
_	Team-building
	'develops teamspirit among employees'

Since the ILO-questionnaire is only recently

developed, some consideration should be given to the measurement and internal consistency of the subscales used. First, using the same procedure as Den Hartog (1997), the missing values, if less than 25% in a scale, were replaced with the rounded mean of the respondent's score on the other items in the a-priori scale. Next, exploratory factor analysis was done to assess whether the same factor structure as described by Den Hartog (1998) would be found. Principal component factor analyses were done, and the factor structure found after varimax rotation was interpreted. Although five factors had an eigenvalue larger than one, both the three and the four factor solutions were well interpretable. In this study, we interpreted the three-factor-solution for further use in our analyses, because differences between this solution and the four a-prior subscales could theoretically be explained. The three empirical factors found are reported in Table 3. Three items of which the difference between factor loadings on two factors was less than .20 were discarded.

Factor 1 is identical to the a-priori scale for teambuilding. Factor 2 is almost identical to the a-priori scale for active management-by-exception. The only difference is that the item 'expects a lot from us' is dropped. This item also had the lowest item-rest correlation in Den Hartog's study and was not part of the a-priori scale for active management-by-exception in Den Hartog's study. Den Hartog placed it in this scale because it correlated higher with this scale than with the a-priori scale it was in originally. Finally, the third factor found in our data contains items from the other two a-priori scales, i.e. demonstrating trust and confidence in subordinates and individualized consideration. These two factors were highly correlated in Den Hartog's study (r=.76, p<.01) and loaded on a single factor, combined with the participation subscale that was not measured in our study, in a second-order factor analysis of the inspirational scales (see Den Hartog, p.126). Collapsing those two subscales into a single scale seems therefore justified.

Finally, Cronbach's alphas, average inter-item correlations and corrected item-rest correlations were calculated to test the internal consistency of the three empirical scales. The criteria were that Cronbach's alpha should be >.70, and item-rest correlations and average inter-item correlation should be >.30.

#### 3.4 Data Analysis

To test the differences between the perceived (real) and preferred (ideal) emphasis on the 17

items of evaluative style (proposition 1), we conducted paired samples t-tests. Proposition 2, stating that we expected more agreement among subordinates on the preferred importance scores of aspects of evaluative style than on the perceived importance scores, was not actually tested, but as an indication we looked at the standard deviations of the 17 items of perceived and preferred evaluative style.

To test whether there were significant differences between leaders in mean scores on the items of perceived evaluative style (prop. III), on the items of preferred evaluative style (prop. V), and on the three leadership subscales (prop. IX) Ftests using one-way ANOVA were done. To test the extent to which subordinates within groups agreed in their ratings of the perceived importance of aspects of evaluative style (prop. IV), preferred importance of aspects of evaluative style, and leadership (prop. X), Eta squared was calculated. Eta squared is the proportion of variance in the dependent variable that is explained by differences among groups. It is the ratio of the betweengroups sum of squares and the total sum of squares. The value of Eta squared could serve as a measure of the homogeneity of observations within classes, relative to between classes. The higher the value of Eta squared, the more variability in style is accounted for by variation among leaders. Finally, to explore our expectations about the pattern of relationships between leadership subscales and evaluative style items (prop. VI and prop. VII), Pearson correlations between the 17 evaluative style items and the three leadership subscales were calculated. Pearson correlations were also used to test the relationship between leadership and preferred evaluative style (prop. VIID.

#### 4 Results

#### 4.1 Perceived and Preferred Evaluative Style

The results of the paired samples t-test to test for differences in perceived and preferred evaluative style are reported in the left-hand side of Table 4. Table 4 shows that all differences but one are significant at the 5% level. Surprisingly, of the 16 significant differences only two differences indicate that in the actual situation a higher emphasis is placed on this item than preferred. These two items are 'short term goals' and 'his/her intuition'. All other items on average receive less emphasis than preferred. However, from our expectations that perceived evaluative style varies across leaders, but preferred evaluative style does not, it follows that we expect differences between

Factor 1: Team-building (team)	factor loading	item-rest correlation
Encourages employees to be 'team players'	.90	.85
(Spoort medewerkers aan om 'teamspelers' te zijn)		
Develops teamspirit among employees	.88	.83
(Ontwikkelt teamgeest bij medewerkers)		
Gets the group to work together for the same goal	.86	.81
(Krijgt de groep tot samenwerking voor hetzelfde doel)		
Works at creating a climate of trust among members of	.85	.83
the management team		
(Tracht een klimaat van vertrouwen te creëren onder de leden		
van het management-team)		
Breaks down barriers to communication between work groups	.75	.73
(Doorbreekt communicatie-barrières tussen werkgroepen)		
5 items Cronbach alpha = .93 average inter-item correlation 7	2 mean = 3.67 science	d = .90
Factor 2: active management-by-exception (ambe)		
Points it out to me when my work is not up to par	.83	.76
(Wijst mij erop wanneer mijn werk onder de maat is)		
Monitors performance for errors needing correction	.82	.73
(Let op fouten in de prestaties die correctie behoeven)		
Shows his or her displeasure when my work is below	.73	.60
acceptable standards		
(Toont zijn/haar ongenoegen wanneer mijn werk onder		
aanvaardbare normen is)		
Focuses attention on irregularities, exceptions and	.73	.67
deviations from what is expected of me		
(Vestigt de aandacht op onregelmatigheden, uitzonderingen		
en afwijkingen van wat van mij verwacht wordt)		
Would indicate disapproval if I performed at a low level	.70	.59
(Zou afkeuring laten blijken als ik op een laag niveau zou presteren	n)	
Focuses attention on errors I make	.68	.60
(Vestigt de aandacht op vergissingen en afwijkingen die ik bega)		
Keeps careful track of mistakes	.66	.53
(Houdt fouten goed in de gaten)		
7 items , 1 item dropped; Cronbach alpha = .87 average inter-it	em correlation .48	$mean = 3.22 \ sd = .68$
Factor 3: supportive (support)		
Treats me as an individual rather than just a		
member of the group	.73	.46
(Behandelt mij als een individu, in plaats van als zomaar		
een lid van de groep)		
een lid van de groep) Listens to my concerns	.67	.64
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn)	.67	.64
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals	.67	.64
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen	.67	.64 .39
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen)	.67 .66	.64 .39
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen) Shows confidence in my ability to contribute to the goals of	.67 .66	.64 .39 .61
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen) Shows confidence in my ability to contribute to the goals of this unit < organisation >	.67 .66 .65	.64 .39 .61
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen) Shows confidence in my ability to contribute to the goals of this unit < organisation > (Toont vertrouwen in mijn vermogen bij te dragen aan de	.67 .66 .65	.64 .39 .61
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen) Shows confidence in my ability to contribute to the goals of this unit < organisation > (Toont vertrouwen in mijn vermogen bij te dragen aan de doelen van deze eenheid)	.67 .66 .65	.64 .39 .61
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen) Shows confidence in my ability to contribute to the goals of this unit < organisation > (Toont vertrouwen in mijn vermogen bij te dragen aan de doelen van deze eenheid) Demonstrates total confidence in me	.67 .66 .65 .64	.64 .39 .61 .57
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen) Shows confidence in my ability to contribute to the goals of this unit < organisation > (Toont vertrouwen in mijn vermogen bij te dragen aan de doelen van deze eenheid) Demonstrates total confidence in me (Toont een volledig vertrouwen in mijn)	.67 .66 .65 .64	.64 .39 .61 .57
een lid van de groep) Listens to my concerns (Luistert naar zaken die voor mij van belang zijn) Allows me a strong hand in setting my own performance goals (Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen) Shows confidence in my ability to contribute to the goals of this unit < organisation > (Toont vertrouwen in mijn vermogen bij te dragen aan de doelen van deze eenheid) Demonstrates total confidence in me (Toont een volledig vertrouwen in mijn) Looks out for my personal welfare	.67 .66 .65 .64 .40	.64 .39 .61 .57 .33

# Table 3: Leadership Subscales: Results from Factor Analysis (n = 56)

importance ratings on perceived and preferred evaluative style to vary across leaders. To investigate whether the result that on average less emphasis is given to aspects of evaluation than preferred by subordinates holds across leaders, we compared the differences between perceived and preferred also for each of the groups. The results are reported in the right-hand side of Table 4. The finding that preferred emphasis is higher than perceived emphasis holds in most of the groups too. Again, the two items, 'short-term goals' and 'his/her intuition' are exceptions, indicating that for those aspects of evaluation subordinates prefer restricted rather than expanded use of these two instruments.

Furthermore, the scores on the items in the ideal situation tend to have a lower standard deviation than the scores on the same items in the actual situation, indicating that differences in opinion about the ideal style are smaller than the perceived differences in actual evaluative style. This result supports proposition two, and supports the use of preferred evaluative style as a measure of norms and values.

Finally, if preferred evaluative style reflects norms and values, the absolute scores in Table 4 suggest

that in this particular organisation subordinates feel that the boss's intuition, financial information, and short-term goals are the least preferred performance evaluation instruments, scoring 3.36, 3.45 and 3.61 respectively, whereas possibilities to improve/develop your performance, long-term goals and performance of your unit are the three mostly valued instruments of performance evaluation (respective scores 4.29, 4.16 and 4.02).

# 4.2 Level of Analysis

Whether there were significant differences in subordinates' evaluations of their superiors between leaders on items of evaluative style and leadership subscales (prop. III, V and IX) was assessed through an F-test using one-way ANOVA. The results for leadership, perceived and preferred evaluative style are reported in Table 5. However, one of the assumptions of one-way ANOVA is homogeneity-of-variances, which we tested in this paper using Levene's test. If the Levene statistic is significant, the assumption (null hypothesis) of homogeneity-of-variances is rejected (see Newton and Rudestam, 1999). For the three leadership subscales the assumption was supported (see section 4.3). Of the 17 items on perceived evaluative style, for nine items the null hypothesis of equal

	Difference between perceived and preferred for individual respondents $(n = 56)$ Number of groups reporting low equal or higher perceived emphasis than preferred $(n = 12)$		g lower, emphasis				
	Perceived Mean (sd)	Preferred Mean (sd)	t-value	lower	equal	higher	
short-term goals	3.84 (.78)	3.61 (.62)	2.436*	2	2	8	
long-term goals	3.50 (.83)	4.16(.63)	-6.238*	10	1	l	
financial information	2.98 (.86)	3.45 (.69)	-4.560*	10	1	L	
non-financial information	3.45 (.74)	3.78 (.60)	-3.632*	9	3	0	
deviations from agreed performance	3.14 (.82)	3.86(.70)	-5.756*	9	2	1	
your explanation of your performance	3.41 (.87)	4.00 (.69)	-5.337*	9	2	1	
objectively measurable performance	3.41 (.83)	3.86(.67)	-4.536*	11	1	0	
his/her own intuition	3.79 (.85)	3.36 (.90)	3.028*	2	0	10	
performance delivered in the past year	3.78 (.82)	3.91 (.68)	-1.188	6	2	4	
your plans for the next year	3.42 (.85)	3.91 (.93)	-3.203*	8	2	2	
positive aspects of your performance	3.57 (.74)	3.96(.74)	-3.311*	8	3	1	
negative aspects of your performance	3.27 (.84)	3.91 (.77)	-5.877*	11	0	1	
performances of your unit	3.48 (.95)	4.02 (.59)	-4.126*	10	1	1	
personal, individual performance	3.66 (.79)	3.91 (.67)	-2.512*	8	3	1	
causes for failings in performance	3.02 (.86)	3.93(.78)	-7.253*	11	0	1	
possibilities to improve your performance	e 3.16 (.95)	4.29 (.76)	-8.648*	12	0	0	
information from conversation	3.23 (.85)	3.93 (.71)	-5.368*	10	2	0	
* t-value is significant at the 0.05 level	(2-tailed)					-	

Table 4: Average Perceived and Preferred Importance of Aspects of Evaluative Style

variances could not be rejected; for the other eight items the assumption of homogeneity-of-variance was not supported. Of the 17 items on preferred evaluative style only six items did not meet the assumption of homogeneity-of-variance.

The F-test to assess whether significant differences in means existed between different groups (grouped by leader) was done only for the variables that met the assumption of homogeneity of variance. This analysis showed that for the nine items measuring perceived evaluative style none of the means differed significantly between leaders. In contrast, for preferred evaluative style, a significant difference in means between leaders on one item, performance delivered in the past year, was found. These results were not as we expected, since we expected that scores on preferred evaluative style would not be affected by group-membership, while scores on perceived evaluative style would (prop. IV and V). However, a closer look at Table 5 reveals that, although no significant differences in means on the items of perceived evaluative style were found, the probability scores (p-values) for most items indicate that differences between means do exist. P-values for the items of preferred evaluative style are much higher, indicating that it is more likely that the assumption of equal means across different leaders holds for preferred evaluative style items than for perceived evaluative style items, which is as expected. Thus, although we were not able to find support for our expectations in propositions IV and V, our results do support our expectation that scores on perceived evaluative style items would show more variation across leaders than scores on preferred evaluative style items.

As an indication of the extent to which subordinates within groups agree in their ratings of evaluative style, the value of Eta squared is reported in Table 5 too. The results for perceived evaluative style indicate that for most aspects of evaluative style between 20 to 30% of the variability in importance scores is accounted for by differences between leaders. Although 70 to 80 % of the remaining variance reflects individual differences in subordinate's perception of evaluative style, this is an indication that ratings of subordinates within a group are more similar than ratings of subordinates in different groups, and these results support proposition IV. Yet, evaluative style items in this study show both a substantial variation within and between groups.

## 4.3 Evaluative Style and Leadership Style

Turning to leadership, we will first discuss the

results on the relationship between leadership and perceived evaluative style. Table 6 reports the Pearson correlations between 17 evaluative style items and the three leadership subscales. Positive correlations are found between active management-by-exception and six evaluative style items. These six items are either referring to quantitative information or focus on shortcomings in performance, which confirms our expectation (prop. VII).

Furthermore, support is positively correlated to evaluative style items which refer more to qualitative, interpersonal aspects of evaluation, although support is also positively correlated with objective measurable information. Finally, the teambuilding subscale of leadership does not seem to be of high significance in explaining differences in evaluative style. It is positively correlated with only three items, and two of these items correlate with all three leadership subscales. Therefore, these correlations could be explained through supportive leadership. To test this, partial correlations were calculated between all 17 items and teambuilding, controlling for supportive leadership. When supportive leadership is partialled out, teambuilding is only significantly and positively correlated with the items 'possibilities to improve/develop your performance' and 'causes of failings in performance'. The results for the support leadership subscale confirm our expectations (prop. VI), but the results for teambuilding, which is an aspect of inspirational leadership too, do not support proposition VI.

A further look at table 6 reveals that none of the leadership subscales is correlated with the long term/short term items. This could indicate that emphasis on short term/long term goals is indeed a separate dimension of evaluative style, as argued by Otley and Fakiolas (see section 2.1). Further analysis reveals that the two items, shortterm goals and long-term goals, are significantly negatively correlated (r = -.350, p = .01), but not correlated with any of the other items, which suggests it is one single dimension.

Secondly, the relationship between leadership and preferred evaluative style is explored. In contrast with our expectation of no relationship (prop. VII), the right-hand side of table 6 shows that active management-by-exception is significantly positively related to four items of preferred evaluative style: objectively measurable performance, explanation of performance, negative aspects of performance, and possibilities to improve/develop performance. These results suggest that subordinates working under a supervisor characterised as high on active management-by-exception, think

#### Table 5: Differences between Groups (F-tests) and Agreement within Groups (Eta squared) (n = 12)

	Leadership					
	Levene statistic	F	Eta squared			
ambe	1.153	3.147**	.440			
team	1.377	2.524**	.476			
support	1.394	3.551**	.387			

	Perceived evaluative style		Preferred	yle		
	Levene	F	Eta	Levene	F	Eta
	STATISTIC 2 747*		squarea_	Mansuc	1.225	
snort-term goals	2.307*		.412	1.536	1.225	.234
long-term goals	1.568	1.416	.261	4.293*		.109
financial information	1.401	1.204	.231	3.085*		.189
non-financial information	2.065*		.156	1.990	.486	.111
deviations from agreed performance	1.036	1.515	.275	1.146	.489	.109
your explanation of your performance	2.680*		.193	1.173	.636	.137
objectively measurable performance	.813	1.371	.255	1.567	1.090	.214
his/her own intuition	2.552*		.290	.836	.720	.152
performance delivered in the past year	1.440	1.292	.253	1.179	2.242**	.359
your plans for the next year	1.642	1.375	.260	2.291*		.197
positive aspects of your performance	2.909*		.263	.940	1.201	.231
negative aspects of your performance	1.856	.916	.186	1.883	.773	.162
performances of your unit	2.170*		.246	2.641*		.228
personal, individual performance	1.177	2.009	.334	2.048*		.220
causes for failings in performance	2.019*		.353	1.327	1.416	.261
possibilities to improve your performance	3.888*		.364	.627	.434	.098
information from conversation	1.535	1.749	.304	1.366	.509	.113

\* Levene's test is significant, indicating the null hypothesis of homegeneity-of-variances is rejected.

\*\* F-test is significant at the 0.05 level

#### Table 6: Pearson Correlations between Leadership Subscales and Items measuring Evaluative Style (n= 56)

	Leadership				
	AMBE	SUPPORT	TEAM		
AMBE	000.1	.195	.185		
SUPPORT	.195	1.000	.416(**)		
TEAM	.185	.416(**)	1.000		

	Perceived evaluative style			Preferred evaluative style			
	AMBE	SUPPORT	TEAM	AMBE	SUPPORT	TEAM	
short-term goals	.101	068	.099	.242	.259	.155	
long-term goals	.108	.056	024	108	.235	.076	
financial information	.317*	.079	041	.095	.142	078	
non-financial information	.108	.210	087	.034	.161	.032	
deviations from agreed performance	.480**	020	117	.198	087	064	
your explanation of your performance	.212	.498**	.157	.339*	.201	.178	
objectively measurable performance	.477**	.403**	.062	.541**	.181	.156	
his/her own intuition	.091	.029	214	.193	.437**	.095	
performance delivered in the past year	.303*	.163	007	.074	.165	062	
your plans for the next year	.110	.382**	.124	.065	079	014	
positive aspects of your performance	.257	.526**	.096	.259	.058	.047	
negative aspects of your performance	.423**	030	113	.291*	.186	.050	
performances of your unit	.385**	.347**	.090	023	.166	030	
personal, individual performance	.245	.384**	105	.209	.172	.011	
causes for failings in performance	.476**	.464**	.331*	.127	.187	094	
possibilities to improve your performance	e .396**	.393**	.426**	.312*	.248	058	
information from conversation	.234	.57()**	.391**	.173	.120	.025	

\* Correlation is significant at the 0.05 level (2-tailed). \*\* Correlation is significant at the 0.01 level (2-tailed).

that when their performance is evaluated, their boss should not only point out their shortcomings, but that he should give them a chance to explain those shortcomings and help them learn from their mistakes.

Finally, we explored whether significant differences existed between groups of subordinates. To test whether leadership subscale scores differed significantly between leaders, an F-test using oneway ANOVA was done. The results are reported in Table 5. As discussed earlier, all three scales met the assumption of homogeneity-of-variance. The results indicate that significant differences in the mean scores on all three leadership subscales do exist across different leaders. This supports proposition IX. To test the extent too which subordinates reporting to the same superior agree in their ratings of leadership style, Eta squared values were calculated (table 5). For all three leadership subscales, about 40% of the variability is accounted for by differences between leaders, which is more than we were able to show for perceived evaluative style items. These findings do support our expectation that there is considerable agreement within groups on leadership subscale scores (prop. X). Taken together, the support for proposition IX and X indicate that although differences in perception of leadership exist, the final score reported by subordinates does reflect actual differences between leaders too.

## **5** Discussion

In this paper we reported results from a questionnaire survey measuring 17 aspects of perceived evaluative style that were thought to be important in exploring and understanding differences in how superiors evaluate the performance of their subordinates and four aspects of these superiors' leadership style. In the analyses we explored three issues regarding our understanding of the variability and meaning of performance evaluation style of superiors. First, we looked at the difference between perceived (real) and preferred (ideal) evaluative style. After that we studied the issue whether perceived evaluative style is an individual or group level phenomenon. Finally, we looked at the relationship between leadership and aspects of performance evaluation style.

Our results suggest that the 17 aspects we measured were indeed important aspects of performance evaluation, since we also asked subordinates to indicate how much emphasis they think each of these aspects should be given by their superior when evaluating their performance, i.e. preferred evaluative style. On average, subordinates indicate that they feel that almost every aspect that is measured deserved more attention than it presently received. Exceptions are short-term goals and boss's intuition. The most preferred instruments were paying attention to possibilities to improve/ develop your performance (score 4.29 on a 5 point scale), long-term goals (4.16) and performance of your unit (4.02). With respect to the role of financial information, this suggests that the motivational and appraisal role of budgetary targets and budgetary control in this organisation may be limited. This would be in line with Marginson's (1999) recent observations in a large British organisation (see section 2.1), and should be investigated in more depth when analysing the interview data.

Perceived and Preferred (Ideal) Evaluative Style Since disagreement with evaluation criteria has been argued to be important in understanding the behavioural consequences of evaluative style, in this study we measured both perceived and preferred evaluative style. A difference between the two would indicate that subordinates disagree with how their performance is evaluated. The use of preferred evaluative style as a measure of norms and values held by participants within the organization is supported by our findings that there is less variation in the preferred importance ratings of aspects of evaluative style as compared to the variation in the perceived importance ratings of the same aspects, both within groups (subordinate level) and between groups (group level). This suggests that subordinates' opinion about what should be important when their performance is evaluated is, at least partly, influenced by organizational culture.

Comparing average preferred importance scores of aspects of evaluative style to average perceived importance scores, the results indicate that on average subordinates prefer a higher emphasis on most aspects of evaluation than the emphasis his/her superior actually lays on those aspects, which was not what we expected. Possibly, in this organization disagreement with evaluation does not arise because bosses place too much emphasis on certain items, but because bosses neglect certain items. These results are interesting since research on performance evaluation style has focused attention mainly on what is emphasized when evaluating performance, rather than on what is neglected. The practical implication is that, if disagreement with evaluation criteria is important in understanding the behavioural consequences of evaluative style, future research should not only focus on what superiors emphasize when evaluating performance, but also on what they neglect.

#### Individual or Group Level Phenomenon

Concerning the level of analysis, our results indicate that evaluative style items in this study show substantial variation within and between groups. Some of the variability in perceived evaluative style is explained by differences in superiors, meaning that ratings of subordinates within a group are more similar than ratings of subordinates between different groups. This suggests that differences in ratings of perceived evaluative style partly reflect actual differences between how superiors evaluate their subordinates. But although there were differences between the mean scores of evaluative style items among different groups, none of these differences were significant. A possible explanation for this result is that we only looked at single items. When the scores on several items would be summed, because they are part of a single construct, it is more likely to find significant differences in mean scores on those summed scores across leaders. Another explanation is that, although superiors may differ from each other, superiors may also evaluate the performance of different subordinates in different ways. So, within group variability could be caused by differences in perception on the part of the individual subordinates, but could also reflect actual differences in how a superior evaluates subordinates' performance. Hopefully, the interview data can shed some light on the extent to which differences in perception and actual differences may explain the substantial within group variability.

Leadership and Performance Evaluation Style Finally, the relationship between leadership and evaluative style was explored. Although leadership too showed substantial variability within and between groups, the results indicate that significant differences in the mean scores on all three leadership subscales do exist across different leaders, and that there is considerable agreement within groups on leadership subscale scores. Furthermore, the results indicate that leadership and evaluative style are related. Different aspects of leadership relate to different aspects of evaluative style. Active management-by-exception is positively correlated with an emphasis on quantitative information or shortcomings in performance when performance is evaluated, while supportive leadership is positively correlated with more qualitative, interpersonal aspects of evaluation. Although teambuilding is a subscale of inspirational leadership, as is supportive leadership, teambuilding is not as important in explaining differences in evaluative style. However, supportive leadership and teambuilding differ from each other in that supportive leadership concerns the leader's approach to individuals, as against his/her approach towards a group of individuals in the teambuilding subscale. Since performance evaluation concerns the evaluation of individuals' performance, this could explain why supportive leadership is positively correlated with aspects of evaluative style, while teambuilding is not.

Although leadership and perceived evaluative style are related, none of the leadership subscales is correlated with either short-term goals or longterm goals. However, it is likely that aspects of leadership that measure a leader's concern with the future, such as vision, are related to this dimension of evaluative style. Since the leadership subscales that we used in this study focused more on the interpersonal aspects of leadership than on vision, this may explain why we do not find a relationship between leadership and the emphasis on short/long term goals when evaluating performance.

In contrast to the relationship between leadership and perceived evaluative style, as expected, there was no relationship between preferred evaluative style and leadership subscales, except for active management-by-exception. Active managementby-exception is positively correlated with preferred emphasis on objectively measurable performance, explanation of performance, negative aspects of performance, and possibilities to improve/develop performance. We did not expect a relationship between preferred evaluative style and leadership, since we assumed that preferred evaluative style reflected norms and values held by individuals across the organization. But maybe some norms and values are set or at least influenced by leaders, which could explain our results. Another explanation for this result could be that active management-by-exception has the potential danger that subordinates perceive it as a negative approach, since active management-by-exception means that a superior focuses attention on errors. irregularities, underperformance, et cetera. Subordinates might be willing to accept an active management-by-exception type of leadership, and a focus on negative aspects of performance, as long as subordinates feel that their superior wants them to learn from their mistakes, and is willing to listen to their explanation.

As in any study, the results from this study should be considered in light of its limitations, some of which have been mentioned in the paper and on which we will not elaborate at this point. Not withstanding those limitations, overall, this study indicates that differences exist in how superiors evaluate the performance of subordinates, and in how they fulfil their role of being a leader. Since many studies suggest that subordinates' behaviour is not only affected by the adequacy of systems within the organisation, such as performance evaluation systems, but as much by how managers use these systems, such differences could have important behavioural consequences at the subordinate level. Certain ways of evaluating subordinates, and certain leadership characteristics, may be more effective in influencing people in intended directions than other ways. Further research is necessary to identify those ways and characteristics, and the circumstances under which they are most effective in achieving intended results. Therefore, evaluative style is still a potentially important construct that deserves researchers' attention.

#### REFERENCES

- Brownell, P., (1985), Budgetary Systems and the Control of Functionally Differentiated Organizational Activities, *Journal of Accounting Research*, Vol. 23, No. 2, pp. 502-512.
- Brownell, P. and M.K. Hirst, (1986), Reliance on Accounting Information, Budgetary Participation and Task Uncertainty: Tests of a Three-Way Interaction, *Journal of Accounting Research*, Vol. 24, No. 2, Autumn, pp. 241-249.
- Hartmann, F.G.H. and F. Moers, (1999), Testing contingency hypotheses in budgetary research: an evaluation of the use of moderated regression analysis, *Accounting*, *Organizations and Society*, Vol. 24, No. 4, pp. 291-315.
- Hartog, D. N. den, (1997), *Inspirational Leadership*, Ipskamp, Enschede.
- Hopwood, A.G., (1973), An Accounting System and Managerial Behavior, Saxon House.
- Kaplan, R.S. and D.P. Norton, (1992), The Balanced Scorecard – Measures that Drive Performance, *Harvard Business Review*, January-February, pp. 71-79.
- Kaplan, R.S. and D.P. Norton, (1993), Putting the Balanced Scorecard to Work, *Harvard Business Review*, September-October, pp. 134-142.
- Kaplan, R.S. and D.P. Norton, (1996), Using the Balanced Scorecard as a Strategic Management System, In: *Harvard Business Review*, January-February, pp. 75-85.
- Lynch, R.L. and K.F. Cross, (1991), *Measure Up! Yardsticks for Continuous Improvement,* Blackwell, Cambridge.
- Marginson, D.E.W., (1999), Beyond the budgetary control system: towards a two-tiered process of

management control, *Management Accounting Research*, Vol. 10, No. 3, pp. 203-230.

- Newton, R.R. and K.E. Rudestam, (1999), Your Statistical Consultant: Answers to Your Data Analysis Questions, SAGE, Thousand Oaks.
- Otley, D.T., (1978), Budget Use and Managerial Performance, *Journal of Accounting Research*, Vol. 16, No. 1, Spring, pp. 122-149.
- Otley, D.T. and B.J. Pierce, (1995), The control problem in public accounting firms: an empirical study of the impact of leadership style, *Accounting*, *Organizations and Society*, Vol. 20, No. 5, pp. 405-420.
- Otley, D.T. and A. Fakiolas, (2000), Reliance on accounting performance measures: dead end or new beginning?, *Accounting, Organizations and Society*, Vol. 25, No. 4-5, pp. 497-510.
- Ross, A., (1994), Trust as a Moderator of the Effect of Performance Evaluation Style on Job-Related Tension: A Research Note, *Accounting*, *Organizations and Society*, Vol. 19, No. 7, pp. 629-635.
- Ross, A., (1995), Job Related Tension, Budget Emphasis and Uncertainty: A Research Note, *Management Accounting Research*, Vol. 6, No. 1, pp. 1-11.

#### ΝΟΤΕ 5

1 Since the ILO-questionnaire contained 92 items, only four of the 13 a-priori subscales listed in Table 1 were used to limit the total length of the questionnaire.