

Diversity of the Quality Criteria in Qualitative Research in the Health Sciences: Lessons From a Lexicometric Analysis Composed of 133 Guidelines

Marie Santiago Delefosse, Christine Bruchez, Amaelle Gavin & Sarah L. Stephen

Key words:
qualitative
research
evaluation; quality
criteria; health
sciences;
lexicometric
analysis; Alceste®

Abstract: A review of health sciences literature shows a substantial increase in qualitative publications. This work incorporates a certain number of research quality guidelines. We present the results of the Alceste® lexicometric analysis, which includes 133 quality grids for qualitative research covering five disciplinary fields of the health sciences: medicine and epidemiology, public health and health education, nursing, health sociology and anthropology, psychiatry and psychology. This analysis helped to cross-check the disciplinary fields with the various objectives assigned to the different criteria in the grids examined. The results obtained with Alceste® show the variability of the objectives sought by the authors of the guidelines. These discrepancies are not directly associated to disciplinary fields, and appear to be more closely linked to different qualitative research conceptualizations within the disciplines, and with essential qualitative research validation criteria. These conceptualizations must be clarified to help users better understand the objectives targeted by the grids, and promote more appreciation for qualitative research in the health sciences.

Table of Contents

- [1. Introduction](#)
- [2. Problem: A Variety of Qualitative Approaches in the Health Sciences](#)
- [3. Methodology](#)
 - [3.1 Make-up of the corpus: A complex bibliographical research](#)
 - [3.2 Presentation of Alceste® and relevance of the lexicometric analysis](#)
 - [3.3 Preparation of the corpus for Alceste® coding and initial observations](#)
- [4. Results](#)
 - [4.1 Distribution of the corpus into six classes](#)
 - [4.2 Significant terms related to the six classes and dendrogram](#)
 - [4.3 Analysis of the classes according to semantic content and significant terms](#)
 - [4.3.1 Quadrants C + D \(Classes 1, 5 and 4, clockwise direction\)](#)
 - [4.3.2 Quadrants B + A \(Classes 2, 6 and 3, clockwise direction\)](#)
- [5. Discussion](#)
- [6. Conclusion](#)
- [Acknowledgments](#)
- [References](#)
- [Authors](#)
- [Citation](#)

1. Introduction

Starting in the 1990s, publications on how to conduct and evaluate qualitative research increased significantly in the human and social sciences, in medicine and nursing, and in public health (BRITTEN, JONES, MURPHY & STACY, 1995; CÔTÉ & TURGEON, 2005; GUBA & LINCOLN, 1981; POPE & MAYS, 1995). Ample literature on the validity and possible use of qualitative health research emerged, and numerous proposals of guidelines and evaluation grids¹ to evaluate qualitative research were published, leading to a wealth of instructions, sometimes hard to compare, which contributes to the perception that "quality in qualitative research is a mystery to many health services researchers" (MAYS & POPE, 2000, p.50). These guidelines are geared toward researchers, but also journal editors and National Research Fund experts, as many of them state that they are ill-informed on the criteria needed to evaluate qualitative research (BLAXTER, 1996; MAYS & POPE, 1996; PICKLER, 2007). [1]

As part of an ongoing research², we examined 133 evaluation grids in five disciplinary fields of the health sciences: medicine and epidemiology, public health and health education, nursing, health sociology and anthropology, psychiatry and psychology. This document is an excerpt of this research and focuses on the results of the comparative analysis of the quality criteria included in these grids. [2]

In the first section, we briefly outline the status of the matter in the health sciences research field. In the second, we present our methodology. In the third, we present the results of the lexicometric analyses of the 133 grids examined. It highlights the variability and diversity of the concepts forming the structure of the grids, within the disciplines they represent. The last section discusses the inherent limits of the various research conceptualizations that support the grid conceptualization. [3]

2. Problem: A Variety of Qualitative Approaches in the Health Sciences

Debates on the relevance of establishing specific criteria or not for qualitative methodologies have continued to evolve, and in the 2000s, various demonstrations (seminars, networks, websites, etc.) focused on qualitative work and its quality. Similarly, various journals, namely the *Journal of Advanced Nursing*, *British Medical Journal*, *Canadian Journal of Public Health* and *Sociology of Health and Illness*, as well as publishers such as Blackwell enacted

1 We use the words "grids," "criteria grids," and "guidelines" as synonyms. They refer either to very structured grids set out in a quality criteria table format (a relatively long checklist) or to more detailed texts that explain each criterion separately (guidelines). Generally speaking, the terms "criteria grids" and "grids" include "guidelines" and "checklists."

2 This document represents part of a larger research funded by the Swiss National Research Foundation (2011-2014): "Quality of Qualitative Research in the Health Sciences: Which Evaluation Criteria?," whose main applicant is Prof. Marie Santiago DELEFOSSE, co-applicants Prof. Lazare BENAROYO and Dr. Alain KAUFMANN, University of Lausanne. Further information on the research is available on our website: <http://www.unil.ch/qualityofqualitativeresearch> [Accessed: March 7, 2015].

their own criteria for the evaluation of manuscripts, both for authors and experts. Numerous evaluation grids have also been published by renowned researchers in the field of qualitative methodologies (CRABTREE & MILLER, 1999; GUBA & LINCOLN, 1981; MILES & HUBERMAN, 1994; MORSE, BARRETT, MAYAN, OLSON & SPIERS, 2002; SANDELOWSKI & BARROSO, 2002; SILVERMAN & MARVASTI, 2008; YARDLEY, 2008). Universities, public health organizations (e.g. the [Critical Appraisal Skills Programme](#)), websites dedicated to qualitative methodologies, and various health institutes, also published their own evaluation grids. Although the dissemination of qualitative research work is increasing, researchers stress that the diversity of the criteria grids fosters a feeling of confusion and renders the evaluation difficult for expert evaluators and users (BLAXTER, 1996; ILG & BOOTHE, 2010; MAYS & POPE, 1996, 2000; PICKLER, 2007). And despite this series of work, qualitative research is still often evaluated using criteria applied to quantitative and experimental research. This leads to publication and/or recognition refusals. Discrepancies regarding the criteria proposed by the various grids indicate persistent difficulty in building minimal consensus on the criteria that would allow for the evaluation of qualitative research quality and that could be transmitted to the various reviewers of this work. [4]

It is in this regard that we undertook the task of collecting and analyzing the content of 133 existing guidelines to identify the reasons for this variability. Such an analysis will help us better understand what accounts for the diversity of the criteria in the grids, and if this diversity is related or not to the disciplinary fields. This is the work we are presenting below. We are limiting ourselves to the grids that we were able to collect in the health sciences, which are our specialty first and foremost and for which recognition of research and publications on qualitative work remains the most problematic. [5]

3. Methodology

3.1 Make-up of the corpus: A complex bibliographical research

An investigation using the keywords "health AND "qualitative research," qualitative research" AND "health" AND "assessment" or "appraising," "quality criteria" AND "guidelines," "peer review process" AND "guidelines," "evaluation" or "standards" AND "health" AND "quality criteria," was conducted in the [MEDLINE](#), [PsycInfo](#), [CINAHL](#), [PérUnil](#), [ScienceDirect](#) and [Web of Science](#) databases. This investigation aimed to find published grids offering quality criteria for researchers or/and reviewer's use. We examined data from various health sources: 1. instructions for authors of scientific journals; 2. articles designed to improve qualitative work for the use of authors and peer researchers; 3. articles designed for expert reviewers and describing the peer-review process of qualitative work; 4. theoretical articles discussing quality research validity issues; and 5. chapters from qualitative methodology manuals, regarding the validity of qualitative research. [6]

This literature helped us to examine 133 grids (checklists or guidelines), with the following inclusion criteria: 1. grids published for health scientists (medicine, nursing, health psychology, sociology of health, anthropology of health, public health, etc.); 2. grids that were sufficiently developed and commented; and 3. written work published by journals, health institutes or organizations, whose scientific quality has been validated. The material examined is essentially meant to be used as a support tool during the evaluation process. It contains lists of criteria to meet to produce "good qualitative research" that is solid and reliable. Our complete corpus includes the grids examined, namely tables, lists of criteria provided by the authors with texts specifying their meanings and uses, and additional texts from the same authors explaining their criteria (when the criteria grids were not sufficiently explicit). Most grids were published between 1993 and 2011, corresponding to the period during which many debates on the evaluation of qualitative research were held from the 1990s onwards. Table 1 shows the corpus and its distribution according to the five disciplinary fields from which the grids stem.

	Disciplines	Number of grids per discipline
Corpus: Grids, guidelines (tables, lists, box) and additional texts	Medicine	45
	Nursing	39
	Research methodology	20
	Public health	19
	Psychology, psychiatry	10

Table 1: Corpus description (133 guidelines/grids) [7]

Note that these grids were mostly edited by medical (e.g. CÔTÉ & TURGEON, 2005; KUPER, LINGARD & LEVINSON, 2008) and nursing journals (e.g. CESARIO, MORIN & SANTA-DONATO, 2002; COBB & HAGEMASTER, 1987). There are fewer methodology (e.g. CRESWELL, 2003; FLICK, 2006) and public health grids (e.g. DIXON-WOODS, SHAW, AGARWAL & SMITH, 2004; O'CATHAIN, MURPHY & NICHOLL, 2008). The disciplines that published the least are psychology/psychiatry (e.g. ELLIOTT, FISCHER & RENNIE, 1999; YARDLEY, 2000). [8]

3.2 Presentation of Alceste® and relevance of the lexicometric analysis

We used an analysis produced by [Alceste®](#) to process the contents of this corpus. Alceste® is textual or statistical data analysis software, originally designed by Max REINERT (1990, 1993) of the [Centre National de la Recherche Scientifique](#) (CNRS) in France. Its use spread to the human and social sciences field in the 1990s. It functions by way of frequency vocabulary count and helps to obtain analysis units that are based on formal criteria. It uses an inductive and recursive approach and helps to identify co-occurrences, or word associations in

a sentence, using a treatment that is based on word resemblances and differences. Technically, Alceste® breaks the corpus into fragments that are relatively similar in size, referred to as "elementary context units."³ These fragments are then reclassified statistically and split into classes that are as differentiated as possible in terms of their specific vocabulary level. This classification is meant to split statements into classes marked by the contrast of their vocabulary, and thus opposed to one another (KALAMPALIKIS, 2003). These classes are called "lexical worlds" and are deemed to present an "idea" of the representations contained in the text as well as the main ideas and themes of the corpus (GARRIC & CAPDEVIELLE-MOUGNIBAS, 2009). Then, Alceste® defines a "mapping" of what the software developers called "contextual variables."⁴ These variables serve to identify texts and are related to their content. They are introduced by the researchers according to their relevance to the research questions. The classes highlighted by the software must then be examined and linked to the co-occurrences to give them meaning and explain their differences (AUBERT-LOTARSKI & CAPDEVIELLE-MOUGNIBAS, 2002). [9]

We thought that using this software was relevant given its textual classification abilities. Our hypothesis is that the differences between the grids are linked to the differences between the disciplinary fields in the health sciences. If it is validated, we will be able to locate these fields in the lexical analysis, indicating a clear class separation between the various fields. [10]

3.3 Preparation of the corpus for Alceste® coding and initial observations

Using Alceste® requires prior coding of the texts to indicate to the software which variables must be identified. To locate the variables that we deemed most relevant, in addition to the "disciplinary field" variable, we conducted a content analysis of our corpus (BRAUN & CLARKE, 2006; ROBERT & BOUILLAGUET, 2002). The results were triangulated and discussed collectively to come to an inter-judge agreement. This content analysis allowed us to identify the variables that should be pinpointed in the grids: other than the disciplinary fields, we identified differences in the objectives for the criteria in the grids. This led us to develop a "typology of the content of the criteria" present in the corpus, or the objectives targeted by each type of criteria (evaluate the implementation of a research, evaluate the methodology, evaluate the skills of the researcher, etc.). Each grid excerpt was thus subject to an analysis, a consultation and a codification, according to this typology. Table 2 presents the typology, as well as its codification for Alceste®.

3 *Elementary context units* (e.c.u.) defined by Alceste® are the smallest statistical units created by the software, based on a compromise between the syntactic form (proper punctuation) and the statistical constraints (these units must be of comparable size).

4 Contextual variables are data elements known by the researchers and used as instructions for the software so that the lexical analysis can be conducted with these variables in mind.

Criteria focused on	Description of the category	Code ⁵
Criteria (taken from quantitative methods)	General description of the criteria to include in a research plan	crigen
	Detailed description of the criteria usually included in a research plan (usually the same as the quantitative ones)	cridet
Criteria designed for qualitative research	General description of the criteria that are methodologically and explicitly designed for qualitative research	crigq
	Detailed description of the criteria that are methodologically and explicitly designed for qualitative research	cridq
Theories	General description of the theories and aspects of qualitative research, with no description of the methodological criteria <i>per se</i> (description geared toward the processes and epistemology of the research)	thegen
	Detailed description of the theories and aspects of qualitative research, with no description of the methodological criteria <i>per se</i> (description geared toward the processes and epistemology of the research)	thedet
Content	General description of the content to appear in an article (with no description of the methodological criteria <i>per se</i>).	cogen
	Detailed description of the content to appear in an article (with no description of the methodological criteria <i>per se</i>).	codet
Qualities of the researcher	General description of the qualities of the researcher (his "moral" and "ethical" position, his training, etc.)	chgen
	Detailed description of the qualities of the researcher (his "moral" and "ethical" position, his training, etc.)	chdet
Research procedure	General description of the research procedure	regen
	Detailed description of the research procedure	redet

Table 2: Typology of the criteria in the grids used to code the criteria for Alceste® [11]

The need to conduct a content analysis of the criteria in the grids to identify the major categories for Alceste® led to initial observations. During this analysis, we noticed strong variability in the objectives targeted by the authors of the grids, and somewhat less attention given to certain types of criteria. This shows that the authors do not give the same importance to the same criteria when it comes to determining the quality of a qualitative research. According to the grids, the priority criteria are:

⁵ The data in the "Code" column identify the criteria for the lexicometric software.

- general or specific and very detailed;
- adapted or not to qualitative research;
- focused on general or specific and detailed theories;
- focused on general or detailed contents;
- focused on the general or very detailed qualities of the researcher;
- focused on general or detailed research procedures. [12]

We deemed that the differences in objectives defining the criteria in the grids examined were significant enough to keep them as contextual variables to cross-check the disciplinary fields with (Table 3). We wanted to see if the differences in the criteria's focal point were specific to the disciplinary fields (e.g. greater emphasis put on the qualities of the researcher in certain fields compared to others, etc.), or if these differences were external to the disciplinary fields themselves. Set out below are the results that cross-tabulate the "disciplinary field" contextual variables and "type of contents in the quality criteria."

Five disciplinary fields	Methodology (meth); medicine (med); nursing sciences (nurs); public health (pubh); psychology, psychiatry (psy)
Twelve types of content (cf. Table 2)	crigen; cridet; crigg; cridq; thegen; thedet; cogen; codet; chgen; chdet; regen; redet

Table 3: Contextual variables kept for Alceste® coding [13]

4. Results

4.1 Distribution of the corpus into six classes

Processing the corpus with Alceste® resulted in its distribution into six distinct classes, integrating 67% of the total corpus. The factorial design splits the grids according to the following six classes (numbered from 1 to 6) and indicates the contextual variables most often related to them. The most significant terms for each class are also included. To make it easier to interpret the results, we numbered the four quadrants of the factorial design (A, B, C and D).

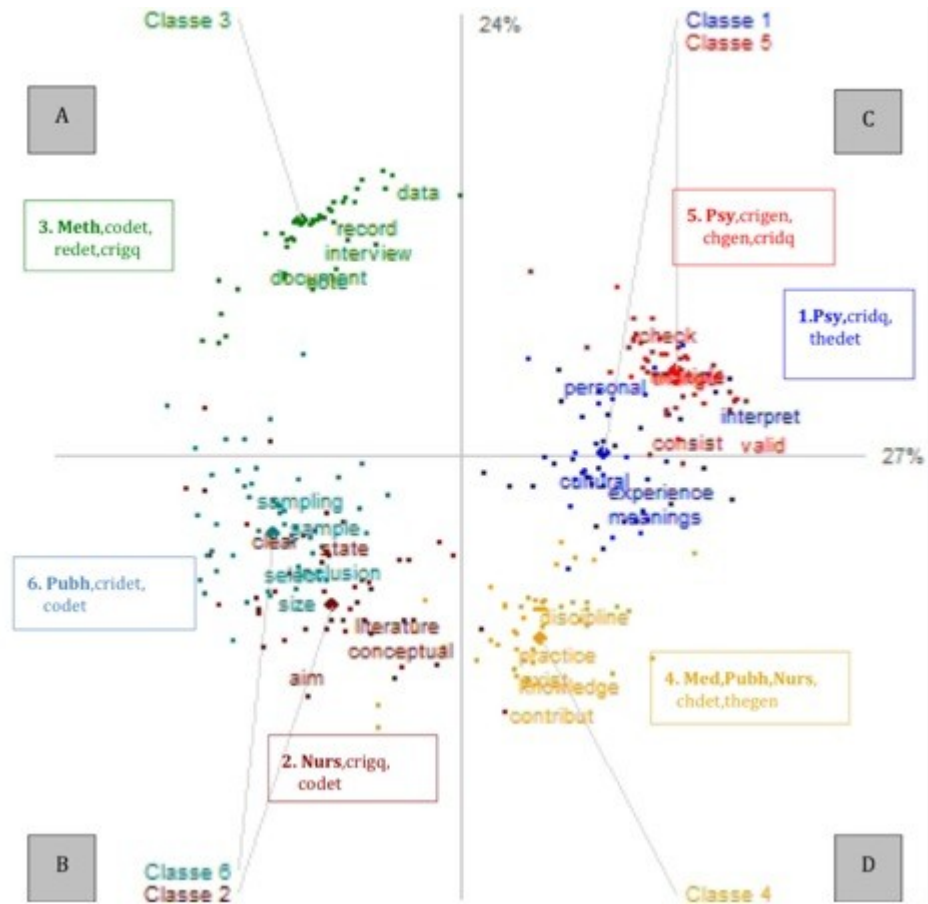


Illustration 1: Factorial design of the distribution of the six classes and the contextual variables most often related to them [14]

Illustration 1 shows that there are "independent" classes and others that are more closely linked. Classes 3 and 4 are well identified and relatively independent. Others are more intertwined, namely Classes 1 and 5, and Classes 2 and 6. To this combination we must add that certain disciplinary fields appear in two different classes, which indicates a split at the heart of certain disciplines. For example, there are two different groups of psychology/psychiatry grids. A similar division appears in the nursing and public health grids, which are split between Classes 2, 6 and 4. [15]

The factorial design also shows a bipartite distribution between two types of criteria content, on the one hand focused on theoretical aspects (thedet, thegen) and the qualities of the researcher (chgen, chdet) for Quadrants C and D, and on the other hand on the criteria relative to research methods (codet) and the general/detailed criteria of qualitative research (crigq, cridet) for Quadrants A and B. [16]

4.2 Significant terms related to the six classes and dendrogram

To explain the particularities identified in the factorial design and identify the specificities of each class, we must examine the semantic features specific to each class. We conducted an analysis of the significant terms (chi2) in each class, which was useful in highlighting each one's identity and better understanding the differences and/or similarities between the grids, and the various distributions of the content typologies. The dendrogram (Illustration 2) shows how the six classes organize themselves, with the five most significant words for each.

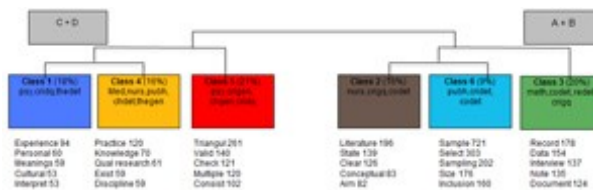


Illustration 2: Dendrogram: Distribution of the corpus into six classes with significant terms (chi2). Please click [here](#) for an increased version of Illustration 2. [17]

The classes are split into two main branches grouping together Classes 1, 4 and 5 (Quadrants C + D), and Classes 2, 6 and 3 (quadrants A + B). In the first branch, two separate groups of psychology grids have different types of criteria content: for Class 1 (18% of the corpus), they focused on the experience, the meaning of things and comprehension, and for Class 5 (21%) on criteria linked to validation. Class 4 (16%) is more mixed and groups together part of nursing, medicine and public health. It is also more focused on practices, care, discipline and clinic. At the opposite end of the dendrogram (Quadrants A + B), Class 2 (16%) groups together another part of the nursing grids and is more focused on the analysis of existing literature, the workframe and research process; Class 6 (9%) includes a small group of public health grids, which are very different from those in Class 4 and are focused on sampling methods and inclusion/exclusion criteria; finally, Class 3 (20%) includes a group of methodology grids, with their content focused on practices and tools. [18]

These initial results show that the characteristics of the grids do not necessarily cover a classification based on disciplinary fields: Instead, what seems to connect or differentiate them relates more to the type of content of the criteria in the grids, or their "target content" (Table 4).

Criteria linked to the researcher and his/her creativity. Meaning analysis and practical interest. Qualitative or mixed methodologies (C+D)	Criteria linked to methodological rigor, namely logical steps, sampling or data and its collection. Quantitative or mixed methodological approaches (A+B)
Class 1: Meaning, culture, and experience of the researcher Class 4: Practice and its related skills Class 5: Qualitative methodology and criticism	Class 2: Literature review, detailed objectives Class 3: Data collection tools, discussions, documents, notes Class 6: Sampling, population

Table 4: Classification of the grids according to the semantic content of the criteria [19]

This content requires careful examination to specify its links to the disciplines concerned. We present a more detailed analysis of the six classes based on their elementary context units⁶ (semantic content), representative of their "lexical world"⁷, from the contextual variables (discipline, types of criteria content) processed by Alceste®, to better understand their importance in the differentiation of the grids. [20]

4.3 Analysis of the classes according to semantic content and significant terms

4.3.1 Quadrants C + D (Classes 1, 5 and 4, clockwise direction)

Class 1: This class groups together part of the grids stemming from the psychology/psychiatry field (chi2=60) as well as grids of detailed criteria adapted specifically to qualitative research (cridq, chi2=33), and detailed criteria linked to theoretical aspects (thedet, chi2=10). On the lexical front, we notice the presence of words such as *effect, self, understand, change, live, preconception, reflexivity, power, affect, instrument*, etc. Analyzing the elementary context units in which they appear reveals a semantic field that is guided by "meaning," by the "lived experience," the context and culture, suggesting interpretations and empowerment as a potential action on the reader or participants:

"uce n° 2198 Khi2 = 20 uci n° 109 : *aut_Stiles *year_1993 *clas_psy *c1_cridq⁸
 if (they) are (thereby) (empowered), (they) (will) (change) (by) (taking) more control of
 (their) (lives); (as) (an) interesting corollary, (it) (is) in scientists professional interest
 to (empower) (their) (research) participants, (as) a (way) of validating (their)
 (interpretations)." [21]

This class also includes the issue of comprehension and interpretation, which must be linked to the researcher's assumptions. Specific to the qualitative

6 See Note 3.

7 See Section 3.3.

8 The first line refers to the contextual variables, and the following paragraph is an excerpt of text treated by the software Alceste® and based on word associations.

approach, the criteria in this class are based on the importance of the researcher's reflective attitude. The latter is considered to be an integral research tool, and can thus affect the quality of the research as well as its participants and their attitudes and behaviors, hence the importance granted to the research criteria that report on it. The quality criteria deemed important in qualitative research are linked to meaning analysis and the research's practical interest. One can conclude that the research conceptualization, in the grids found in this group, is that of a research meant to serve participants to transform themselves and to produce new and creative ideas and theories. [22]

Class 5: It includes a second group of psychology/psychiatry grids ($\chi^2=3$). As for the types of criteria content, they mostly include the general criteria of a research plan (crigen, $\chi^2=65$), general descriptions of the researcher's qualities (chgen, $\chi^2=41$), and detailed criteria specific to qualitative research (cridq, $\chi^2=24$). Although the disciplines present may be the same as those in Class 1, the grids in this group are different in that they focus more on research bias validation and reduction ideas. So, the software highlights the importance of words such as *member*, *establish*, *different*, *confirm*, *external*, *corroborate*, *truth*, *using*, *negative*, or *source* and *audit*. A review of the elementary context units helps to shed light on the group's main objectives: establish a level of trust using various triangulation techniques, and achieve adequate transparency by explaining these methods. This can be attained by hybridizing several perspectives (data, sources, data stemming from the literature, methods or researchers) by using general research validation techniques. The research can also be strengthened through means that are more specific to qualitative research, such as iterative analysis evidence, the search for negative cases or the sharing of results with participants (member-checking) to see if the interpretations are coherent, or through the saturation/recurrence of themes presented in the results. Generally speaking, it is the multitude of data, from the investigators and auditors, and the multitude of viewpoints on a given phenomenon, which can serve to convince:

"uce n° 2135 $\chi^2 = 35$ uci n° 108 : *aut_Stiles *year_1999 *clas_psy *c1_crigq
(replication:) did (multiple) investigators who were (familiar) (with) the observations,
(members) of the research (team), (external) reviewers or (auditors), (find) the
(proposed) interpretation convincing? Were the (conclusions) based on (formal) rules
of (evidence)?" [23]

The criteria present in this group seem to promote research that is mostly based on the cross-referencing of data, hence the importance of the researchers' collaboration and the various forms of triangulation. This collaborative form of validating results and interpretations includes researchers and research participants (confrontation of opinions), whom are all considered research instruments (a common aspect with the group in Class 1). So, research is therefore a part of a collective co-construction whose facilitator is the researcher. The validity and meaning of the research stem from deliberation. The content of

the criteria of this group refers to the research's conceptualization as a co-construction involving various stakeholders. [24]

Class 4: This heterogeneous group is relatively close to the psychology/psychiatry group in Class 5 and includes a lot of medical grids ($\chi^2=10$) and a few public health ($\chi^2=5$) and nursing ones ($\chi^2=2$). The type of criteria content focuses on are detailed descriptions of the researcher's qualities (chdet, $\chi^2=19$) and general descriptions of the theoretical aspects of qualitative research (thegen, $\chi^2=11$). On the lexical front, the software indicates the weight of words such as *contribute, health, care, adapt, policy, patients, useful, social, evaluate, new, or practitioner, world and standard*. Examining the class' significant terms and elementary context units reveals criteria that are linked to pragmatic methodologies. The practical use of care research and/or future research is considered to be a quality criterion in itself. Criteria regarding applicability are often valued in the clinical practice:

"uce n° 492 Khi2 = 39 uci n° 30: *aut_Emden *year_1998 *clas_nurs *c1_crigq
(quality) of (qualitative_research) is about an (emphasis) (on) the (practical) (utility) of
research because (nursing) is (evaluated) (on) (its) relationship (to) (real) (practice)
problems more (than) researchers (tend) (to) think." [25]

Other criteria highlight the use of bringing new knowledge and thus contributing to the development of disciplines:

"uce n° 2593 Khi2 = 22 uci n° 127 : *aut_jan *year_2011 *clas_nurs *c1_crigen
*c2_codet
for example, this (new) (knowledge) could (contribute) (to) (new) conceptualizations
or question (existing) ones; it could lead (to) the (development) of (tentative),
(substantive) theories, or even hypotheses, it could (advance), question (existing)
theories or provide (methodological) (insights), or it could provide data that could lead
(to) (improvements) (in) (practice)." [26]

Generally speaking, the criteria proposed are concerned about research serving the development of the discipline (especially in nursing) and/or existing theories through new knowledge. The criteria listed therefore highlight the use of the research and its results, in terms of their contribution toward practices and knowledge. The validity of research is mostly based on existing knowledge, but it must also be original and meaningful, while remaining open to interdisciplinarity. One can conclude that the research conceptualization of this class is more focused on its use and contribution of new knowledge than on its methodological aspects. [27]

Overall, the quality criteria of qualitative research in the grids of Quadrants C and D are geared toward the evaluation of situations focused on the subjects, meaning and comprehension in practical and clinical situations. Hence, "technical" criteria do not characterize the grids of these three groups as much as criteria looking to evaluate the researcher's qualities, research forms giving a

voice to actors and concrete situations that foster contribution to practice and knowledge. [28]

4.3.2 Quadrants B + A (Classes 2, 6 and 3, clockwise direction)

Class 2: This class includes a group of grids from the nursing field ($\chi^2=11$), whose major lexical content has words such as *framework*, *question*, *paper*, *study*, *review*, *reference*, *discuss*, *purpose*, *discern*, or *rationale* and *explain*. The types of criteria content are general criteria linked to qualitative research (crigq, $\chi^2=24$) and criteria detailing the "experimental" plan of a research (codet, $\chi^2=17$). A review of the significant terms shows that the class seems focused on the research's anchoring ideas in a clear and proper theoretical framework, and research validation through a comprehensive literature review, explicit and coherent objectives, or research links to the clinic, or impact on health practices and policies (a common aspect with Class 4). Detailed analysis of the elementary context units shows a focus on classic semi-quantitative criteria such as the relevance of the literature review, the clarification of concepts, the theoretical framework or the ideological orientation that could influence the researcher or the conduct of the research:

"uce n° 110 Khi2 = 58 uci n° 6: *aut_Beck *year_2009 *clas_nurs *c1_codet
*c2_cridet

(literature) (review:) (report) adequately (summarizes) (the) existing (body) of knowledge (related) to (the) (problem) (or) (phenomenon) of interest; (provides) (a) solid basis (for) (the) new (study). (conceptual) (underpinnings:) (key) (concepts) (defined) (conceptually); (philosophical) basis, (underlying) (tradition), (conceptual) (framework) (or) (ideological) (orientation) is (explicit), (and) appropriate (for) (the) concerning (the) conduct of (a) (study), (method) is (stated) (or) implied" [29]

This class is characterized by a need to set out and document all steps of a "classic" research clearly, to then be able to re-duplicate it, to ensure comparability in context and transferability. In this background, the quality criteria of qualitative research involve specifying the logical steps of the research to demonstrate the validity and reliability of the reasoning; the reader must be able to identify the theoretical framework and its relationship to the research question. The work must be anchored in (relevant and comprehensive) existing literature, and specify the contributions of the study. The quality of the research is characterized by an explicit method that is carefully monitored and well detailed, contributing to existing science through the addition of new theories/models. It is based on technical and methodological rigor. [30]

Class 6: This class is closely linked to Class 2, which we just looked at; however, it differs through much of its criteria content. It includes a group of public health grids ($\chi^2=13$) composed of criteria different from those found in Class 4 (med, nurs, pubh). These grids are more focused on methodology and sampling issues, and on the detailed description of the technical components of a research (codet, $\chi^2=19$), with significant words or word roots such as *purposive*, *case*, *character*,

exclusion, convenience, settings, choose, age, extreme, or recruit- and redund-. The criteria are focused on the quality of the methodological procedures such as a broad and diverse sampling that is justified based on the study, the selection processes used for participants, the inclusion criteria, or the characteristics of the population studied:

"uce n° 1087 Khi2 = 48 uci n° 63: *aut_Long *year_2006 *clas_pubh *c1_cridet
is the (sample) (appropriate) to the aims (of) the study? is the (sample) (appropriate)
in (terms) (of) (depth), (intensity) (of) data collection in (individuals), (settings) and
(events), and width across (time), (settings) and (events), (example), to (capture) key
(persons) and (events)." [31]

This focus on technical procedure criteria, at times closely linked to statistical treatment, can be explained by the criteria's focus on the transferability of results. Special attention is paid to the criteria establishing the validity of the sample, including its size and selection process. The final sample should be adequate and suited to the research in terms of the depth and intensity of the individual data collection, i.e. sufficiently broad and diversified to meet research objectives. It is thus important to detail the profile of the selected populations and the case inclusion criteria to compare the objectives with the samples obtained, or the population. Generally speaking, this class favors criteria linked to the rigor of the methodology, the logic of the steps, the sampling and data collection. The grids focus on the rigor, deemed to be the proper use of the sampling technique. It presents a "supervisory" approach to research, based on the suitability of the selected population. [32]

Class 3: This last class is well defined and does not overlap with any other. It includes many grids stemming from methodological articles and/or publications (chi2=94). The type of criteria content includes detailed descriptions of research content (codet, chi2=73), detailed descriptions of research conduct (redet, chi2=27) and general qualitative research criteria (crigq, chi2=22). The software indicates the importance of words such as *collect, consent, transcript, site, technique, procedure, audio, file, observe, analysis, or confidential, computer, and protect*. Comprehensive review of these terms and of the classes' elementary context units shows the importance given to some elements, as the description of the data collection tools (interview, notes, various documents), the description of the data processing method, and the ways to store and/or secure (ethics) the data:

"uce n° 1606 Khi2 = 33 uci n° 84: *aut_Plochg *year_2002 *clas_meth *c1_cridq
*c2_codet
new (name) for (cleaned) (up) (files). (data) (storage:) (securing) (data) against loss.
(storage) of (transcript) and (cleaned) (up) (files:) the (primary) (file), the (transcript)
and the (processed) (files) (should) always (be) (stored) in different places; ideally the
(primary) (file), with a good (description), (should) (be) (locked) (away)." [33]

This class is more concerned with classic methodological criteria, and remains very focused on criteria linked to the "evidence": evidence of what was done, the

relevance of the data collected, its existence and clarification of the data collection method:

"uce n° 516 Khi2 = 37 uci n° 32: *aut_Fitzpat *year_1994 *clas_pubh *c1_cridet *c2_cridq

are (careful) (records) of (data) (kept)? (audio), (video) (recordings) and (fieldnotes) which can (be) (independently) (inspected). (data) (analysis:) are the processes of (data) (analysis) (adequately) described? an account of how (data) were (processed) and interpreted." [34]

Research quality is linked to the researcher's pedagogical skills, which explain his/her data and its collection method, and presents them to peers for evaluation and proof of their relevance. The criteria grids of this class focus on the data's quality control criteria (storage methods, faithful transcription, etc.) and on criteria that help to ensure that the researcher has been trained on the methods used (existence of a logbook, data storage and protection, etc.). There are also criteria focused on ensuring that the research meets its objectives, and on the presence of data in publications (*verbatim*). While in the other classes ethics was not an explicit quality criterion, here it is an established and intrinsic criterion, beyond the researcher's values. It deals with the consent of participants as well as their information on the objectives and the research conduct and data use. [35]

In total, the qualitative research quality criteria in the grids of Quadrants A and B are essentially focused on the evidence and rigor in the methodological steps, more than on the subject or meaning. These criteria are therefore less focused on the contribution of the practice and clinical data, and more focused on the rigor of logical thought, the evidence of the careful monitoring of the various steps and the contribution to theories and models. [36]

5. Discussion

The Alceste® lexicometric analysis confirms the presence of major discrepancies between the quality criteria grids examined. Many conclusions can be drawn from this analysis.

1. Despite the disciplinary origins identified in the evaluation grids, the various underlying quality conceptualizations of qualitative research do not necessarily include the disciplines to which these grids belong. This is why we find different content and orientations within the same discipline, or similar content and orientations within different disciplines. Therefore, the criteria grids do not represent a division by disciplinary field that would explain their diversity. This diversity actually seems to represent a split within certain disciplinary fields. Therefore, two groups stemming from psychology/psychiatry (Classes 1 + 5) can be identified, as well as two groups of guidelines stemming from nursing (Classes 2 + 4) or public health (Classes 4 + 6). In contrast, the methodology grids in Class 3 represent a conceptualization of the research quality that is specific and different from the rest. Our results thus show that the diversity of

the quality criteria grids of qualitative research in the health sciences is mainly due to the possible contradictions within the disciplinary fields themselves, contradictions linked to different research conceptualizations. In this case, a major difficulty lies in the fact that, on the one hand, these research conceptualizations are usually implicit in the guidelines, and on the other hand, their discrepancies seem to appear unbeknownst to the authors themselves, which could explain the difficulty in perceiving the consistency between the various criteria grids. The six classes stemming from the lexicometric analysis seem to reflect at least six conceptualizations that are rather different from the "essential" criteria to ensure the validity of qualitative research. Yet, these conceptualizations are not necessarily compatible.

2. Despite the discrepancies observed however, two major trends are noticed and should be exploited. On the one hand, a first group of grids seems to focus on the importance of criteria linked to methodological rigor. These grids favor the anchoring in a theoretical framework, methodological rigor, a proper use of research tools, the ethical compliance through the consent of participants, the use of logical steps in research conduct, the anchoring in literature review, the detailed description of sampling, target populations, etc. On the other hand, a second group of grids focuses on criteria linked to the researcher and his/her creativity, and to the research use value. Those grids detail the mission of the researcher, who is presented as a teacher explaining his data, a specialist of their representativeness, a research instrument, a facilitator of the collective co-construction of the meaning and interpretations, as well as a caution regarding the practical use of the research, namely the empowerment of participants. This split remains very "global" however, and must be examined further as it is important to bear in mind that various conceptualizations of research quality can be found within the same grid.
3. The dispersion of grid components in different classes shows a lack of internal homogeneity, both in the structure and the content of the grids⁹. As such, parts of one grid can be found in two opposite classes. This lack of homogeneity within the grids can be explained in various ways: on the one hand, it reflects the discrepancies between the expectations and the priorities given to the grids by their authors, and on the other, it can also reveal the various objectives within one grid, thus revealing the lack of a clear guideline on the specific criteria on which the quality of a qualitative research is based. Moreover, this heterogeneity can also be linked to a lack of consistency in the construction of the grids themselves. The heterogeneity of the associated criteria explain, in part, why it is so hard to agree on the quality criteria of qualitative research, as the grids can represent different objectives, evaluate various levels, and not address the same audience. [37]

In light of these results, one can consider that: 1. the authors should specify their research conceptualizations and the grid objectives; 2. they could better specify and define the terminology used, which remains scattered and not sufficiently

⁹ It is important to note that the Alceste® analysis was conducted on the broader corpus group, and that consequently, the content of the elementary context units is composed both of the criteria themselves and the explanatory texts about these criteria. This helps to support our findings by applying it to more consequential material compared to simple criteria lists.

informative; and 3. it would be a good idea to test grids that are being conceptualized on a concrete field to evaluate their feasibility and use value. This set of procedures would help to make the grids more homogenous and comparable. Indeed, many grids were obviously built through the addition or subtraction, or both, of terms taken from other grids, thus creating confusion and a lack of internal consistency. Finally, the fact that a split between two major criteria groups was identified can also serve as a guide for the construction of a consensual grid composed of the most important criteria in the two sectors (rigorous methodology vs. creativity/use validity). [38]

6. Conclusion

Our research results show how much quality criteria grids for qualitative research remain heterogeneous between the various disciplinary fields, within the fields, and in the conceptualization of the grids themselves. Consequently, it is hard to group together "essential and consensual" quality criteria allowing for the *in abstracto* evaluation of qualitative research. Existing grids do not seem to be representative of a discipline but more of a position regarding research and objectives inherent to each of their designers. Although these positions can be found in two major groups, they remain global and implicit, especially, as they are never mentioned in the grids themselves. In this division, we first notice the grids whose essential quality criteria focus on the "technical and methodological procedures" (monitoring of the quantitative type experimental plan, the predominance of the evidence and contributions to the models, and lack of criteria focused on the researcher and meaning), and second the grids whose essential quality criteria are more focused on the "meaning production conditions" (researchers and their position, values, epistemology and practical contributions). [39]

These results confirm the need to proceed to a more in-depth analysis of the quality criteria grids for qualitative research, which should help to highlight the influence of the various conceptualizations of qualitative research, beyond the disciplinary fields. Completing our results with the identification of the underlying paradigmatic and theoretical positions for the division of the grids should improve comprehension of this diversity. This could shed light on their differences and origins as well as on the ways we could regroup them according to specific objectives. Indeed, as long as conceptualizations deemed major by researchers are not identified and the common definitions of these conceptualizations clearly specified, the grids will only be of limited assistance, especially in the recognition of qualitative research. Although this work still remains to be done, a large part of the grids remains very useful in helping researchers identify the various parts of a qualitative research, article or report. [40]

Acknowledgments

The authors acknowledge the funding provided by the Swiss National Research Foundation (CR3211-132259), and would like to thank Valérie CAPDEVIELLE- MOUGNIBAS, from the University of Toulouse, France, for his valuable help in processing the data for Alceste®.

References

- Aubert-Lotarski, Angeline & Capdevielle-Mougnibas, Valérie (2002). Dialogue méthodologique autour de l'utilisation du logiciel Alceste en sciences humaines et sociales: "lisibilité" du corpus et interprétation des résultats [Methodological dialogue on the use of Alceste software in social and human sciences: "readability" of the corpus and interpretation of results]. *6èmes Journées Internationales d'Analyse Textuelle* [6th International Conference on Textual Analysis], March 2002, Saint-Malo, France.
- Blaxter, Mildred (1996). Criteria for the evaluation of qualitative research papers. *Medical Sociological News*, 22, 68-71.
- Braun, Virginia & Clarke, Victoria (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.
- Britten, Nicky; Jones, Roger; Murphy, Elizabeth & Stacy, Rosie (1995). Qualitative research methods in general practice and primary care. *Family Practice*, 12(1), 104-114.
- Cesario, Sandra; Morin, Karen & Santa-Donato, Anne (2002). Evaluating the level of evidence of qualitative research. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, 31(6), 708-714.
- Cobb, Ann Kuckelman & Hagemaster, Julia Nelson (1987). Ten criteria for evaluating qualitative research proposals. *Journal of Nursing Education*, 26(4), 138-143.
- Côté, Luc & Turgeon, Jean (2005). Appraising qualitative research articles in medicine and medical education. *Medical Teacher*, 27(1), 71-75.
- Crabtree, Benjamin & Miller, William (Eds.) (1999). *Doing qualitative research* (2nd ed). Thousand Oaks, CA: Sage.
- Creswell, John W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Dixon-Woods, Mary; Shaw, Rachel L.; Agarwal, Shona & Smith, Jonathan A. (2004). The problem of appraising qualitative research. *Quality & Safety in Health Care*, 13(3), 223-225.
- Elliott, Robert; Fischer, Constance T. & Rennie, David L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, 38, 215-229.
- Flick, Uwe (2006). *An introduction to qualitative research*. London: Sage.
- Garric, Nathalie & Capdevielle-Mougnibas, Valérie (2009). La variation comme principe d'exploration de corpus: Intérêts et limites de l'analyse lexicométrique interdisciplinaire pour l'étude du discours [Variation as a principle to explore the research corpus: Interests and limitations of the interdisciplinary lexicometric analysis to study discursive material]. *Corpus*, 8, 105-128.
- Guba, Egon G. & Lincoln, Yvonna S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.
- Ilg, Stefan & Boothe, Brigitte (2010). Qualitative research in psychology: What does a good publication contain?. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 11(2), Art. 25, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1002256> [Accessed: March 11, 2015]. .
- Kalampalikis, Nikos (2003). L'apport de la méthode Alceste dans l'étude des représentations sociales [The contribution of Alceste method in the study of social representations]. In Jean-Claude Abric (Ed.), *Méthodes d'étude des représentations sociales* [Methods for studying social representations] (pp.147-163). Paris: Erès.
- Kuper, Ayelet; Lingard, Lorelei & Levinson, Wendy (2008). Critically appraising qualitative research. *British Medical Journal*, 337, 687-689.
- Mays, Nicholas & Pope, Catherine (1996). *Qualitative research in health care*. London: BMJ Books.

Mays, Nicholas & Pope, Catherine (2000). Qualitative research in health care: Assessing quality in qualitative research. *British Medical Journal*, 320, 50-52.

Miles, Matthew B. & Huberman, A. Michael (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Morse, Janice M.; Barrett, Michael; Mayan, Maria; Olson, Karin & Spiers, Jude (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1(2), 13-22,
<http://ejournals.library.ualberta.ca/index.php/IJQM/article/view/4603/3756> [Accessed: February 20, 2014].

O'Cathain, Alicia; Murphy, Elizabeth & Nicholl, Jon (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research & Policy*, 13(2), 92-98.

Pickler, Rita H. (2007). Evaluating qualitative research studies. *Journal of Pediatric Health Care*, 21, 195-197.

Pope, Catherine & Mays, Nicholas (1995). Qualitative research: Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *British Medical Journal*, 311, 42-45.

Reinert, Max (1990). Alceste. Une méthodologie d'analyse des données textuelles et une application: Aurélia de Gérard de Nerval [Alceste. A methodology for analyzing textual data and its related application: Aurélia by Gérard de Nerval]. *Bulletin de méthodologie sociologique*, 26, 24-54.

Reinert, Max (1993). Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars [The "lexical worlds" and their "logics" through the statistical analysis of a corpus of stories of nightmares]. *Langage et société*, 60, 5-39.

Robert, André D. & Bouillaguet, Annick (2002). *L'analyse de contenu* [Content analysis]. Paris: Presses Universitaires de France.

Sandelowski, Margarete & Barroso, Julie (2002). Reading qualitative studies. *International Journal of Qualitative Methods*, 1(1), Art. 5,
http://www.ualberta.ca/~iiqm/backissues/1_1Final/html/sandeleng.html [Accessed: September 25, 2012].

Silverman, David & Marvasti, Amir (2008). *Doing qualitative research. A comprehensive guide*. London: Sage.

Yardley, Lucy (2000). Dilemmas in qualitative health research. *Psychology & Health*, 15(2), 215-228.

Yardley, Lucy (2008). Demonstrating validity in qualitative psychology. In Jonathan A. Smith (Ed.), *Qualitative psychology. A practical guide to research methods* (2nd ed., pp. 235-251). London: Sage.

Authors

Marie SANTIAGO DELEFOSSE has been a full professor in health psychology at the University of Lausanne (Switzerland) since 2003. She was a practicing general hospital psychologist for a decade. In 1992, she was appointed as a senior lecturer in the French university system, where she worked until 2003. Her empirical research (fertility, HIV, chronic pain, etc.) has led her to propose an alternative theorization of the biopsychosocial model. Her work pays particular attention to the history of ideas and to epistemology in psychology; it is mostly qualitative and based on a historico-cultural and phenomenological approach. She is the author of more than 100 articles and several books, including "Psychologie de la santé: Perspectives qualitatives et cliniques" [Health Psychology: Qualitative and Clinical Perspectives], 2002, Brussels: Mardaga, and "Méthodes qualitatives en psychologie" [Qualitative Methods in Psychology], ed. by SANTIAGO DELEFOSSE and ROUAN, 2001, Paris: Dunod.

Contact:

Marie Santiago Delefosse

Faculty of Social and Political Sciences
Institute of Psychology
Quartier UNIL-Mouline, Bâtiment Géopolis
CH-1015 Lausanne

E-mail: Marie.Santiago@unil.ch

URL: <http://www.unil.ch/cerpsa>

Christine BRUCHEZ has been a teaching assistant, then research coordinator, in health psychology at the University of Lausanne since January 2004. She holds a master's degree in arts (1995) and a master's degree in psychology (2002). She has also completed a DESS (Advanced Graduate Diploma) in clinical psychology (University of Geneva) and is currently a PhD candidate under the supervision of Prof. Marie SANTIAGO DELEFOSSE. Her DESS thesis (2005) has led to her collaboration on the validation of assessment frames in the field of health-related quality of life. Trained in computer-assisted text analysis (Alceste® software), she also received in-depth training in discourse analysis (modern French literature and linguistics). Her research interests in epistemology and methodology have led her towards the comparison of qualitative methods and the experimenting with new research tools in psychology, as well as the analysis of modes of communication on Internet health forums. She has also carried out a preliminary study on the dissemination of qualitative research in psychology on the Internet.

Contact:

Christine Bruchez

Faculty of Social and Political Sciences
Institute of Psychology
Quartier UNIL-Mouline, Bâtiment Géopolis
CH-1015 Lausanne

E-mail: Christine.Bruchez@unil.ch

URL: <http://www.unil.ch/cerpsa>

Amaelle GAVIN has been research coordinator in health psychology at the University of Lausanne since October 2012. She obtained her master's degree in psychology in 2013 and is currently training in clinical sexology at the University of Geneva. Since February 2015, she is a PhD candidate under the supervision of Prof. Marie SANTIAGO DELEFOSSE at the University of Lausanne. Her different trainings allowed her to develop skills and interests in various subjects and domains, such as sexology, couple counseling, ethics, critical health psychology, qualitative research among others.

Contact:

Faculty of Social and Political Sciences
Institute of Psychology
Quartier UNIL-Mouline, Bâtiment Géopolis
CH-1015 Lausanne

E-mail: Amaelle.Gavin@unil.ch

URL: <http://www.unil.ch/cerpsa>

Sarah STEPHEN worked as a Junior SNSF (Swiss National Science Foundation) researcher in an SNSF-funded project (2011-2014), directed by Professor SANTIAGO DELEFOSSE at the Faculty of Social and Political Sciences at the University of Lausanne. Since 2011, she is also a PhD student based at the Faculty of Business and Economics at the University of Lausanne. Her interests include quality of quantitative and qualitative research and interdisciplinary research methods, as well as paradigms underlying research.

Contact:

Sarah Stephen

Faculty of Business and Economics
Quartier UNIL-Mouline, Bâtiment Internef
CH-1015 Lausanne

E-mail: SarahLilian.Stephen@unil.ch

Citation

Santiago Delefosse, Marie; Bruchez, Christine; Gavin, Amaelle & Stephen, Sarah L. (2015). Diversity of the Quality Criteria in Qualitative Research in the Health Sciences: Lessons From a Lexicometric Analysis Composed of 133 Guidelines [40 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 16(2), Art. 11, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1502110>.