Hanne Foss Hansen

Organizing for Quality – A Discussion of Different Evaluation Methods as Means for Improving Quality in Research

In research and science it is difficult to evaluate quality and develop systems of quality control. *Firstly*, the time perspective makes evaluation difficult. Today's uninteresting research results can tomorrow become important and sensational. *Secondly*, cognitive and organizational structures are differentiated between different fields. That is the characteristics of the focus of evaluation differs. *Thirdly*, the concept of quality itself is unclear. Is quality correctness, novelty, stringency, relevance, etc. (Hemlin & Montgomery, 1990) Most researchers agree that all aspects should be included in the concept of quality. But this agreement does not make evaluation an easy task.

Do these reasons make it reasonable to avoid and refuse evaluation? For me, the answer is no, specially when we talk about research financed by public funds. It is reasonable that society makes demands to researchers and research organizations about value for money. The problem is given the fact that evaluation is reasonable and difficult: How do we do it?

In the following I will discuss the literature of organizational effectiveness in order to define and specify the concept of quality and quality control. The problem is: Has research policy and research organization something to learn from organizational evaluation in general?

It is important to stress that what I discuss is organizational effectiveness. That is I do not discuss evaluation either at the product level (an article for publication, a Ph.D. thesis, etc.) or at the personal, individual level but at the level of a research group, a department, a research institute, a subfield or even a discipline. The discussion is not delimited to formal organizations. Also more loosely coupled organizations, informal organizations, networks and programmes can be analyzed using the different perspectives in the literature.

Science Studies, Vol. 8 (1995), No. 1, 36–43.

Organization effectiveness: theoretical perspectives

If we look at the literature about organizational effectiveness we find that it is rich and differentiated. Several theoretical perspectives are competing. Very many criteria for effectiveness are being discussed and many arguments concerning which factors determine or influence effectiveness are presented (for reviews of the literature, see for example Scott, 1987; Cameron, 1986a and 1986b; or in a Nordic context Foss Hansen, 1991). Thus, there is no agreement on how to evaluate, whether or not an organization is effective.

The literature can be divided into six very different theoretical perspectives: the goal attainment perspective, the internal process perspective, the systems perspective, the constituency perspective, the symbolic perspective and the paradox perspective.

Goal attainment

The goal attainment perspective is the classical perspective of organizational effectiveness. It builds upon classical organizational theory perceiving organizations as machines (Morgan, 1986). The basic idea is that an organization is effective if it through its activities attains its goals. To evaluate effectiveness therefore simply means: 1) to clarify what are/were the goals? 2) to evaluate if the goals are/have been attained.

It seems simple. But it may not be as simple as imagined. To evaluate effectiveness according to the goal attainment perspective implies that: 1) goals are clear; 2) different actors agree upon goals; 3) goals are formulated "running forward", that is goals have been guiding activities.

Goal attainment evaluation is very much used in practice. It is very familiar with basic rationalities of the western societies. Therefore this perspective may be difficult to avoid even when it is not suitable.

Is it then suitable in respect to research organization and research activities? Probably very seldom. From the sociology of science (e.g. Whitley, 1984; and the review in Foss Hansen, 1986) we know, that the basic assumptions which goal attainment rests upon are not typically found in research organizations. Research organizations are very complex organizations. Frequently there are conflicts concerning goals. Therefore goals very often are formulated in broad terms (e.g. "we wish to contribute to the development of the area x"). Goals are also very often formulated "running backwards". That is goals are descriptions of activities going on or even finished. In such cases goals of course are unsuitable for evaluation.

However, using the goal attainment perspective may uncover the ability of the research organization to formulate goals. This may be important if goal attainment thinking is an aspect of research policy, as it seems to be in many European countries in the 1990's.

Internal process perspective

In the internal process perspective an organization is perceived as a closed, natural organism. Human relations, human resources, motivation etc. are important keywords.

According to the internal process perspective an organization is effective if the internal processes of production, communication, collaboration and management develop promptly without conflicts. Unplanned production stops, strikes, and in research flight into other activities such as administration, teaching, consulting etc. are symptoms of ineffectiveness.

According to this perspective, evaluation deals with appraising if processes are organized appropriately, including taking the pulse at organizational actors and their commitment in research activities.

However, in the research system conflicts are often the motors which shape and push the development of knowledge. That is conflicts are desirable. On the other hand most researchers are aware of fields of research

where there are nearly no dialogue and collaboration because competition and conflicts are so immense, that people are not able to communicate. Therefore, the internal process perspective is not without relevance in respect to research evaluation.

The weakness of the internal process perspective is that it is build upon an assumption about clear and well known technology, that is complete knowledge about the flow between activities and results. The problem is of course, what a high level of processes and commitment is worth if activities develop on a wrong track, if the direction people are working in is wrong?

In research, knowledge about technology is typically not complete. Uncertainty and conflicts about which technology, defined as which methods to use, are most often the case. On the contrary research results gained from working on the wrong track is not worthless. Wrong track results often give important insights into which track to follow next.

Despite severe limitations evaluations based on the internal process perspective are important e.g. in order to give input to discussions about if working-time is spent appropriately, if reward systems are appropriate, if collaboration is organized in a way which increase motivation etc.

The systems perspective

The systems perspective (see for example Yuchtman & Seashore, 1967) views organizations as systems transforming an input to an outcome. According to this perspective, the organization is effective, if it survives as a system, that is if it manages to maintain integration and organizational borders towards the environment.

Figure 1 presents a picture of a research organization according to the systems perspective. The frame of reference is beyond the general systems perspective inspired by the work of Scharioth and Gizycki (1986).

According to the model the research organization transforms an input (tasks, economic resources, knowledge, etc.) to an outcome. Diffusion of knowledge through publications, patents, lectures, education of new generations of researchers and so on constitute outcome. Production goes on through several types of processes: theoretical and empirical analysis, experiments, reading, writing, discussing, participation in conferences, traveling abroad, etc.

Production processes are rooted in internal and external structures, in the structures of qualifications and collaboration, in the professional and economical/political networks, as well as in the norms and traditions of the discipline and the organization.

In the short run these structures constitute the conditions for production. In the long run the structures can be changed according to professional and political constellations in and around the organization. Thus, the model is not a classical input-output model, implying that the organization is viewed as a black-box, neither it is a contingency model implying that structure is the only variable determining production processes.

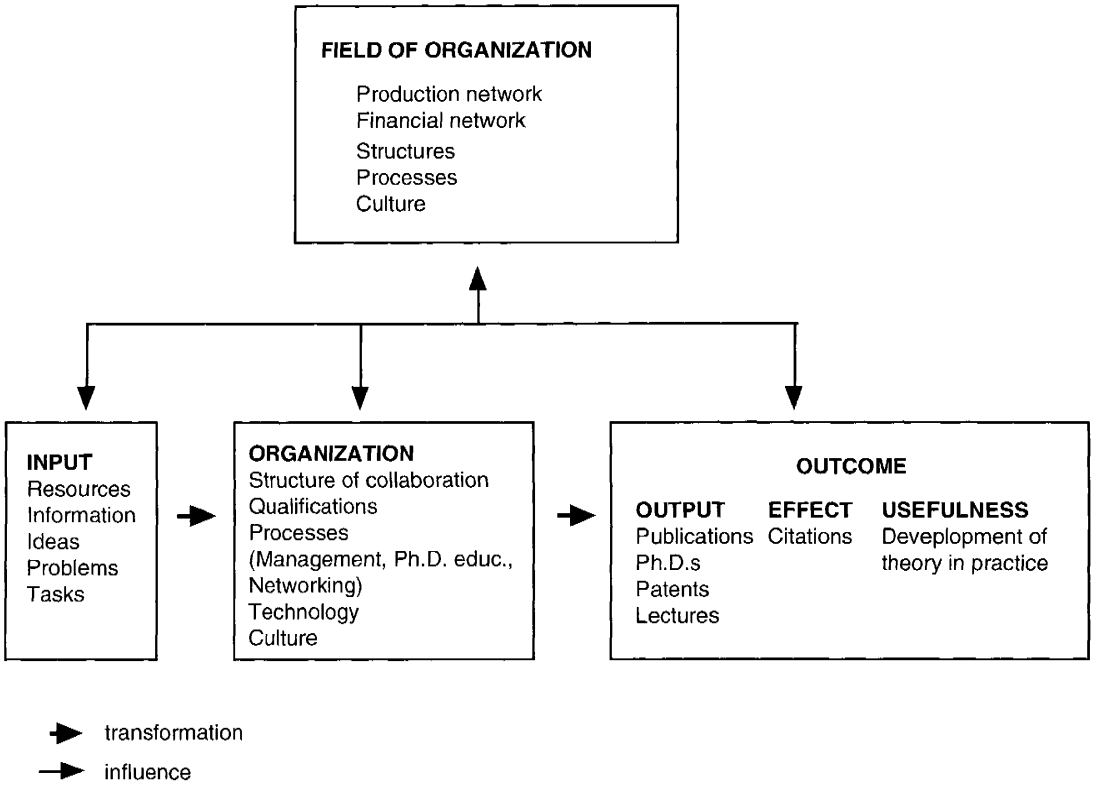
Evaluating an organization in this perspective, one can choose between evaluating the input side, the transformation and/or the outcome side. An evaluation on the input side has focus on the organizational abilities to attract scarce resources from the environment. In other words, the organizational position of negotiation in respect to the environment is appraised.

An evaluation on the transformation side focuses on organizational structures (collaboration structures, networking, equipment, etc.) as well as on processes, just as mentioned above in the discussion of the internal process perspective.

Finally, an evaluation on the outcome side focuses on quality, quantity and impact of the products produced. Bibliometric analysis, the counting of publications and citations, are examples of evaluation practices based on outcome evaluation.

The systems perspective is based on the assumption that there is a one-to-one connection between input, transformation and outcome, that technology is well known and

Figure 1. A simple model of a research organization.



organizational slack non-existing. As mentioned earlier technology is typically not well known considering research organizations. And from organizational theory in general (e.g. Galbraith, 1977), we know that many organizations are characterized by existence of organizational slack, that is growth in input, in quality or quantity, is not automatically followed by a corresponding growth in outcome. If carried out in periods where slack is being "stored" or used, evaluation runs the risk of producing wrong statements.

Although limitations, it may be reasonable to use the systems perspective in research evaluation. Especially two kinds of use may be valuable. Evaluation on the input side is in my experience often a neglected aspect. Good conditions for research (resources, personnel etc.) is a necessary but not an adequate condition for carrying out interesting research. Outcome evaluation may be carried out either as a kind of minimum check of the quantity of production or combined with input- and transformation evaluation in order to understand the research organization in question in a more holistic perspective.

The constituency perspective

Where the perspectives presented above are perspectives of harmony, the constituency perspective is a perspective of politics and conflict. It is based on a more complex understanding of the internal and external organization. The constituency perspective (Conolly, Conlon & Deutsch, 1980) perceives the organization as constituted by different actors, both actors working in the organization (researchers), actors collaborating or competing with the organization (other re-

searchers, industry, etc.), and actors involved in the organization in a more general way (e.g. candidates working with the field in question in practice, research politicians, research councils, media, etc.)

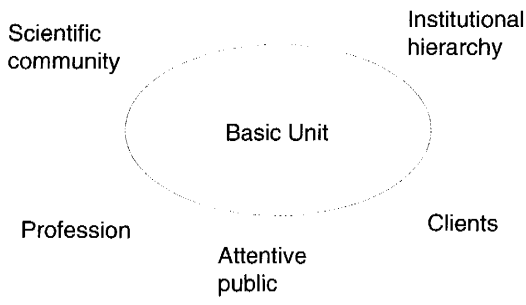
In the constituency perspective it is not at all possible to talk about effectiveness as a absolute phenomenon. Instead we talk about different evaluations of effectiveness. The idea is that one actor may appraise a research group, a department or even a cluster of departments within a given field, as excellent at the same time as another actor appraises the organization in question either negatively or just not interesting.

A general model of a research organization and its surroundings is pictured in figure 2. The model lists actors generally involved in research, actors which from time to time try to influence the organization, and actors which the organization is depending on concerning resources, information, equipment, etc.

As indicated in figure 1, it is fruitful analytically to distinguish between different kinds of networks surrounding the organization, a production network and a financial network. The production network is constituted by other researchers working within the field and nearby fields, as well as of candidates using the field in practice in public administration and industry. The financial network is constituted by potential resource financing actors.

A central task for a research organization is to make the two networks play together,

Figure 2. The research organization and the environment.



that is seen from an organizational view, to develop a financial network which makes involvement in tasks, in first order interesting from a research point of view, possible.

Using the constituency perspective in research evaluation makes possible also evaluation and development of research management. Using the constituency perspective one may develop organizational consciousness about interests, wishes, and assessments of different actors.

The symbolic perspective

According to the symbolic perspective (Gaertner & Ramnaryan, 1983) an organization is effective, if it has the ability to represent itself as effective in respect to internal and external actors. Thus, the criteria of effectiveness changes character from an objective, substantial criteria to a symbolic criteria. Evaluation no longer is a "measure" of effectiveness. It has become a question about creating a picture, image, or myth that the organization in question is effective. And it has become a question about assessing the organizational ability to legitimate existence and activities. Whether the picture produced is real or not, is within the symbolic perspective without importance. What is important, is the picture created.

Using the symbolic perspective in respect to research evaluation means appraising the organizational ability to communicate with the environment. Questions about communication become important: Are research results spread and made visible and available through appropriate channels? Is the appropriate "language" used to the receiving actors? Is the layout attractive? etc.

In addition, using the symbolic perspective also include appraising organizational ability to understand, adapt to and influence the perspectives of effectiveness used by other central actors, e.g. research politicians and financial actors. If central actors use the perspective of goal attainment in their evaluation of the research organization, it may be important also for the research organi-

zation to use this perspective regardless of the fact that it is unsuitable.

The weakness of the symbolic perspective is that it is questionable whether symbolic behavior is durable in the long run. Symbolic behavior is not without importance but what counts in the long run probably (hopefully) is substantial results.

The strength of the symbolic perspective is that it puts focus on the communication and promotion aspects of research organizations. Many research organizations, and in general professional organizations, are not very good at communicating with others than colleagues.

The paradox perspective

The paradox perspective is a rather new theoretical perspective, not developed as well as the perspectives already discussed. The idea is that effective organizations are characterized by contradictory characteristics in respect to tasks, processes, structures, etc. To give an example organizations are at the same time centralized and decentralized; they are adaptive but hold on to old traditions and cultures, and they handle both substance and symbols.

Contradictions are tensions which keep organizations breathing and alive, because tensions release energy and thus improve performance and effectiveness.

From the sociology of science and surrounding fields we know that there are very many contradictory demands addressed towards research organizations (Elzinga, 1986; Premfors, 1986; Hemlin, 1991).

Below, I have listed some of the most important contradictory demands:

1) In research organizations both norms of elitism and norms of egalitarianism are important. Differences in production in the past imply differences in status, not only between young and established scientists but also between groups, departments, etc. And differences in status have important impact on the distribution of recognition, resources and so on. At the same time there are pressures for evenness, probably especially in the Nordic countries. All universities within subfield x ought to have part of the resources of the research council y. All scientists in department z ought to have xx resources for traveling each year etc.

Both elitism and egalitarianism are visible in organizational structures. At the same time research organizations are hierarchical and anarchic. Elitist norms give room for meritocracy, hierarchies and collegial interdependence, while egalitarian norms make everybody his or hers own master.

2) In research organizations both international and local integration is important. International integration ensures that the research organization keeps ajour with the research frontier, while local integration creates the daily working environment and research circles, which are very important, especially for the younger researchers. Being internationally integrated means a lot of traveling and staying abroad, while being locally integrated means staying in the laboratory/office, taking part in laboratory life and discussions at seminars and around the lunch table.

In the sociology of science this has been described at the individual level by the concepts of cosmopolitans and locals. Having very many cosmopolitans, a research organization runs the risk of attaining no local integration. On the other hand having many locals, it runs the risk of having no or too little touch with the research frontier.

3) Renewal and at the same time sticking to old paradigms and problems are claims of the research organization. If not renewing, only old knowledge is repeated. If not sticking to paradigms and problems, necessarily analytical depth is not achieved.

In studies of evaluation procedures these contradictory claims are found also on the individual level. Thus Montgomery and Hemlin (1991) found that both specialization and breadth were important criteria assessing candidates for professorships.

Probably more paradoxes could be discussed, e.g. the importance for motivation of having both a challenging and confident

research environment, the conflict in some research fields between extra- and intrascientific relevance, etc. Reading science studies in general and especially the sociology of science having the paradox perspective in mind probably would uncover even more paradoxes.

What is then the core in the paradox perspective concerning the discussion about effectiveness and evaluation of effectiveness? According to the paradox perspective the contradictions create energy and improve performance, in other words they ensure effectiveness. But is this always the case? And is it typically the case in research organizations?

Having psychological theories of stress in mind, one can ask whether organizational contradictions imply energy or stress? Or more specifically, where the organizational thresholds for stress are situated? It is not possible to answer the question, because we have very little empirical evidence concerning these problems.

In spite of the limitations of the paradox perspective there is no doubt that it is fruitful in order to understand the character and development of research organizations. However, much more work is needed both on a theoretical and empirical basis in order to develop and enrich this perspective.

Organizational effectiveness and research evaluation

Above I have discussed several different theoretical perspectives which may be used for evaluating research organizations. Figure 3 sums up the discussion.

Figure 3. Perspectives of organizational effectiveness.

Perspective:	Uncover if used:
Goal attainment	Organization ability to formulate and work towards goals
Internal process	Level of internal processes, e.g. communication, collaboration, education of new generations of scientists Symptoms of ineffectiveness, e.g. overburden of other activities such as teaching, consulting; flight into other activities
Systems	Ability to attract input, ability to transform input to outcome, process level of transformation, quantity and quality in outcome
Constituency	Interest groups, conflicting demands towards research profile, the assesment of the interest groups of the organization
Symbolic	External communication and legitimation
Paradox	Existence of paradoxes, ability to manage and weight paradoxes

In the introduction I asked: Have science studies, research policy and research organization something to learn from organizational evaluation in general? The answer seems to be yes. Today evaluation practice is primarily built on tradition, professional norms and common sense. The need for a more theoretically based evaluation practice is great. And the perspectives borrowed from the general discussions of organizational effectiveness constitute a proposal for such a theoretical basis.

Surely, there are problems. Using one theoretical perspective is seldom sufficient for specific evaluations. Instead several perspectives must be used in order to get a holistic assessment. However, we still need some principles for choosing and combining perspectives.

Furthermore, the perspectives presented are more suitable as basis for evaluations aiming at learning and improving research organization than as basis for evaluations aiming at controlling. This, I think, is a great advantage.

REFERENCES

Cameron, Kim S.
1986 "A study of Organizational Effectiveness and its Predictors", Management Science 32 (January): 87–112.

Cameron, Kim S.
1986 "Effectiveness as Paradox: Consensus and Conflict in Conceptions of Organizational Effectiveness", Management Science 32 (May): 539–553.

Conolly, Terry, Conlon, Edward J. & Deutsch, Stuart Jay
1980 "Organizational Effectiveness: A Multiple-Constituency Approach" Academy of Management Review 5: 211–217.

Elzinga, Aant
1986 Kreativitet, Paradigmeteori och Social Epistemologi – ett vetenskapteoretisk diskussionsinlägg om kreative forskning miljöer. Göteborg: Institutionen för Vetenskapsteori, rapport 147.

Foss Hansen, Hanne
1986 Organisering og Styring af forskning (Organization and Control of Research). Copenhagen: Nyt fra Samfundsvidenskaberne.

Foss Hansen, Hanne
1988 "Control and Organization of Research: A Case of Humanities Departments in Denmark" Science Studies 1/88: 15–24.

Foss Hansen, Hanne
1991 Organisatorisk effectivitet og økonomistyring (Organizational effectiveness and management accounting), in Foss Hansen & Neergaard: Organization og Økonomistyring. Copenhagen: Samfundslitteratur.

Gaertner, Gregory H. & Ramnarayan, S.
1983 "Organizational Effectiveness: An Alternative Perspective", Academy of Management Review 8: 97–107.

Galbraith, Jay R.
1977 Organizational Design. Reading Massachusetts: Addison-Wesley Publishing Co.

Hemlin, Sven
1991 Quality in Science. Researchers Conceptions and Judgments. Department of Psychology, University of Göteborg.

Hemlin, Sven & Montgomery, Henry
1990 "Scientists' conceptions of scientific quality. An interview study", Science Studies 1/90: 73–81.

Montgomery, Henry & Hemlin, Sven
1991 Judging Scientific Quality. A Cross-Disciplinary Investigation of Professiorial Evaluation Documents. Göteborg Psychological Reports 21, no. 4.

Morgan, Gareth
1986 Images of Organizations. Beverly Hills: Sage.

Premfors, Rune
1986 Forskningsmiljön i Högskolan. En kunskapsöversikt. University of Stockholm. Group for the Study of Higher Education and Resaerch Policy, Rapport 36.

Scharioth, Jürgen & Gizycki, Rainald V.
1986 Voraussetzungen und Möglichkeiten von Wissenschaft und Forschung (im universitären und ausseruninversitären Breich) für die Entstehung hochentwickelter Industrietechnologien, in Der Bundesrepublik Deutschland und für eine Intensivierung der Kooperationen mit Wirtschaft und Industrie in TI, transferinformation, Wissenschaft, Vermittlung, Praxis, Geselschaft zur Förderungen des Wissenschafttransfers. Frankfurt an Main (heft 1/1986).

Scott, W. Richard
1987 Rational, Natural and Open Systems. New Jersey: Prentice-Hall.

Whitley, Richard
1984 The Intellectual and Social Organization of the Sciences. Oxford: Claredon Press.

Yuchtman, Ephraim & Seashore, Stanley
1967 "A System Resource Approach to Organizational Effectiveness", American Sociological Review (December): 891–903.

Hanne Foss Hansen
Institute of Political Science
University of Copenhagen
Rosenborggade 15
1130 Copenhagen K
Denmark

Sven Hemlin
Pirjo Niemenmaa
Henry Montgomery

Quality Criteria in Evaluations: Peer Reviews of Grant Applications in Psychology

The authors of this paper have performed previous research on judgments of scientific quality by asking researchers about quality criteria (Chase, 1970; Hemlin & Montgomery, 1989; Hemlin, 1993) and by analysing documents containing quality judgments (Montgomery & Hemlin, 1991; Hemlin & Montgomery, 1993). The latter approach is followed up in this study. In contrast to Montgomery and Hemlin (1991) and Hemlin and Montgomery (1993) who analysed the evaluations of candidates for professorships (ex post evaluation), this investigation deals with evaluations of applications for research grants (ex ante evaluation).

Investigations of the reliability of peer review for grant submissions are scarce according to Cicchetti (1991) who reported only two studies showing surprisingly low reliability measures between judges. However, van den Beemt and le Pair (1991) demonstrated higher consensus when comparing the judgments of two independent juries who were evaluating research proposals in technology.

In previous psychological research less interest has been given to the values used in evaluations of scientific quality. Billig (1991) argued that researchers in social cognition have forgotten the social aspect of thinking and decision making. According to Billig the values and principles common to a group or society influence the language and arguments used by single individuals belonging to this group or society when making decisions or defending opinions and standpoints. On other occasions it has been argued that the members of the scientific community are expected to commit to a superordinate set of values and objectives characteristic for doing science. Kuhn (1970) claimed that scientific discourse takes place within paradigmatic frameworks. However, within the general scientific community there are smaller communities following their own intraparadigmatic orientations but all share the superordinate principles that unify all scientists as scientists (Merton, 1973). In a previous study by Montgomery and Hemlin (1991) researchers in different areas were

Science Studies, Vol. 8 (1995), No. 1, 44—52.

asked about scientific quality criteria. The results showed that some of the criteria were the same across the areas while other criteria differentiated mainly between soft and hard sciences. Another study (Hemlin & Montgomery, 1993) concerned which criteria were referred to when actually selecting a candidate for the professor position. Also in this study it was shown that some of the criteria were the same across disciplines. In our studies we found that Stringent or Correct Methods, New Results and New or Stringent Problems, in that order, were the three most frequently occurring criteria in peer judgments of scientific quality.

The purpose of the present study is to examine the evaluative criteria used in judgments of research grant applications. The purpose is to demonstrate the criteria of scientific quality used in the decisions of approved and rejected applications and to analyse the differences in judgments between approved and rejected applications. Another purpose is to investigate if the quality criteria regarded as important by researchers and the criteria actually used in evaluations coincided. The criteria of scientific quality used in evaluations of research applications in the current study were compared with the results in previous studies performed by Montgomery and Hemlin (1990) dealing with the quality criteria regarded as important by Swedish researchers in different research areas. An additional purpose is to compare the structure of criteria associated with acceptance and rejections with previous research on how people justify selection and rejection decisions (Beach & Mitchell, 1987; Westenberg & Koele, 1992).

Method

Documents

Peer review protocols of applications for research grants in psychology at the Swedish Council for Research in Humanities and Social Sciences (HSFR) were analysed. The protocols covered a period of six years between 1988 and 1993. Review protocols were divided into four groups in accordance with the decisions made. The first group (N = 145) consisted of applications which were approved with continued grants. The second group consisted of applications which were approved with new grants (N = 86). The third group (N = 92) consisted of applications which were approved but not granted due to the budget limits of the council, and, finally, the fourth group (N = 90) consisted of applications which were rejected. The length of the review protocols varied between 2 and 33 written lines. Decisions were made by the Council's review group of psychology consisting of 6 alternating senior researchers in psychology. Each group member usually took part in decisions on three successive years. Each application was reviewed by one member of the group. Reviewers were chosen by the chairman of the review group for psychology. However, decisions were made by the group collectively.

Procedure

Quality criteria. Each protocol was analysed according to a procedure developed by Montgomery and Hemlin (1991) in previous research on professorial evaluation documents, i.e., this procedure means that all value statements containing a positive, negative, or a neutral evaluation were marked in each protocol, i.e., all judgments in which a positive, negative, or neutral evaluation was made of the previous or actual research concerning the application or the researcher him/herself. Also, descriptive statements regarding the qualifications of the applicant were marked (e.g., teaching experience, number of supervised graduate students reaching a Ph.D., number of scientific articles, number of citations). This means that the statements were used as a unit of the analysis. Each delimited statement was coded into the four overall categories;

(a) the *object* of the judgment (the actual application, a single paper/ article/ re-

search effort, or the research/ researcher as a whole),
(b) the *aspect* of the research judged (problem, method, theory, results, reasoning, writing style, budget, or no specific aspect),
(c) the *attribute* connected to the judged aspect (for example, stringency, novelty, beauty). This category also included various descriptive statements such as the number of papers and published articles,
(d) the *value* of the statement, which could be positive, negative, neutral, or missing. For example, "high quality" was coded positively, while "method for the analysis and processing of data are not described" was coded negatively, and "a very much applied project" was coded neutral.

The coding procedure was applied in accordance with a manual in which each coding category was defined and exemplified (Montgomery & Hemlin, 1991). The codings for the first three years, between 1988 and 1990, and for the three last years, between 1991 and 1993, were performed independently by different judges (the first and second author, respectively). Approximately 30 percent randomly chosen codings of the three last years were also coded by the first author. The agreement of judgments in the codings performed by two independent authors was 79 percent.

Results

Criteria of scientific quality in evaluations

Values. To begin with, the percentage of different values in statements was studied. The direction of the value in the statements was categorized as positive, negative, neutral, or as no value. The results are displayed in Table 1. It can be seen that in the total sample the percentage of positive evaluations was greatest. About 66 percent of the evaluations were positive and about 32 percent were negative. Only about 2 percent of the evaluations were neutral, or a statement was missing. The results are in almost perfect accordance with the golden section hypothesis (Benjafield & Adams-Webber, 1976), which postulates a 63/37 "normal" distribution between positive and negative evaluations.

However, the distribution of positive and negative statements varied greatly in the different decision categories. For applications with a continued or new grant the percentage of positive evaluations was greater than 80 percent. For applications which were approved but not granted the percentage of positive statements was about 61 percent, and of negative statements about 36 percent. The distribution of positive and negative evaluations in this decision category came nearest the golden section. Interestingly, in rejected applications the distribution of positive and negative statements showed a reversed order. The percentage of neutral statements was so low that in the further analyses only the distribution of positive and negative statements was taken into consideration. Because of this, and, partly, because of the used method, according to which the judged objects, aspects, and attributes were coded separately and independently, both the absolute frequency of statements and the percentage of positive and negative evaluations of all evaluations varied to some extent in different analyses.

Objects. The percentage of different objects mentioned in the statements, and the distribution of positive and negative statements for different objects were studied. The results and also the number of statements are shown in Table 2.

The results show that in every decision category the overwhelming majority of judgments concerned the application. Rather few statements concerned the researcher or his/ her merits. Positive statements about individual research work were more frequently mentioned in judgments of applications with continued grant than in judgments of other decision categories. These judgments usually referred to the efforts made during the first year(s) of the project. Positive statements about the total research production/

Table 1. Percentage of judgments related to judged values.

Valence	Approval continued grant n = 755	Approval new grant n = 302	Approval no grant n = 394	Rejected n = 431	Total n = 1882
Positive evaluations	81.2	84.1	60.7	30.9	65.8
Negative evaluations	17.4	14.9	35.8	67.7	32.4
Neutral statements	.5	1.0	2.3	1.4	1.2
No value statements	.9	.0	1.3	.0	.6
Sum	100.0	100.0	100.0	100.0	100.0

Table 2. Percentage of judgments related to positively and negatively judged objects.

	Approval continued grant n = 755		Approval new grant n = 299		Approval no grant n = 394		Rejected n = 425		Total n = 1873	
Object	+	−	+	−	+	−	+	−	+	−
Individual research work	14.0	2.3	5.6	1.0	1.3	1.3	1.6	1.2	5.6	1.5
Total research/researcher	4.1	.0	9.3	.3	7.1	.8	1.4	2.8	5.5	.9
Application	63.0	15.1	69.2	13.6	52.3	33.8	27.8	63.8	54.7	31.6
Sum	81.2	17.4	84.1	14.9	60.7	35.8	30.9	67.7	65.8	34.0

Note. A few statements were not possible to categorize in the positive or negative judgment category and therefore the judgments do not always sum up to 100%.

the researcher were most frequently mentioned in judgments of applications with new grants.

Aspects. Most of the judgments, in each decision category, did not contain any specified aspect. (Table 3) However, when aspects were specified, the most emphasized positive aspects in the total sample were Theory, Method, and Problem, in that order. The most emphasized negative aspects in the total sample were Method and Theory. In the four different decision categories this pattern did not follow the same order. The results also show that in the judgments of granted and rejected applications the same aspects were not stressed.

Positive statements about Theory were most frequently made in judgments of applications with new grants. In judgments of the applications with continued grants positive statements about Results were more frequently occurring than other aspects. This is not surprising since applicants in this category are likely to have performed research within the granted project. In judgments of approved but not granted applications the percentage of negative statements was greatest concerning Method. In rejected applications the percentage of negative statements was greatest concerning Method and Theory, in that order. From these results it appears that the Method and Theory aspects were used to sort out applications of lower quality.

Attributes. The next analysis dealt with the distribution of positive and negative attributes in the judgments. (Table 4) In the total sample the most frequently used attribute concerning both positive and negative evaluations was Stringency (e.g. clear reasoning), followed by Novelty (e.g. original ideas) and Importance (e.g. essential problems) concerning positive evaluations, and by Correctness concerning negative evaluations. In judgments of applications with continued and new grants the most frequently used posi-

Table 3. Percentage of judgments related to positively and negatively judged aspects.

	Approval continued grant n = 755		Approval new grant n = 302		Approval no grant n = 394		Rejected n = 429		Total n = 1873	
Aspect	+	−	+	−	+	−	+	−	+	−
Problem	9.1	1.2	6.0	.3	8.4	2.8	4.4	4.4	6.9	2.2
Method	9.7	3.3	8.9	2.6	8.9	10.2	4.0	15.2	7.9	7.8
Theory	5.8	1.6	15.6	2.0	7.6	7.6	4.2	11.2	8.3	5.6
Results	9.8	.8	5.0	.7	6.3	1.3	1.9	1.9	5.6	1.0
Reasoning	2.5	.5	.7	.0	1.0	.5	.7	2.8	1.0	1.0
Writing style	.9	.5	2.3	.3	1.5	.5	.7	2.1	1.4	.9
Budget	5.6	3.6	1.0	3.6	2.5	1.5	.9	2.3	2.5	2.8
No aspect specified	37.7	5.8	44.7	5.3	24.4	11.4	14.2	27.7	32.5	12.6
Sum	81.2	17.4	84.1	14.9	60.7	35.8	31.0	67.6	64.3	33.9

Note. A few statements were not possible to categorize in the positive or negative judgment category and therefore the judgments do not always sum up to 100%.

tive attribute was Stringency, followed by Novelty. In judgments of applications which were approved but not granted and in judgments of rejected applications, Stringency and Correctness were the most frequently used negative attribute. Novelty, Importance, Extrascientific relevance, International relations, Breadth, and General Evaluative statements were more frequently mentioned as positive than negative attributes in every decision category.

Negative evaluations of Relevance of Sub-

Table 4. Percentage of judgments related to positively and negatively judged attributes.

	Approval continued grant n = 731		Approval new grant n = 294		Approval no grant n = 376		Rejected n = 424		Total n = 1825	
Attribute	+	−	+	−	+	−	+	−	+	−
Correctness	3.8	2.0	4.4	2.0	4.9	8.5	2.3	12.1	3.9	6.2
Importance	7.6	.1	10.1	.0	10.8	1.0	3.3	1.4	8.0	.6
Novelty	14.2	.8	13.5	.3	10.0	1.3	7.2	.9	11.2	.8
Stringency	17.1	4.1	19.5	6.4	9.7	15.4	5.8	18.6	13.0	11.1
Intrascient. relevance	3.9	.8	3.7	.3	4.4	2.3	.7	5.1	3.2	2.1
Extrascient. relevance	3.0	.0	3.0	.0	2.1	.3	2.8	.0	2.7	.1
Internat. relations	.5	.0	2.4	.0	.0	.0	.7	.0	.9	.0
Relevance of subject	.0	.9	.3	.3	.0	1.3	.2	8.8	.1	2.8
Breadth	.4	.0	1.3	.3	.3	.0	.5	.5	.6	.2
Depth	.3	.0	.3	.0	.0	.5	.0	.5	.2	.3
Productivity/activity	6.2	2.6	2.0	.3	2.1	.8	.7	1.9	2.8	1.4
Research ethics	.1	.0	.0	.0	.0	.0	.0	.2	.0	.1
Knowledge of subject	.1	.0	2.0	.0	1.5	.0	.5	.7	1.3	.2
Leadership of research	.0	.0	.0	.0	.0	.0	.0	.2	.0	.1
Research competence	1.2	.0	2.7	.0	2.1	.0	.5	.9	1.6	.2
General eval. statements	6.5	.5	7.7	.0	3.3	.5	3.0	.7	5.1	.4
Various	6.0	2.0	6.1	1.3	6.2	1.5	1.6	8.4	5.0	3.3
No attribute specified	11.1	2.9	4.7	3.7	3.3	2.6	1.2	6.7	5.1	4.0
Sum	71.0	14.0	79.1	11.5	57.2	33.3	29.7	61.0	59.6	29.9

Note. A few statements were not possible to categorize in the positive or negative judgment category and therefore the judgments do not always sum up to 100%.

ject, i.e. if the application concerned a problem relevant in psychology, was a frequently used attribute concerning rejected applications but not the other decision categories. Intrascientific Relevance, i.e. if changes in or development of theories are likely to occur because of the research proposed, was used more frequently as a positive than negative attribute concerning all decision categories except for rejected applications. Productivity/ activity, i.e. if the applicant had produced a lot of papers and/or were active in research in other ways, was more frequently used as a negative attribute in judgments of rejected applications than other applications.

In sum, approved and granted applications were connected with positive, or at least equally positive and negative evaluations of all attributes except for Relevance of Subject. Rejected applications were connected with negative, or equally negative and positive evaluations of most of the attributes. Stringency and Novelty were the positive attributes most frequently stressed concerning applications with continued and new grant, while lack of Stringency and Correctness were the attributes most frequently mentioned concerning applications which were approved but not granted and rejected applications. In rejected applications Novelty was less frequently mentioned as a positive attribute than in other decision categories but it was, however, not an important negative attribute in this decision category.

Quality criteria regarded as important compared with criteria actually used

Aspects. In previous studies by Montgomery and Hemlin (1991) and Hemlin (1993) criteria of scientific quality regarded as important by researchers from several research areas were studied. The researchers made free reports of the aspects and attributes they regarded as important. These results were compared with the results of this study where evaluations of the scientific quality in research applications in psychology were analysed. Thus, in the first study the results referred to criteria regarded as important, and in this second study to criteria actually used by the judges. In Table 5, in rank order, the most frequently mentioned positive aspects in the total group in this study are shown together with the aspects most frequently mentioned by researchers within different research areas. The category called No aspect specified was deleted in the table to promote clarity.

According to results (Table 5) Problem and Method were the aspects frequently used and mentioned in every group. The importance of Theory and Result varied greatly between the groups. Theory was more frequently mentioned than Result by the judges at HSFR and also regarded as more important by the researchers in social sciences and humanities, while Result was regarded as more important than Theory by researchers in natural, medical, and technical sci-

Table 5. Ranking of different aspects by researchers in different areas and the ranking of aspects used in evaluations by the judges at HSFR.

Aspect	HSFR	Hum	Med	Nat	Soc	Tech
Problem	3	2	1	3	3	1
Method	2	1	2	1	1	1
Theory	1	3	5	5	2	5
Result	4	5	3	2	4	3
Reasoning	6	5	6	6	6	3
Writing style	5	4	4	4	5	6

Note. HSFR = Swedish Council for Research in Humanities and Social Sciences, Hum = humanities, Med = medical sciences, Nat = natural sciences, Soc = social sciences, Tech = technical sciences. Rankings are based on the proportion of occurences in the present investigation and in a former one reported in Montgomery and Hemlin (1991) and Hemlin (1993).

Table 6. Ranking of the most frequently mentioned attributes by researchers in different areas and the most frequently mentioned (positive) attributes in evaluations by the judges at HSFR.

Attributes	HSFR	Hum	Med	Nat	Soc	Tech
Correctness	3	3	2	3	4	2
Novelty	2	2	1	1	1	1
Stringency	1	1	3	2	2	2
Intrascientific relevance	5	4	6	5	4	6
Extrascientific relevance	4	4	4	6	3	2
International relations	6	8	5	4	6	5
Breadth	7	6	7	7	5	7
Depth	8	6	7	7	7	8

Note. HSFR = Swedish Council for Research in Humanities and Social Sciences, Hum = humanities, Med = medical sciences, Nat = natural sciences, Soc = social sciences, Tech = technical sciences. Rankings are based on the proportion of occurences in the present investigation and in a former one reported in Montgomery and Hemlin (1991) and Hemlin (1993).

ences. Thus, the difference was greater between soft sciences and hard sciences than between the aspects regarded and actually used as important aspects within soft sciences.

Attributes. In Table 6 the rank order of the importance of the attributes in the two samples are displayed. Only those attributes regarded as important at least by 10 percent of one or several research groups are shown. Concerning the attributes mentioned by the judges at HSFR only the most frequently mentioned positive attributes in the total group are used in comparison.

The results show that some of the attributes were both regarded and actually used as important criteria of scientific quality in all. Novelty and Stringency were the attributes most frequently mentioned by the researchers both in social and other sciences, as well as by the judges at HSFR. However, concerning some of the criteria the results did not coincide. Extrascientific relevance was regarded as a more important criteria than intrascientific relevance by the researchers in social sciences, while the order was reversed in the evaluations made by the judges at HSFR.

The results in Table 5 and Table 6 show that there was an agreement about the important criteria of scientific quality between the evaluations made by judges at HSFR and the criteria reported as important by researchers in different research areas. However, there were some differences between the groups, but they mainly dealt with the differences between soft and hard sciences.

Discussion

The present study gives additional support to the notion that there are some objects, aspects, and attributes characteristic to judgments of scientific quality. By and large, the distribution of these criteria agreed with distributions found in an earlier study where scientists were asked to characterize qualities of good research (Hemlin, 1993). The agreement was especially good with judgments from representatives of social science, i.e., the main area into which psychology traditionally is subsumed. The same types of objects, aspects, and attributes have also been found in studies of professorial evaluation documents (Hemlin, Johansson & Montgomery, 1990; Hemlin & Montgomery, 1993), as well as in interviews with scientists on how they view scientific quality (Hemlin & Montgomery, 1991).

The distribution of important aspects and attributes within different decision groups suggests that the evaluators to some extent used different criteria for different types of decisions. The selection of high quality proposals was largely based on the Theory aspect and on the Stringency and Novelty attributes, whereas rejection decisions largely were based on the Method aspect and the

attributes of Correctness, Stringency, and also relevance of Subject. In the same vein, Montgomery and Hemlin (1991) found that the selection of the No. 1 candidate for a professorship partly was based on criteria that were different from the criteria used for screening out non-top candidates. It is interesting to compare these findings with results from psychological decision research. Several studies have demonstrated that it is fruitful to distinguish between screening decisions (rejection of nonsatisfactory options) and selection decisions (selection of good (the best) options) (Beach & Mitchell, 1987; Dahlstrand & Montgomery, 1984; Montgomery & Svenson, 1989; Westenberg & Koele, 1990, 1992; Wright & Barbour, 1977). Screening decisions tend to focus more on negative features and to be less compensatory than is true for selection decisions. In addition, Cicchetti (1991) reported that reviewers are much more in agreement on rejection than on acceptance of scientific documents. However, we have found no research on the extent to which different criteria (attributes or aspects) are focused in the two types of decisions. The present results as well as the findings in the Montgomery and Hemlin (1991) study suggest that it is worth-while to conduct research on this topic.

The present documents showed how the review group was influenced by some shared values in their evaluations and probably also how the group wished to justify their recommendations to the applicants and to members of the Board of the Social Science Research Council who make the final decisions. Presumably, the evaluation documents examined in the present study represent rhetorical accomplishments, which also may be true for the professorial evaluations documents analysed in the Montgomery and Hemlin (1991) study. The distribution of positive and negative evaluations in different decision groups may be interpreted from a rhetorical perspective. All approved proposals including those which were ranked quite low (approved but not granted) were much more positively than negatively evaluated. In sharp contrast, rejected applications were much more negatively than positively evaluated. It may be speculated that the need to write convincing justifications may have induced the evaluators to sharpen the differentiation between approved and non-approved applications.

It may be asked how much the evaluation documents reflect the total background to the recommendations made. Rhetorical purposes as well as notions about what is socially expected of the evaluative protocols in the scientific community may reveal only certain aspects of the criteria used in decisions. There are also other criteria (Merton, 1973) and counter-norms (Mitroff, 1974), such as the applicant's position in the scientific community, which may have a more or less direct influence on the evaluations made by the peers. The rather low figures for the evaluation of the researcher and/or his/her full record (as appeared in Table 2) might also reflect this idea. In a following study, we will examine how factors besides those analyzed in this study are related to the evaluators' evaluations and recommendations.[1]

NOTES

1. The authors are grateful for a number of valuable suggestions on the manuscript from an anonymous referee. The study was supported by a grant to the first and third author from the Council for Studies of Research and Higher Education.

REFERENCES

Beach, L.R., & Mitchell, T.R.
1987 Image Theory: Principles, Goals, and Plans. Acta Psychologica, 66: 201–220.

Benjafield, J., & Adams-Webber, J.
1976 The golden section hypothesis. British Journal of Psychology, 67: 11–15.

Billig, M.
1991 Ideology and Opinions. Studies in Rhetorical Psychology. London: Sage.

Chase, J.M.
1971 Normative criteria for scientific publication. The American Sociologist, 5: 262–265.

Cicchetti, D.V.
1991 The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. Behavioral and Brain Sciences, 14: 119–135.

Dahlstrand, U., & Montgomery, H.
1984. Information search and evaluative processes in decision making: A computer based process tracing study. Acta Psychologica, 56: 113–123.

Harnad, S.
1985 Rational disagreement in peer review. Science, Technology, & Human Values, 10: 55–62.

Hemlin, S., & Montgomery, H.
1993 Peer judgments of scientific quality: A cross-disciplinary document analysis of professorship candidates. Sciences Studies, 6: 19–27.

Hemlin, S.
1993 Scientific quality in the eyes of the scientists. A questionnaire study. Scientometrics, 27: 3–18.

Hemlin, S., & Montgomery, H.
1991 Scientists' conceptions of scientific quality. Science Studies, 1: 73–81.

Hemlin, S., Johansson, U.-S., & Montgomery, H.
1990 Professorstillsättningar i arkitekturämnen. På vilka grunder sker besluten? (Decisions on chairs in architecture. On what grounds are decisions made?) Stockholm: Byggforskningsrådets vetenskapliga nämnd, BVN, Skriftserie 1990:1.

Kuhn, T.S.
1970 The Structure of Scientific Revolutions (2nd. ed.). Chicago, Ill: The University of Chicago Press.

Merton, R.K.
1973 The normative structure of sciences. In N. W. Storer (ed.), The Sociology of Science: Theoretical and Empirical Investigations. Chicago, Ill: University of Chicago Press.

Mitroff, I.I.
1974 Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. American Sociological Review, 39: 579–595.

Montgomery, H., & Hemlin, S.
1991 Judging scientific quality. A cross-disciplinary investigation of professorial evaluation documents. Göteborg Psychological Reports, 21.

Montgomery, H., & Svenson, O.
1989 A think aloud study of dominant structuring. Pp. 135–150 in H. Montgomery & O. Svenson (eds.), Process and Structure in Human Decision Making. Chichester: Wiley.

Van den Beemt, F., & Le Pair, C.
1991 Grading the grain: Consistent evaluation of research proposals. Research Evaluation, 1: 3–10.

Westenberg, M.R.M., & Koele, P.
1992 Response modes, decision processes, and decision outcomes. Acta Psychologica, 80: 169–184.

Westenberg, M.R.M., & Koele, P.
1990 Response modes and decision strategies. Pp.159–170 in K. Borcherding, O.I. Larichev, & D. M. Messick (eds.), Contemporary Issues in Decision Making. Amsterdam: Elsevier.

Wright, P., & Barbour, F.
1977 Phased decision strategies: Sequels to an initial screening. Stanford, CA: Stanford University, Graduate School of Business, Research Paper No. 353.

Sven Hemlin
Department of Psychology
Göteborg University.

Pirjo Niemenmaa
Department of Psychology
University of Stockholm.

Henry Montgomery
Department of Psychology
University of Stockholm.