Organizational Research Methods
Volume 11 Number 3
July 2008 510-540

10.1177/1094428107308855
http://orm.sagepub.com
hosted at
http://online.sagepub.com

The Emergence of Implicit Criteria Actually Used by Reviewers of Qualitative Research Articles

Case of a European Journal

Henri Savall
Véronique Zardet
Marc Bonnet
ISEOR, Université Jean Moulin Lyon 3
Michel Péron
ISEOR, Université Paris III–Sorbonne Nouvelle

This article brings to light the actual evaluation practices of reviewers when assessing qualitative manuscripts. The authors conducted the first empirical research entirely based on reviewer reports for a journal on management sciences over a 28-year period. Content analysis of 474 reviewer reports written by 56 reviewers identified 19 critical points and 10 criteria, making up a synthetic table of 190 possible cells, 51 of which proved to be actually used by reviewers. These findings are then compared with the quality criteria identified in the extant U.S. literature. Comparison reveals some shared quality criteria but also throws new light on a number of discrepancies. Analysis over time shows marked evolution from emphasis on internal validity criteria over the first 10 years toward emphasis on external validity criteria in the past 10 years. Factors ensuring reliability and replicability of the results of this research are discussed at length.

Keywords: *academic review process; qualitative content analysis; implicit criteria; method; replicability; validity*

The object of this article is to identify the pattern of criteria used by reviewers when they are called on to judge qualitative manuscripts. The methodology employed is based on the content analysis of 474 reports written by 56 reviewers over a period of 28 years at a European journal. Nineteen themes or key points, referred to here as ''critical points,'' and 10 quality criteria were inductively developed. These themes are compared to the dimensions identified by several U.S. commentators on what makes good qualitative research (e.g., Gephart, 2004; Locke & Golden-Biddle, 2002; Suddaby, 2006). This research is specifically devoted to the empirical identification of what criteria were actually used (although implicitly) to judge qualitative manuscripts. Furthermore, it identified some interesting findings for qualitative reviewers and editors. It showed that some dimensions identified through comparative analysis

are not found in the American literature, that some reviewers changed their criteria over time, and that most of the critical points/quality criteria remained nevertheless stable over time. This article provides the first empirical look at what qualitative reviewers use (the ''intangible cues''; Daft, 1995) in judging qualitative manuscripts and consequently is an important contribution to our knowledge in that area.

Given that there is no a priori evaluation grid of quality criteria for qualitative research articles at the studied journal entitled *Sciences de Gestion—Management Sciences—Ciencias de Gestión (RSDG)*, the research question is: What implicit criteria are actually used by reviewers in their evaluations of articles employing qualitative research methods submitted to the journal? Our explanatory hypothesis (Peirce, 1903) is that it is possible to bring out implicit criteria through a rigorous content analysis of a large number of reviewer reports written by a significant number of diversified reviewers. This hypothesis can be broken down into several questions that oriented our content analysis:

- What are the critical points that attract the attention of reviewers?
- What are the actual quality criteria implicitly used by reviewers in their evaluation, which were written without recourse to an a priori grid of explicit criteria?
- Have reviewers systematically used or, on the contrary, only partially used a set of quality criteria? Are homogeneity or dispersion observable in the practices of the different reviewers?
- Do the criteria resorted to differ with the admissibility categories of articles: admitted ''as is,'' minor modification, major modification, rejected?
- Does observation reveal a certain stability of criteria over time or, on the contrary, a significant evolution over the long period of reviewer reports under study?
- Are the critical points and quality criteria evoked in recent American literature identical to those of reviewers at the European journal? What are the differences?
- Is the method followed in this study replicable to other journals, in the perspective of a cautious generalization?

We thus embarked on an in-depth study of implicit criteria actually used by reviewers in their written reports. We analyzed the content of these evaluations written for articles submitted to *RSDG* since its creation in 1978. This substantial body of raw data, made up of 969 reviewer reports, is, in our opinion, an original means by which to detect the qualitative criteria actually used, beyond normative discourses on what ''authentic'' criteria should be. This article deals exclusively with the reports on articles employing qualitative methodologies.

Following a study of recent American literature on qualitative criteria in qualitative research, the empirical material processed for this article will be presented. A detailed explanation of the content analysis method that we designed and applied will be discussed and compared with methods of two recognized authors in content analysis (Krippendorff & Neuendorf). Finally, we will present and analyze the results obtained, namely, the quality criteria that emerged. We will compare them with the criteria proposed in U.S. literature. We conclude this article with a discussion on the interpretation and analysis of our findings and their limits.

Quality Criteria in Qualitative Research

A survey of recent literature published in the United States on qualitative research indicates a fairly wide consensus among American scholars for major criteria to be used in reviewing qualitative research papers. Gephart, Lee, Golden-Biddle, and Locke, for instance, insist on the necessary link that should be established between criteria actually used, on one hand, and theoretical developments and methodological processes, on the other (Gephart, 2004; Lee, 1998; Golden-Biddle & Locke, 1997). A precise definition of the analytical methods and procedures should be expanded on (Aguinis, 1993; Daft, 1995; Easterby-Smith, Thorpe, & Lowe, 2002; Gephart, 2004; Hammersley, 1992; Lee, 1998).

Qualitative research should be clearly focused and the key concepts used should be properly defined and explained (Daft, 1995; Gephart, 2004; Lee, 1998). It should be possible to assess "how the data were used to generate key conceptual categories" (Suddaby, 2006, p. 640). Daft (1995) significantly underlines the necessity to categorize our reviewing criteria, as does Lee (1998), to attain some sort of common identification principles (Daft, 1995) in well-written scholarly papers (Sutton, 1997).

Data collecting and recording are to be made clear so as to ensure traceability (Gephart, 2004; Lee, 1998; Seale, 1999). Data by itself has no meaning; it must be infused with meaning (Boje, Oswick, & Ford, 2004; Callon & Latour, 1991; Gephart, 1988; Lee, 1998; Weick, 1989). What Daft calls "insufficient rationale design" sums up a recurring deficiency in qualitative papers. "The researchers' theoretical sensitivity" and "the skill with which they combine literature, data and experience" (Suddaby, 2006, p. 640) should be taken into consideration (Glaser & Strauss, 1967).

Our objective, which is to elicit implicit criteria from reviews of qualitative papers, may be in keeping with what Daft (1995) calls "intangible cues" (p. 165). He finds them instrumental in the acceptance or rejection of papers, without giving any further explanation. We believe they probably belong to the implicit criteria we attribute to the specific idiosyncrasy of the reviewer.

Even though American commentator dimensions were not used in our content analysis (see Method section), we present them in Table 1, for the sake of comparison.

Empirical Research Material

As empirical material for our research, we have chosen reviewer reports on articles submitted to a peer-reviewed journal on management sciences. *RSDG*, first published in 1978 and without interruption ever since, is neither dominantly quantitative nor dominantly qualitative as it was meant by its cofounders, the world-renowned economist François Perroux and Henri Savall, to be ecumenical. Focusing our research on one journal alone may appear restrictive. Our findings could therefore be considered as partial, but in our opinion, the data collected on reviewers' implicit criteria, as well as on the reviewers themselves, indisputably form a solid basis on which to work.

Table 1
Comparative Table of Quality Criteria for Qualitative Research According to the Literature

	Gephart (2004)	Lee (1998)	Golden-Biddle & Locke (1997)	Suddaby (2006)	Seale (1999)	Seale (1999, Annex A)	Daft (1995)	Krippendorff (2004)	Neuendorf (2002)	Content Analysis of *RSDG* Reports	Savall & Zardet (1996, 2004)[a]
Theoretical or conceptual basis											
Paper embedded in ongoing research projects	p. 459										
Rigor of the research process								p. 212 / p. 363		Table 6	p. 298
Consistency between theory, method, and results	p. 457	p. 174		p. 640		p. 189	p. 167	p. 513			p. 278
Relevance of the methodology	p. 455					p. 189	p. 171	p. 214		Table 6	p. 218
Literature review											
Explicit connection between bibliography and research theme	p. 460	p. 174				p. 189		p. 313	p. 50	Table 6	p. 291
Critical references cited on the theme		p. 174								Table 6	
Critical references cited on the methodology	p. 460	p. 174	p. 82								
Conceptual development											
Precise use of concepts	p. 460			p. 640			p. 168		p. 50 / p. 103	Table 6	p. 28
The research intends to solve a distinct managerial problem or to generate new theory	p. 460	p. 174	p. 122							Table 6	p. 267
Illuminating new ways in which to handle problems			p. 122								p. 163
Context											
Precise research goals	p. 460	p. 174				p. 189		p. 363			p. 38
Pointing to a lacuna in the literature that the study can address	p. 460		p. 122						p. 107		p. 287

(continued)

Table 1
(Continued)

	Gephart (2004)	Lee (1998)	Golden-Biddle & Locke (1997)	Suddaby (2006)	Seale (1999)	Seale (1999, Annex A)	Daft (1995)	Krippendorff (2004)	Neuendorf (2002)	Content Analysis of *RSDG* Reports	Savall & Zardet (1996, 2004)[a]
Tackling issues that matter to the readers as well as to the people in the field			p. 5						p. 50		p. 134
Author's capacity for scientific commitment										Table 6	
Data											
Clarification of the origin and the kind of data	p. 460	p. 175							p. 102		
Clarification of the record data procedure		p. 175		p. 636		p. 190		p. 216	p. 141		
Data convey compelling human stories			p. 65								
Showing data representing different points of view	p. 460				p. 46	p. 190			p. 56		
Clear accounts of the criteria used for the selection of subject matter for study and of the data collection			p. 122	p. 636		p. 191				Table 6	
Analysis											
Clear accounts of the procedures for analysis	p. 458	p. 175		p. 636		p. 190	p. 169	p. 18 / p. 213	p. 50	Table 6	p. 278
Clarification of the origin of the results (traceability)	p. 460	p. 176				p. 191		p. 212	p. 51	Table 6	p. 283
Linkage between theoretical developments and field data	p. 458	p. 176	p. 70	p. 636		p. 191	p. 172				p. 195

(continued)

Table 1
(Continued)

	Gephart (2004)	Lee (1998)	Golden-Biddle & Locke (1997)	Suddaby (2006)	Seale (1999)	Seale (1999, Annex A)	Daft (1995)	Krippendorff (2004)	Neuendorf (2002)	Content Analysis of *RSDG* Reports	Savall & Zardet (1996, 2004)[a]
Clear distinction between the data and their analysis	p. 458	p. 176		p. 636		p. 190		p. 25			
Depicting the quality of the content analysis				p. 638				p. 213	p. 52		
Verification											
Presenting criteria of development and evaluation of hypothesis	p. 460	p. 175		p. 639	p. 38	p. 191		p. 313	p. 112		
Controls to discount bias		p. 176						p. 21			
Evaluation of the study by the participants		p. 175			p. 9	p. 191					p. 254
Analysis comprehensible and relevant to the participants			p. 122		p. 9	p. 191					p. 259
Discussion											
Evaluating the meaning of the findings	p. 461	p. 176					p. 172	p. 18	p. 167	Table 6	p. 345
Critical discussion and analysis of the results		p. 176	p. 102		p. 43	p. 190		p. 363		Table 6	
Clarification and relevance of recommendations and operational implications		p. 176			p. 9			p. 363		Table 6	p. 361
The researcher examined his or her influence on the research (in particular, the relationship with subject matter)		p. 175		p. 640	p. 44	p. 191		p. 21	p. 50		p. 141
Discussion of limits		p. 176						p. 363		Table 6	

(continued)

Table 1
(Continued)

	Gephart (2004)	Lee (1998)	Golden-Biddle & Locke (1997)	Suddaby (2006)	Seale (1999)	Seale (1999, Annex A)	Daft (1995)	Krippendorff (2004)	Neuendorf (2002)	Content Analysis of *RSDG* Reports	Savall & Zardet (1996, 2004)[a]
Demonstrating an educative awareness of the consequences of particular methodological decisions	p. 460			p. 640	p. 33			p. 25			p. 317
Contribution of generic knowledge		p. 176			p. 44			p. 18			
Allowing the manuscript to ripen naturally							p. 180				
Listening to the reviewers			p. 104				p. 179				
Ethical issues considered		p. 175		p. 640		p. 192					p. 322
Presentation											
Data presented in a way that proves the participation of the researcher in the field	p. 460	p. 175	p. 84					p. 363			
Attractive style			p. 87				p. 170				
Consider the norms for evaluating quantitative research	p. 456		p. 74								
Capacity of the study to emancipate its readers			p. 123		p. 9		p. 181				
No overengineering	p. 460						p. 172				
The paper is as concise as possible, and the author doesn't exaggerate						p. 191	p. 180				p. 344
Concentration on macrostructure	p. 460		p. 26				p. 169			Table 6	p. 199
Clarity and relevance of the conclusion	p. 460	p. 176				p. 192				Table 6	

a. See Discussion section.

The Research Corpus

Qualitative research articles account for 48% of all articles submitted for publication to the journal since its creation. The articles published so far address strategic management (36%), financial theory and management (22%), human resource management (14%), methodology and epistemology (10%), accounting theory and management (7%), marketing (5%), information systems management (5%), and industrial systems (1%).

Given the ecumenical nature of the journal, the editorial advisory board is made up of 56 reviewers so as to represent a wide variety of problematics, methodological options, paradigms, epistemological preferences, and institutional membership. Every field of specialization is represented by 2 to 10 members of the board. The corpus is made up of reports written in French. Articles written in English and Spanish and reviewed by English-speaking and Spanish-speaking reviewers since 2004 were not included in the content analysis we carried out, which eliminates the bias connected to linguistic interpretation.

From its inception, the reviewing policy of the journal editors has been to require at least two double-blind reports for each submitted paper from members of the editorial advisory board designated according to their specialization, with the essential objective of providing genuine pedagogical contributions to scientific dialogue between authors and reviewers. As Golden-Biddle and Locke (1997) aptly put it, "Engagement with the manuscript seems to predispose reviewers to want to become involved with, and to actively participate in, directing the revision of the work" (p. 99). *RSDG* reviewers provide a perfect illustration of this engagement, because they deem it their responsibility not only to assess the article but to make substantial comments, thus enriching the authors' contributions with a variety of standpoints. The editor's constant preoccupation is to avoid any distortion in reviewer reports, attributable to the fact that "qualitative contributions run the risk of being evaluated by reviewers resorting to quantitative methodology" (p. 74).

To make sure the reviewing process is efficiently carried out, a modification-control procedure is initiated when the authors submit their revised article, according to a modus operandi fairly similar to *Organizational Research Methods* proceedings. Each author is requested to return the revised version of his or her article along with a chart displaying a checklist of all remarks, comments, and suggestions by reviewers and a checklist of amendments by the author with their precise location in the article. Conflicts or tensions between reviewers inevitably arise. It behooves the editor to make the final decision and notify the authors. This notification is usually aligned with the most demanding of the reviewers. This explains why the average frequency of publication was 2.5 years over the first 20 years.

The object of this research is twofold: to identify the acceptance or rejection criteria applied to qualitative articles and to examine the specific nature of the implicit methodological and epistemological criteria presiding over the reviewer's final decision (Daft's intangible cues). The considerable time span covered by the journal was a decisive factor in selecting the research material because it enabled the apprehension of any changes in the criteria actually used by reviewers over the years.

Method

Codification of Material

All the reports have the same structure: descriptive information followed by a freely composed text exposing the reviewer's appreciations *without criteria explicitly defined a priori*. Every report includes the following descriptive information, which is codified to permit ulterior analysis:

- the article's proposed title, replaced by an article code;
- the names of the two reviewers, replaced by a reviewer code;
- the management domain in which the article was classified by the reviewers, according to a preestablished list of 13;
- the reviewer's recommendation for the article, according to a four-position scale: admitted as is, minor modification, major modification, or rejected; and
- the final decision of the journal's editor, following examination of recommendations submitted by the two reviewers,[1] on the same scale as above.

Reports do not include confidential appreciations that would be exclusively addressed to the editor.

The content of each reviewer report is made up of sentences composed by the reviewer in response to an open indication: "Detailed comments are indispensable. Please adopt, whenever possible, a formulation that will make it possible to transmit comments to the author (reviewer is anonymous) without requiring 'translation' by the editor of the Journal."

Content Analysis Method of Reports

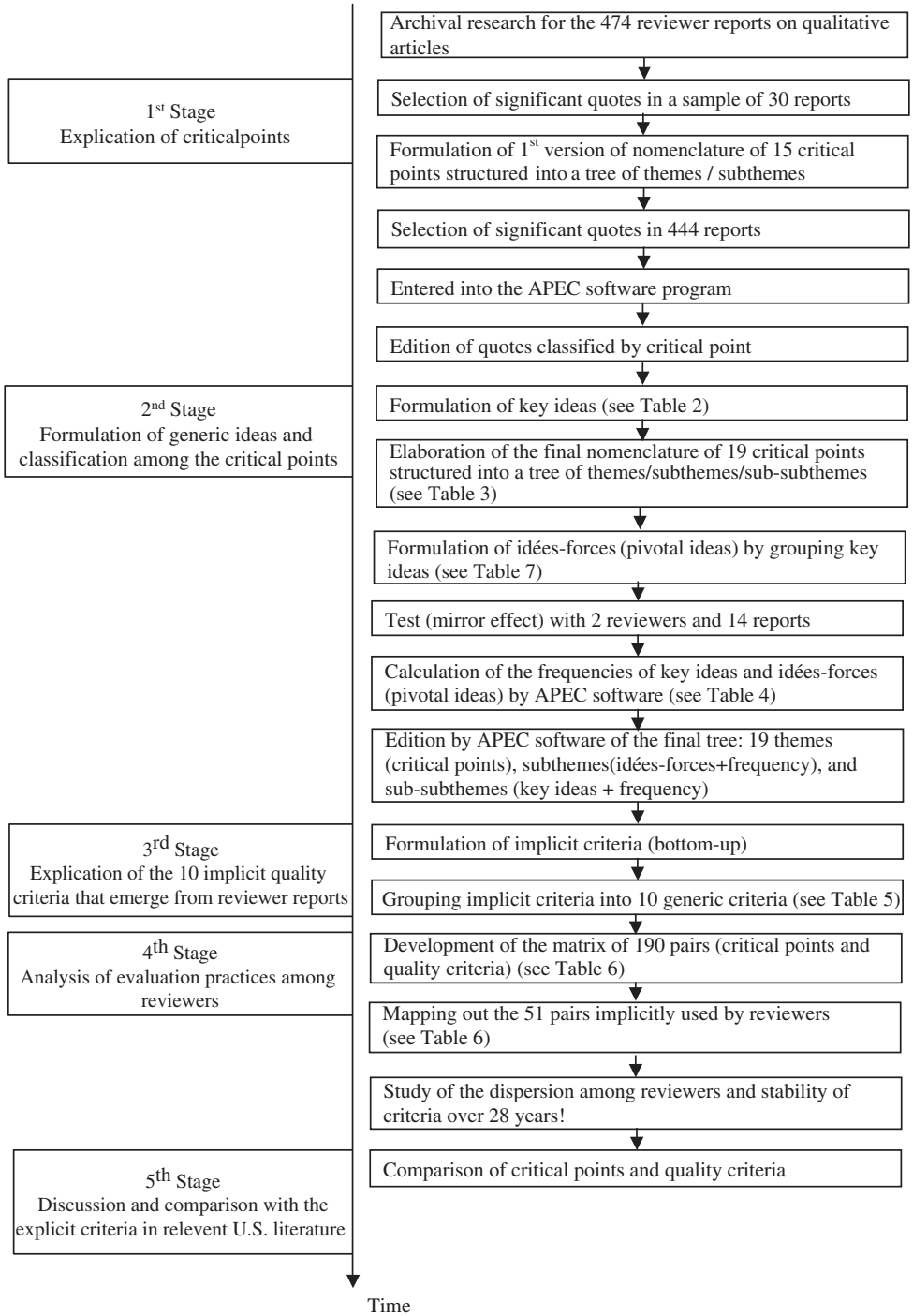
A sequential model, depicted in Figure 1, starting with specific reviewer reports and working toward a generic model, was used.

Content Analysis of Reviewer Reports

Content analysis requires reliable techniques (Krippendorff, 2004). We endeavored to ensure such reliability through the following protocol, the essential stages of which are summarized below.

Reviewer reports were analyzed by a team of researchers directed by the first two authors of this article, responsible for the conception and supervision of content analysis. The two coders have several years of content analysis experience, and both carried out concomitantly the content analysis of the same 30-report subgroup during the exploratory phase. The start of the analysis included selecting the significant quotes maintained in the specific wording of each reviewer that express, more or less explicitly, a positive or negative critical appreciation of the article. Each of the two coders thus entered all the sentences that he or she judged relevant, each sentence being coded with reference to three elements: number of the article, number of the reviewer, and number of sentence order inside the report.

Figure 1
The Sequential Process of the Content Analysis



> Example: 69804, "You should have taken advantage of comparisons between typical case study methods and narrative research methods," a: 698, r: 034.

This sentence concerns article number 698 submitted to the journal; it is an exact quote from the report written by reviewer number 034 and represents the fourth sentence (04) of significant appreciation selected from the body of the reviewer's report.

The first author compared the work of the two coders and selected quotes from the first 30 test reports, based on those proposed by the two coders. In most cases, the number of quotes selected from each report represented the addition of the two coders' proposals. Thus, the number of significant quotes from reviewer reports corresponded with an average five key ideas per report.[2]

A meeting with the researchers made it possible to formulate precise coding instructions for the ongoing work. Suddaby (2006) aptly advocates the necessity of a clear articulation of the "process of data analysis, including coding techniques and category creation" (p. 636). After reading the significant quotes from the first 30 test reports, the researcher supervising content analysis selected 15 keywords representing the critical points extracted through content analysis from the reports on evaluated articles. These 15 keywords made up the first version of the nomenclature of critical points.

Instructions provided to coders thus became more precise than they were during the exploratory phase: The selection of significant quotes should, henceforth, be made in reference to the 15 keywords of the previously identified critical points, or eventually by identifying significant ideas not using these keywords but that contribute to completing the first nomenclature of 15 critical points, structured into a tree of themes and subthemes.

The two computer specialists in the research team, who have for many years maintained an expert-system software application for performing dysfunction diagnostics in organizations (Buono & Savall, 2007; Savall, Zardet, & Harbi, 2004), completed the first application of the analysis program "Academic Publication Evaluation Criteria" (APEC), a provisional tree that corresponded with the initial 15 critical points. Table 3 presents the final version, which includes 19 critical points, after analysis of the entire corpus of 474 reports.

The 444 remaining reports were then divided among the coders who selected the significant quotes from the appreciations and entered them into the APEC program. This enabled editing all quotes of quality expressed by reviewers in their spontaneous, specific wording and classifying them into themes and subthemes according to the nomenclature of critical points. Through verification of the classification of significant quotes within the critical-point nomenclature, some final corrections were made.

Quotation excerpts from reports, corresponding to 235 article projects employing a qualitative methodology, were thus coded, entered, and classified according to the provisional tree of themes and subthemes. The entire corpus of material was then classified according to the critical points, without interpreting, at this stage, the meaning of quotations.

The following stage included formulating (generic) key ideas of appreciations of the quality of the submitted articles. The coders were asked to propose a first expression of these (generic) key ideas, based on a study of the meaning of each significant quote. This entailed, on the basis of highly contextualized knowledge, the formulation of generic ideas that could be reused in other contexts, such as other articles, other reviewers, perhaps even

Table 2
Example of the Creation of a (Generic) Key Idea

Theme	Methodology (see Table 4)
Subtheme	Methodological choices (see Table 3)
(Generic) key idea	Unclear explication of methodology and methodological choices (see Table 6, cell C6)
Quote 1, a: 106, r: 005	"The practical details of the research carried out are not precisely articulated."
Quote 2, a: 116, r: 006	"The problematics is not clarified, i.e., the objective pursued and the method employed are not explicitly stated."
Quote 3, a: 149, r: 006	"No information regarding the research method is provided."
Quote 4, a: 427, r: 009	"Specify the nature of your methodological choices."
Quote 5, a: 550, r: 031	"Very little information is provided on the methodological aspects of the research."
Quote 6, a: 552, r: 029	"The methodology and findings of the prior research upon which this article is based are not properly accounted for."
Quote 7, a: 698, r: 034	"You should have taken advantage of comparisons between typical case study methods and narrative research methods."

other journals. The key ideas were created and entered into the APEC program, inserted between the critical-point tree and the significant quotes initially expressed by each reviewer. For example, seven significant quotes from seven reports on different articles (106, 116, 149, 427, 550, 552, and 698), written by six different reviewers (005, 006, 006, 009, 031, 029, and 034), were classified under the same (generic) key idea, "unclear explication of methodology and methodological choices" (see Table 2).

Three types of significant quotes appear in appreciations of one critical point and one quality criterion:

- A positive appreciation, for example, "The article deals with an important topic."
- A negative appreciation, for example, "The thesis presented is not original."
- A recommendation, which we usually equate with an implicitly negative appreciation, for example, "Define the problematics more precisely."

For a given theme, subtheme, and quality criterion, positive key ideas (e.g., relevant problematics) and negative key ideas (e.g., problematics devoid of interest) were formulated.

At the end of this content interpretation and analysis operation on significant quotes, the "translation" of all significant quotes into key ideas was verified and corrections were proposed to coders. This eliminated redundancy and resulted in more precise and better-adapted keywords that would facilitate reading by coders and analysts. This work of controlling quality, coherency, and homogeneousness of analysis was carried out four times over the course of the period of reviewer-report content analysis, which lasted 4 months.

A test was carried out with the participation of two reviewers who represented 14 different article reviews produced over the 28-year period. The test compared the classification of significant quotes extracted from the reviewers' reports by coders, using the software program (critical points and generic key ideas), with what the reviewers had originally written in their reviews.

Idées-forces (pivotal ideas) were then formulated, based on all key ideas, each idée-force being devised to synthesize all key ideas of a given family (see Table 7).

Finally, a qualimetric approach, which is a hybrid between qualitative and quantitative approaches (Savall & Zardet, 2004), enabled calculating the frequency of key ideas and idées-forces (pivotal ideas). It is possible to measure the frequency of expressions at each level of the tree: themes, subthemes, sub-subthemes; calculating the frequency of idées-forces; and key ideas made it possible to identify the most frequent appreciations.

Fundamental Principles of Content Analysis and Their Application in Evaluating the Quality of our Analysis

To evaluate the rigor of our method and benchmark our respective criteria, we used the norms recommended by Krippendorff (2004) and Neuendorf (2002). These recognized authors in content analysis emphasize the fact that the content analysis process includes the contribution that content analysts make in selecting what counts as content (Krippendorff, 2004; Neuendorf, 2002); thus, there is no right way to define a construct, and every researcher is "the boss" (Neuendorf, 2002, p. 50).

The robustness of content analysis depends on two major quality criteria: the reliability of the techniques used and the validity of the results. With reference to the content analysis we practice, we think we have established protocols that ensure the reliability of the analysis (Krippendorff, 2004). Stability over time is monitored by the computer system that memorizes the nomenclature of themes, subthemes, and key ideas, permitting future researchers to use the same nomenclature, thus reducing the subjectivity of each analyst. Careful attention was devoted to the reproducibility of the process by means of a precise protocol for coding, classifying, and formulating key ideas. The protocol was defined, written out step by step, and transmitted to different researchers, each one with several years of experience in the domain. They worked as a team under the responsibility of one of the authors of this article who supervised the entire content analysis phase, then the interpretation of the findings. As for the database, it was established and has been operated since 1978 by the person in charge of the journal's copy desk, thus totally independent from us.

The validity of our findings perhaps calls for more precise comments (Krippendorff, 2004). Facial validity and social validity seem satisfactory to us, from our viewpoint, in light of the results obtained and the evolution of evaluation practices they have incited. With regard to empirical validity, it would probably prove worthwhile to pursue our investigations by collecting new data from reviewers, after presenting them, in "mirror effect," with the results derived from their own reports. These research tracks are developed below.

Types of Results Derived From Content Analysis

The software program APEC enabled us to draw up a nomenclature of critical points and quality criteria actually used by reviewers and a detailed list of all (generic) key ideas and associated quotes. It contributed to highlighting convergent key ideas common to several reports and key ideas specific to a given report, to test hypotheses of criteria convergence or dispersion. Various sorting methods (by article, review, or management domain) made it possible to test dispersion among reviewers for a given article, to check the stability of evolution of criteria over a long period, and to check discriminating criteria in a given domain. But this goes beyond the scope of our objective, which is to elucidate the implicit quality criteria in use relative to qualitative management research.

Findings

The Implicit Criteria That Emerge

Content analysis brought to light two essential dimensions: the critical points (which designate the content areas of an article) and the actual quality criteria. Study of the meaning of sentences contained in the reports enabled us to pinpoint the critical points on which reviewers spontaneously express appreciations, positive or negative. Then, by studying the nature of reviewers' appreciations, we progressively identified the real quality criteria. For example, with regard to the problematics of an article (critical point), appreciations emerged relevant to its originality, its relevance, its explications, its justification, and its delimitation. These five quality criteria can, in turn, be broken down by the positive or negative appreciations, expressed in the form of (generic) key ideas.

Theoretically, one could consider, through logico-deductive reasoning, that each critical point can be evaluated in terms of each quality criterion identified, according to a double-entry table where all cells could be filled by a reviewer. In reality, the criteria actually used by reviewers are fewer than the entire group of all possible criteria. This is due to the fact that reviewers, not having a normative evaluation grid with a list of criteria and scale of appreciations, spontaneously stress the criteria that appear essential to them. We present below the critical points identified and then the quality criteria actually used by reviewers.

The Critical Points for Evaluation of Qualitative Articles

Nineteen critical points emerged from the analysis of the entire corpus of reports. They are presented in Table 3.

Of course, in a given report, the reviewer expresses himself or herself referring to only part of these 19 critical points and never makes a systematic evaluation of all points. The critical points about which reviewers most often express themselves are 7 in number out of the possible 19 points: problematics and research issue (A), methodological choices (C), state of the art and bibliographical analysis (H), theoretical and conceptual analysis, mobilized concepts and theories (I), positioning with respect to the research question in academia (M), generic knowledge elicited (N), and field materials (J).

The expression of reviewers, all articles and all periods combined, is distributed in roughly balanced fashion among the different families of critical points (see Table 4). Note that the critical points were then structured into seven families of critical points through a bottom-up content analysis process: problematics and hypotheses (A+B), methodology (C+D+E+F), bibliographical references and state of the art (G+H), theoretical and conceptual analysis (I), field work and method of exploitation (J+K), findings (L+M+N+O+P), and rationale and conclusions (Q+R+S).

Identification and Definition of Quality Criteria for Qualitative Research

In studying the nature of reviewers' critiques for each report, we first classified their critiques into one of the 19 critical points and then identified a (generic) quality criterion. Through incremental procedure, and after much trial and error, 10 actual quality criteria

Table 3
The 19 Critical Points for Evaluation of Articles Submitted to *Sciences de Gestion—Management Sciences—Ciencias de Gestión* (*RSDG*)

A	problematics and research issue
B	research hypothesis(es)
C	methodological choices
D	epistemological choices
E	underlying ideologies
F	methodological process
G	bibliographical materials and references
H	state of the art and bibliographical analysis
I	theoretical and conceptual analysis, mobilized concepts and theories
J	research field materials
K	field investigation method, exploitation of materials and data
L	interpretation
M	positioning with respect to the research question in academia
N	generic knowledge elicited
O	recommendations and operational implications
P	discussion, critical analysis of findings, and exposition of limits
Q	communication capacity (of the article)
R	guiding thread, line of reasoning, outline
S	conclusion and loop back to title

Table 4
Distribution of Reviewer Comments Among the Seven Families of Critical Points

Total	100.0%
Problematics and hypotheses	16.6%
Methodology	9.3%
Bibliographical references and state of the art	17.0%
Theoretical and conceptual analysis	6.4%
Fieldwork and method of exploitation	17.2%
Findings	20.4%
Rationale and conclusion	13.1%

emerged, either through identification of a keyword employed by the reviewer that we had established as a criterion or through proximity of meaning. We present below the semantic proximity and classification we implemented.

Example: The significant quote from report number 211, ‘‘Subject of current interest, original approach,’’ was processed as follows:

- Classified in critical point A *problematics and research issue*,
- Linked to two quality criteria: relevance (‘‘current interest’’) and originality (‘‘original approach’’).

Table 5
The 10 (Generic) Quality Criteria

Rigor	This scientific quality criterion applies to the degree of semantic precision, the meticulous conduct of the methodological process, and the in-depth treatment of the theoretical concepts mobilized. Through proximity of meaning, we associated superfluous theoretical detours, digressions, and useless rhetoric.
Formulation	Rigor as applied to the form, to the writing, and to the choice of words with incidence on the clarity and readability of the publication.
Coherency	This criterion refers to the articulation between several critical points in the article. Thus, explication of the connections between the research issue, the methodology chosen, and the results obtained makes up a criterion of coherency. In the literature analysis, superficial recourse to concepts poorly integrated into the demonstration is another criterion of coherency. Through semantic proximity, we associated the critiques expressed in terms of dispersion or heteroclicity, as well as tautological reasoning.
Originality	This criterion refers to innovation on the part of the researcher in terms of problematics, literature analysis, and fields chosen; in contrast to practices of replication and standardized research.
Relevance	Reviewers evaluate here whether the chosen problematics are relevant, either in terms of their current interest or in terms of their usefulness for research or business practices, or whether the scope of results is promising. Relevance also applies to the methodological options, the problematics and concepts chosen, and the methods of investigation chosen.
Explication	Numerous criticisms address insufficient explication on the part of the author, lack of traceability in his or her research, and whether this pertains to the problematics, the hypotheses, the methodological options, the concepts, or the operational or managerial recommendations.
Positioning	Reviewers expect an article to adopt a position with regard to the relevant environment, the economic and social context, the existing literature, or empirical and fieldwork data. We classified here through proximity the criteria of anchorage and contextualization of the research.
Contribution	Is the publication, with regard to one or two critical points, a contribution and value-added for the academic community? Or, is the article's contribution, in terms of contributed knowledge or synthetic value-added of the literature, significant?
Rationale	Reviewers are interested here in the justification of the findings presented by the outline and structure of the article, the demonstrative reasoning used, and the argumentation chosen. The conclusion is essentially a loop back to the title; it also pertains to this criterion.
Delimitation	Reviewers are concerned here with evaluating whether the problematics, the object, and the theoretical framework are sufficiently circumscribed, delimited, and focused; whether the researcher's openness is sufficient or excessive; and whether the criteria of completeness, density, and concision of the article are satisfactory.

The 10 quality criteria that emerged from this iterative, bottom-up, content analysis process are of equivalent value, in terms of the overall objective of quality in qualitative research. These criteria are therefore given in random order in Table 5.

Quality Criteria Actually Used by Reviewers

The 19 critical points and 10 quality criteria potentially represent 190 pairs of "quality criterion per critical point," through systemic crossing on a double-entry table of 19 lines

of critical points and 10 columns of quality criteria. However, in practice, the study of the corpus of reports shows that only a subset of 51 critical points of quality criteria were selectively used, consciously or not, by the reviewers of the articles submitted to the journal, in the form of positive or negative remarks. Table 6 shows the synthetic results of quality criteria employed, in reality, by reviewers.

At the intersection of a line and a column, one or several appreciations appear, usually in the form of pivotal ideas (see section "Content Analysis of Reviewer Reports"), mostly negative or subject to qualification. Table 7 was elaborated through bottom-up inductive analysis leading to the identification of several key ideas whose combination brings to light an idée-force. The latter features at the intersection of one critical point and one quality criterion.

The quality criteria most frequently used are, in the following order (see Table 6), relevance (used in 393 reports out of 474), explication (314), contribution (287), positioning (108), delimitation (107), rigor (104), originality (76), formulation (66), rationale (61), and coherency (43). The first three criteria (relevance, explication, contribution) are used in a high proportion of reports: 393, 314, and 287, respectively, out of a total of 474. Conversely, certain critical point/quality criterion pairs are only used in two or three reports (see Table 6).

These results show that reviewers endeavor to evaluate, on one hand, the internal validity of the research (explication—delimitation—rigor—formulation—rationale—coherency) and, on the other hand, its external validity with regard to the academic community and, perhaps as well, toward management practitioners (relevance—contribution—positioning—originality). Concerning the criterion of explication for validity, for example, reviewers emit critiques on the lack of explication touching 12 critical points (cf. Table 6): problematics, unformulated implicit hypotheses, insufficient explication of the methodological choices, the epistemological choices, the author's underlying ideologies, the concepts mobilized, relationships between the theories and the authors on the studied theme, the research terrains used, the concrete method applied to fieldwork material, the author's contribution with respect to existing work, the managerial implications of the research, and the limits of the research and of the reasoning.

We analyzed the four categories of reports—as is (W), minor modification (X), major modification (Y), and rejected (Z), hereafter designated as W, X, Y, and Z—to compare the frequency of each pair (critical point/quality criterion) and to highlight significant differences. The examination of frequencies for each category showed some results that were foreseeable. Thus, half of the appreciations (25 cells out of 51) were consistent and mainly concerned relevance, explication, formulation, contribution, and rigor, because their frequency increases from category W through category Z when the appreciation is negative, and decreases in the opposite case.

Examination also showed surprising results. The same frequency of appreciations for all categories (W, X, Y, and Z) revealed a lack of sensitivity on the part of reviewers to those criteria when frequency was very low (one report per category). When frequency was higher and identical for all four categories, the results revealed neutrality with regard to the decision proposed by the reviewer. Thus, those criteria had no incidence on the decision proposed by the reviewer. This last case concerns the criteria *unclear explication*

Table 6
Pairs of Critical Points and Quality Criteria Actually Used by *Sciences de Gestion—Management Sciences—Ciencias de Gestión* (*RSDG*) Reviewers as Revealed by our Content Analysis

Quality criteria 1559	1—Rigor (Semantic Precision) 104	2—Formulation (Clarity, Readability) 66	3—Coherency (Integration, Homogeneousness/ Dispersion, Explication of Relationships, Tautological Reasoning) 43	4—Originality (Innovation) 76	5—Relevance (Comprehension, Usefulness, Scope, Current Implications) 393	6—Explication (Traceability) 314	7—Positioning (Anchorage, Contextualization Sources) 108	8—Contribution (Value-Added, Contribution, Knowledge Contributed) 287	9—Rationale (Demonstration, Argumentation, Justification) 61	10—Delimitation (Selectiveness/ Openness, Completeness, Density, Concision) 107
Critical (key) points 1559										
A—Problematics and research question 230		Unformulated or poorly formulated problematics 30	Coherent analytical field and situation 5	Interesting problematics 70	Relevant problematics 34 Irrelevant problematics 41	Insufficiently explained problematics 2	Origin of the problematics and positioning insufficiently justified 28		Insufficiently justified problematics 2	Problematics need to be refocused; object poorly delineated, outlined, and defined 18
B—Research hypothesis(es) 27		Misrepresented hypothesis 25				Absence of formulated hypothesis 2				
C—Methodological choices 103					Relevant investigation methods 2	Unclear explication of methodology and methodological choices 36 Significant explication of methodology 2		Lack of methodological contribution 18 Good contribution to the methodology 45		
D—Epistemological choices 5					Relevant epistemological options 2	Unconvincing explication of epistemological choices 3				
E—Underlying ideologies 2						Insufficiently articulated ideological explication 2				

(continued)

Table 6
(continued)

Quality criteria 1559	1—Rigor (Semantic Precision) 104	2—Formulation (Clarity, Readability) 66	3—Coherency (Integration, Homogeneousness/ Dispersion, Explication of Relationships, Tautological Reasoning) 43	4—Originality (Innovation) 76	5—Relevance (Comprehension, Usefulness, Scope, Current Implications) 393	6—Explication (Traceability) 314	7—Positioning (Anchorage, Contextualization Sources) 108	8—Contribution (Value-Added, Contribution, Knowledge Contributed) 287	9—Rationale (Demonstration, Argumentation, Justification) 61	10—Delimitation (Selectiveness/ Openness, Completeness, Density, Concision) 107
F—Methodological process 34	Rigorous conduct of methodological process 22 Weak methodological process 10								Insufficiently justified methodology 2	
G—Bibliographical materials and references 114					Defective state of the art 70 Missing bibliographical references 12 Hypertrophic bibliographical references 7					Insufficient open-mindedness; incomplete literature survey; total ignorance of an important school of thought 25
H—State of the art and bibliographical analysis 150		Hypertrophic state of the art 9		Original literature analysis 2			Relevant positioning with respect to the literature 3 Insufficient positioning 74			Interesting and complete state of the art 62
I—Theoretical and conceptual analysis, mobilized concepts and theories 100	Rigorous, relevant treatment (precision, density of concepts used) in the theoretical reflection 28		Superficial mobilization of poorly integrated concepts in the demonstration or argumentation 2		Relevant choice and treatment of concepts 2	Lack of explication of relationships between the studied topic, theories, and authors inadequate explication of concepts 3		Significant value-added of theoretical and conceptual approach; poor value-added of theoretical and conceptual approach 60	Excessive description; insufficient analysis with regard to the hypotheses 3	Superfluous theoretical detours, digression, esotericism, useless rhetoric 2

(continued)

Table 6
(continued)

Quality criteria 1559	1—Rigor (Semantic Precision) 104	2—Formulation (Clarity, Readability) 66	3—Coherency (Integration, Homogeneousness/ Dispersion, Explication of Relationships, Tautological Reasoning) 43	4—Originality (Innovation) 76	5—Relevance (Comprehension, Usefulness, Scope, Current Implications) 393	6—Explication (Traceability) 314	7—Positioning (Anchorage, Contextualization Sources) 108	8—Contribution (Value-Added, Contribution, Knowledge Contributed) 287	9—Rationale (Demonstration, Argumentation, Justification) 61	10—Delimitation (Selectiveness/ Openness, Completeness, Density, Concision) 107
J—Field materials 74			Absence or lack of fieldwork, erroneous choices of terrains 33	Untapped terrains or materials 4	Relevant choice of terrains 16 Defective choice 16	Clear and precise presentation of the terrains 2	Absence of anchorage in empirical data or fieldwork 3			
K—Field investigation method, exploitation of materials and data 194					Relevant methodology 18 Irrelevant methodology 56	Well-specified concrete methodology 8 Insufficiently articulated concrete methodology 23		Low value-added and implications of results 37 High synthetic value-added 52		
L—Interpretation 44	Misinterpretation of concepts and theories 44									
M—Positioning with respect to the research question in academia 40					Relevant, convincing, and committed demonstration 14 Irrelevant demonstration 26					
N—Generic knowledge elicited 131						Unclear author's contribution 47 Significant author's contribution 9		Valuable contribution of generic knowledge (concepts, theories, models) 75		

(continued)

Table 6 (continued)

Quality criteria 1559	1—Rigor (Semantic Precision) 104	2—Formulation (Clarity, Readability) 66	3—Coherency (Integration, Homogeneousness/ Dispersion, Explication of Relationships, Tautological Reasoning) 43	4—Originality (Innovation) 76	5—Relevance (Comprehension, Usefulness, Scope, Current Implications) 393	6—Explication (Traceability) 314	7—Positioning (Anchorage, Contextualization Sources) 108	8—Contribution (Value-Added, Contribution, Knowledge Contributed) 287	9—Rationale (Demonstration, Argumentation, Justification) 61	10—Delimitation (Selectiveness/ Openness, Completeness, Density, Concision) 107
O—Recommendations and operational implications 34						Absence of managerial recommendations and implications 27 Positive managerial recommendations 7				
P—Discussion, critical analysis of findings and exposition of limits 72					Unsatisfactory discussion and analysis of results 32 Relevant discussion 4	Biases, limits, difficulties not properly dealt with 32 Proper exposition of limits 4				
Q—Communication capacity (of the article) 5		Sloppy style; spelling, typing errors; inadequate presentation format 2	Insufficient guidelines; weak structure 3							
R—Guiding thread, line of reasoning, outline 159						Clear and understandable reasoning 47 Confused reasoning 58			Unconvincing argumentation 54	
S—Conclusion and loop back to title 41					Clear and relevant conclusion 14 Unclear and irrelevant conclusion 27					

Note: Each cell indicates the number of reports where the pair (critical point/quality criterion) was expressed. The sum of 1559 corresponds with the number of key ideas mobilized in the 474 reports (i.e., an average of 3 idées-forces [pivotal ideas] per report).

Table 7
Example of Idée-Force (Pivotal Idea) Formulation Through Grouping Together Key Ideas

Five key ideas emerged from the analysis of significant quotes:

1. Explicitly state the source of the problematics and their mode of emergence.
2. Justify that the studied problematics are situated within the field of management sciences.
3. Expose the upstream problematics.
4. Situate with regard to important recent developments in the discipline and to empirical work by other authors (or point out their verified absence).
5. The stakes are unclear with regard to the nature of the article (Research? Pedagogy? Report on an intervention in an enterprise?).

They are then encapsulated within an idée-force, ''origin of the problematics and positioning insufficiently justified,'' mentioned in cell A7 of Table 6.

of methodology, *rigorous conduct of methodological process*, and *hypertrophic state of the art*.

Other results are even more surprising or atypical, such as more frequent negative appreciations for ''as is'' articles than for ''rejected'' ones or, on the contrary, more frequent positive appreciations for ''rejected'' and ''major modification'' articles than for ''as is'' or ''minor modification'' articles (e.g., low value-added and implication of results; absence or lack of fieldwork; missing bibliographical references; origin of the problematics insufficiently justified; lack of methodological contribution; incomplete literature survey; insufficient positioning with respect to literature; rigorous, relevant treatment in the theoretical reflection; interesting problematics; relevant choice of terrains; significant value-added of theoretical and conceptual approach; high synthetic value-added).

Convergence and Specificity of Quality Criteria Made Evident by the Literature

In light of the state of the art presented at the beginning of this article and of the results obtained through content analysis of the reviewer reports, it is interesting to ask whether our study sheds new light and converges with the authors chosen to represent the literature on qualitative research methodologies. For the analysis that follows, we are obliged to pose two hypotheses: on one hand, that all critical points and all quality criteria are equally important and, on the other hand, that most readers are familiar with the literature we have chosen to study. Thus, the sum of all criteria is presumed to be an implicit model in the mind of the reviewer.

Results Converging With the Literature

We reexamined the quality criteria identified in the literature in light of our findings. In particular, we measured the number of critical points per quality criteria mobilized in the literature to determine if they were superior, inferior, or equivalent to our own findings. We furthermore determined the proportion of converging and diverging results in the

literature with our own findings. Then, we attempted to appreciate each criterion's degree of "density" (i.e., the proportion of each criterion in the entire group of all critical points). This enabled us to identify the critical points and quality criteria most often mobilized by *RSDG* reviewers or, on the contrary, by the authors studied in the literature. Convergences apply to approximately one third of the results (18 common points out of 51).

This reveals that the common criteria most often mobilized on the same critical points are, in the following order:

- the explication (6 common points): of key concepts (Gephart, 2004), of research goals (Gephart, 2004; Lee, 1998), of the outcomes and their implications (Gephart, 2004; Lee, 1998), of their recommendations and implications related to issues in management (Golden-Biddle & Locke, 1997), of the researcher's operational method in the field (Gephart, 2004; Golden-Biddle & Locke, 1997; Lee, 1998), and of empirical and fieldwork material (Locke & Golden-Biddle, 2002; Seale, 1999);
- the relevance (2 common points) applied to the problematics (Golden-Biddle & Locke, 1997) recommending tackling research issues and problematics of interest to both universities and practitioners, to methodological choices (Daft, 1995; Gephart, 2004; Seale, 1999), to the bibliographical choices in reference to the studied subject (Daft, 1995; Gephart, 2004; Seale, 1999), and to the findings and conclusions presented (Seale, 1999);
- the rationale (2 common points) concerning the storyline and demonstration logic and the line of reasoning that should be clear for the reader (Daft, 1995; Gephart, 2004; Golden-Biddle & Locke, 1997);
- finally, five criteria with one single common point among the studied authors: the delimitation and completeness of the bibliographical references (Lee, 1998), the contributions and value-added of the theoretical and conceptual analysis (Gephart, 2004; Locke & Golden-Biddle, 2002), the originality of the problematics, which could entail solving a distinct managerial problem identified or bringing insights to new ways of approaching problems or generating new theory (Gephart, 2004; Golden-Biddle & Locke, 1997; Lee, 1998), the coherency of the researcher's positioning and scientific commitment (Seale, 1999), and the rigor of the theoretical and conceptual analysis, in particular the definition and explication of key concepts (Gephart, 2004; Rynes, 2004).

Quality Criteria Proposed by the Literature but not Identified in our Report Analyses

Nine criteria proposed by the literature are never found in our content analysis of reports.

1. Inscribe one's research in a long-term program (Daft, 1995).
2. Show that one is conscious of the consequences of one's methodological choices (Gephart, 2004; Seale, 1999).
3. Allow one's manuscript to ripen naturally (Daft, 1995).
4. Listen to one's reviewers (Daft, 1995; Golden-Biddle & Locke, 1997).
5. Take into account the ethical questions (Lee, 1998; Seale, 1999).
6. Introduce data that attest to the researcher's presence in the field (Gephart, 2004; Golden-Biddle & Locke, 1997; Lee, 1998).
7. Do not allow the description of the methodology to become overly important (Gephart, 2004; Lee, 1998).

8. Develop an analysis that is comprehensible and significant for the participants of the study (Golden-Biddle & Locke, 1997; Seale, 1999).
9. Obtain validation of the study from the participants (Lee, 1998; Seale, 1999).

The authors of this article, as teacher-researchers and, as such, directors of research and participants in the reviewing process of articles submitted to European and American journals other than *RSDG*, consider that these criteria, with the exception of point 7, are entirely relevant as shown in Table 1. They use them on a regular basis in their evaluation work (Savall & Zardet, 2004). Concerning point 7, our experience has shown that the most frequently encountered problem is insufficient, and not excessive, methodological explication, taking into account that we are equally demanding in terms of the relevance of the findings.

Thus, we recognize the influence that may have exerted the explicit or implicit model harbored by the researchers who constructed the model of content analysis used in this article. Certain critical points (such as *underlying ideologies* and *guiding thread, line of reasoning, outline*) were included in the content analysis nomenclature even though they were rarely used by reviewers. Nevertheless, this content analysis method enables us to avoid a great deal of bias in report analysis, by adopting a bottom-up, iterative process. If it is true that certain outcomes of the corpus analysis surprised us, with regard to certain points for which we anticipated finding more critical appreciations, then again comparison with the studied literature brought to light that we share numerous criteria present in the literature, for we use them in our practice of article review for European and American journals. The quality criteria most often mobilized in the literature are relevance (7 critical points out of 19), explication (6/19), rigor (6/19), delimitation (5/19), and formulation (5/19).

In addition to these common points, the criteria that were more stressed by the literature than those revealed by our study are rigor (6 additional critical points), formulation, delimitation (4 additional critical points), and contribution and implications of the research (3 additional critical points). The greater importance lent to those criteria is perhaps attributable to the more normative character of the articles and monographs we chose to study in the literature. Indeed, those publications are more useful as recommendations for authors in qualitative research and their goal does not seem to be the study of actual evaluation practices of reviewers.

Criteria Used by Reviewers and Rarely Highlighted by the Literature

Our analyses detected six criteria used by the reviewers that were not highlighted in the literature. They are ideologies of the article's author, methodological process followed, quality of bibliographical analysis, interpretation made by author, author adopting a position concerning the state of the issue, and generic knowledge elicited. However, with regard to the critical point N generic knowledge elicited (see Table 6), which is a contribution and value-added, one could question whether they are assimilated to the notion of "reproducible results" proposed by the studied literature. The quality criterion explication is applied to six more critical points in our study of reviewer reports than in the literature; the criterion contributions to five additional criteria; the criteria relevance, coherency, positioning, and delimitation to three additional critical points. Reviewers of *RSDG* attributed more importance to those criteria.

Discussion

Critical Analysis of the Findings

A significant result is the dispersion of reviewers concerning the criteria and the critical points mobilized. In the entire corpus of 474 reports, emanating from 56 different reviewers over a long period, very few points of convergence are to be found among the reviewers, and even fewer points of strong convergence for the same reviewer over time.

Several explanations are possible. On one hand, reviewers had no stabilized evaluation grid in their minds, explicit or implicit, when they evaluated an article and thus followed their reflex of focusing attention and expression on the critical points and quality criteria that were particularly striking, either for their high quality or, on the contrary, their weaknesses. First, reviewers, over the course of their cognitive progression, endeavored to selectively identify, on one hand, the article's strong points and, on the other hand, its principal weak points. Our critical analysis also shows that rather than using all the critical points and all the quality criteria, reviewers focus on certain critical points and certain quality criteria when highlighting the positive or negative points of the articles they reviewed. Table 6 displays the 51 cells most often used by reviewers out of the 190 possibilities. Moreover, if one examines the 51 cells in closer detail, in terms of frequency, one cannot help but notice that 22 are sparsely used (in only 2 or 3 reports), which brings to question the degree of interest of reviewers for those criteria. Furthermore, instability of criteria used over time by the same reviewer can be interpreted either as instability resulting from an implicit evaluation grid or as the evolution of the criteria considered important in the eyes of the reviewer, with regard to his or her personal evolution and professional experience, a learning-curve effect. Thus, among reviewers with the most experience, 28 years of reviewing articles for the studied journal (*RSDG*), the following evolution appeared:

- more internal validity criteria: coherency, explication, literature review, and so on at the beginning of their career (1978–1987);
- more external validity criteria in the third decade, 1998–2005: relevance, value-added, positioning, and so on.

This evolution could correspond with an evolution in the conception of management research on the part of professors in the second part of their careers, and/or to the maturity of the discipline of management in the history of social sciences.

To better understand the implications, evaluate the robustness, and appreciate the phenomenon of quality criteria emerging over time, we conducted a comparative study of reviewer reports over the course of three consecutive periods, deriving the idées-forces for the periods 1978–1987, 1988–1997, and 1998–2005. The most striking outcome is that 44 idées-forces out of the total 51 were common to the combined three periods, that is, they appeared in the first period 1978–1987 and were maintained over the course of the two following periods. In contrast, three specific idées-forces disappeared in the following two periods. They concerned two positive idées-forces—well-adapted method and relevant analysis of results—and one negative idée-force—lack of semantic precision. However,

they represented less than 5% of reviewer expression over the period 1978–1987. Similarly, in the second period, 1988–1997, two new positive idées-forces appeared and disappeared in the third period: rigorous conduct of the methodology and relevant bibliographical references. They appeared in less than 5% of the reports. In the third period, no new idées-forces appeared.

One can conclude that certain hardcore criteria exist, which we referred to above as "critical points/quality criteria pairs," used by the entire group of reviewers over a long period of time and despite the growing number of reviewers: 20 reviewers from 1978 to 1987, plus 17 new reviewers from 1988 to 1997, plus 19 new reviewers from 1998 to 2005. This illustrates the principle of generic contingency (Savall, 2003; Savall & Zardet, 1996), where contingency factors (numbers of reviewers, a given historical period, the evolution of the number of reviewers per domain, etc.) did not prevent the creation of quality criteria generic hardcore for qualitative research.

With regard to the proportion of key ideas per critical points during the three periods, one also remarks that the critical points maintained their importance, with only minor fluctuations. The volume of key ideas and number of criteria used by the reviewers remained stable over the course of the three periods concerning the critical points: problematics, bibliographical references and state of the art, theoretical analysis, field material, and methods of exploiting field. It increased concerning the critical points research methodology and findings. The critical point R *guiding thread, line of reasoning, outline* was the only one whose frequency decreased. It went from 15% to 9%, whereas critical point P findings and limits went from 13% to 16%, and critical point C methodological choices went from 4% in the first period to 10% in the second and third periods. According to us (see first page of Discussion section), the marked increase in references to critical point C testifies to the greater and greater attention paid by *RSDG* reviewers to the proper articulation of methodology as a quality criterion of articles.

Our study revealed another interesting aspect of reviewer practices. Fewer evaluation criteria were employed to recommend articles for admission "as is" and for "rejection." Reviewers mobilized substantially more criteria to justify decisions for "minor modification" and "major modification."

Limits of the Findings

This study is based on a large volume of homogeneous material, for the format of reports and the instructions given to reviewers have remained stable over the entire period studied, from 1978 to 2005. Full-text archival preservation and inscription in a chronological logbook of all articles received, all designated reviewer reports, their recommendation, and the editor's final decision have provided optimal methodological conditions for ensuring the homogeneousness of the data being processed.

The first major limit of this research probably stems from the fact that work was focused on one single journal, with specific objectives with respect to other European and American journals. It is a multidisciplinary journal in management sciences, welcoming articles from all other subdisciplines in management (strategy, finance, human resources, accounting, marketing, etc.), in addition to articles specialized in epistemology and on qualitative and quantitative methodology. Although the number of reviewers had regularly

increased over the long period, attaining today 56 reviewers for articles employing qualitative methodologies, all of whom are solicited at least once a year, it would be interesting to conduct a comparative study of other journals. Another more realistic and modest approach could be to conduct interviews of reviewers of *RSDG* to evaluate whether their assessment practices differ when they consider articles for other academic journals. Indeed, *RSDG* is recognized for its high quality standards. It should be kept in mind that only 8% of articles are admitted as is; major modifications are required for 56%; minor modification for 26%; and 10% are rejected. Thus, the limit to one single journal could, it would seem to us, be surmounted through further research.

The second limit concerns the form of the empirical material analyzed. Indeed, the material is made up entirely of documents written by reviewers without explicit a priori criteria and is thus selective with respect to the multitude of ideas that reviewers have in mind when they evaluate an article. The free-text part of the report, from one to two pages long, should be considered a synthesis, and thus a selection of priority points in the eyes of the reviewer. Hence, these texts contain a portion of *non-dit* (unvoiced comments). This term refers to appreciations that are not stated or not made explicit by reviewers, either for lack of time or for reasons related to the dissemination of written reports that might have led them to silence or attenuate certain criticisms. This interpretation is supported by the relatively low average number of five appreciations per report that are significantly different in nature.

The third limit could stem from the provenance of reviewers. These are essentially European reviewers, marked by a tradition of qualitative research that is probably better established and more widespread than in the United States. Furthermore, North American influence on management research in Europe introduced quality criteria in European reviewer reports that are probably not found among North American reviewers. For example, with regard to the critical points *relevance* and *delimitation of bibliographical references*, the following comments appear: (a) ''lack of reference to French-speaking authors,'' (b) ''lack of reference to English-speaking authors,'' and (c) ''excessive English-speaking references.''

This contingent character stems from the provenance of the reviewers. However, an in-depth examination of the detailed results of the content analysis permits us to confirm that the presence of contextualized key ideas does not exceed 1.5%. The reproducibility outside of Europe of the results obtained inside the European context would merit a comparative study with American or Asian reviewers. In the meantime, we may point out that since the inception of our content analysis, the authors of this article have organized in Lyon, France (Savall, Bonnet, & Péron, 2004; Savall, Bonnet, Zardet, & Péron, 2007), four international conferences bringing together more than 900 contributions in partnership with the Academy of Management (Research Methods Division, 2004 and 2007, Social Issues in Management Division, 2005, and Organizational Development and Change Division, 2006). The relevance of the interpretation grid of the 19 critical points and the 10 quality criteria can thus be considered as confirmed.

The fourth limit of our findings is inherent to bias introduced by the method of content analysis used. Content analysis is sometimes disturbed by semantic ambiguity in the terms employed by the reviewers. For example, the word *field* takes on various significations, often implicit, among the different reviewers. For some reviewers, *field* designates an

enterprise where the researcher was physically present, conducting interviews, participating in meetings, and so on. For others, *field* designates data, most often qualitative data, collected through a medium (questionnaire, Internet), without the presence of the researcher in the enterprise nor his or her direct contact with the actors of the enterprise. The implicit model of the authors of this article, teacher-researchers in management and, as such, themselves directors of research and reviewers for other journals, probably influenced the structuring of critical points and quality criteria, resulting perhaps, in certain cases, in a hypertrophic or, on the contrary, atrophic representation of certain criteria. Indeed, the method of elaborating key ideas authorizes, in certain cases, a double interpretation due to semantic proximity of certain keywords.

Let us cite two examples: the critical point E *underlying ideologies* formulated in the thematic nomenclature as "explication of ideologies" was encountered in only very few reviewer comments. Is it relevant to maintain this critical point? We think it is, insofar as it represents, in our eyes, an authentically critical point that corrupts the epistemological quality of publications on management and social sciences (Perroux, 1970), and we fully assume this part of subjectivity inherent in the content analysis. The second example concerns the quality criterion *rigor*, which we identified because the keyword *rigorous* regularly appears in reviewer appreciations. However, one could consider that explication, coherency, and rationale are three characteristic qualities of rigor. Thus, it is probably a more synthetic quality criterion than the other three, yet at the same time, less precise under the pen of reviewers.

Finally, a limit to our findings results from the absence of a "mirror effect" (Savall & Zardet, 2004) of all reviewers reacting to the obtained analysis of their reports. We have, however, selectively performed this test on two reviewers. The dispersion and the inconsistent frequency of critical points/quality criteria pairs (see the 51 cells of Table 6) suggest the following hypothesis: the distribution and frequency of appreciations would probably have been different if reviewers had had at their disposal the grid of 19 critical points and 10 quality criteria because report content would probably have been rendered more homogeneous.

Conclusion

Following the discussion of our findings and their implications, we may once again underline what is the novel contribution of our study, as compared to what is typically admitted as reliable criteria for judging the quality of qualitative research. In contrast to the normative approach of our American colleagues, our research is the first and only empirical study of the implicit criteria actually employed in reviewer practices.

The content analysis of 474 reviewer reports on articles using qualitative methodologies written by 56 diversified reviewers identified 19 critical points, of which 8 are particularly mobilized by *RSDG* reviewers in their evaluations, and 10 actual quality criteria related to weaknesses detected by the reviewers, 3 of which are very frequently used (relevance, explication, and contribution).

Not all of these critical points and quality criteria are systematically mobilized by the reviewers. Reviewers use only certain critical points and certain quality criteria,

sometimes to indicate the article's major shortcomings, sometimes to highlight its positive qualities. This is clearly visible in the 51 cells actually used by reviewers out of the 190 possible cells of Table 6. Furthermore, among the 51 cells actually used, 22 were only rarely used (in only 2 or 3 reports), revealing a certain degree of indifference to those criteria on the part of reviewers. This accounts for the dispersion of reviewers, probably due to the fact that they had no preestablished evaluation grid at their disposal.

Most of the critical points and quality criteria have remained stable over the course of the 28-year period of reports on qualitative research. However, a marked evolution can be observed from emphasis on internal validity criteria for the first 10 years toward emphasis on external validity criteria in the past 10 years.

Out of the four admission categories (as is, minor modification, major modification, rejected), we noticed that for articles admitted "as is" or "rejected," reviewers resorted to fewer quality criteria to sustain their judgment.

This article explains with precision the content analysis method used. The method is replicable and could be reproduced by other journals if one wished to generalize (external validation) the results obtained for *RSDG*.

The comparison with extant literature reveals that some major quality criteria are shared: explicitness, relevance, and rationale. In fact, our analysis shows that 18 criteria out of the 51 mentioned in Table 6 feature both in U.S. literature and in *RSDG* reports. Nine criteria mentioned in U.S. literature never appear in the studied reviewer reports. Six criteria used by those reviewers are never mentioned in U.S. literature.

The limits of our study stem from the inevitable subjectivity of our content analysis and the fact that our study is based on a single European journal (Kwesiga & Pattie, 2006).

An important question is raised concerning the use of the results of this research: What status is to be attributed to the grid of the 19 critical points and 10 quality criteria? Is it a tool, to the degree that the objective could be systematizing the usation of these criteria or, on the contrary, could it be construed as simply a reference framework permitting an autonomous, flexible use by reviewers, who could mobilize only the variables that seem the most relevant to them? The answer corresponds to a choice of editorial policy by journals and can depend on the epistemological position that has been adopted vis-à-vis standardization versus flexibility, as a factor for improving the scientific quality of articles. We leave this question open to debate.

Notes

1. A third opinion was solicited in seven cases of qualitative articles.
2. The entire corpus of reviewer reports contained 1,559 idées-forces (pivotal ideas) and 5,234 quotes, among which featured 2,376 insightful reviewer appreciations and 2,858 key characteristics describing domain, type of research, and so on.

References

Aguinis, H. (1993). Action research and scientific method: Presumed discrepancies and actual similarities. *Journal of Applied Behavioral Science*, *29*, 416-531.

Boje, D. M., Oswick, C., & Ford, J. D. (2004). Language and organization: The doing of discourse. *Academy of Management Review*, *29*, 571-577.

Buono, A., & Savall, H. (2007). *Socio-economic intervention in organizations: The intervener-researcher and the SEAM approach to organizational analysis*. Charlotte, NC: Information Age.

Callon, M., & Latour, B. (1991). *La science telle qu'elle se fait* [Science in the making]. Paris: La Découverte.

Daft, R. L. (1995). Why I recommended that your manuscript be rejected and what you can do about it. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences* (pp. 164-182). Thousand Oaks, CA: Sage.

Easterby-Smith, M., Thorpe, R., & Lowe, A. (2002). *Management research*. London: Sage.

Gephart, R. (1988). *Ethnostatistics: Qualitative foundations for qualitative research*. Newbury Park, CA: Sage.

Gephart, R. (2004). Qualitative research and the Academy of Management Journal. *Academy of Management Journal*, *47*, 454-462.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine.

Golden-Biddle, K., & Locke, K. (1997). *Composing qualitative research*. Thousand Oaks, CA: Sage.

Hammersley, M. (1992). Deconstructing the qualitative-quantitative divide. In J. Brannen (Ed.), *Mixing methods: Qualitative and quantitative research* (pp. 127-143). Aldershot, UK: Avebury.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Kwesiga, E., & Pattie, M. (2006). Is management science built on a shaky foundation? The case for intersubjective certifiability. *Revue Sciences de gestion—Management Sciences—Ciencias de Gestión*, *50*, 33-68.

Lee, T. (1998). *Using qualitative methods in organizational research*. Thousand Oaks, CA: Sage.

Locke, K., & Golden-Biddle, K. (2002). An introduction to qualitative research: Its potential for industrial/organizational psychology. In S. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 99-118). Malden, MA: Blackwell.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Peirce, C. S. (1903). *The essential Peirce: Selected philosophical writings* (Vol. 2). Bloomington: Indiana University Press.

Perroux, F. (1970). Les conceptualisations implicitement normatives et les limites de la modélisation en économie [Implicitly normative conceptualizations and the limits of modelization in economics]. *Economies et Société*, *4*(12), 461-516.

Rynes, S. (2004). Qualitative research and the Academy of Management Journal (editor's note). *Academy of Management Journal*, *47*(4), 454.

Savall, H. (2003). International dissemination of the socio-economic method. *Journal of Organizational Change Management*, *16*(1), 107-115.

Savall, H., Bonnet, M., & Péron, M. (Eds.). (2004). *Crossing frontiers in qualitative and quantitative research methods* (Proceedings of the AOM-RMD/ISEOR International Conference). Lyon, France: ISEOR.

Savall, H., Bonnet, M., Zardet, V., & Péron, M. (Eds.). (2007). *Innovative research methods to create valid and operational knowledge* (Proceedings of the AOM-RMD/ISEOR International Conference). Lyon, France: ISEOR.

Savall, H., & Zardet, V. (1996). La dimension cognitive de la recherche-intervention [The cognitive dimension of intervention-research]. *Revue Internationale de Systémique*, *4*(1-2), 157-189.

Savall, H., & Zardet, V. (2004). *Recherche en sciences de gestion. Approche qualimétrique: Observer l'objet complexe* [Research in management sciences. Qualimetric approach: Observing the complex object]. Paris: Economica.

Savall, H., Zardet, V., & Harbi, N. (2004). Analyse transversale de diagnostics socio-économiques: Approche qualimétrique appliquée à des données qualitatives [Transversal analysis of socio-economic diagnostics: Qualimetric approach applied to qualitative data]. In *Crossing frontiers in qualitative and quantitative research methods. Proceedings of the AOM-RMD/ISEOR International Conference* (pp. 571-588). Lyon, France: ISEOR.

Seale, C. (1999). *The quality of quantitative research*. London: Sage.

Suddaby, R. (2006). What grounded theory is NOT. *Academy of Management Journal*, *49*, 633-642.

Sutton, R. E. (1997). The virtues of closet qualitative research. *Organization Science*, *8*(1), 97-106.
Weick, K. (1989). Theory construction as disciplined imagination. *Academy of Management Review*, *14*, 551-561.

Henri Savall received his PhD in economic sciences and management sciences from the University of Paris. He is a professor at the Institut d'Administration des Entreprises, Université Jean Moulin Lyon 3, and founder and director of the ISEOR. He is the director of the Centre d'Enseignement EUGINOV (Ecole Universitaire de Gestion Innovante) at the Institut d'Administration des Entreprises and of the Socio-Economic Management Master's program. His current research interests are qualimetric methodology (hybrid of qualitative and quantitative), socio-economic theory, and tetra-normalization. In 2001, he received the Rossi Award from the Academy of Moral and Political Sciences (Institut de France) for his work on integration of social variables into business strategies.

Véronique Zardet is a professor at the Institut d'Administration des Entreprises, Université Jean Moulin Lyon 3, and the codirector of the ISEOR Research Center. She heads the Research in Socioeconomic Management Master's program. She holds a PhD in management sciences from the University of Lyon. Her research is centered on the conduct of strategic change and the improvement of socio-economic performances in private enterprises and public services. In 2001, she received the Rossi Award from the Academy of Moral and Political Sciences (Institut de France) for her work on integration of social variables into business strategies.

Marc Bonnet is the deputy director of the ISEOR and a professor at the Institut d'Administration des Entreprises, Université Jean Moulin Lyon 3, where he heads the ''Industrial Security—Environment-Certification'' Master's program. He holds a PhD in management sciences from the University of Lyon and carries on his research, in particular, in the fields of hybrid methodology, hidden costs, and occupational disorders.

Michel Péron is professor emeritus at the Université Paris III–Sorbonne Nouvelle. He received his PhD from the Université Lumière Lyon II. He is a researcher at the ISEOR and the CERVEPAS. His research interests lie in hybrid methodology, cross-cultural management, corporate ethics, and the history of economic ideas. He has authored and coauthored a number of books and articles on civilization issues with a focus on corporate development on a global scale.