Sven Hemlin and Henry Montgomery

Scientists' conceptions of scientific quality[1] An interview study

What is scientific quality? A definitive answer to this question will perhaps never be given. Indeed, a definitive answer may not even be desirable since scientific progress is linked to developments in the prevailing conceptions of scientific quality. Moreover, disagreements about the meaning of scientific quality may be a natural ingredient in the competition between scientists from different camps, a competition which may be a prerequisite for scientific progress.

Nevertheless, it is meaningful to examine what might be meant by the term "scientific quality". It is of interest to clarify different aspects of scientific quality, which is an enterprise different from asserting the "true" meaning of that concept. It is also of interest to examine how researchers in different disciplines look upon scientific quality. To what extent do scientists have similar conceptions of scientific quality? The present article aims at shedding light on these questions.

Laudan (1984) suggested that empirical studies should be carried out in order to find out how research is guided by different values (internal scientific values)[2]. This proposal is in line with our view and the empirical study reported here is an attempt in this direction.

A conceptual framework for describing scientific quality

The basic idea underlying our conceptual framework is that judgments about scientific quality are related to seven interacting factors, which in turn could be characterized in a number of respects. These factors are: quality indicators (which could be subdivided into quantitative indicators such as citation frequencies, number of awards, etc., and more qualitative indicators such as judgments emanating from peer reviews), the research effort (which could vary in "size" from a single research paper to the research carried out by a whole nation), the researcher (i.e. his or her personality and competence), the research environment (physical and social characteristics), intrascientific effects (i.e. contributions to the knowledge state of a discipline or a research programme), extrascientific (societal) effects (e.g. contributions to peace and disarmament, remedies for cancer) and research policy (i.e. the policy conducted by the society and the scientists themselves through various media like funding committees).

These seven components form a network in which different meanings of the concept "scientific quality" may be traced. The exact meaning of "scientific quality" in a given context will depend on which of the listed factors are emphasized. For example, the importance of using appropriate quality indicators may be stressed and certain indicators may be considered to be especially important. One may stress certain components of the research effort. One may take the researcher's general competence into account. The importance of intrascientific and societal goals may be more or less explicitly considered. And so on. Below, we describe how a sample of Swedish scientists looked upon and stressed these factors. We will in this paper concentrate on statements concerning the research effort, which perhaps is the key component of scientific quality. As a background to the interview results, we first describe a content analysis of the criteria of scientific quality that have been suggested in the literature.

Content analysis of quality criteria in the literature

We found five lists based on empirically derived criteria from judgments of articles in scientific journals (Chase, 1970; Frantz, 1968; Lindsey & Lindsey, 1970; Rowney & Zenisek,1968; Wolff, 1970). The number of criteria in each of these studies was 10, 14, 6, 15, and 12 respectively. As an example consider the list of criteria put forward by Chase (1970). She developed her criteria for scientific publication from three sources in the literature and had them ranked by 313 scientists from different disciplines. The highest ranked criteria were the following: "logical rigor", "replicability of research techniques", "clarity and conciseness of writing style", "originality", "mathematical precision", "coverage of significant existing literature", "compatibility with generally accepted disciplinary ethics", "theoretical significance", "pertinence to current research in the discipline" and "applicability to 'practical' or applied problems in the field".

Normative criteria were reported in three papers (Kuhn, 1977; Michael, 1967; NSF, undated). Kuhn (1977) claimed that theories should be "accurate", "consistent", "simple", "fruitful" and have "scope". Yet another proposal was made by Michael (1967) concerning criteria for evaluating research reports, articles or theses. Detailed instructions were stated in Michael's manual regarding title, problem, design and methodology, presentation and analysis of data, summary and conclusion. The criteria of scientific quality proposed by the National Science Foundation of the USA are used for evaluation of research proposals. They consist of four criteria, namely "research performance competence", "intrinsic merit of the research", "utility or relevance of the research" and "effect of the research on the infrastructure of science and engineering", to be judged on a rating scale of five steps (excellent, very good, good, fair and poor).

A closer look at the total of eight sets of criteria shows that they usually could be analysed into two components. For example, the criterion "new, statistical method" refers, on the one hand, to a certain aspect of a research effort, viz. Method, and, on the other hand, to a desirable attribute, viz. New. We shall call these two components aspects and attributes. Aspects and attributes were often but not always combined, i.e. "new, original theory" (an example from Rowney and Zenisek, 1980) and "objectivity in reporting results" (proposed by both Frantz (1968) and Wolff (1970).

Table 1 shows a classification in terms of aspects and attributes of all criteria for scientific quality found in the eight lists.[3] Six aspects are represented in the table, viz. Problem, Method, Theory, Results, Reasoning, and Writing style. Eight attributes were common, viz. Correctness, Novelty, Stringency, Intrascientific effects, Extrascientific effects, Utility in general, Breadth, and Competence. As can be seen, it was usually possible to categorize the criteria in terms of aspects and attributes, although in some cases only one component was specified, i.e. either the aspect or the attribute.

The most frequent combination in Table 1 were Stringent Method (5 items). New Theory, Stringent Theory and Correct Reasoning shared 3 items each.

Interviews with researchers

Method

Twenty-two researchers, all employed at Swedish universities as full professors (21 persons) or as associate professor (1 person), were interviewed. The sample was chosen to represent scientists experienced in evaluating research. Names of professors were randomly drawn from staff catalogues of seven universities in five university cities of Sweden. The respondents represented all major faculties, i. e., medicine, natural science, humanities, theology, law, social science, and technology, as well as two inter-disciplinary fields (the "theme" subjects at the University of Linköping, Sweden)[4]. The interviews were carried out orally by two persons, namely one of the authors and a trained psychologist. One of the interviewers asked the questions and the other took notes. The tasks of asking questions and taking notes were interchanged. The interviews were anonymous, semi-structured and referred to each of the factors described above. One question directly concerned the features of good science in the respondent's discipline ("What characterizes good and bad research within your area in Sweden?"). This question was given early on in the interview (after a "warming-up" question about current research the interviewee was doing). For a few questions, examples of possible answers were given. The interviews lasted approximately two hours and were conducted in the respondent's office or home.

Results and discussion

The results of the interviews are summarized in Figure 1. The numbers in the figure show the number of interviewees who endorsed a particular view of a certain factor. Frequencies of particular answers were also computed for different groups of disciplines, viz. natural science (including medicine and technology), humanities (including theology), and social science (including law), and "theme" subjects. The number of respondents in each of these groups were, respectively, 8, 7, 5, and 2. Due to the low number of respondents in each group, only very tentative conclusions can be drawn regarding differences between views in the groups.

Aspects and attributes. In Figure 1 the numbers in the box labelled "Research effort" denote the frequency of interviewees who mentioned each combination of aspects and attributes in response to the question that concerned characteristics of good (and bad) research.

The most common aspects were Problem, Method, Results, Theory, and Reasoning. An example of a statement dealing with the Problem aspect was: "Clearness of problem statement", uttered by a "theme" professor. The importance of Method was described by a social scientist: "Clear understanding of which empirical data are needed to produce answers to the problem statement".

"Pioneering contributions, previously not studied areas" was another answer from a professor in humanities that concerned the Result aspect. A social scientist focused on the Theory aspect: "...development of theories". "Proposal of a new problem statement or theory..." is another example stressing both the Problem and the Theory aspects.

The aspect we called Reasoning is examplified in the answer of a professor in inter-disciplinary research. "The most important things are logic and consequence. The arguments of different chapters should not contradict each other".

The most common attributes were, Novelty, Correctness, Stringency, and Intrascientific and Extrascientific effects. As examples of these attributes consider the following statements.

"Originality is important. It is of general interest that the researcher contributes with something new and that the novelty will have a general interest" (social scientist). This statement illustrates the Novelty attribute.

"The figures should be correct. There should not be miscalculations or any doubt of what a

Table 1. Categorization of criteria for scientific quality in the literature.

			Aspects
Attributes	Problem	Method	Theory
Correctness		Appropriateness of the statistics (M) Replicability of research techniques (C)	Accurate theories (K)
Novelty	Topic of interest (R & Z)	New statistical methods (R & Z)	New theory ... (R & Z). New thought (R & Z) Creativity of ideas (L & L)
Stringency	Clear problem ... (M)	Clarity of tabular material (F) Clarity of tables (W) Simplicity of methods (L & L) Clear description of samples (M) Mathematical precision (C)	Precise... hypotheses (M) Consistent theory (K) Simple theory (K)
Intra-scientific effects			Theoretical relevance (L & L) Theoretical significance (C)
Extrascientific effects			
Utility in general	Importance of the problem area (M)		Fruitful theory (K)
Breadth			Scope of theory (K)
Competence			
Other attributes or no attribute specified	Topic selection (F; W) Relevance to journal (L & L)	Research design (W) Design (F) Statistical analysis (F; W) Grasp of design (L & L)	Coverage of significant literature (C) Theoretical model (F; W)

Note. C=Chase, F=Frantz, K=Kuhn, L & L=Lindsey & Lindsey, M=Michael, N=NSF, R & Z=Rowney & Zenisek, W=Wo

Result	Reasoning	Writing style	No aspects specified
Conclusions other... can verify (M)	Objectivity in reporting results (F; W) Appropriately discussed... (M)	Objective... presentation (M)	Thorough and complete (S & R)
Original empirical evidence (L & L)			Originality (C)
	Logical and orderly expositions... (M) Appropriate caution in drawing conclusions (M)	Clarity and conciseness of writing style (C) Unambiguous statement of... (M)	Logical rigor (C)
Contribution to knowledge (F; W)			Pertinence to current research in the discipline (C) Intrinsic merit of the research (N) Suggestions for future research (N)
			Effect of the research on the infrastructure (N) ... implementation of findings (M) Applicability to "practical" problems (C) Practical implications (F; W) Value to social life (L & L) Value of findings (L & L)
			Utility or relevance of the research (N)
Conclusions at a scope (M)			
			Research performance competence (N) Author's status (W) Reputation... (F; L & L; R & Z) Scholarship (L & L)
		Writing style (F; W) Length... (F; R & Z; W) Punctuation (F; W) Entertainment... (L & L)	Compatibility with... ethics (C) Ethical sense (L & L) Institutional affiliation (F; W) Review of literature (F; W)

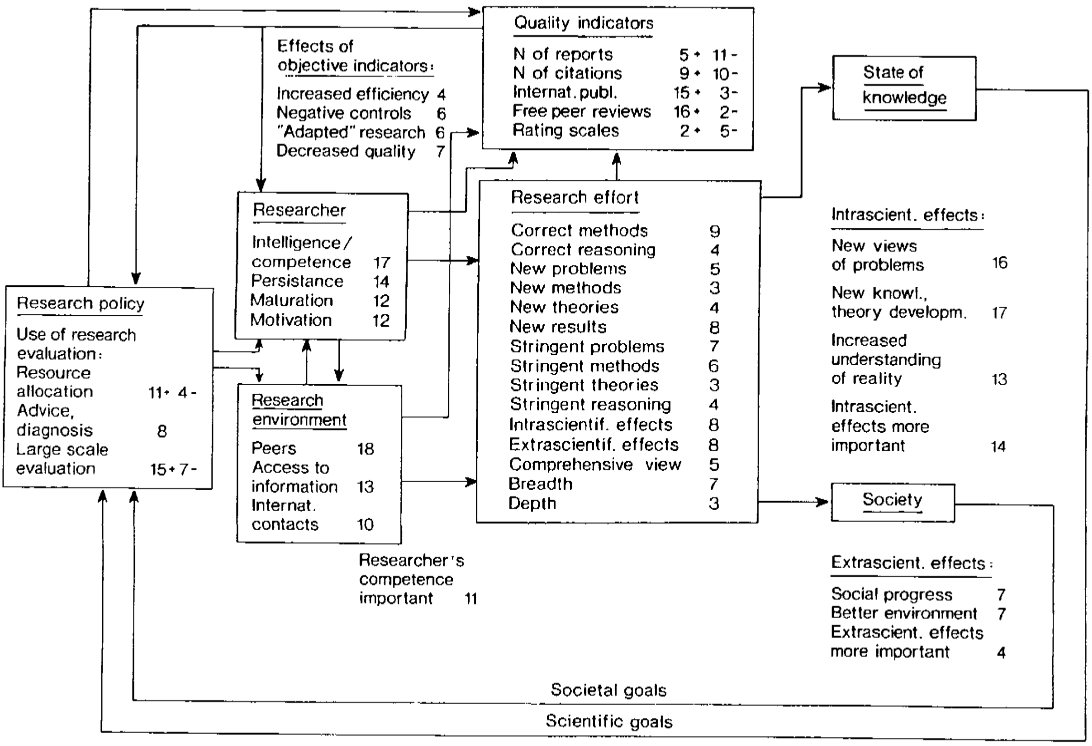

Figure 1. Conceptions of factors related to scientific quality according to interviews with 22 researchers. Only the most frequent answers are shown. +=positive view, —=negative view or de-emphasizing.

figure stands for"; "Bad science is recognized by the incompleteness of the data...and the incorrect systematization of the data" are two utterances coded into the Correctness category.

Stringency was often mentioned, e.g. when a scientist in the humanities answered that good science should be "...stringent, concise and clear".

An example of an utterance referring to intrascientific effects is: "It is important how research influences other research and education" (natural scientist). Sometimes the answers concerned Societal or Extrascientific effects: "...how shall we escape being destroyed...how can people in the developing countries attain a more tolerable living" (humanistic researcher).

Breadth of research was more often stressed than Depth. "When evaluating candidates for an academic position the breadth of the person is important", was mentioned by a social scientist.

Aspects and attributes were normally combined in a number of ways such as Correct Methods, New Results and Stringent Problems (the three most common combinations).

In general, the answers were rather evenly spread across the four different groups of disciplines. For example, the importance of Correct methods were emphasized by representatives for natural science, humanities, social science, and the "theme" subjects. However, the importance of Theory was only mentioned by one natural scientist as compared to seven persons from the other groups. Natural scientists, in general also had less to say about what they meant by good science as compared to the other groups of scientists (M=5.6 statements for natural scientists, M=9.7

for humanists, M=8.0 for social scientists, and M=10.0 for "theme" scientists).

The attributes shown in Table 1 fit our empirical results quite well. Kuhn's (1970) attribute "accuracy" is linked to Correctness, "consistency" and "simplicity" with Stringency, "scope" with Breadth and "fruitfulness" with Utility in general. The NSF norms focus on Intra- and Extrascientific effects while Michael's (1967) list covers all of the aspects we have found. Michael's list also included several attributes such as Correctness, Stringency and Novelty.

The favoured combinations were also in line with many of the criteria that we found in the literature. The three highest ranked criteria in the studies of Chase (1970), Frantz (1968), Rowney and Zenisek (1968) and Wolff (1970), emphasized combinations involving methods and research design. Most frequent attributes in the literature were combinations including Correctness and Novelty, attributes that were frequent also in our empirical findings.

Other components in the framework. The intelligence and competence of the researcher seem to be an important factor in judgments of scientific quality inasmuch as 17 respondents said that they took this factor into account when evaluating research quality.

It was often stressed that the eminent researcher should have a certain degree of intelligence, but lots of creativity. He or she should be persistent and have a strong motivation for research. This is all in line with the three factors that Rushton, Murray and Paunonen (1983) found best predicted creative research, viz. achievement, motivation and ambition. According to Albert (1975) intelligence ceases to explain the variance between scientists at an IQ-level of 120. Other variables explain the variance better (Mahoney, 1979).

The most important feature of a good research environment according to a majority of our respondents was the possibility of exchanging ideas with peers. The importance of good international contacts was also emphasized. Access to information via books or computers were also regarded as important by several respondents.

Other physical aspects such as good premises and sufficient research money were mentioned only by a few respondents or they were explicitly de-emphasized.

A majority of the respondents regarded intrascientific effects as more important than extrascientific effects when evaluating scientific quality. However, 4 of the 7 humanists regarded extrascientific effects as equally or more important than intrascientific effects.

Most of the respondents favoured results from qualitative peer reviews as indicators of scientific quality. The attitude to quantitative indicators such as citation frequencies and number of reports were mainly negative with an exception for the natural scientists who had a more positive attitude to such indicators.

International publication, however, was generally regarded as a good quality indicator. As to the effects of using different types of indicators, negative effects of quantitative indicators were mentioned much more often than positive effects. It was feared that the use of such indicators may lead to shallow and "adapted" research.

In answering questions about large-scale evaluations (e.g, of a nation-wide evaluation of a particular discipline) we received positive replies from the majority of the scientists. They were not negative to evaluations with purposes such as resource allocation or advisory and diagnostic functions. Half (11) of our respondents asserted that research evaluation should be used for resource allocation, and 8 respondents stressed the advisory or diagnostic function of research evaluation. Those researchers who did not approve of large-scale evaluations were also the same persons who rejected the resource allocation purpose.

General discussion

Below, we discuss what the results suggest with respect to the existence of common themes in conceptions of scientific quality in the academic community.

The results seem to support our conceptual framework for describing scientific quality. Our respondents' statements suggest that behind the notion of scientific quality there is a complex network of ideas of causal and conceptual relations. There are ideas of the relationship between quality indicators and features of the research effort, of the effects of using different types of indicators, of intrascientific and societal effects of science, of the importance of the researcher and of an appropriate research environment, etc. It should be noted that these ideas often were offered spontaneously in response to our general question concerning the features of good science. Of particular interest is the finding that the respondents often described good research in terms of combinations of aspects and attributes (e.g., Correct methods, New results). In the literature we also found normative or empirically generated criteria of scientific quality. Almost without exception the criteria were connected to the notion of the research effort which could be analysed in terms of aspects and attributes. An awareness of the existence of this structure in the language of scientific evaluations might be helpful in attempts to make systematic and comprehensive research evaluations.

Is there a more specific theme in our respondents' statements? Let us choose a few common keywords from these statements: new problems and results, correct methods, stringent problems, methods, and reasoning, stress on intrascientific effects, a strong motivation for research, exchange of ideas with peers, qualitative peer reviews. Behind these keywords, one may get the impression of a search of knowledge for its own sake conducted by a brotherhood of strongly motivated individuals who are anxious to keep a maximum of freedom in their enterprise (the favouring of free, qualitative peer reviews). This may appear to be a somewhat idealistic picture of the academic world. However, the picture is based on the words from persons many of whom have spent half their lives in academia. It may be interesting to find out in future research how well this picture describes the research practice and how research actually is evaluated.

It is important, however, to note that the respondents, although restricted to a small sample of professors, were fairly united in their views of scientific quality. The picture given above was largely true for natural scientists, humanists, and social scientists. This indication of consistency in views from different disciplines goes together with other empirical results like Maini's finding (1980) that natural and behavioural scientist were similar in their judgments of critical incidents in the research process. On the other hand, there are some indicators of different views concerning the role of theory in natural sciences vs. other sciences. Although we cannot make any claim on differences from our restricted data, there might be a tendency for natural scientists not to discuss the value of theory. This finding might indicate that natural scientists regard theories as such a fundamental aspect of research that they tend to take them for granted. It might be interpreted in terms of natural science having reached the state of "normal science", that is a state in which few new theories are invented or questioned (Kuhn, 1962). In accordance with this tendency is Chase's (1970) and Laudan's (1984) proposal that natural scientists agree more about theories and methods in their area than social scientists who often dispute the value of a theory or a research technique. Therefore it will be important for social scientists and researchers in the humanities to lay special emphasis on the need for good theories.

In conclusion, we wish to stress the need for supplementing interview data with other types of data to uncover scientist's conceptions of scientific quality. One approach would be to look at written documents concerning evaluations of research papers (e.g. referee statements). Another possibility is to conduct cognitive process tracing studies (e.g. think aloud reports) of how scientists evaluate research efforts or research proposals.

NOTES

1. This study was supported by a grant from The Board of the Chancellor of the Swedish Universities. The authors have contributed equally to this paper. We

wish to thank Ulla-Stina Johansson for her interviews with researchers.

2. McMullin (1982) proposed that a distinction ought to be made between epistemic and non-epistemic values in science. The former are internal values of science and the latter are pragmatic, e.g. political and moral values. We are especially concerned with the quality of the research effort which seems to be equivalent to McMullin's epistemic values.
3. The quotations in the table cover all criteria proposed in the lists with an exception for Michael's list which sometimes included different number of specifications of an attribute. In the latter case, one of these specifications are examplified in Table 1.
4. Professors of the following disciplines were chosen: Medicine (histology, pediatrics), Dental faculty (oral radiology), Natural science (computer science, inorganic chemistry, numerical analysis), Agriculture (food hygiene), Humanities (computational linguistics, history, history of literature, musicology, philosophy, philosophy of science), Theology (dogmatics), Law (jurisprudence), Social science (business economics, political science, psychology, sociology), Technology (steel and timber structures), Inter-disciplinary subjects or "themes" (technology and social change, water in environment and society).

REFERENCES

Albert, R. S.
1975 "Toward a Behavioral Definition of Genius". American Psychologist, 35: 140—151.

Chase, J. T.
1970 "Normative Criteria for Scientific Publication". The American Sociologist, 5: 262—265.

Frantz, T. T.
1968 "Criteria for Publishable Manuscripts". Personnel and Guidance Journal, 47: 384—386.

Kuhn, T. S.
1962 The Structure of Scientific Revolutions. Chicago: The University of Chicago Press.
1977 The Essential Tension. Chicago: The University of Chicago Press.

Lindsey, D. & Lindsey, T.
1970 "The Outlook of Journal Editors and Referees on the Normative Criteria of Scientific Craftsmanship: Viewpoints from Psychology, Social Work and Sociology". Quality and Quantity, 12: 45—62.

Laudan, L.
1984 Science and Values. Berkeley and Los Angeles: University of California Press.

The National Science Foundation (NSF).
Undated Information for Reviewers. Xerox copy.

Mahoney, M. J.
1979 "Psychology of the Scientist: An Evaluative Review", Social Studies of Science, 9: 349—375.

Maini, S. M.
1980 Explorations in the Psychology of the Research Process, Lund: CWK Gleerup.

McMullin, E.
1983 "Values in Science." Pp. 3—28 in P. Asquith and T. Nickles (eds.), PSA 1982, 2, East Lansing, Michigan: Philosophy of Science Association.

Michael, W. B.
1967 Criteria for Evaluation of a Research Report, Article, or Thesis. University of Southern California. Xerox copy from unknown source.

Rowney, J. A. & Zenisek, T. J.
1968 "Manuscript Characteristics Influencing Reviewers' Decisions". Canadian Psychology/Psychologie Canadienne, 21: 17—21.

Rushton, J. P., Murray, H. G, & Paunonen, S. V.
1983 "Personality, Research Creativity, and Teaching Effectiveness in University Professors", Scientometrics, 5: 93—116.

Wolff, W. M.
1970 "A Study of Criteria for Journal Manuscripts". American Psychologist, 25: 636—639.

Sven Hemlin and Henry Montgomery
Göteborg University
Department of Psychology
Box 14158
S-40020 Göteborg
Sweden