PURPOSIVE SELECTION AND THE QUALITY OF QUALITATIVE IS RESEARCH

Completed Research Paper

Attila Marton
Department of IT Management
Copenhagen Business School
Howitzvej 60, 4
2200 Frederiksberg, Denmark
am.itm@cbs.dk

Abstract

As qualitative research has found broad acceptance within the IS community, the methodological discourse has turned its attention to questions concerning the quality of qualitative research mostly emphasizing the development of how-to guidelines for good practice. By contrast, criteria for the evaluation of qualitative research, although equally important, have not received equal attention. Drawing on literature from behavioral and social science, this paper discusses methodological concepts of evaluating qualitative research by focusing on techniques to purposefully select data for analysis. In particular, the technique of corpus construction will be introduced, which was specifically designed as an evaluation criterion for qualitative research. Adapted from linguistics, corpus construction offers an alternative that is functionally equivalent to statistical sampling techniques in terms of the quality of empirical research. Hence, the paper contributes the concept of purposive selection as an evaluation criterion to the methodological tool-box and terminology of IS research.

Keywords: qualitative research methodology, quality criteria, purposive selection, purposive sampling, corpus construction
Introduction

Qualitative research has become an accepted research approach for the study of information systems for quite some time now. As a consequence, the methodological discourse has shifted from the need to justify and defend the scientific rigor of qualitative research towards a focus on quality. An equivalent shift can be found in the behavioral and social sciences in general that has brought questions concerning the conduct of proper qualitative research or good practice to the forefront. Since the 1980s, the behavioral and social sciences have extended and complemented the notion of good practice by questions concerning criteria for the demonstration and evaluation of the quality of qualitative research in order to enable readers of research proposals and reports, such as funding bodies, reviewers or fellow academics, to make an informed assessment. In IS research, evaluation criteria are mainly addressed by scholars exclusively in terms of specific research traditions and the topic has yet to receive due attention in terms of the broader methodological discourse. This paper is an attempt to raise awareness on the central role of evaluation criteria in qualitative research in general and to trigger further discussion in the IS community. Ultimately, the aim is to introduce a more balanced discourse on the quality of qualitative research that takes good-practice as well as evaluation into account as two sides of the same coin.

In particular, the paper takes a core quality criterion as an example for addressing notions of good practice as well as evaluation – selecting data for analysis. Be it statistical sampling in terms of quantitative or, as it will be referred to later in this paper, purposive selection in terms of qualitative research, the structure or rationale according to which units of data are collected is the foundation of any empirical study as it connects methodological issues of quality with practical issues of conducting fieldwork. Against this backdrop, the paper focuses on the technique of corpus construction as an example for conducting as well as demonstrating high quality qualitative research. As corpus construction is hardly known in the IS community, the paper also seizes the opportunity to introduce the technique as an alternative qualitative selection rationale and provides a detailed example from the author's research in information systems and information management.

Corpus construction will be compared with statistical sampling and theoretical sampling in order to elaborate on the central role of selecting data for analysis as a quality criterion. Since corpus construction enables researchers to convey a structured plan on how data is going to be selected, the technique is expected to be of particular interest to novice qualitative researchers who struggle with convincing stakeholders and funding bodies of the rigor and feasibility of their research proposals (Conboy et al. 2012). By the same token, corpus construction will also be employed as an illustration for a specific approach towards the quality of qualitative research that offers an alternative to established approaches, which derive from the controversy and divide between quantitative and qualitative research. Referred to as functionally equivalent criteria by Gaskell and Bauer (2000), this alternative approach seeks to develop criteria that are different from but functionally equivalent to the established criteria of quantitative research. Thus conceived, the argument put forth is meant to help bridge the divide between quantitative and qualitative research.

The paper is structured as follows. The first section gives a brief overview over the general methodological discussion on the quality of qualitative research in the behavioral and social sciences as well as IS research by identifying three different schools of thought: 1) conventionalism, 2) counter-conventionalism and 3) functional equivalence. Since the first two are well known in the IS community, the main focus will be on the third approach, which serves as the foundation for the discussion of corpus construction. After a closer look at the data selection process as one specific quality criterion in section three, the paper explains the technique of corpus construction as a selection rationale and elaborates on the reasons why it is very different from but functionally equivalent to statistical sampling in terms of demonstrating the quality of a research project. The paper concludes with a discussion of the relationship between specific purposive selection strategies and the three approaches towards the quality of qualitative research. Finally, the argument is put forth to complement the canon of good practices established in IS research by adding evaluation criteria according to the approach of functional equivalence.
The Quality of Qualitative Research

The quality of empirical research entails two aspects; how to conduct research - good practice - and how to enable stakeholders (funding bodies, reviewers, editors, readers etc.) to assess a research proposal or project - evaluation. In the behavioral and social sciences, however, it was not until the 1980s that the aspect of evaluation took an increasingly prominent role in the broader methodological discourse (Flick 2007; Golafshani 2003; Packer et al. 1989). In their seminal publication, Lincoln and Guba (1985:289-301) define the quality or, as they call it, trustworthiness of scientific research with respect to answering the four basic questions of truth value, applicability, consistency and neutrality, summarized in Table 1. They further propose that, according to the conventional approach towards the quality of empirical research, these questions are addressed in terms of internal validity, external validity, reliability and objectivity. While internal validity refers to the degree of confidence with which a relationship between dependent and independent variables can be inferred, a key criterion for experiments and case studies (Diekmann 2001:296-304; Yin 2003:36), external validity refers to the generalizability of the findings beyond the collected data, as is the case with clinical trials or surveys studying samples of rather than entire populations (Campbell et al. 1966; King et al. 2005). Reliability denotes whether an instrument consistently reports the same value after repeated measurements under the same or similar conditions (Gaskell et al. 2000). Finally, objectivity focuses on the researcher's bias or rather influence on the empirical study tackled by techniques such as double-blinding a clinical trial (Kaptchuk 2001).

<table>
<thead>
<tr>
<th>Quality Criteria</th>
<th>Question</th>
<th>Conventional Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth value</td>
<td>How can one establish confidence in the “truth” of the findings?</td>
<td>Internal validity</td>
</tr>
<tr>
<td>Applicability</td>
<td>How can one determine the applicability of the findings in other contexts?</td>
<td>External validity</td>
</tr>
<tr>
<td>Consistency</td>
<td>How can one determine whether the findings are replicable in the same or a similar context?</td>
<td>Reliability</td>
</tr>
<tr>
<td>Neutrality</td>
<td>How can one determine the degree of bias introduced by the inquirer?</td>
<td>Objectivity</td>
</tr>
</tbody>
</table>

Quantitative research has aimed to approximate these four, conventional ideals of empirical research by various, highly formalized techniques, such as randomized trials in terms of internal validity or representativeness through statistical random sampling in terms of external validity, to mention but a few (Diekmann 2001). In the approach, referred to as conventionalism in this paper, the very same ideals are implemented into qualitative research accepting validity, reliability and objectivity as the basic quality criteria of any kind of empirical research (Kirk et al. 1986). A problematic approach, since qualitative research cannot deliver the kind of numerical formalization these criteria derive from. Hence, softer versions of conventionalism propose different techniques for conducting and demonstrating quality, which fit the rationale of qualitative research. Silverman (2005), for instance, suggests non-numerical procedures such as deviant case analysis and proper documentation. By the same token, Yin (2003), although establishing case studies as a different research strategy than experimentation, derives the quality criteria for case studies from laboratory experiments (cf. Campbell et al. 1966). Case studies have to demonstrate internal validity by, for instance, addressing rival explanations and external validity by analytical generalization rather than representational sampling (cf. Lee 1989; Lee et al. 2003).

By contrast, counter-conventionalism emerged out of a movement against the perceived hegemony of conventional research methodology and positivist epistemology. As qualitative research is studying the meaning social actors give to the world they live in, the positivistic perspective of studying an objective reality does not apply (Packer et al. 1989), which is reflected by a substantially different approach towards quality criteria (Corbin et al. 2008). Accepting qualitative research as an umbrella term for a variety of ways of studying social phenomena (Madill et al. 2000; Reicher 2000), counter-conventionalism leads to
a considerable degree of fragmentation, since different sets of quality criteria had to be developed, which address the specificities of each tradition in terms of approach and research focus (Corbin et al. 1990) – be it hermeneutic research (Packer et al. 1989) or ethnomethodology (Titscher et al. 2000), to name but a few. It is against this backdrop that the criteria suggested for grounded theory by Charmaz (2006) will be used as a typical example of counter-conventionalism. Her proposal of credibility, originality, resonance and usefulness (see Table 2) are exclusively meant to relate to the main rationale for conducting grounded theory – to generate theory grounded in empirical data (Glaser et al. 1967). Other criteria, such as providing well documented data for replicating a research project or testing theories, which would be a valid rationale for a case study (Yin 2003), are not considered, because they do not fit the procedures and aims of grounded theory (Suddaby 2006). As a consequence, quality criteria developed for one qualitative research strategy are of little or even of no help to another (Corbin et al. 1990).

<table>
<thead>
<tr>
<th>Quality Criteria</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credibility</td>
<td>Claims are supported by sufficient data. Strong logical links between observations and analysis are evident.</td>
</tr>
<tr>
<td>Originality</td>
<td>Research offers a fresh insight and contributes to the existing body of knowledge.</td>
</tr>
<tr>
<td>Resonance</td>
<td>The grounded theory reveals taken-for-granted meanings and makes sense to the participants.</td>
</tr>
<tr>
<td>Usefulness</td>
<td>The analysis is helpful in people’s everyday life and sparks new research questions.</td>
</tr>
</tbody>
</table>

The developments outlined above are closely matched by the IS field. The notable increase in the number of qualitative research publications in leading IS journals (Cheon et al. 1993; Galliers 1993; Mingers 2003) gives testimony to the fact that qualitative research is an accepted methodology (Walsham 1995a). As is the case in the behavioral and social sciences in general, the rise of qualitative research has been the outcome of vocal criticism championing epistemological and methodological plurality against the hegemony of positivistic research (Davison 2011; Galliers 2011; Lee 1991; Orlikowski et al. 1991). With respect to the quality of qualitative research, the IS field is mostly focused on how-to guidelines and good practices developed for specific research strategies and the epistemologies they are based upon (cf. Mingers 2004). Hence, the IS discourse reflects the degree of fragmentation found in counter-conventionalism in general. For instance, one can find guidelines for good practice designed for hermeneutical research (Walsham 1995b; 2006), grounded theory (Bryant 2002; Suddaby 2006; Urughart et al. 2009), case study research (Benbasat et al. 1987; Bygstad et al. 2011; Eisenhardt 1989), action research (Baskerville et al. 2004; Baskerville et al. 1996; Mathiassen et al. 2012) or critical research (Cecez-Kecmanovic 2011; Myers et al. 2011). While celebrating the variety of qualitative research, the proposed techniques and good practices only refer to specific perspectives and frames of references and are not meant to be employed for other research traditions.

By the same token, evaluation criteria and guidelines for demonstrating quality apply only within the same tradition as the good practices the criteria are supposed to complement. For instance, Benbasat et al. (1987) recommend evaluation criteria exclusively for case study research. In a similar fashion, Klein and Myers (1999) derive seven principles for evaluating research, which are based on hermeneutics and, as the authors warn, only refer exclusively to interpretive field studies. In more general terms, the developed guidelines for good practice and evaluation focus on specific domains, such as specific research strategies or epistemological traditions. While the majority of qualitative IS research seems to subscribe to the approach of counter-conventionalism, there is, however, a noteworthy deviation from the situation found in the behavioral and social sciences. Since IS as a field is situated between the worlds of academia and practice, the community also discusses the tension between scientific rigor and practical relevance as two quality criteria for IS research in general, which, obviously, also applies to qualitative research (Constantinides et al. 2012). For instance, Benbasat and Zmud (1999) criticize that IS research has paid too much attention to scientific rigor in order to be accepted as an academic field at the cost of producing output relevant for IT professionals who should be able to implement findings into their respective
organizational settings. Caught between academia and consultancy (Davenport et al. 1999), IS research should, therefore, be evaluated according to the degree its findings strike a balance between rigor and relevance rather than according to scientific quality criteria alone. This discourse, however, addresses the IS community as a whole. By contrast, the call for the development of explicit evaluation criteria in qualitative IS research (cf. Conboy et al. 2012; Lee et al. 1995) has been answered by the community according to the approach of counter-conventionalism, thus, taking the specificities of and the substantial differences between epistemological traditions and research strategies into account. Exceptions that elaborate on more general evaluation criteria for qualitative research irrespective of epistemological considerations and research tradition are hard to find and will be discussed in the following section.

**Functional Equivalence**

Conventionalism and counter-conventionalism disagree on the question whether there are fundamental differences between quantitative and qualitative research in the ways quality criteria can and should be defined. The third approach of **functional equivalence** shifts the focus of the quality of qualitative research from substantial differences, as is the case in counter-conventionalism, to procedural commonalities (for similar categorizations see also Flick 2007; Gaskell et al. 2000). As it will be discussed in more detail further below, qualitative research can be seen as a cyclical process of exploring, describing or explaining social phenomena based on their variety rather than distribution (Bauer et al. 2000a; Flick 2006; Krohn et al. 1989) – a common procedure that applies irrespective of epistemology or research strategy. In terms of good practices, the procedural approach can be found in guidelines addressing specific aspects of a research project, such as proper documentation (Flick 2007; Gibbs 2007), qualitative interviewing (Kvale 1994; Leggewie 1987) or authoring an authentic research report (Elliott et al. 1999; Golden-Biddle et al. 2007). Examples can be found in IS research as well. Myers and Newman (2007), for instance, develop general guidelines for conducting qualitative interviews based on the dramaturgical model of renowned sociologist Erving Goffman. By the same token, Bygstad and Munvold (2011) elaborate on the role of the informant to corroborate or even criticize the findings of a qualitative IS research project – a practice that will be referred to as communicative validation further below (cf. Gaskell et al. 2000; Kvale 2002; Lanzara 1991). As these examples illustrate, procedural guidelines cover a wide range of issues and topics; although they may have been derived from different research traditions, they do not remain exclusive to a particular epistemology or research tradition. Bygstad and Munvold’s (2011) paper is a case in point; despite the fact that they introduce communicative validation for interpretive case studies, the notion can be used for other epistemological traditions and research strategies thus providing a procedure for qualitative research in general.

While procedural guidelines as such may be useful for any type of research, functional equivalency goes one step further by observing the differences between quantitative and qualitative methods as equivalent ways of studying social phenomena and, hence, should demonstrate quality in different but equivalent ways as well. The first to employ the term functional equivalence in this context, Gaskell and Bauer (2000) identify two basic functions of quality criteria, which they term confidence and relevance. While confidence refers to criteria demonstrating that findings were not simply made up, relevance refers to criteria demonstrating utility and importance in the sense that data is linked to theory or reveal surprising findings. Thus conceived, internal validity, for instance, fulfills the function of confidence, while external validity fulfills the function of relevance. By contrast, qualitative research fulfills the same two functions in an equivalent way by demonstrating good practices such as triangulation in terms of confidence, communicative validation in terms of relevance or thick description in terms of confidence and relevance (for a comprehensive comparison see Gaskell et al. 2000:344). Another set of criteria for qualitative research, which will be used as the representative example for functional equivalency for the remainder of the paper, is developed by Lincoln and Guba (1985) in response to the four basic questions they pose concerning the truth value, applicability, reliability and objectivity of research in general (see Table 1). They propose for qualitative research the criteria of credibility, transferability, dependability and confirmability. Equivalent to 1) internal validity, qualitative research demonstrates credibility by, for instance, providing documentation of the collection and analysis of diverse data sources (data triangulation). The same applies to 2) external validity and transferability by means of, for instance, quoting extensively from the collected data (thick description), 3) reliability and dependability through, for instance, peer reviews as well as 4) objectivity and confirmability by, for instance, checking preliminary findings through communicative validation (see Table 3).
Table 3. Functional Equivalent Quality Criteria according to Lincoln and Guba (1985)

<table>
<thead>
<tr>
<th>Quality Criteria for Qualitative Research</th>
<th>Explanation</th>
<th>Equivalent to...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credibility</td>
<td>Production of credible findings and interpretations (e.g. prolonged engagement, triangulation, negative case analysis).</td>
<td>Internal validity</td>
</tr>
<tr>
<td>Transferability</td>
<td>Provision of the researcher’s database to enable other researchers to transfer the findings to another setting (e.g. thick description).</td>
<td>External validity</td>
</tr>
<tr>
<td>Dependability</td>
<td>Verification of the research process and its product by an “inquiry audit” (e.g. peer review).</td>
<td>Reliability</td>
</tr>
<tr>
<td>Confirmability</td>
<td>Scrutiny of raw data, notes, reconstruction and synthesis products, and preliminary developmental information through a “confirmability audit” (e.g. communicative validation).</td>
<td>Objectivity</td>
</tr>
</tbody>
</table>

Methodological arguments that resemble the approach of functional equivalency can be found in IS research as well. The starting point of functional equivalence, in general, is to separate methodological considerations from epistemology. In the IS literature, this point is, for instance, made explicitly by Klein and Myers (1999:69) in terms of distinguishing between qualitative research and interpretive research, as they argue that qualitative research can be based on positivistic, interpretive as well as critical assumptions. Klein and Myers (1999), however, continue to focus on interpretivism, which is the reason why their contribution is categorized as counter-conventional further above. By contrast, functional equivalency goes down the path of focusing on qualitative research in order to compare and contrast it with quantitative research. It is in this sense that Lee (1989) elaborates on a methodological argument as to how IS case study research, which he argues is typically based on qualitative data and analysis, can employ quality criteria according to the natural science model of scientific research. Based on the analysis of an exemplary IS case study, he concludes that quality standards of quantitative research, such as laboratory or statistical controls, are not the objectives of scientific research but rather means to accomplish the objectives of scientific research, such as falsifiability, replicability or logical consistency. Case studies accomplish these objectives only through different means. Generalizability is a case in point as the case study research strategy is based on the same rationale as laboratory experimentation; hence, both accomplish the same objectives of scientific research but through different means. Issues of generalizability are further developed by Lee and Baskerville (2003) in their discussion of four different ways of achieving generalizability in IS research. While generalizability is accomplished through statistical random sampling in descriptive statistics, Lee and Baskerville propose a different type of generalizability in terms of, for instance, grounded theory or case study research, which are, differences put aside, based on the rationale to ground theory in properly collected and coded data rather than on representative samples. An exemplary approach of functional equivalency, the four types of generalizability can be seen as different but equivalent ways of providing generalizable findings or, to be more precise, to accomplish applicability as proposed by Lincoln and Guba (see Table 1).

Given the discussion of the three approaches towards the quality of qualitative research, the distinctions between these approaches are, obviously, not clear cut and present some grey areas. However, the categorization itself is useful as it provides a quite abstract overview of the various discourses and traditions (see Table 4). At a first glance, the differences may seem to be the outcome of terminological nuances. After all, most researchers propose similar, if not the same, good practices, such as proper documentation, thick description or triangulation. However, when it comes to the evaluation of the quality of the research project, the seemingly superficial differences become more pronounced. For instance, the quality of a grounded theory research project (counter-conventionalism) is assessed according to its originality and resonance with the readership (Charmaz 2006). By contrast, a qualitative
case study may be assessed according to criteria of replicability or falsifiability (conventionalism) (Lee 1989; Suddaby 2006). Functional equivalence attempts to develop a third alternative by bridging the differences between conventionalism and counter-conventionalism. In order to further explore these differences, the paper will focus on a key practice of empirical research – selecting data for analysis – and discuss how the choice of a rationale for selecting data is connected to considerations of evaluation.

<table>
<thead>
<tr>
<th>General Quality Criteria</th>
<th>Conventionalism</th>
<th>Counter-Conventionalism (Grounded Theory)</th>
<th>Functional Equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Internal Validity</td>
<td>Credibility</td>
<td></td>
</tr>
<tr>
<td>- Truth Value</td>
<td>External Validity</td>
<td>- Originality</td>
<td>- Credibility</td>
</tr>
<tr>
<td>- Applicability</td>
<td>Reliability</td>
<td>- Usefulness</td>
<td>- Transferability</td>
</tr>
<tr>
<td>- Consistency</td>
<td>Objectivity</td>
<td>- Resonance</td>
<td>- Dependability</td>
</tr>
<tr>
<td>- Neutrality</td>
<td></td>
<td></td>
<td>- Confirmability</td>
</tr>
<tr>
<td>(Lincoln and Guba 1985)</td>
<td></td>
<td></td>
<td>(Lincoln and Guba 1985)</td>
</tr>
</tbody>
</table>

The Selection Rationale of Qualitative Research

The process of selecting the proper sources for data collection and analysis, usually referred to as sampling, is a crucial element of any behavioral and social science as it addresses the quality of research as well as issues of feasibility and logically links hypotheses or research questions with data elicitation and analysis. Kemper et al. (2003), for instance, develop six general guidelines;

1. The sampling strategy should stem logically from the conceptual framework as well as from the research questions being addressed by the study.
2. The sample should be able to generate a thorough database on the type of phenomena under study.
3. The sample should at least allow the possibility of drawing clear inferences from the data; the sample should allow for credible explanations.
4. The sampling strategy must be ethical.
5. The sampling plan should be feasible.
6. The sampling plan should allow the research team to transfer or generalize the conclusions of the study to other settings or populations.

These guidelines address various aspects of the selection process that apply to quantitative as well as qualitative research. The sixth guideline is of particular interest as it refers explicitly to the criterion of applicability, which will be the focus for the remainder of the paper. At this point, it may be necessary to clarify that the notion of applicability and the rationales for selecting data sources are discussed in terms of selecting units of data collection, not units of analysis. The selection of a unit of analysis, such as populations of nation states, groups, organizations, settings and so forth, derives from the research interest formulated as a set of hypotheses or research questions. The methodological justification for choosing a specific unit of analysis is the expectation to find answers and, thus, to increase the body of knowledge of a discipline or field. It is in this sense that one may, for instance, select a specific deviant group in order to study the use of mobile phones for illegal purposes, a specific organization as an extreme case for the impact of cloud computing on organizational structures or a specific population of a nation state to study the level of computer literacy (Abbott 2004). However, this paper discusses selecting units of data collection and its contribution towards demonstrating applicability. In contrast to units of analysis, units of data collection refer to sources of data to be elicited by data collection methods, such as randomly sampled citizens to be questioned by a questionnaire or purposefully selected employees of an organization for interviewing. The population of a nation state and the organization are the unit of
analysis; the citizens and employees are the units of data collection (Yin 2003:75-76).

The evaluation of applicability by means of measuring external validity (see Table 1) may be less of a challenge for quantitative research (this is not to say that it is easy to execute the selection plan) since there is an established body of formalized procedures. For instance, sampling techniques are an important aspect for achieving and demonstrating external validity (Diekmann 2001). Representativeness through random sampling allows for the generalization of findings to a given population and hence enables the evaluation of applicability on the part of quantitative research. By contrast, qualitative research employs a process of selecting units of data collection (e.g. interviewees, documents etc.) referred to as purposive sampling. The underlying rationale is to purposefully select data that promises to enrich a researcher's concepts, categories and theories rather than to represent a population. In this sense, Flick (2006) further differentiates between a priori and gradually determined sampling structures. A priori determined sampling structures are developed in advance before the research project is conducted; that is, the criteria according to which units of data collection are selected for analysis are defined beforehand. Typical examples would be demographic categories such as age or gender. Second, a relationship between the selection criteria and attributes of the studied phenomenon is inferred from theory or other empirical studies. A thus conceived sampling structure may, for instance, be based on an inferred relationship between seniority and perceived effectiveness of newly implemented information systems leading to the selection of equal numbers of junior and senior members of an organization.

The rationale of a priori determined sampling structures is not that different from the rationale of statistical sampling. In particular, it resembles the technique of quota sampling - a technique of sampling a predetermined number of respondents according to demographic dimensions (usually gender and age), which is preferably used in market research (Diekmann 2001). Indeed, one can argue that quota sampling is as purposive as a priori determined sampling. As a consequence, the divide between quantitative and qualitative selection strategies becomes highly questionable, the differences being gradual rather than categorical. It is in this sense that Teddlie and Tashakkori (2003) propose to conceptualize the various selection rationales as parts of a continuum (see Figure 1). At the one end, they position statistical random sampling. At the center of the continuum, one finds quota sampling but, given above argumentation, one could find Flick’s a priori determined sampling structures just as well. At the other end of the continuum, Teddlie and Tashakkori see theoretical sampling or, in Flick’s terminology, gradually determined sampling structures. As the name already suggests, gradually determined sampling structures are not based on selection criteria, which were decided upon beforehand. Quite the contrary, a researcher sets out with a broad research interest or question and develops a sampling structure during the course of the empirical study (Corbin et al. 2008). The underlying rationale is to step into the research field ideally without preconceptions in order to explore social phenomena as they are lived and made meaningful by the native inhabitants of the field. The researcher remains open towards unexpected findings and surprises, which would have been overlooked or ignored, if the project was based on a predetermined sampling structure developed to answer a specific research question or hypothesis. It is due to this possibility to follow up on surprises during the same research project why qualitative research, based on gradually determined sampling structures, is the preferred choice for exploratory research.

A broadly used technique and exemplary illustration for gradually determined sampling structures is, of course, theoretical sampling. The origin of theoretical sampling - grounded theory - was developed in the spirit of providing a contrasting position to “theory generated by logical deduction from a priori assumptions” (Glaser et al. 1967:3), which would also include having a sampling structure in place before starting the research. Thus conceived, theoretical sampling is based on what the researcher discovers in
the field. After the analysis of a set of data, the researcher sets out to find new sources of data which would further enrich the initial categories and their dimensions (Corbin et al. 2008). This procedure is supposed to keep the analysis and the developed theories close to the data as well as to focus the research on the concepts rather than on the distribution of certain variables across a given population, which makes theoretical sampling into a very different approach than statistical sampling. In more detail, the sampling structure of statistical sampling is known; that is, the distribution of specific variables (gender, age etc.) throughout a population is known and needs to be reflected by the sample in order to be confident about its representativity. A key point to observe is that the sampling structure of theoretical sampling is unknown, thus enabling a research project to discover what is yet unknown as well. While opening a multitude of avenues for in-depth studies that stay close to the complexities of social interactions, theoretical sampling becomes quite a challenge, when it comes to demonstrating the quality of a grounded theory research project; even more so, when it comes to research proposals. In the words of Corbin and Strauss (2008:146), “human subjects and dissertation committees want to know in advance what persons or groups will be sampled.” In this respect, theoretical sampling cannot deliver, which may be difficult to explain and to justify, especially for novice researchers. Accordingly, the quality criteria developed by Corbin and Strauss (1990) refer to the decisions and justifications that guided theoretical sampling to be assessed after the fact.

Theoretical sampling has become a widely accepted selection rationale in qualitative research in general (Silverman 2005), which is not unproblematic. While theoretical sampling is an excellent practice for exploration and theory development, it does not work as an evaluation criterion for research strategies other than grounded theory (Suddaby 2006). Obviously, these observations reflect the reasoning of counter-conventionalism, which grounded theory has significantly helped to establish (Corbin et al. 1990). Theoretical sampling is simply not meant to be an evaluation criterion the way random sampling is (Charmaz 2006). Thus conceived, the initially superficial differences between conventionalism, counter-conventionalism and functional equivalence in terms of good practice turn into quite significant differences in terms of conceptualizing the quality of qualitative research and its evaluation. Assessing qualitative research based on theoretical sampling, for instance, only makes sense for a grounded theory research project. Applicability, by contrast, is of a lesser or even no concern for grounded theory; for other research strategies, such as case studies, it may be a key criterion. These differences have a tremendous impact on the research project as a whole to be discussed after a short introduction into an alternative selection rationale based on a mix of known and unknown selection structures – corpus construction.

Corpus Construction

Originally developed for linguistics (Atkins et al. 1992; Biber 1993) and adapted into qualitative social science methodology by Bauer and Aarts (2000a), corpus construction introduces a different terminology and, thus, clears up some of the conflation with respect to sampling and qualitative research. As already stated above, a sample represents a given population constructed according to the distribution of known variables; hence, the population needs to be known and, therefore, limited (Diekmann 2001). The researcher needs to be confident that the distribution of specific variables used to describe the population is mirrored in the sample. Of course, there are limits in time and resources leading to the development of sophisticated methods to minimize the effort while maximizing the confidence, as is the case with random sampling. Language, on the other hand, is an open system that is constantly changing providing an unlimited set of possibilities for human communication. The population of language is, therefore, unknowable. In this case, the statistical rationale is neither valid nor viable. Since qualitative social science research studies similar phenomena, such as social (inter-)actions, meaning(s), conversations or speech acts, purposive selection is to be seen in a different light. The reason for selecting purposefully is not the infeasibility of, say, conducting enough qualitative interviews in order to achieve representativeness of a population (cf. Kemper et al. 2003) but rather lies in the very nature of the phenomena qualitative research usually deals with. The studied phenomenon may be unknowable like meanings or attitudes. Others may require qualitative exploration as its population is yet unknown like actions and situations (Bauer et al. 2000a).

As a consequence, the selection rationale needs to take the peculiarities of these kinds of unknown and unknowable populations into account. In other words, declaring a number for how many interviews a researcher should conduct in order to achieve high quality qualitative research is not only arbitrary but
defies any methodological basis. It would be an attempt “to fit the sampling rationale […] like the choice of a false analogy” (Bauer et al. 2000a:19). Instead, purposive selection is an attempt to analyze the variety of a phenomenon and not its distribution. Given this contrast, it may be more appropriate to refer to “purposive selection” rather than “purposive sampling” when discussing the selection rationale and strategies of qualitative research. In what follows, the paper will provide a very brief introduction into how to conduct corpus construction and address its relevance to the functional equivalence approach towards the quality of qualitative research, in general, as well as to the quality criterion of transferability as a functional equivalence for external validity, in particular (see Table 3). Finally, the paper will conclude with a discussion based on a comparison of a priori determined selection structures, theoretical sampling and corpus construction.

Constructing a Data Corpus

A corpus is a collection of data that may result from simply collecting text, such as transcribed interviews and documents, to any kind of symbolic token. Depending on the research interest, a social scientist or linguist collects the relevant data to represent the whole range of the variety of ways a phenomenon is expressed in the field. Hence, the aim is not representativeness derived from proportional demographic sampling but rather the inclusion of as much variation as possible (Biber 1993). The usual outcome is a typology developed for the phenomenon of interest. Against this backdrop, Bauer and Aarts (2000a) propose four basic rules on how to construct data corpora in qualitative research (see Table 5). The rules will be explained briefly followed by guidelines for good-practices on how to organize data into corpora and how to demonstrate the quality of the selection procedure. Further below, these rather abstract rules and principles will then be illustrated by an example taken from the author’s research in order to convey the fundamentals of corpus construction as they were applied in actual IS research.

| Table 5. The four Rules of Corpus Construction (Bauer and Aarts 2000) |
|--------------------------|------------------------------------------------------------------------|
| Rule 1 | Proceed stepwise: select, analyze, select again. |
| Rule 2 | Selection based on known categories (e.g. gender, age, social status) precedes unknown variety (variety of representations of the studied phenomenon). |
| Rule 3 | Characterizing the unknown variety of representations has priority over anchoring them in existing, known categories. |
| Rule 4 | Maximize the unknown variety of representations by extending the range of the known categories until saturation is achieved. |

Since qualitative researchers deal with phenomena being unknown or unknowable in their variety, the problem is not only where to start but also how to proceed. A qualitative researcher does not know what is to be found in the field and has to expect the unexpected. Rule 1, therefore, prescribes a cyclical procedure of selection and analysis to increase the knowledge of a certain phenomenon step-by-step. The collection of data – the corpus – grows over time. A key difference between corpus construction and theoretical sampling is the point of departure, which is addressed in Rule 2. To begin with, the research project is approached according to two dimensions - known categories and the unknown variety of the studied phenomenon. In contrast to theoretical sampling, the selection strategy is initially planned beforehand based on what is known (known categories) in order to have a preliminary rationale that guides the study of the phenomenon of interest or rather its unknown variety. For linguistics, Biber (1993) proposes known categories such as primary channel (e.g. written, spoken, scripted) or setting (e.g. public, private) that can guide the initial selection of incidents of human communication. By the same token, Bauer and Aarts (2000a) suggest well established social categories such as gender, age or education. Similar known categories can be proposed for IS research such as organizational size (e.g. SMEs and large corporations) or by distinguishing between non-digital and digital data sources (Hedman et al. 2013), to name but a few.

While Rule 2 addresses the difference between corpus construction and other gradually determined selection strategies, such as theoretical sampling, Rule 3 introduces the distinction between corpus construction and quota sampling. As it was already established above, quota sampling is based on the inference of a correlation between the selection structure (e.g. distribution of age and gender across the
entire population) and the phenomenon of interest (e.g. appeal of the packaging of a new product). The sample needs to represent a given population, which can be established by, for instance, sampling 50% female and 50% male respondents, if one wants to represent the general population of a typical nation state. By contrast, it is not necessary to establish any kind of correlation in order to conduct corpus construction, because the selected data does not represent a known population. The aim is to discover as much variety in the ways a phenomenon is expressed. Thus conceived, the known categories only serve as structuring devices and guidelines for selecting data; they are not part of the explanation of the findings. As a consequence, the selection criteria can be extended by further known categories, which were not considered initially, if they promise to reveal more variety. In this sense, Rule 4 establishes corpus construction as a gradually determined selection structure. Known categories have to be added as selection criteria until no more variety is discovered and saturation is accomplished.

Based on Roland Barthes’ semiology, Bauer and Aarts (2000a) further complement the four rules of corpus construction by three basic principles for the collection and organization of data into corpora - relevance, homogeneity and synchronicity. Relevance refers to having a thematic focus when collecting data. Each corpus should be relevant for answering the research question from one point of view only. As a consequence, collected data needs to be grouped accordingly in order to form relevant corpora for comparison. Homogeneity proposes to organize data according to the method they were collected with. For instance, data collected through interviews may be complemented with project documentation or archival material each forming a distinct corpus for comparison. Finally, synchronicity introduces time into the data corpus. The studied phenomenon will most likely change over time – sometimes quite radically. As a consequence, a researcher needs to be sensitive towards significant changes in the field and to construct corpora containing data from each phase before and after the change – referred to as cycles in corpus construction. Cycles can be used for further comparisons in order to discover more variety.

Given these three principles as well as the four rules of corpus construction, a researcher can rely on a set of good practices on how to design a selection strategy in order to construct relevant, homogeneous and synchronized corpora. Corpus construction, however, also addresses issues of evaluation as a research project is required to demonstrate the quality of the selection process by documenting 1) the substance of materials (text, images, sounds etc.), 2) the characterization of the research topic, 3) the modalities of the stepwise extension of the open corpus, 4) the known categories that were used at entry into the field, 5) the known categories that were added later, 6) evidence for saturation, 7) the timing of the cycles of data collection and 8) place of data collection. More importantly, corpus construction enables a researcher to plan the selection process beforehand. Making informed decisions on the known categories, which are used as a starting point for collecting data, as well as explicitly describing the selection rationale, which guides the modification and refinement of the selection strategy as the project unfolds, allows project stakeholders to evaluate the quality and rigor of a proposed project (Conboy et al. 2012). Before these issues will be further discussed, however, the paper provides an example how corpus construction may be applied in IS research.

*The Digitization of Memory Institutions – An Example from IS Research*

The following example is taken from seven years of research in the digitization of memory institutions (libraries, archives, museums) and the changes it brings about in terms of institutionalized collective memory practices. In particular, this section will focus on a revelatory single case study on the Europeana initiative. Financed by the European Commission (EC) and led by the Conference of European National Librarians (CENL), Europeana is a unique project of bringing the digitization efforts of European libraries, archives and museums under a single framework in order to grant access to their digital repositories through the online portal [www.europaeana.eu](http://www.europaeana.eu). The broader research interest was to understand how Europeana defines itself as an emergent memory institution in the wider digital ecosystem in contrast to related services such as Google Books. Thus conceived, the main focus was on how Europeana observes itself as a functional digital memory institution, combining libraries, archives and museums, and how it fulfills its societal mandate as a caretaker of cultural heritage and curator of digitized cultural heritage artifacts. The conceptual framework was mostly based on social systems theory (Luhmann 1998), which will not be further elaborated here as it is irrelevant for corpus construction and the argument put forth in this paper. Given the nature of the research domain, the research project resulted in contributions to the IS field, information management, digital memory studies as well as sociology (cf. Kallinikos et al. 2013a;
Kallinikos et al. 2013b; Marton 2009). In terms of methodology, the Europeana project was defined as the unit of analysis and corpus construction was employed for the selection of units of data collection. Due to practical reasons, it was not feasible to interview members of the Europeana project team; hence, the main body of data consists of documents, such as project documentation, memos or project deliverables, amounting to a total of 137 documents. It was only in the final phase of the research project that Europeana established a dedicated organization – the Europeana Office - whose employees were interviewed for communicative validation of the findings (see further above).

Given the focus of the research project on how Europeana observes itself as a digital memory institution, the known categories used at entry into the field were based on the distinction between project internal documents (how does Europeana observe itself) and external documents authored for the outside world (how does Europeana observe itself through the eyes of relevant outsiders such as the European Commission). The categorization was carried out by assigning documents written by Europeana project members for project members to the corpus of internal documents and documents written by project members for non-project members to the corpus of external documents. As expected, it became apparent very soon that the initial known categories were too broad and did not capture the variety found in the field. For instance, the known categories were not sensitive towards documents authored by non-project members for Europeana, such as project evaluations by the EC, which have a major influence on how Europeana observes itself through the eyes of significant others. Hence, the known categories were restructured into a 2x2 matrix distinguishing between project members and non-members as receivers as well as authors of documents or rather, as they were referred to in the research project, as address and addressee of mediated communication (see Figure 2).

![Figure 2. Gradual Determination of Selection Structure and Data Corpora on Europeana](image)

As data collection and analysis progressed, the non-members were gradually subdivided into more granular categories to capture even more variety within the data. Directly involved stakeholders of the project, such as the EC, were differentiated from, for instance, external experts and media in order to be able to account for differences in the ways Europeana was observed by the community of librarians expressed in journal publications or how project members explained the role of Europeana in an interview conducted by a journalist for the general public. The final selection structure, shown in Table 6, was a 5x6 matrix. Each cell of the matrix was a defined corpus to which documents were assigned according to their address and addressee. In other words, the rationale of address and addressee was the organizing principle for the formation of homogeneous corpora. The second principle of corpus construction – homogeneity – was, therefore, accomplished in a different way than recommended by Bauer and Aarts (2000a). Their suggestion to organize data according to the methods of collection did not fit the research design and was replaced by the distinction between address and addressee borrowed from document analysis (Prior 2003).

In detail, the known categories entailed 1) members of the Europeana project, 2) external project...
stakeholders, such as the EC, the European Parliament or the CENL, 3) members of the media, such as journalists, 4) the public at large and, finally, 5) experts in digital libraries, archives and museums. The last row, the researcher, was added during the final stage of the project as employees of the then founded Europeana Office were interviewed for communicative validation. The interviewer, i.e. the author of this paper, was treated like any other addressee of a document in order to form an additional corpus for interview data.

As Table 6 illustrates, the two-dimensional unfolding of the selection structure according to address and addressee resulted in a vast array of data sources, which was not feasible to elicit and to analyze entirely within the given timeframe of the research project. Hence, data collection required more focus according to the first principle of corpus construction - relevance. As stated above, the main goal of the research project was to study the variety of ways Europeana observed itself as a digital memory institution internally as well as through the eyes of others, which was translated into empirical fieldwork by selecting only documents whose address or addressee was Europeana (see grey cells in Table 6) for further analysis. The other corpora (see white cells in Table 6) were disregarded as evidence supporting the findings of the study and were only used to contextualize the case when necessary. Finally, the last principle of corpus construction – synchronicity – was accomplished quite easily. In case of Europeana, a defining event was the launch of a prototype version of the online portal www.europeana.eu as a proof of concept. After the launch, the project entered a new phase focusing on the development of Europeana as an operational service. Hence, the collected data was ascribed to two cycles – before and after the launch of the prototype – for comparative analysis.

Following the technique of corpus construction resulted in a set of relevant, homogeneous and synchronized data corpora, which formed the foundation for a comparative analysis of the various ways Europeana observed itself as a digital memory institution. This was accomplished by a cyclical procedure of selecting and analyzing documents in order to approach an unknown phenomenon step-by-step (Rule 1). Selection criteria were developed based on established concepts in document analysis (Prior 2003) by differentiating between address and addressee of documents as known categories (Rule 2). The known categories were then used to capture variety in the ways Europeana is observed and, hence, to guide and

<table>
<thead>
<tr>
<th>Address of Document</th>
<th>Project Members</th>
<th>External Project Stakeholders</th>
<th>Media</th>
<th>Public</th>
<th>Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Members</td>
<td>E.g. memo, project documentation</td>
<td>E.g. project deliverables to the EC</td>
<td>E.g. press release</td>
<td>E.g. project website</td>
<td>E.g. journal publication</td>
</tr>
<tr>
<td>External Project Stakeholder</td>
<td>E.g. funding contract</td>
<td>E.g. project evaluation</td>
<td>E.g. press release</td>
<td>E.g. public speech</td>
<td>E.g. project report</td>
</tr>
<tr>
<td>Media</td>
<td>E.g. newspaper interview</td>
<td>E.g. newspaper interview</td>
<td>n/a</td>
<td>E.g. newspaper article</td>
<td>E.g. newspaper interview</td>
</tr>
<tr>
<td>Public</td>
<td>E.g. user feedback</td>
<td>E.g. petition</td>
<td>E.g. letter to the editor</td>
<td>E.g. online discussion</td>
<td>E.g. user feedback</td>
</tr>
<tr>
<td>Experts</td>
<td>E.g. advisory report</td>
<td>E.g. advisory report</td>
<td>E.g. letter to the editor</td>
<td>E.g. blog</td>
<td>E.g. general guidelines for digital libraries</td>
</tr>
<tr>
<td>Researcher</td>
<td>Interviews</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
organize the collection of data (Rule 3). Finally, the known categories were gradually extended and refined during the course of the study in order to address the unexpected peculiarities of the field and to increase the variety of the studied phenomenon in the collected data (Rule 4). Despite the specificities of the research project, these descriptions and explanations should have conveyed the rather abstract notions and concepts of corpus construction to researchers, who are not familiar with this technique, since the goal was to outline the fundamentals in such a way that would allow applying the selection rationale in IS research.

**Discussion**

Corpus construction does not subscribe to a certain epistemology, research tradition or research strategy but rather proposes a procedure that is applicable for qualitative research in general. From the perspective of purposive selection, qualitative research can be described as a cyclical process of selecting and analyzing data – a perspective that offers a procedural basis for qualifying qualitative research (Flick 2006). Be it the hermeneutic circle (Packer et al. 1989), the concept driven oscillation between data collection and analysis through theoretical sampling (Corbin et al. 1990; 2008) or the step-by-step cycle of corpus construction (Bauer et al. 2000a), a qualitative research project is based on a feedback loop between analysis and selection rather than on a linear process of testing hypotheses. It is from this perspective that corpus construction joins the approach of functional equivalence based on the procedural commonalities of rather than epistemological differences between qualitative research strategies. Thus conceived, corpus construction can be seen as a method for demonstrating transferability – the functional equivalence to external validity (see Table 3). As a reminder, transferability is defined by Lincoln and Guba (1985) as the provisioning of the researcher's database to enable other researchers to transfer the findings to another setting. Based on the rationale of studying the variety in the ways a phenomenon is expressed, another researcher may build upon previous findings and attempt to discover more variety by introducing known categories not considered by previous projects. By the same token, the selection structure and selection criteria, which were developed during one empirical study, may be re-used for another project dealing with a similar research question in another setting. For instance, the known categories, used to construct data corpora for studying Europeana, can be re-used in another case study on digital libraries to allow for their comparison. It is in this sense that corpus construction applies to the criterion of transferability and can be seen as the functional equivalence to random sampling. The underlying rationales are very different; corpus construction aims for maximizing variety of an unknown or unknowable population while random sampling aims for representativity of a known population. From a procedural perspective, however, both enable stakeholders to evaluate the quality criterion of applicability before and after the research project – be it in terms of external validity in a quantitative or transferability in a qualitative study. Thus conceived, purposive selection can demonstrate the quality of qualitative research in a different but equivalent way to statistical sampling methods.

The problem of evaluating qualitative research proposals is not that stakeholders want to know in advance what persons or groups will be sampled but rather how a researcher proposes to select persons, groups, settings, cases, documents and so forth. It is the procedure that matters. After all, in random sampling, the researcher does not know in advance who or even how many will end up in the sample either. Corpus construction formalizes the procedure of selection, not the selection itself. It offers an underlying rationale of maximizing the variety of an unknown or unknowable population by reference to known social categories. Consequently, a researcher is able to outline the external categories expected to guide the data collection. In corpus construction, the selection steps can, to some extent, already be planned in advance. Of course, a researcher may change the selection plan during the actual research, but potential funding bodies, supervisors or gate keepers can at least get an initial understanding what a researcher will do while being in the field. This is due to the two-dimensional approach of known categories and unknown variety.

Thus conceived, the three main purposive selection rationales introduced in this paper can be further categorized based on the quality criterion they apply to. A priori determined selection structures can be seen as the qualitative version of quota sampling, thus leaning towards conventionalism and the criterion of external validity by sampling a known population. The selection structure is known. By comparison, theoretical sampling only deals with unknown variety or rather an unknown or unknowable population of phenomena. The selection structure is unknown. If at all, known categories are used only as a starting
Marton / Purposive Selection and Quality

Given above discussed points, the crucial role of selecting units of data collection becomes apparent. It is not only a key element in connecting methodology and conceptualization but also addresses the two main controversies on the quality assessment of qualitative research. One controversy refers to the fundamental differences among qualitative methods, which renders a unified approach in terms of quality and evaluation impossible (Corbin et al. 1990; Madill et al. 2000; Reicher 2000). However, the cyclical research process by means of purposively selecting units of data collection offers a procedural basis for qualifying qualitative research (Flick 2006). Qualitative research can be seen as the study of variety rather than of a representation of a known population. In other words, qualitative selection is an alternative way of selecting units of data collection based on a rationale that is not defined as anything that is non-quantitative or non-statistical. Quite the opposite, the justification for conducting purposive selection is founded on the rationale of studying unknown or unknowable populations of social phenomena and the variety of ways they are represented.

The second controversy addresses the problem of benchmarking (Flick 2007) and refers to qualitative research as non-formalizable due to being very close to social phenomena and the unique settings the phenomena are embedded into. Qualitative research does not measure a phenomenon by means of scaled instruments. Since there are not even basic units of measurement, qualitative research cannot be

<table>
<thead>
<tr>
<th>Selection Rationale</th>
<th>Quality Approach</th>
<th>Complementary Quality Criteria</th>
<th>Selection Structure</th>
</tr>
</thead>
</table>
| A Priori Determined Sampling Structure and Quota Sampling | Conventionalism | - Internal Validity
- External Validity
- Reliability
- Objectivity | Known |
| Theoretical Sampling | Counter-Conventionalism (Grounded Theory) | - Credibility
- Originality
- Resonance
- Usefulness | Unknown |
| Corpus Construction | Functional Equivalence | - Credibility
- Transferability
- Dependability
- Confirmability | Known / Unknown |
Research Methods and Philosophy

measured in a standardized way that allows for comparisons. Hence, it is impossible to benchmark the results of qualitative research projects and, as a consequence, to assess their quality in a comparative fashion (Patton 2002). By contrast, quantitative research achieves a necessary degree of abstraction by means of quantifying social phenomena that enables the formalization of research and, therefore, the comparison of proposed research projects irrespective of research topic and hypothesis. The critique of benchmarking qualitative research, however, conflates numerical/non-numerical research with the degree of formalization (Bauer et al. 2000b). Quantitative research delivers numerical results, but it takes an additional step of abstraction to formalize the quality of these results, as is the case with the assessment of representativity of sampling techniques based on theories of probability. In other words, benchmarking of quantitative research does not come with the numbers but rather is an accomplishment by researchers and the statistical models they apply to further qualify the numbers (Jensen 1991). By contrast, qualitative research may lack the degree of formalization necessary for benchmarking, but that is not the case because the results are non-numerical. Quite the opposite, the formalization of qualitative research is not only possible but a necessary step to advance its methodological basis and rigor.

As the introduction into corpus construction demonstrates, gradually determined selection can be formalized in a non-numerical way. In more detail, the selection procedure is formalized – how selection is going to be conducted rather than who or what is going to be selected. Hence, a qualitative researcher can already convey a selection strategy in a research proposal allowing its evaluation by stakeholders according to Teddlie and Tashakkori’s (2003) six guidelines discussed above. Be it the logical steps from concept and research question to selection structure and analysis, the feasibility of the selection strategy or transferability of the findings, stakeholders can make an informed assessment based on the justifications given by the researcher for selecting specific known categories as a starting point and the rationale for selecting more known categories during the project itself. This should be helpful for novice researchers, in particular, to convince evaluation committees of their selection plan without abandoning notions of gradually determining selection structures during the actual fieldwork. In other words, demonstrating the quality of qualitative research according to functional equivalences offers alternative evaluation criteria that can be applied irrespective of epistemological traditions and research strategies, thus, contributing towards a formalization of the evaluation of qualitative research.

Conclusion

The paper discussed the central issues that revolve around the quality of qualitative research as a combination of good practices and evaluation criteria by comparing the three different approaches of conventionalism, counter-conventionalism and functional equivalence. While the first attempts to accommodate the conventional quality criteria of empirical research, i.e. validity, reliability and objectivity, the second proposes a fundamentally different conceptualization of the quality of qualitative research – a conceptualization that considers the different epistemological foundations and specificities of each qualitative research strategy. Functional equivalence derives from the comparison of quantitative and qualitative research as different but equivalent ways of studying social phenomena. Hence, qualitative research should develop different but equivalent quality criteria to quantitative research.

In order to discuss the quality of qualitative research in-depth, the selection of units of data collection was taken as an example that addresses notions of good practice as well as evaluation. Consequently, three qualitative selection rationales – a priori determined selection structures, theoretical sampling and corpus construction - were introduced, which illustrate different methodological frameworks for selecting data for analysis. By the same token, each of these three rationales fits one of the approaches towards the quality of qualitative research and, therefore, addresses very different criteria for evaluating qualitative research. Focusing on corpus construction, a selection rationale hardly known in IS research, the paper demonstrated how qualitative research can address the quality of the selection procedure in a different but equivalent way to statistical sampling techniques by referring to the quality criterion of transferability as the equivalent to external validity. Thus conceived, the paper contributed an outline on how purposive selection can be used to demonstrate quality qualitative research in IS irrespective of the epistemology, research strategy or tradition an empirical study is based upon. As a consequence, the paper raised awareness to a broader methodological discussion on evaluation criteria that complement good practices but, more importantly, apply across epistemological traditions and research strategies and can be seen as a step towards the formalization of the evaluation of qualitative research without proposing to formalize
qualitative research itself.

References


