

Thematic Analysis Approach: A Step by Step Guide for ELT Research Practitioners

Dr. Saraswati Dawadi

Abstract

The author, having drawn examples from her doctoral study, describes how a thematic analysis approach was employed to interpret raw data in her doctoral study which explored the impacts of a high-stakes test on students and parents in Nepal. As the main purpose of this paper is to provide some guidelines for English language teaching (ELT) practitioners and early career researchers to rigorously apply a thematic analysis approach, this paper presents a step-by-step guideline for the application of the approach. It also presents detailed examples of the processes the author followed during the analysis of her data (from familiarising with the data to identifying initial codes to preparing a final report) to reveal how analysis of the raw data (from interviews and oral diaries transcripts) progressed towards the identification of overarching themes that captured the nature of the test impacts in the Nepalese context described by participants in the study.

Keywords: rigor, flexibility, thematic analysis, qualitative research

Introduction

Thematic analysis is a qualitative research method that researchers use to systematically organise and analyse complex data sets. It is a search for themes that can capture the narratives available in the account of data sets. It involves the identification of themes through careful reading and re-reading of the transcribed data (King, 2004; Rice & Ezzy, 1999). A rigorous thematic analysis approach can produce insightful and trustworthy findings (Nowell, Norris, White & Moules, 2017). However, there is no clear guidance for early career researchers to conduct rigorous thematic analysis. Braun and Clarke (2006) argue that thematic analysis is theoretically flexible for identifying, describing, and interpreting patterns (themes) within a data set in great detail. It fits well with any qualitative study which attempts to explore complex research issues. Indeed, it is so flexible that it “can be incorporated into any epistemological approach” (Chamberlain, 2015, p.68). Nevertheless, there is a potential limitation of thematic analysis, that is., its methodology is not often clearly reported although it has been widely used in

qualitative studies. Highlighting the benefits of using thematic analysis in a qualitative study, Braun and Clarke (2006) argue that this approach makes the analysis more valid because of its accessibility, transparency, and flexibility.

Thematic analysis can be made in both deductive (top-down) and inductive (bottom-up) way (Braun & Clarke, 2006). In the inductive analysis, the data is coded without trying to fit the themes into a pre-existing coding frame or the researcher's preconceptions about the research (Brown & Clark, 2006). So, themes emerge through the data itself without paying attention to the themes included in other studies. Themes are strongly linked to the data instead of the researcher's theoretical interest in the topic. On the other hand, the deductive approach is explicitly researcher-driven allowing the researchers to analyse the data in relation to their theoretical interest in the issues being investigated (Braun & Clarke, 2006). The researcher using this approach usually begins the analysis with the themes that are identified by the researcher through a literature review.

In order to maximise the overall depths of the analysis, both deductive and inductive approaches can be utilised. A deductive approach can be used as the starting point which allows analysing data in relation to the themes that have emerged through the review of literature done for the study or the research questions designed for the study. However, each of the interesting or relevant information (themes) emerging through the data can also be considered. Even the unexpected themes can be taken into consideration for a better understanding of the phenomenon in question. Therefore, a large number of inductive codes may emerge when analysing data.

Thematic analysis is a constant-comparative method that involves reading and rereading the transcripts in a systematic way (Cavendish, 2011) and the most important aspect in the thematic analysis is that the analysis process should be systematic so that the final product is of good quality. In order to maintain necessary rigour in the analysis process, a study can adopt the six-phase process as proposed by Braun and Clarke (2006); each of which is discussed below. Nevertheless, those analytic procedures are not a linear series of steps but rather an iterative and reflective process; it involves a constant moving back and forward between phases.

The application of a thematic analysis approach sounds challenging for early career researchers and/or ELT practitioners. Therefore, this paper aims at presenting a step-by-step guideline for the application of the approach. In the following section, the author describes the steps or the processes that the author followed during the analysis of qualitative data in her doctoral study which explored the impacts of the Secondary Education Examination (SEE) English test on English as a second language (ESL) learners (15 to 16 years old) and parents in Nepal. It is worth pointing out that ELT and testing have a very long history in Nepal but very little is known about testing practices in the Nepalese context. Therefore, the study explored the issues around the SEE English test which is conducted at the end of 10-year school education in Nepal. The study employed a mixed-methods approach and the qualitative data in the study included oral weekly diaries (n=72) intermittently recorded by six students over a year,

and semi-structured interviews with six students and their parents (n=24, one parent for each).

Phase One: Familiarisation with the Data

The first phase (familiarisation with the data), as its name suggests, begins with researchers' interest in familiarising themselves with their data. This phase helps them to figure out the type (and number) of themes that might emerge through the data. Indeed, the phase is crucial as it guides further steps that the researcher may have to carry out to analyse the data in an appropriate way. The following section describes what the current author did during the first phase of data analysis in her doctoral study. At first, all the oral diaries and interviews were transcribed in full to have a sense of how the participants reacted to the test impact issues raised in this study. All the transcripts were transferred into NVivo 10 for the analysis. Then, a repeated careful reading of the transcript was made to read the transcripts as 'things in themselves' (Denscombe, 2007, p.77) and to avoid the influence of the author's prior knowledge and experience in the field. While reading the transcripts, all the interesting information was highlighted; 507 points of interest in total were detected and cross-referenced against the Research Questions (see Table 1). The main purpose of going through all the data in such a way was to become fully immersed in the whole dataset and collect initial points of interest (Chamberlain, 2015). Thus, this step informed the author well about the depth and breadth of the content.

As indicated in Table 1, there were 5 uncertain points of interest. Those points were accepted after the discussion about the points with the supervisors; two were subsumed into the theme of career development, two into test preparation, and one into parental involvement.

Table 1: Phase one: Familiarisation of Data - Points of interest linked to Research Questions

Research questions	Codes	Initial points of interest
What are students' and parents' pre-test and post-test attitudes towards the SEE English test?	Test quality, Test fairness, test accuracy	84
Does the test motivate students to learn English? If yes, how does it affect students' motivation to learn English in the pre-test and post-test context?	Motivation, test preparation strategies	103
How do students prepare themselves for the SEE English test?	Test preparation	113
Do students and parents suffer test pressure and anxiety? If yes, what sorts of pressure and anxiety do they suffer?	Test pressure, test anxiety	101

How do parents involve themselves in preparing their children for the test?		
	Parental involvement	49
What are the impacts of the test on students' career and educational development?	Study at grade 11, learning English after the test, career development	52
Uncertain points of interest		5
	Total	507

Phase Two: Generating Initial Codes

The first phase of the data analysis (i.e. familiarisation with the data) allowed the richness of the initial findings to emerge. However, the importance of rereading the transcripts before creating codes was considered. Therefore, she reread the transcripts carefully and coded all the data. The NVivo coding feature, which is efficient, enabled multiple codes to be applied by selecting phrases or sentences/ paragraphs that were of interest. All the transcripts were coded after reading the transcripts carefully for several times. A large number of codes (n=116) emerged, some containing just one phrase and others containing one or more sentences. Table 2 presents a few examples of how codes were applied to short segments in the data set.

Table 2: Data Extracts and Codes

Data extracts	Coded for
My mom is very much worried about me and she has a hope that I can do well on the test.	Test anxiety on parents
I am scared of the test.	Test anxiety on students
I am also worried that there might be some carelessness when checking our answer sheets.	Test anxiety
Test quality	
If I do well on the test, people think that I am a smart girl and I will be praised by them. All the people in my village will know that I have done well on the test. So, the way they look at me will be different. I also think that they will present me as an example to other students for encouraging them to work hard and do well on the test.	Test and social prestige
I understand that I must try my best to learn English and do well on the test. My parents also always tell me that I must practise hard for the test. So, I am working hard these days.	Motivation to learn English.
Test pressure	
I have told her that the SEE is an iron gate for her. If she cannot do well on the test, her future will be dark.	Test importance- parents' view

My mother has also guessed some of the important questions, especially essay topics, for the SEE and she has asked me to write the answers of those questions.	Parental support
I memorised a lot of answers for the test.	Memorisation

The above table indicates the sorts of operations involved in the data coding process. In order to have an overall picture of the codes, all the 116 codes, along with some relevant extracts, were exported from the NVivo and presented on a table. The table supported her to further understand the nature of the data in the study.

Phase Three: Searching for Themes

This phase, as suggested by Braun and Clarke (2006), began with a long list of the codes that were identified across the data set. The main purpose of this phase was to find out the patterns and relationships between and across the entire data set (Chamberlain, 2015). The codes had to be analysed considering how different codes could be combined to form an overarching theme (Braun & Clarke, 2006). In other words, the major focus in this step was on the analysis at the broader level of themes, rather than codes. As Brown and Clarke (2006) point out “a theme captures something important about the data in relation to a research question and represents some level of patterned response or meaning within the data set” (p.10). Therefore, it was important to conceptualise those codes as the building-blocks and combine similar or multiple codes to generate potential themes in relation to the research questions (Ansari, 2015).

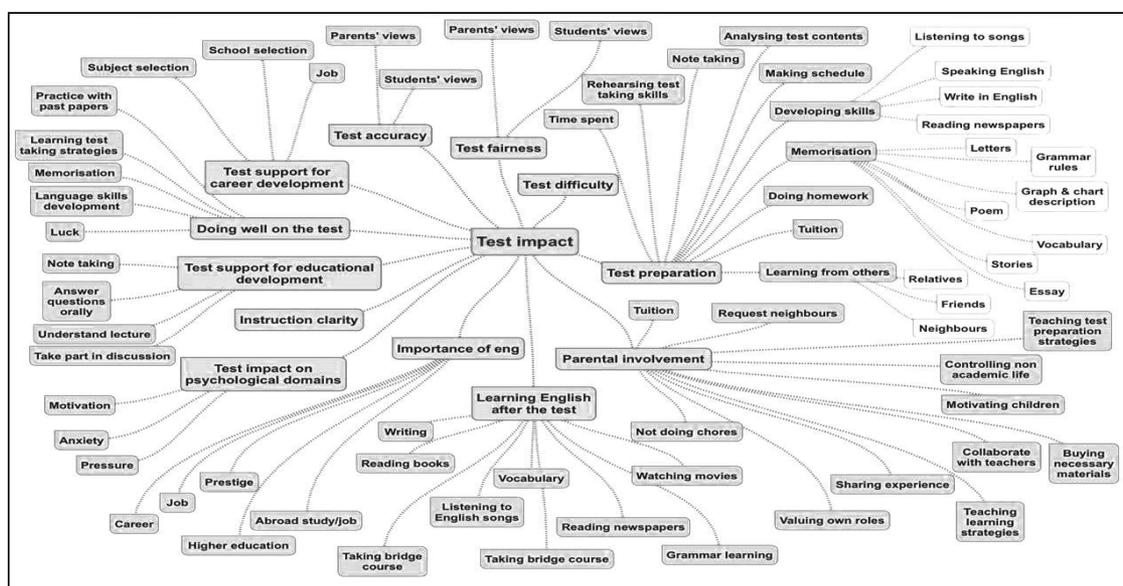
This phase was the most difficult phase in the analysis process. In order to ease the process, following Braun and Clarke’s (2006) suggestions, a list of the codes was prepared on a separate piece of paper and then they were organised into theme-piles which reflected on the relationship between codes and themes. Because of the explorative nature of the study, it was also important to return to and re-read all the transcripts before clustering codes according to the themes. Thus, the transcripts were re-read and different codes were combined into potential themes, collating all the relevant coded data extracts within the identified themes. When developing the themes, the author could bring in the concepts and issues that she had previously identified in her literature review. She found that some of the themes from the literature review were truly meaningful and some codes could be subsumed under them.

Braun and Clarke (2006) suggest that themes in a study should be prevalent in most or all of the data items. However, any sort of relevant information, though it appeared in a few sources, was considered in this study. In order to cluster all the codes, a thematic map was initially created (displayed in Figure 1) which contained 12 overarching themes (namely: test fairness, test accuracy, test difficulty, test support for educational development, test support for career development, doing well on the test, instruction clarity, psychological domain, importance of English, learning English after the test, test preparation, and parental involvement). As the main purpose of creating main

themes or categories was to capture the essence of the clustered codes, the main code would include all the related codes. For instance, the main code, test preparation would contain all the codes and sub-codes aimed at capturing students' strategies to prepare themselves for the test preparation. It was found that all the codes were somehow connected to one of the main codes.

As seen in Figure 1, the first thematic map was huge because it included 12 main themes and 54 sub-themes along with their 14 lower-level codes that initially emerged through the data. All these initial themes were further refined at the next stage of the analysis. The process of refinement in the phase of the analysis is explained in the next section.

Figure 1: Initial thematic map



Phase Four: Reviewing Themes

At this stage, all the themes (master themes, main themes and sub-themes) were intentionally brought together as it was aimed at the refinement of those initially grouped themes and presentation of those themes in a more systematic way. Braun and Clarke (2006) suggest that themes must be checked for internal homogeneity (coherence and consistency) and external heterogeneity (distinctions between themes).

This stage consisted of two levels. At level one, all coded extracts relevant to each initial theme were extracted from the NVivo file and pasted into a Microsoft Word document to facilitate cross-referencing of coded extracts with the themes and to carry out the retrieval, comparison and organisation of coded extracts and themes in a meaningful way. The author reread all the collated extracts for each theme, clustered all the themes and sub - themes to check whether they could form a coherent pattern. All the codes

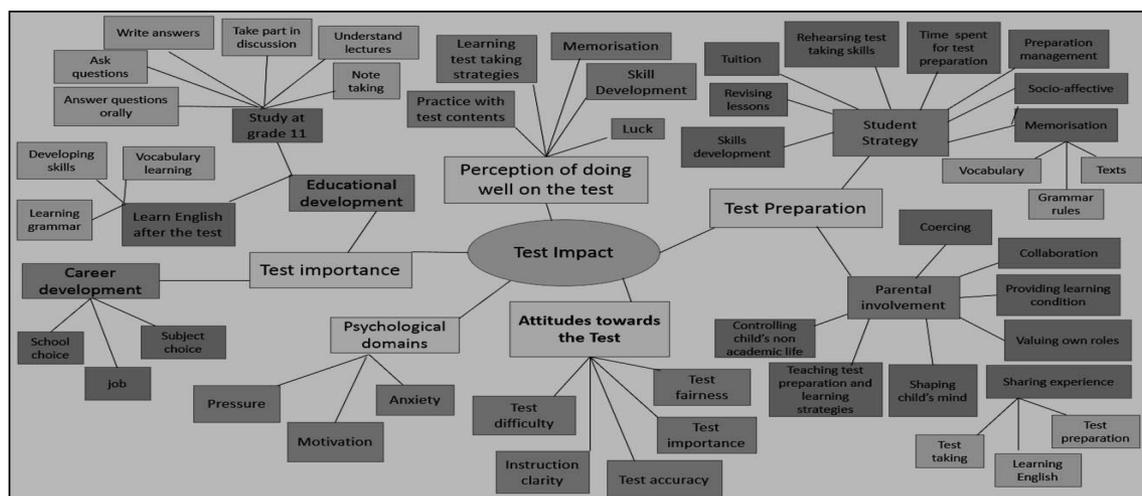
and themes along with the collated extracts were considered to see whether they could form a coherent pattern adequately capturing the contours of the coded data.

At level two, a similar process was followed but in relation to the entire data set. At this level, the validity of individual themes in relation to the data set was considered. It was very important to ascertain that the “thematic map ‘accurately’ reflects the meanings evident in the data set as a whole” (Braun & Clarke, 2006, p.91). Therefore, all the transcripts were reread, (where appropriate, the extracts were also recoded) to ensure that the themes ‘work’ in relation to the entire data set. Some new codes emerged at this stage. Nonetheless, the last few codes did not add anything substantial. That is why, it was decided to stop recoding the data. Then, all the themes were put back together and the thematic map was refined which could reflect on the type of themes developed for the study, how the themes in the study fit together and the overall story the themes tell us about the data.

During the reviewing process, many of the themes or sub-themes were either merged with other (main) themes or discarded. For instance, ‘taking a bridge course’ did not appear to belong to any thematic category. Similarly, the theme ‘importance of English’, which contained five elements, was later considered not to be directly related to the objective of this study. Therefore, those two themes (taking a bridge course and importance of English) were later deleted on the ground that they were not directly relevant to the study. Similarly, ‘making schedule’ was found to have little data to stand as a separate sub-theme. There was only one student who made a daily schedule following her parents’ suggestion, so it was merged with ‘time spent’. Some new themes were also introduced to merge related themes. For instance, one new theme ‘memorising text’ was introduced to include four sub-themes: memorising stories, essays, letters, and description. Furthermore, five of the main themes: test fairness, test importance, test accuracy, instruction quality and test difficulty were merged in a new theme ‘perceptions of the test’. Moreover, since two of the main themes ‘educational development’ and ‘career development’ showed similar patterns, they were brought together within a new name as ‘test importance’.

Other themes and/or sub-themes were also reviewed, renamed, discarded or merged in the same way. The outcome of the whole process of revision is set out in Figure 2. Having clustered the themes together, five different categories/codes emerged: Test preparation, perceptions of doing well on the test, psychological domains, test importance and parental involvement (see Figure 2). Those five themes (in yellow colour in the Figure) were the master themes and 17 main themes (blue colour) subsumed under those master themes. Among them, four had several sub-themes (brown colour) and four of the sub-themes had also some lower level codes (green colour). Figure 2 captures all of them.

Figure 2: Revised thematic map



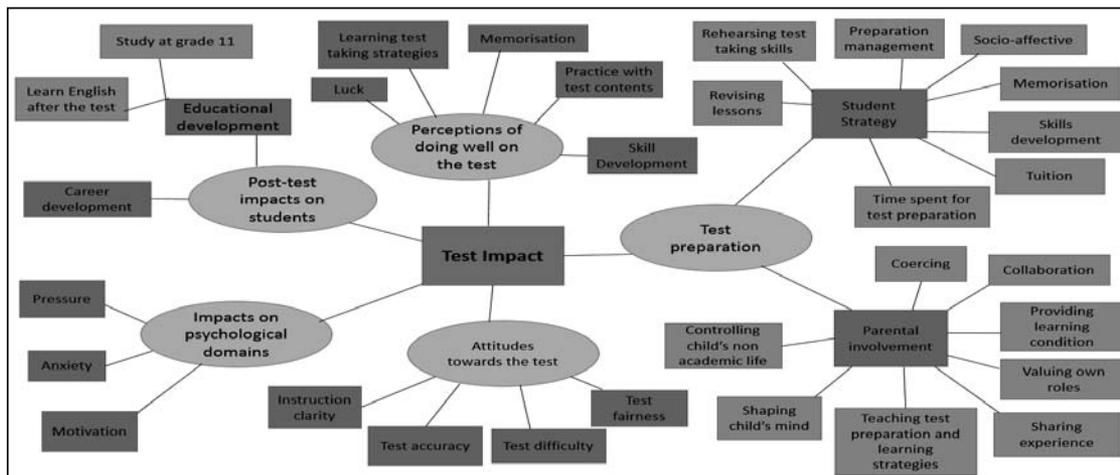
Phase Five: Defining and Naming Themes

This phase began with an aim of further refining and defining the themes, that is, “identifying the essence of what each theme is about (as well as the themes overall), and determining what aspect of the data each theme captures” (Braun & Clarke, 2006, p.92). Braun and Clarke argue that a theme cannot be too diverse and complex. Therefore, the author went back to collated data extracts for each theme and organised all the themes into a coherent and consistent account. Careful attention was paid to identify the ‘story’ that each theme told, and how it fitted into the broader overall ‘story’ that she wanted to tell about her data in relation to the research questions and to ensure that there was not too much overlap between the themes. The specifics of each theme were refined carefully.

The themes were further refined by reading through all the main themes and sub-themes, codes and extracts. Then, final name along with its definition was assigned to each theme to tell a story about the data. In this stage, some of the lower level themes were merged with higher-order themes as it was realised that those lower level themes would make the thematic map more complex and also add little to the story told by the data. For instance, the three lower level themes (job, school choice and subject choice) of the sub-theme ‘career development’ were merged in it. Furthermore, one of the sub-themes (i.e. tuition) of the theme ‘student strategy’ was considered to be a common sub-theme of the two main themes, student strategy and parental involvement, as it was found that both the students and parents followed this strategy. Similarly, the sub-theme ‘time spent on test preparation’ was considered to be a part of student test preparation strategy. However, after reading the extracts, it was realised that ‘time spent on test preparation’ was not clearly a strategy for the test preparation, rather it was related to the amount of time spent for the test preparation. So, it was treated as a

separate theme. The final mind-map for the entire dataset resulted from this phase has been displayed in Figure 3. This has been interpreted to report the qualitative findings in her thesis.

Figure 3: Final thematic map



Phase Six: Writing Report

The final phase of the analysis was to write down the report of the findings. Braun and Clarke (2006) state that report of a thematic analysis must convince the readers of the merit and validity of the analysis. Therefore, a great effort was made to provide a concise, coherent and logical account of the story that the data represented within and across themes by providing sufficient evidence and particular examples and/or extracts which could capture the essence of the point the author was demonstrating. The examples and extracts were embedded within the analytic narrative in such a way that they could make an argument in respect of the research objectives, besides illustrating the story being told.

Practical Implications

Teachers in this present era are (need to be) research oriented. It is crucial that they “have the necessary pedagogical skills and competency to improve not only students’ performance but also their ability to critically think, generate new knowledge and innovations” (Basaffar Almasri, & Almasri, 2017, p.171). In other words, their classroom practices have to be informed (or guided) by their research. As rightly pointed out by Rosenshine (2012), teachers can only come up with innovative teaching techniques or strategies if they have adequate research skills to critically analyze their own teaching strategies/practices and identify problems. This might indicate that teachers have to frequently conduct research (particularly action research) to identify the learning needs of their students and introduce remedial teaching (or to solve teaching related problems

in their local contexts). Indeed, the present scenario indicates that action research has a momentum in ELT; more and more teachers have been making efforts to make research-based remedial teaching plans in order to promote students' learning. However, many teachers lack research skills (Sahan & Tarhan, 2015) and they usually find it difficult to analyse their classroom data. Indeed, analysing qualitative data can present a big challenge even to experienced researchers, let alone early career researchers and/or ELT practitioners. It is hoped that ELT practitioners (and early career researchers) might find this paper useful for them as it provides them a step-by-step guideline to analyse data in a qualitative study.

References

- Basaffar, Almasri & Almasri (2017). EFL university teachers' professional development through a research-oriented training program. *Journal of Education & Social Policy*, 7 (1), 171-177.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Cavendish, L. M. (2011). *Stories from international teachers: A narrative inquiry about culturally responsive teaching*. An unpublished PhD thesis, The University of Iowa, Iowa.
- Chamberlain, L. (2015). *Exploring the out-of-school writing practices of three children aged 9 - 10 years old and how these practices travel across and within the domains of home and school*. An unpublished PhD thesis, The Open University, England.
- Denscombe, M. (2007). *The good research guide: For small-scale social research projects* (3rd edition). Maidenhead: Open University Press.
- King, N. (2004). Using templates in the thematic analysis of text. In C. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research* (pp. 257-270). London, UK: Sage.
- Nowell, L. S. Norris¹, J. M., White D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16, 1-13.
- Rice, P., & Ezzy, D. (1999). *Qualitative research methods: A health focus*. Melbourne: Oxford University Press.
- Rosenshine, B. (2012). *Principles of instruction: Research-based strategies that all teachers should know*. Spring Publishers. Retrieved from: <https://www.aft.org/sites/default/files/periodicals/Rosenshine.pdf>
- Şahan, H. H. & Tarhan, R. (2015). Scientific research competencies of prospective teachers and their attitudes toward scientific research. *International Journal of Psychology and Educational Studies*, 2(3), 20-31.

Contributor: Dr. Saraswati Dawadi has earned her doctorate in Language Assessment from The Open University, UK. She received the Hornby Trust Scholarship in 2013/2014 to study MA: TESOL at Lancaster University, UK. Her research interest sits broadly within language assessment, work-based learning, second language acquisition and equitable access to quality education.