From data to data: an overview towards qualitative data research reproducibility

Federica Izzo - Alessandra Storlazzi

Abstract

Frame of the research: New methodological approaches and massive amounts of collectible data call for the definition of the state of the art of qualitative research in marketing and management studies.

Purpose of the paper: We provide an overview of qualitative research in the most recent literature in order to detect patterns and shared practices. Our results provide insight to researchers approaching qualitative data analysis.

Methodology: A total of 87 papers, published from 2017 to 2021 in 10 high-ranked international marketing and management journals were collected and analyzed. Information on the sub-components of our analysis was coded, summarized, and re-elaborated to better highlight the research findings.

Results: Following an ideal data flow, our study focuses on specific issues about the data types and sources that were used, the application of analysis techniques, and the sharing of data, thus re-interpreting them in terms of their specific importance to qualitative research.

Research limitations: A major limitation lies in the fact that papers from some top journals were not reviewed. Moreover, some main topics in qualitative research, e.g. research questions, methodologies, and procedures, were deliberately overlooked, as we hope to analyze them in further studies.

Practical implications: We offer input to scholars as we introduce some useful automated and online tools for data collection, analysis, and sharing in qualitative research. The information we produced is of the utmost usefulness for researchers who want to open their study perspective by using less investigated data.

Originality of the study: We raise some challenging questions on the possibility of a synthetic and parsimonious approach to qualitative research and leave scholars with an open question on the evolution of qualitative studies in future research.

Key words: qualitative data analysis; data availability; data analysis techniques

1. Introduction

The ongoing discussion on the methodologies applied in qualitative research, and on its rigor, validity, and generalization, can be examined from a data-driven perspective. By definition, quantitative analysis incorporates a higher level of accuracy because it deals with numerical data and consolidated, shared analysis techniques. In some ways, the quantitative research method is contained within the analysis and the researcher’s subjectivity is restricted, but not eliminated, by data. Conversely, qualitative data analysis, due to its undefinable nature, is still suffering from a lack of
shared methodologies and community-accepted practices on the surface (Corbetta 2014). The massive availability of data of diverse and different nature, type, and format has led the scientific community toward more complex research approaches involving work on qualitative along with, or as an alternative to, quantitative data.

One of the most widely accepted definitions of qualitative research is provided by Wolcott (1994), who identifies it through the three steps of data description, analysis, and interpretation. Wolcott’s research phases can be reinterpreted along a data work continuum, where analysis is a step toward total adherence to the information provided by data and is preparatory for the formulation of a construct or final theory based on data. Nevertheless, it is challenging to give a general definition of qualitative analysis, as this type of work is strongly bound to its research context, as well as to the data. For this reason, Gibson and Brown (2009) prefer to define contextualized analysis as a qualitative research procedure that relates data with conceptual problems in the context of social science and that can trigger different methodological approaches. “The focus on data type is largely a distraction from the more important distinction of inductive versus deductive forms of inquiry (Eisenhardt and Graebner, 2007)” (Nowell and Albrecht, 2019, p. 4).

Qualitative research lays its foundation on three modes of logical reasoning: deduction, which is the formulation of hypotheses that are later tested on data; induction, which is the building of a general theory from particular observations; abduction, which is reasoning toward meaning that explains exceptions in the data with plausible explanations (Given, 2008). The distinction among modes of reasoning in qualitative research is not definitive, as induction may result from a previous deductive approach, while abduction uses both deductive and inductive methods to build and demonstrate theories (Saetre and Van De Ven, 2021). Grounded Theory (GT) initially asserts that theory emerges from data, and considers the researcher a reflexive scientific observer (Glaser et al., 1967). Constructivist grounded theory assumes that the researcher participates in generating data and theories (Timonen et al., 2018). The most recent developments in GT include the alignment of the GT method, which involves a combination of induction, deduction, and abduction in order to gain conceptual clarity about phenomena (Reichert, 2010; Timonen et al., 2018). Nevertheless, some scholars agree that “abduction, rather than induction, should be the guiding principle of empirically based theory construction”, (Timmermans, 2012, p. 167). Abduction is a generative process of new ideas based on the researcher’s hunches and creativity (Saetre and Van De Ven, 2021), where theory neither presides over nor derives from data; rather, research hypotheses and evidence go hand in hand with data to develop a new theory or to expand an existing one (Conaty, 2021). The logical nature of abductive reasoning answers the needs of management scholars against the variability and anomalies typical of a rapidly changing social and organizational context (Saetre and Van De Ven, 2021). Moreover, this approach has proven to be more suitable than others in facing the challenges of the qualitative analysis of multiple and numerous data produced within the infosphere.
The constant connectivity that characterizes our lives allows a continuous interaction of virtual and physical environments (Bresciani et al., 2021) and the production of onlife information, narratives, contents, and communities (Floridi, 2017) that participate in the delineation of so-called Big data. A classical, yet not conclusive, definition of big data is “the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time”, (Chen et al., 2014, p. 173). This data is described through the 4Vs model, whose elements are: the great Volume of available information; the Variety of data types and formats; the Velocity in data generation, acquisition, and management; the Value extracted from the information collected (Chen et al., 2014; AGCOM, 2020). Volume and variety of accessible data present the researchers with “questions [...] about the accessibility, ethics, and utility of big data”, (Mills, 2018). This entails the decision on the quality and scale of information that is needed to understand a phenomenon of interest or to demonstrate a theoretical construct, namely which data has value within the research context. Indeed, “each type of data and how it was collected has different strengths and weaknesses in relation to the research questions and analysis techniques” (Given, 2008). On the other hand, the richness of information can lead to new and comprehensive theoretical developments if researchers are willing to open to the complexity of data and employ their creativity in the data analysis. Moreover, variety is an added value not only as an input but also as a research output, as it carries a perspective in the information that is beyond neutrality. Within an abductive perspective, data availability can trigger a virtuous circle where more information generates more theories, together with new data that can provide hunches and ideas for further research in marketing and management. This scenario, where new methodological approaches and massive amounts of collectible data meet, calls for the definition of the state of the art of marketing and management studies.

This work places itself within the ongoing discussion on qualitative research as a convincing option to quantitative research in business, management, and accounting. Data availability and technical progress provide substantive challenges and opportunities in marketing and management studies (Grewal et al., 2021; Dźwigoł, 2019 and 2020). We provide an overview of qualitative data research through the most recent literature in order to detect patterns and shared practices from a data-driven perspective. Our results provide insight to researchers approaching qualitative data analysis. Starting from the 87 papers that were collected and analyzed, we report data collection methodologies (data source; the number of samples collected; collection time span), the use of dictionaries (dictionary type; name for standardized and previously used dictionaries; validation method), software applied to the analysis of data, and availability and storage of collected data (see Paper supplementary materials). The information we produced is of the utmost usefulness for researchers who want to open their study perspective by using less investigated data.

Findings show an intensive use of data, which is mainly textual and analyzed through dictionaries, together with a trend toward data sharing practices in a research reproducibility logic. We collected papers published
in a 5-year time span from high-ranked journals in marketing and management. Following an ideal data flow, our study focuses on specific issues pertaining to the data types and sources that were used, the application of data analysis techniques such as dictionaries and Computer Assisted Qualitative Data Analysis Software (CAQDAS), the sharing of acquired or resulting data, and re-interprets these in terms of specific importance to research. We also present some automated and online instruments that can support qualitative data research. The rest of the paper is organized as follows: section 2 is about tools and methodologies; section 3 describes the topics of the study in the considered literature; theoretical contributions, limitations, and further studies are presented in section 4.

2. Methodology

In order to provide an overview of qualitative research through recent literature, papers were selected from the highest-ranked international journals in 2020 according to the Scimago Journal & Country Rank\(^1\). A total of 10 journals were shortlisted in the subject area Business, Management, and Accounting and in the Marketing category. The selected ordering criterion was the SCImago Journal Rank (SJR) indicator, described on the Scimago web portal as “a measure of journal's impact, influence or prestige. It expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years”\(^2\) and developed by Guerrero-Botea and Moya-Anegónb (2012). The list was last consulted on 21 October 2021. This is the full list of selected journals, ordered according to the Scimago Journal & Country Rank:

1. Journal of Consumer Research - Oxford University Press
2. Journal of Marketing - American Marketing Association
3. Journal of Marketing Research - American Marketing Association
4. Marketing Science - INFORMS
5. Journal of the Academy of Marketing Science - Springer
10. Journal of World Business - Elsevier

A time span of five years (2017 - 2021) was decided on a priori to focus our efforts on the latest qualitative studies and to limit the amount of research material to analyze. Papers from the selected timeframe were searched in the Google Scholar database by using the advanced research function to find the string “qualitative analysis” in the titles, abstracts, and keywords of each selected journal. In some cases, the publisher was also added to strengthen the correspondence of the results. However, a visual double-check the collocation of a paper into the journal of interest

\(^1\) https://www.scimagojr.com/
\(^2\) https://www.scimagojr.com/journalrank.php?category=1406&area=1400
was also needed. We considered all papers regardless of the nationality of the studies and authors’ affiliation. The Google Scholar database was last consulted on 7 November 2021. A total of 87 papers were collected and the full list of references per journal is displayed in Table 1.

Tab. 1: Full list of collected papers

<table>
<thead>
<tr>
<th>Journal</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Journal of Research in Marketing</td>
<td>Gerrath and Usrey (2021); Deng (2021)</td>
</tr>
<tr>
<td>Journal of Consumer Psychology</td>
<td>Consiglio et al. (2018); Mukhopadhyay et al. (2020); Teeny et al. (2020)</td>
</tr>
<tr>
<td>Journal of Consumer Research</td>
<td>Humphreys and Wang (2018); Netemeyer et al. (2018); Liu et al. (2019); Bellezza and Berger (2020); Coskun-Balli (2020); Melumad and Pham (2020); Borghini et al. (2021); Dinnin et al. (2021); Mimoun (2021); Wieser et al. (2021)</td>
</tr>
<tr>
<td>Journal of Marketing</td>
<td>Nam (2017); Johnson and Matthes (2018); Colm et al. (2019); Molner et al. (2019); Chapman (2020); Singh et al. (2020)</td>
</tr>
<tr>
<td>Journal of Marketing Research</td>
<td>Van Den Bulte et al. (2018); Fournier and Eckhardt (2019); Zyung et al. (2020)</td>
</tr>
<tr>
<td>Journal of Public Administration Research and Theory</td>
<td>Destler (2017); Ulibarri and Scott (2017); Nowell and Albrecht (2018); Whetsell et al. (2020); Lawee (2021); Buntaine et al. (2021); Barnes (2021); Woodhouse et al. (2021)</td>
</tr>
<tr>
<td>Journal of Supply Chain Management</td>
<td>Brito and Miguel (2017); Stolze et al. (2018); Hardy et al. (2020); Marques et al. (2020); Carter et al. (2021); De Vries et al. (2021); Krause and Pullman (2021)</td>
</tr>
<tr>
<td>Journal of the Academy of Marketing Science</td>
<td>Guo et al. (2017); Homburg et al. (2017); Malshe and Friend (2018); Zhang et al. (2018); Hamilton and Price (2019); Arunachalam et al. (2020); Gupta (2020); Nenonen et al. (2020); Pedana (2020); Ahearn et al. (2021); Keller et al. (2021); Peterson et al. (2021)</td>
</tr>
<tr>
<td>Journal of World Business</td>
<td>Borda et al. (2017); Mbalyohere et al. (2017); Peltochori and Yamao (2017); Buchanan and Marques (2018); Gaur and Kumar (2018); He et al. (2018); Khan et al. (2018); Mullner and Puck (2018); Nielsen and Raswant (2018); Santangelo (2018); Teagarden et al. (2018); Tupper et al. (2018); Yakovleva and Vazquez-Brust (2018); Yang and Sonmez (2018); Zeng et al. (2018); Gamso and Nelson (2019); Lunnan and McGaughey (2019); Makela et al. (2019); Dau et al. (2020); De Beule et al. (2020); McWilliam et al. (2020); Outila et al. (2020); Xing et al. (2020); Barnard and Mamabolo (2021); Buciuni and Pisano (2021); Chakravarty et al. (2021); Elo et al. (2021); Kafouros et al. (2021); Knoerich and Vitting (2021); Martin et al. (2021); Nyamrund and Freeman (2021); Pasquali (2021); Tolstoy et al. (2021); Zao et al. (2021)</td>
</tr>
<tr>
<td>Marketing Science</td>
<td>Liu et al. (2018); Timoshenko et al. (2019)</td>
</tr>
</tbody>
</table>

Source: our elaboration

Numerical data allows the detection of publication patterns among the journals. Fig.1 shows an evident discrepancy within our sample on the percentage of papers that deal with qualitative research in each journal. Most of the qualitative studies were found in Journal of World Business

![Fig. 1: Papers on QA research per journal](source: our elaboration)

Anyway, the overall number of published papers per year demonstrates a seesawing yet growing interest in, and application of, qualitative analysis in consumer research, with 27 papers in 2021, 20 papers in 2020, 9 papers in 2019, 22 papers in 2018, and 9 papers in 2017. The pattern of qualitative research papers per year is shown in Fig. 2.

![Fig. 2: Papers on QA research per year](source: our elaboration)

After an exploratory inspection of our research material, we formulated the coding scheme displayed in the Paper supplementary materials. Information on the type and quantity of data that was gathered with the data sources and collection time spans, along with the dictionaries and
automated qualitative analysis software that were used and the acquired or resulting data that was made available for future research was coded in Google Sheets, summarized, and re-elaborated to better highlight research findings. The coding phase consisted of three steps: in the first step, all the journal papers were collected and numerical data, together with abstracts and keywords, were stored; in the second step, each paper was read and analyzed individually to detect our topics of interest; finally, all data was thematically aggregated to show differences, similarities, and patterns in the literature. The whole process was carried out manually by the two authors.

3. Findings and discussion

The purpose of this review is to map qualitative data analysis in marketing and management research from a data-driven perspective. The acquired literature is extensive enough to provide a thorough overview of the various sub-components that constitute the focus of our analysis. Our research path follows an ideal flow of the data from its acquisition, to its analysis techniques, and finally to a resulting sharable output that can provide material for further research. Along this itinerary, we also point out some useful automated and online tools to enhance qualitative research. The introduced research path is mapped in Fig. 3.

### 3.1 Data type and source

The review of recent literature disclosed a highly articulated scenario: the majority of the papers leveraged variate data sources and a multilevel data collection methodology. Against the general criticism of a lack of rigor in qualitative analysis, the data collection methodologies are accurately...
described in their variety of data, sources, and the acquisition procedure.

The concept of qualitative data and its comparison to quantitative data is actually opaque. The definition of qualitative data as all the information that cannot be numerically coded and statistically analyzed is not always verified. Indeed, the two kinds of data are not mutually exclusive in the research, as “it is common for quantitative research to produce some qualitative data [...] and for qualitative research to generate data that can be described numerically and analysed statistically” (Gibson and Brown, 2009, pp. 8-9). Moreover, nonnumerical data can be analyzed statistically, if it is re-coded in a numerical scale (e.g. a Likert scale to rate verbal statements) (Given, 2008). In addition to traditional offline sources of qualitative data such as surveys/interviews, letters, reports, diaries, field notes for verbal data and pictures, maps, print advertisements, diagrams, art, or films for nonverbal data, sources from the digital world, in the form of content creation and sharing platforms, forums for consumers, and social media, need to be considered. In this scenario, text-based data is a central object in social and consumer research. Moreover “researchers can combine the analysis of big data patterns with interviews, focus groups, and ethnographic observations of online users to make the connections between large data trends, and rich complementary data from individual users or cases” (Mills, 2017, p. 598).

As confirmed in the literature and consulted manuals, most qualitative research works with textual data. The importance of text is underlined in Humphrey’s work, where a roadmap for automated text analysis is proposed along with a focus on research questions that lend to text analysis (Humphreys and Wang, 2018). Except in cases when data retrieval is not considered (Humphreys and Wang, 2018, Mukhopadhyay et al., 2020), and in a single case in which text and numerical data are derived from in-store video recordings (Zhang et al., 2018), the totality of our sample of papers gather text as the main type of data. In some cases, text and numerical data are combined in mixed analysis methodologies. The origin of numerical data can be either from the text itself, using word count, occurrences, and other techniques (Peltokorpi and Yamao, 2017; Nielsen and Raswant, 2018; Zeng et al., 2018) or interviews with a Likert scale response (Guo et al., 2017; Zyung et al., 2020; De Vries et al., 2021; Martin et al., 2021), or from unconnected data sources, such as institutional databases, firms’ internal reports, and statistics (e.g. Borda et al., 2017; Martin et al., 2021; Santangelo, 2018; Tupper et al., 2018; Gamso and Nelson, 2019; Pasquali, 2021; Zao, 2021). With the exception of literature reviews, editorials, commentaries, and methodological overviews (e.g. Zao, 2021; Chakravarty et al., 2021; Elo, 2021; Kafouros et al., 2021; Dau et al., 2020; Gaur and Kumar, 2018; Nielsen and Raswant, 2018; Zeng et al., 2018; Teagarden et al., 2018; Hamilton and Price, 2019; Ahearne et al., 2021; Chapman, 2020; Hardy et al., 2020; Carter et al., 2021; Humphreys and Wang, 2018; Mukhopadhyay et al., 2020), interviews, questionnaires, and surveys are the most used methodologies to collect data (62% of our sample). This primary source of data is integrated with different data sources, which are either secondary or on the same level, to collect more text data (e.g. Knoerich and Vitting, 2021; Buciuni and Pisano, 2021; Nymrund and Freeman, 2021; Xing et
al., 2020; Lunnan and McGaughey, 2019; Khan et al., 2018; Yakovleva and Vazquez-Brust, 2018; Mullner and Puck, 2018; Mbalyohere et al., 2017; Peterson et al., 2021; Colm et al., 2020; Molner et al., 2019; Krause and Pullman, 2021; Borghini et al., 2021; Wieser et al., 2021) or numerical data (Martin et al., 2021; Zao, 2021; Santangelo, 2018; Tupper et al., 2018; Borda et al., 2017; Singh et al., 2020; Nam et al., 2017). Other textual data sources include, but are not limited to, institutional, business, or internal databases, archival data, media (tv, radio, books, websites, newspapers), and online (social media, online platforms) sources. A pattern of the data sources that were used in our sample is mapped in Fig. 4.

Fig. 4: Data sources in qualitative research
Source: our elaboration

In some longitudinal studies, archival data may cover a very long time span, e.g. in McWilliam et al. (2020), the collected data covered a 32-year period of time, 31 years in Coskuner-Balli (2020), 30 years in Yang and Sonmez (2018) and Mimoun (2021), 27 years in Zeng et al. (2018), just to cite the most impressive. The data recollection process itself can take more rounds or longer periods to be completed or to fulfill research requirements. This especially happens in experimental studies, e.g. for the experiment in Bellezza and Berger (2020), participants were recruited from 2015 to 2019. As far as experiments are concerned, information is only partially acquired from an existing source, while the primary data emerges during experimental studies (Gupta, 2020; Melumad and Pham, 2020; Van Den Bulte et al., 2018; Consiglio et al., 2018; Teeny et al., 2020). In some studies, and mainly those published in the Journal of Consumer Research, some automated instruments and online platforms, such as Amazon MTurk (Melumad and Pham, 2020; Coskuner-Balli, 2020), Qualtrics or Delicious (Nam, 2017; Borghini et al., 2021; Deng, 2021; Woodhouse et al., 2021), and web scraping techniques (Borghini et al., 2021), were used for data recollection. In Bellezza and Berger's phenomenon-based research on the relationships between low and high status (2020), all the previously mentioned instruments were applied to data recollection.

This overview of data collection procedures in the recent literature, far from being exhaustive, serves as an example of how the scientific...
community applies different strategies to face the challenges of data volume and variety. Indeed, this trend proves the willingness of researchers to reinforce qualitative analysis studies with more data and to generate more comprehensive theory through the aggregation of data from different sources. Despite their careful attention and detailed description of data collection methodologies, lengthy and complicated retrieval processes make research reproducibility difficult to obtain. The delineation of shared practices and the use of automated tools could partially reduce such limitations, notwithstanding the highly contextualized nature of qualitative research.

3.2 Research techniques

Following our ideal data flow, this overview focuses on two techniques in qualitative research, namely the categorization (or coding) phase, with the dictionaries adopted in this process, and the use of CAQDAS for textual analysis. We also provide a brief description of some interesting tools for data analysis that were found in the considered literature.

As pointed out by Grodal et al. (2021), the coding phase is essential for qualitative analysis and preliminary to theoretical and empirical studies applying both top-down and bottom-up approaches. Coding data into categories helps researchers make sense of the variety of retrieved information. Such a process informs theory testing and building, thus organizing data into a certain level of parsimony and consensuality (Corley and Gioia, 2011). This can be seen in the most recent literature in marketing and management as, regardless of the logical reasoning applied, almost the entire selected sample (75.8%) claims to use categorization at least as a first-level analysis. Some exceptions where categorization is not applied, or not declared, are represented by literature reviews (Teagarden et al., 2018; Hamilton and Price, 2019; Ahearne et al., 2021; Carter et al., 2021), methodological overviews (Humphreys and Wang, 2018; Hardy et al., 2020; Borghini et al., 2021; Nowell and Albrecht, 2018), one editorial (Mukhopadhyay et al., 2020), and some theory building (De Vries et al., 2021; Liu et al., 2019 and 2018) and experimental studies (Bellezza and Berger, 2020; Van Den Bulte et al., 2018; Buntaine et al., 2021; Woodhouse et al., 2021). As information is grouped in categories, categories are likewise organized into dictionaries. Fig. 5 shows the distribution of the dictionaries that were applied to categorization in our sample and divided into standardized dictionaries, dictionaries used in previous studies,
dictionaries that were expressly created for an individual search, and mixed approaches with multiple levels of categorization.

Scholars emphasize the importance of carefully building or selecting a reference dictionary. Dictionaries that are created by authors contain categories emerging from data, making them more compliant towards the context, methodology, and purpose of the study, while “generic dictionaries may perform poorly in a given domain” (Chapman, 2020, p. 84). Dictionaries that are retrieved from previous work can contribute to placing a study within a specific branch of research acknowledged by the scientific community. In both cases, to guarantee methodological rigor and scientific validity, the process leading to the choice, merging and removal of categories should be, and usually is, accurately described (Grodal et al., 2021) and validated (Humphreys and Wang, 2018). An alternative or concurrent approach to obtain methodological rigor consists in using a standardized list of categories. Against Chapman’s reluctance (Chapman, 2020), standardized dictionaries cover different domains and, in general, it is advisable to use an available, tested, and scientifically approved dictionary when it fits the dataset and the goals of the research. Nevertheless, a standardized dictionary was used only in 9.2% of our sample.

From the substantial list of standardized dictionaries provided in Humphreys and Wang (2018), only the Linguistic Inquiry and Word Count (LIWC) (Pennebaker, 2015) dictionary was used in our sample (Coskuner-Balli, 2020; Singh et al., 2020; Dinnin et al., 2021; Mimoun, 2021). LIWC is a general dictionary that includes approximately 90 categories (divided into summary language variables, general descriptor categories, standard linguistic dimensions, word categories tapping into psychological constructs, personal concern categories, informal language markers, and punctuation categories) and 6,400 English words (it is also available in languages other than English). It can be easily applied in the analysis of massive textual data through a software that is commercially distributed on the LIWC website³.

³ https://liwc.wpengine.com/
In spite of the difficulties in collecting, managing, analyzing, and interpreting growing amounts of data, the use of automated tools in qualitative research studies is still relatively scarce. Only 22.7% of the analyzed literature leveraged a CAQDAS, even though this kind of automatic tool, in most cases, was specifically developed for human scientists and does not need programming skills. In the section of the New York University Libraries’ website that is dedicated to Qualitative Data Analysis, a list of, and comparison between, CAQDAS is provided\(^4\). Within this 7-item list, NVivo was by far the most used in our sample (used in 12 papers), followed by QDAMiner, Atlas.ti, Dedoose and Leximancer. The usage of automated analysis tools is summarized in Table 2.

<table>
<thead>
<tr>
<th>Software</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>QDAMiner</strong></td>
<td>Buchanan and Marques (2018), Stolze et al. (2018), Arunachalam et al. (2020)</td>
</tr>
<tr>
<td><strong>Single-used CAQDAS</strong></td>
<td>Bellezza and Berger (2020), Coskuner-Balli (2020)</td>
</tr>
<tr>
<td><strong>Other text analysis tools or ML</strong></td>
<td>Nam (2017), Peltokorpi and Yamao (2017), Chapman (2020), Melumad and Pham (2020), Singh et al. (2020), Martin et al. (2021)</td>
</tr>
<tr>
<td><strong>SEM and statistics</strong></td>
<td>Peltokorpi and Yamao (2017), Singh et al. (2020), Whetsell et al. (2020), Deng (2021), Woodhouse et al. (2021), Martin et al. (2021)</td>
</tr>
</tbody>
</table>

Source: our elaboration

---

\(^4\) https://guides.nyu.edu/QDA/qual

\(^5\) https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home/
decisional position while fully exploiting the power of computation: “As NVivo allows researchers to work with a wide variety of data, and adopt whatever methodology best suits their research question, it is advocated that it is the researcher who determines the results found, rather than the software used” (Wiltshier, 2011). Being particularly suitable to the open, axial, and selective coding steps, the use of NVivo is often associated with the Grounded Theory methodology, as found in Mbalyohere et al. (2017), Johnson and Matthes (2018), Nenonen et al. (2020), Krause and Pullman (2021).

Beyond analysis, within a small part of our sample that leveraged a digital instrument to collect data and manage research projects, the before mentioned Mechanical Turk (MTurk) by Amazon is the most commonly used tool (Netemeyer et al., 2019; Dinnin et al., 2021; Humphreys and Wang, 2018; Consiglio et al., 2018; Gerrath and Usrey, 2021; Borghini et al., 2021; Timoshenko et al., 2019). It is a marketplace platform where researchers can hire temporary workers to virtually conduct jobs of different entities, as well as involve sample participants who are willing to answer digital surveys in exchange for a small payment. The advantages of this instrument lay in the possibilities it offers for the management of project tasks and roles, remote work, and data retrieval process in a centered virtual environment.

Over the years, technology has been contributing to effectively empower research potential. Indeed, the ones described, as well as other similar tools, can effectively help to keep track of the research path and data, build unexpected theoretical constructs, boost the reproducibility of a scientific work, and, finally, fix some methodological standards for future research. Automated and online research instruments are slowly spreading in marketing and management studies, but awareness of their existence and potential is increasing, probably encouraged by the challenges of skyrocketing data availability.

3.3 Open data

The last step of our overview considered the life cycle of data after it has been analyzed in qualitative research. Data is a valuable asset whose informative potential should not end within an individual research project, but rather be shared and reused in an open way. In this perspective, 25 papers (28.4%) from our sample partially or fully share their collected and/ or analyzed data. This percentage, although promising, is still far from being satisfactory. In most of the studies (18 out of 25), data was shared as an appendix to the paper. In the same way, McWilliam et al. (2020) share the coding scheme that had been developed in the paper, rather than their own data. Barnes (2021), instead, claims to be willing to share data upon request, as the private nature of some of the information prevents it from being publicly disclosed. A summary of the data sharing options that were adopted is provided in Table 3.
In this study, we highlight the use of online platforms for the storage and sharing of data deriving from scientific research, as we consider these systems to actively contribute to the common good of knowledge.

The Harvard Dataverse⁷ (used in Buntaine et al., 2021 and Woodhouse et al., 2021, both published in the Journal of Public Administration Research and Theory) is a free repository where researchers who are connected or unrelated to the community of the University of Harvard, can archive and share their own data or explore and customize available datasets uploaded by other scholars. A DOI reference identifies all published data in order to obtain academic credit through citations. To date, it collects over 129,700 datasets for 14 subjects (Social Sciences, Arts and Humanities, Earth and Environmental Sciences, Medicine, Health and Life Sciences, Law, Agricultural Sciences, Computer and Information Science, Physics, Engineering, Astronomy and Astrophysics, Business and Management, Chemistry, Mathematical Sciences, Other). In the Business and Management section, 1,050 datasets have been freely available for download since 2007.

The Open Science Framework⁸ (used in Melumad and Pham, 2020 and Mimoun, 2021, both published in the Journal of Consumer Research) is a full-service platform for research projects management. It offers the necessary tools to coordinate collaborators and jobs, track work progresses, and upload, store, and share all the paper preprints, datasets, and all files that are linked to a study. Data is stored within a project folder, so a researcher needs to search for an author’s paper or project name to freely access and download it. Every file that is uploaded in the project folder is identified with a persistent URL for citing and sharing.

In some fields, access to data is still burdensome, time-consuming, and costly for the scholarly community. Thanks to the availability of information, the constant need to engage with the reality of firms and customers, and the increase in methodological and technical awareness, it is our opinion that marketing and management research is developing into what Sawyer would refer to as a data-rich field (2008). Without losing ourselves in the depths of big data, we can certify that if all the studies we

---

⁷ https://dataverse.harvard.edu/
⁸ https://osf.io/
encountered in our review had shared their relatively small, acquired and/or resulting datasets in the same place, we would have had a remarkable dataset that would have been available at no further cost, in accordance with the philosophy of the Open Data Movement (Kitchin, 2014). Indeed, open access to research data could become a requirement in some research realities in the future, just as it is now in the field of public administration.

Reusable data does not come at no cost: policies on data ethics, format definition, and the validation of datasets are needed in order to assure that their use outside the original research context remains meaningful and free of bias (Mills, 2017). Although more data does not necessarily entail better data, retrieving, aggregating, and customizing used datasets can generate new, unexpected value.

To the best of our knowledge, there is no institution in the Italian marketing and management research community that provides storage power and guarantees fairness in data management and sharing. While waiting for an agreement in relation to open data within the scientific community, our contribution to this specific matter can be found in the see Paper supplementary materials, where all the data that was collected for our study is organized and reported.

4. Conclusions and further research

In the present study, we followed a data-driven perspective to provide an overview of data types and sources, analysis techniques, and data sharing in qualitative research. Against the general perception of a lack of rigor, the observations on the recent literature prove that there has been a great evolution in qualitative research in terms of data collection methodologies and scientific approach.

Findings show that almost all of the analyzed papers dealt with textual data. The accurate description of data collection methodologies that was found in most papers demonstrates an attempt towards scientific rigor and, in some cases, motivates the need for fixed procedures. As previously discussed and illustrated in Fig. 5, some shared procedures, such as the use of standardized and previously validated dictionaries and the application of software for data analysis, have already been applied in text analysis. Nevertheless, research reproducibility still seems to be a distant goal, due to the highly contextualized character of qualitative research.

We raised some challenging questions concerning the possibility of a synthetic and parsimonious approach to qualitative research where volume, variety, velocity, and value of data can be channeled to build original and compelling theories. The use of automated and online tools can encourage reproducibility. For this reason, we offered scholars helpful advice by introducing some useful instruments for qualitative research. A wider use of automated and online tools in data collection, analysis, and sharing can empower research potential and boost the consolidation of sharable and replicable practices.

This does not imply renouncing the human factor, which is an added value to qualitative research, but rather finding a balance between the
The suggested strategies are sourced from publications in high-ranked international journals in business, management, and accounting, and this is useful knowledge for researchers who want to approach qualitative analysis. Finally, we detected the urge to plan a data life cycle through the promotion of instruments and practices for sharing and reusing data at a national scientific community level. This matter definitely becomes significant in the current research context, which is promoted and supported, for example, by scientific associations such as the Italian Society of Management (SIMA).

A major limitation of the present work is that papers from some top journals were not reviewed. More time and a larger research team would allow us to include other journals in the fields of business, management, and accounting, both by selecting different parameters (e.g. the H index) on Scimago, or by referring to other databases (Jstor, Web of Science). A different selection of business, management, and accounting journals, based on their topics of interests and sectors (e.g. retailing, services), would also provide some interesting insight on the use of a qualitative approach.

Moreover, certain main topics in qualitative research were deliberately overlooked, as we hope to analyze them in the future. Indeed, studies on this subject could involve an exploration of research questions, analysis methodologies, and procedures, perhaps by extending our list of journals. Further research could also focus on journals that explicitly ask for contributions on qualitative research in their aims and scope, as in the case of the Journal of World Business that, as reported in the methodology section, published the most papers on qualitative research over the last 5 years. Furthermore, by extending the list of sourced journals, qualitative studies could be selected and aggregated according to the authors’ affiliation and nationality in order to map qualitative research also from a geographical perspective.

More complete documentation can contribute to accurately outlining qualitative studies and opening a discussion on data, its characteristics, and its use in marketing and management research. In fact, “Data has long been the fuel that has powered academic marketing research” (Grewal et al., 2021, p. 1028). Previously, quantitative analysis was the preferred method in marketing and management studies, and numerical data was collected and analyzed to prove a priori theoretical constructs. The paradigm is currently shifting from theory to data: the analysis of data can deepen the meaning and usefulness of available information that is either gathered as individual data or aggregated, as in the case of big data. Volume, variety, velocity, and value, along with the properties conveyed by the modality of data (Grewal et al., 2021) seem to stem from a parsimonious and rigorous approach and to easily open up to creativity, which is the moving power of certain qualitative research methodologies like abduction (Saetre and Van De Ven, 2021).

In light of the above discussion, our opinion is that qualitative research is currently experiencing a dichotomy: on the one hand, the multipurpose nature of qualitative studies and the very nature and modality of data encourage the creative process of researchers’ work. On the other hand,
the methodologies and techniques described in this study demonstrate the need to promote a scientific approach towards qualitative research through rigor, parsimony, and reproducibility. This leaves scholars with an open question on which of these two trends will prevail in future research.

References


LAVEE E. (2021), “Who is in charge? The provision of informal personal resources at the street level”, *Journal of Public Administration Research and Theory*, vol. 31, n. 1, pp. 4-20


PENNEBAKER J.W. (2015), *The development and psychometric properties of LIWC2015*, University of Texas at Austin, Austin, TX.


Websites


Amazon mechanical Turk, https://www.mturk.com/

Harvard Dataverse, https://dataverse.harvard.edu

Linguistic Inquiry and Word Count (LIWC), https://liwc.wpengine.com/


NYU Guide to Qualitative Data Analysis, https://guides.nyu.edu/QDA/qual

Open Science Framework, https://osf.io/


Academic or professional position and contacts

Federica Izzo
PhD student of Management
University “Suor Orsola Benincasa”, Naples, Italy
e-mail: federica.izzo@studenti.unisob.na.it

Alessandra Storlazzi
Full professor of Management
University “Suor Orsola Benincasa”, Naples, Italy
e-mail: alessandra.storlazzi@docenti.unisob.na.it