

ASSURING RELIABILITY IN QUALITATIVE STUDIES: A HEALTH INFORMATICS PERSPECTIVE

Hafeez-Baig, Abdul, School of Management and Enterprise, BELA, University of southern Queensland, Toowoomba, Queensland, 4350, Australia. abdulhb@usq.edu.au

Gururajan, Raj, School of Management and Enterprise, BELA, University of southern Queensland, Toowoomba, Queensland, 4350, Australia. gururajan@usq.edu.au

Subrata Chakraborty, School of Management and Enterprise, BELA, University of southern Queensland, Toowoomba, QLD, 4350, Australia.
subrata.Chakraborty@usq.edu.au

Abstract

Assuring the validity and reliability of data is an essential component of data collection. While quantitative studies use certain statistical techniques such as 'Cronbach Alpha' values for a reliability index, in qualitative studies these type of measures are not widely available, and appear to be predominantly subjective. So, many studies, if at all probe this aspect, highlight what is normally termed as 'saturation' and use this as a measure of reliability. While this serves the purpose to some extent, whether researchers can use this as a concrete evidence is questionable. We propose a new approach to assure reliability in qualitative studies, and provide a case study to demonstrate our approach and its benefits. We hope that this serves as a model to many PhD students and other Early Career Researchers who pursue their studies using qualitative approaches.

Keywords Research Methods, Qualitative Approaches, Reliability.

1 INTRODUCTION

In many health studies, qualitative approaches have been used. This approach consists of interviews, focus group sessions and brainstorming sessions, to explore the domain of study. Researchers become part of the study, and the researcher also becomes a surrogate for validity and reliability, as there appears to be limited scientific tools available to establish validity and reliability in qualitative data. However, using text analyses software applications such as NVivo and Leximancer, it is now possible to provide scientific evidence that the qualitative data collected is reliable and valid. In one of our projects, we used a mobile telehealth system to mimic an aged care assessment. The main objective of this project was to explore whether it is possible to use the mobile telehealth systems in an aged care assessment process, and whether the outcomes from both user and client would be satisfactory. We present how we established validity and reliability using some features available in NVivo while exploring qualitative data.

2 WHAT IS QUALITATIVE RESEARCH?

Qualitative research uses a naturalistic approach that seeks to understand phenomena in context-specific settings, such as "real world setting [where] the researcher does not attempt to manipulate the phenomenon of interest" (Patton, 2001, p. 39).

Qualitative research means "any kind of research that produces findings not arrived at by means of statistical procedures or other means of quantification" (Strauss and Corbin, 1990, p. 17) and is instead, the kind of research that produces findings deduced from deduced real-world settings where the "phenomenon of interest unfold naturally" (Patton, 2001, p. 39). While quantitative researchers seek causal determination, prediction, and generalization of findings, qualitative researchers seek illumination, understanding, and extrapolation (Hoepfl, 1997).

Qualitative analysis results in a different type of knowledge due to the fact that the underlying philosophical nature is different to that of quantitative approaches, compatibility of the research methods, and qualitative researchers' involvement and role within the research. Support of the researchers' involvement can be found in Patton (2001), where a qualitative researcher should be present during the changes to record an event after and before the change occurs. Thus, both qualitative and quantitative researchers need to test and demonstrate that their studies are credible. The credibility in qualitative research largely depends on the researcher as the researcher is the instrument (Patton, 2001, p. 14). Therefore, in qualitative studies validity and reliability, refers to the ability and effort of the researcher. While reliability and validity are treated separately in quantitative studies, these terms are seldom viewed separately in qualitative research, the terminology encompasses both.

3 RELIABILITY & VALIDITY TREATMENT IN QUALITATIVE RESEARCH

Reliability

Reliability is mainly a concept used in quantitative research due to the nature of statistical sampling, but the idea is most often used in all kinds of research. The term is synonymously used with 'testing', and if this perspective is adapted then the most important test of any qualitative study is its quality, as this will enable the elimination of any 'confusion' in the study. In qualitative studies confusion can occur at many points – when participants were not clear about the focus of

the study, the information provided may not be relevant. In focus group type studies, individual dominance might digress the conversation from main focus. Therefore, reliability is used in qualitative studies as a concept to evaluate quality with a purpose of generating understanding (Stenbacka, 2001).

Patton (2001) states that validity and reliability are two factors which any qualitative researcher should be concerned about while designing a study, analysing results and judging the quality of the study. In qualitative studies, the term reliability serves as a surrogate for the terms Credibility, Neutrality or Confirmability, Consistency or Dependability and Applicability or Transferability are to be the essential criteria for quality (Lincoln & Guba, 1985). To be more specific with the term of reliability in qualitative research, Lincoln and Guba (1985, p. 300) use “dependability”, in qualitative research which closely corresponds to the notion of “reliability” in quantitative research. The dependability is referred to in the context to which a researcher ‘trusts’ or ‘depends’ on the data to answer the inquiry and lead to consistency. There is support for this notion in Clont (1992) and Seale (1999), who endorse dependability with the concept of consistency or reliability in qualitative research. The consistency of data will be achieved when the steps of the research are verified through examination of such items as raw data, data reduction products, and process notes (Campbell, 1996). In the current time, especially in the context of this paper, these are achieved using qualitative text analyses tools.

Validity

Validity is not a single or universal concept in qualitative studies, rather a wide range of terms for some kind of qualifying check or measure. Creswell & Miller (2000) suggest that the validity is subjective, and affected by the researcher’s perception. Thus, many researchers have developed their own concepts of validity and have often generated or adopted what they consider to be more appropriate terms, such as, quality, rigor and trustworthiness (Davies & Dodd, 2002; Lincoln & Guba, 1985; Mishler, 2000; Seale, 1999; Stenbacka, 2001).

Testing Validity and Reliability

In order for producing credible and consistent results, validity should be maximised, and eventually this may lead to generalizability. This has a great impact on how various components of a study are conceived and documented. Therefore, the quality of a research is related to generalizability of the results and thereby to the testing and increasing the validity or trustworthiness of the research.

Validity in qualitative studies is employed by using various approaches, normally triangulation type approaches. Mathison (1988) stated that Triangulation has assumed an important methodological issue in naturalistic and qualitative approaches to assess and perhaps eliminate control bias and establishing valid propositions because traditional scientific techniques are incompatible with this alternate epistemology (p. 13). Thus it is common to use a range of methodological approaches, mixing paradigms or possibly mixing methods within one paradigm in order to assure confirmation and generalization of a research. Further, Barbour (1998) agrees with this notion and argues for the need to define triangulation from a qualitative research’s perspective in each paradigm. Healy and Perry (2000) see validity and reliability within the realism paradigm which relies on multiple perceptions about a single reality. These views and notions ultimately indicate that several data sources and their interpretations with those multiple perceptions in the realism paradigm can be realised using triangulation.

If this argument is considered valid, then one can consider the ideas and explanations generated by additional researchers studying the research participants, and may include multiple methods of data collection and data analysis. The effect of this notion is that a researcher does not have to fix a method for all the researches. The methods chosen in triangulation to test the validity and reliability of a study depend on the criterion of the research, as should be. If this notion is accepted, then in action type of research studies, each iteration can adopt a completely different method, as the main focus is ‘inquiry into a problem’, and this assumes paramount significance.

4 METHODOLOGY

With the argument presented above, the case study we present here consisted of two academic researchers who are investigating the main question:

4.1. Whether mobile telehealth will yield significant advantages in an aged care assessment?

The academic researchers were supported by adjunct researchers from the organisation where this study was conducted because the organisation was keen to understand how the technology was perceived, whether the proposed technology would be 'usable' and the cost aspects associated with technology implementation.

The 'inquiry' involved a literature scan in the domain of telehealth, organisational benefits, technology adoption and user behaviour. Traditional models such as the TRA, RPB, DoI etc were considered, and using available interview and survey instruments, initial sets of 'instruments' were prepared and presented to the organisation.

The first set of validity and reliability was established by the adjunct researchers validating and providing feedback on the instruments. Due to the domain and the functional aspects, many questions extracted from extant literature were found to be irrelevant to the context, and based on a series of three meetings, the instruments were re-drafted and tested (construct and face validity) with participants.

This enabled the team of researchers to undertake qualitative data collection, which involved a series of three focus groups and 17 interviews. The focus groups were conducted to extract collective views, and the interviews were conducted after each assessment with real participants for aged care assessment. The focus group consisted of an overarching question and the interviews consisted of a series of open ended questions.

It should be noted that the technology was implemented in a real time aged care organisation, and the interview questions were exploring 'real use', NOT perceived use. The Occupational therapists and the field officers conducted the mobile telehealth assessments with clients, and subsequently participated in the data collection providing feedback on the assessment results as a result of using the mobile telehealth system.

5 DATA ANALYSIS & DISCUSSIONS

5.1. Qualitative Data Analysis

The qualitative data analysis consisted of two phases. The first phase involved generating a word frequency cloud using the "Word Cloud" technique to organise and analyse the interview transcripts, and the second phase involved generating in-depth themes called the 'nodes' through "Tree Map" technique. For both phases, NVivo ver 10 was used. Through Nvivo a "World Frequency Cloud" technique was used to analyse the data through Nvivo to establish word count.

5.1. Word Cloud

The Word Frequency Cloud is a technique in NVivo that documents most frequently appearing words in a transcript. For the purpose of this study, the number of most frequent words was fixed at 100, and the transcripts of all interviews were run en-block.

The procedure basically analysed words based on their occurrence, their distance, their context and develops a frequency table of occurrence and percentage. Once this was established, words that did not appear more than 20 times in the overall context were assumed to be less significant in the

overall context of the study and removed from the word frequency run procedure. This procedure was repeated many times to remove any words that did not hold any significance (for example words such as ‘yes’, ‘I’, etc). The following diagram is a screen shot of a words frequency table (also called the ‘word cloud’) to provide an initial view point for further in-depth qualitative analysis.



Figure 1: Initial view point of qualitative analysis

From the above picture, it can be seen that words such as ‘client’, ‘assessment’, ‘technology’, ‘well’, ‘okay’, ‘good’ appear more prominently in the interview. Similarly, words such as ‘difficult’, ‘problems’ are smaller in size and density. Furthermore, through this analysis, the words frequency table, as extracted from the qualitative analysis, is presented below to make things clearer.

No	Word	Length	Count	Weighted Percentage (%)
1	Client	6	216	2.61
2	Able	4	201	2.43
3	See	3	167	2.02
4	Well	4	159	1.92
5	Assessment	10	148	1.79
6	Technology	10	129	1.56
7	Telehealth	10	125	1.51
8	Time	4	123	1.49
9	System	6	122	1.48
10	Good	4	120	1.45
11	Really	6	114	1.38
12	Things	6	107	1.29
13	Using	5	82	0.99
14	Clients	7	80	0.97
15	Face	4	79	0.96
16	Camera	6	78	0.94
17	Different	9	72	0.87
18	Okay	4	71	0.86

19	Comfortable	11	70	0.85
20	Video	5	65	0.79
21	Hear	4	64	0.77
22	Right	5	62	0.75
23	Need	4	61	0.74
24	Field	5	59	0.71
25	Conversation	12	55	0.67
26	Sound	5	50	0.61

Table 1: Summary of word frequency analysis

The above word cloud and word frequency table provided an initial path for further analysis, as the cloud indicated that key words were captured in the interview process, thus indicating ‘reliability’ and ‘validity’. In addition to this, we conducted 17 interviews and the themes were found to be saturated around the 12th interview, indicating that the qualitative process employed in this study was reliable and appropriate. In fact, after the 6th interview, most of the technology related issues such as the connectivity issues and the visual aspect issues (audio and video) were saturated in the interview process, thus indicating a very high level of reliability.

The study also employed another level of reliability. When the organisational participants (who conducted the assessment using the technology) were interviewed, a final question was on their level of overall satisfaction in using the technology. The minimum score on a 10-point scale was 7. This is about 70% of overall satisfaction, especially on the overall process including assessment procedures, technology usage, facilitating conditions and habitual use. Zikmund (2006) indicates that for a new study of this type, 0.7 is a good reliability factor.

By collectively employing these measures, it was inferred that the data aspects were reliable in the study.

Tree Map

The next stage in the data analysis was generating a tree map. The purpose of the tree map was to ascertain that the themes that would be extracted to meet the objectives were adequately represented by the key words. When generating the tree map, the visual provision would indicate the weight of the key words, and this can be inferred by the size of the rectangle allocated to keywords.

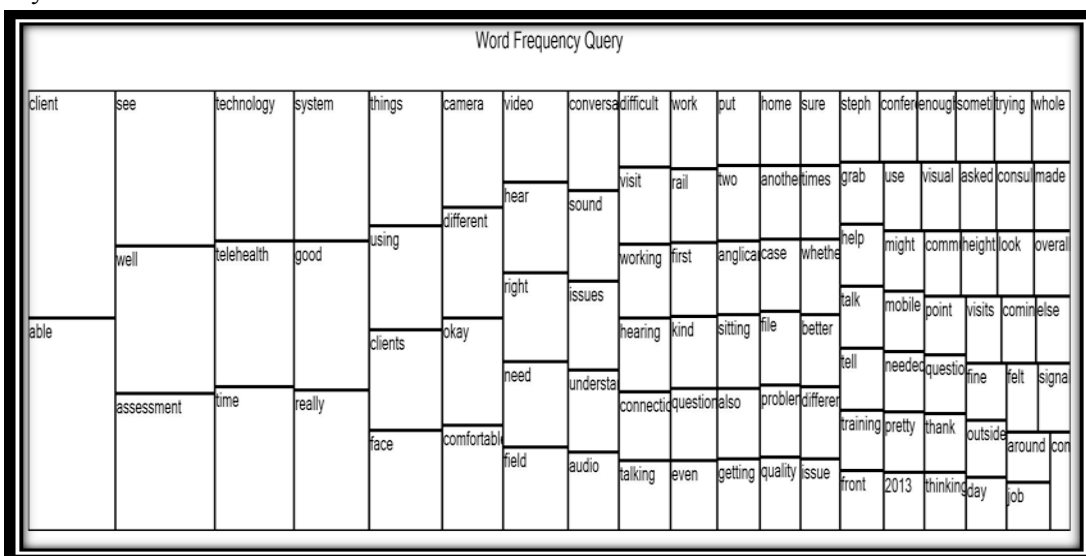


Figure 2: Summary of visual provision of key words

From the above word frequency table, it can be seen that ‘client’, ‘able’, ‘see’, ‘well’, ‘assessment’ and ‘good’ are some of the prominent words appearing on the left side of the picture. As can be seen, these words while indicating frequency and relative importance, also indicate that OTs and FOs were able to perform satisfactory assessment using the given technology in the healthcare setting. As the trial continued, it was evident that the level of comfort has also grown with the technology in the given context.

Therefore, the above data analysis and the two procedures followed gave us further confidence to probe the data in depth. In order to accomplish this, we created themes (called the nodes in NVivo) to slot statements that would most closely match themes that are explored.

The above data analysis clearly provides evidence about the use of technology in the context of “Access Improvement” (AI), “Quality of Services” (QS), “Client Satisfaction” (CS), “Staff Satisfaction” (SS), “Client Safety” (CS), and “Waiting Time” (WT) in the context of themes namely, Technology (T), Organizational (O), and Issues (I). The next stage of this study consisted of the analyses of the qualitative data.

In this study, there were six key objectives, and we created six nodes representing these objectives. These nodes were arbitrarily titled as:

- Access Improvement
- High Quality Service
- Client Satisfaction
- Staff Satisfaction
- Client Safety
- Waiting Time Reduction

Once these nodes were created in NVivo, we manually analysed the interviews and slotted statement that most closely match the nodes. In choosing the statements for each node, we chose different interviews conducted with different OTs, to eliminate bias. Once this was accomplished, a Cluster Analysis was performed.

The main reason for performing the Cluster Analysis was to ascertain that words that belong to similar themes are grouped together in the same source level. This was accomplished through a dendrogram, as shown below:

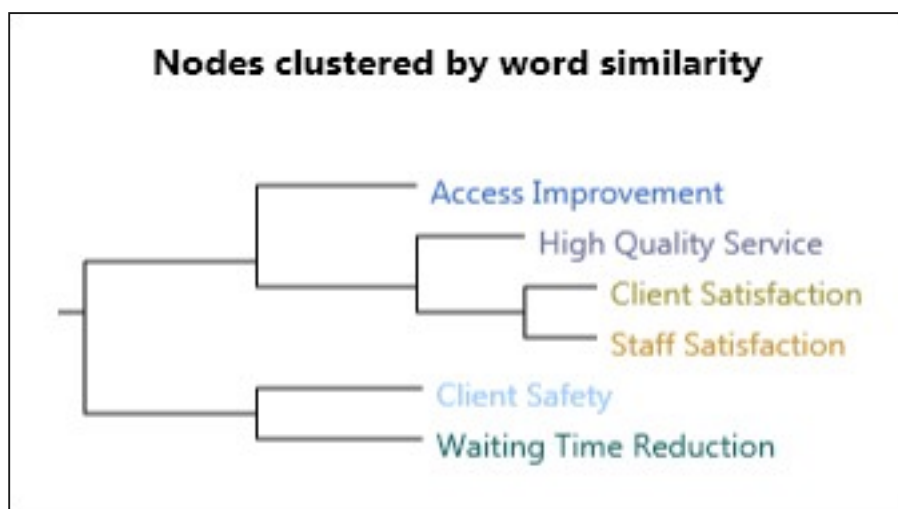


Figure 3: Summary of visual dendrogram of key words

The above diagram is further proof that the data collected for this study is reliable as it can be seen that client and staff satisfaction were grouped into the same cluster. This is as a result of High Quality Service offered to both groups. In the case of clients, qualified allied health staff offered the service, and for these allied health staff, ICT provided the facilitating conditions for using the technology. Similarly, Client Safety and Waiting Time Reduction are grouped together indicating that if the waiting time is reduced to make minor modifications, client safety will improve as there is a direct correlation between these two processes. Access Improvement subsumes High Quality Service as access is important to offer services. In this case, telehealth has provided such access to facilitate services.

6 CONCLUSION

This research paper was written to provide clarifications and guidance in terms of reliability and validity concepts with a case study approach. It demonstrated through a research project conducted in a healthcare organization the established reliability and validity of the data collected in quantity and qualitative methods. Through this research project a new approach was introduced and tested to establish the reliability and validity of qualitative data. This research is unique in its nature and has accurately demonstrated the qualitative data reliability and validity through actual use of technology and the perceived use of the technology.

7 LIMITATIONS AND FUTURE RESEARCH

This is a research of an exploratory nature to demonstrate the reliability and validity of qualitative data in a healthcare environment. Further research is required to establish the findings of this research in order to apply the proposed methodology in healthcares and other domains.

8 ACKNOWLEDGEMENT

We thank Ms Alison Little and her colleagues of Anglicare Southern Queensland (ASQ) for their support and participation in this study. The study was initiated by the ASQ, and their ethics committee approved the data collection techniques employed in this study.

References

- Babour, R. S. (1998). Mixing qualitative methods: Quality assurance or qualitative quagmire? *Qualitative Health Research*, 8(3), 352-361.
- Campbell, T. (1996). Technology, multimedia, and qualitative research in education. *Journal of Research on Computing in Education*, 30(9), 122-133.
- Clont, J. G. (1992). The concept of reliability as it pertains to data from qualitative studies. Paper Presented at the annual meeting of the South West Educational Research Association. Houston, TX.
- Creswell, J. W. & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124-131.
- Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative Health research*, 12(2), 279-289.
- Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research*, 3(3), 118-126.
- Hoepfl, M. C. (1997). Choosing qualitative research: A primer for technology education researchers. *Journal of Technology Education*, 9(1), 47-63. Retrieved February 25, 1998, from <http://scholar.lib.vt.edu/ejournals/JTE/v9n1/pdf/hoepfl.pdf>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17(2), 13-17.

- Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Seale, C. (1999). Quality in qualitative research. *Qualitative Inquiry*, 5(4), 465-478.
- Stenbacka, C. (2001). Qualitative research requires quality concepts of its own. *Management Decision*, 39(7), 551-555
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage Publications, Inc.