
6-22-2022

The Craft of Using NVivo12 to Analyze Open-Ended Questions: An Approach to Mixed Methods Analysis

Jane Elliott

University of Exeter, UK, Jane.Elliott@exeter.ac.uk

Follow this and additional works at: <https://nsuworks.nova.edu/tqr>



Part of the [Family, Life Course, and Society Commons](#), [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#), and the [Social Statistics Commons](#)

Recommended APA Citation

Elliott, J. (2022). The Craft of Using NVivo12 to Analyze Open-Ended Questions: An Approach to Mixed Methods Analysis. *The Qualitative Report*, 27(6), 1673-1687. <https://doi.org/10.46743/2160-3715/2022.5460>

This How To Article is brought to you for free and open access by the The Qualitative Report at NSUWorks. It has been accepted for inclusion in The Qualitative Report by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.



The Craft of Using NVivo12 to Analyze Open-Ended Questions: An Approach to Mixed Methods Analysis

Abstract

The aim of this paper is to explore and explicate a method for using NVivo12 to analyze the responses to open-ended questions in surveys. It focuses on the features of NVivo12 that facilitate this particular type of mixed-methods research. Open-ended questions are used within surveys across a wide variety of disciplines and provide opportunities for participants to provide their own perspectives on a topic of interest. However, they can be challenging to analyze; if the survey has a large sample size, a very considerable corpus of text will be generated, yet the text associated with any single individual is likely to be modest and lack the context and richness usually associated with qualitative research. The paper builds on a recent contribution by Feng and Behar-Horenstein (2019) and demonstrates how the new features of NVivo12 specifically assist mixed-methods research by adding a new Crosstab query function and making it easier to export and import data directly from SPSS.

Keywords

NVivo, qualitative, surveys, analysis, CAQDAS, future

Creative Commons License



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Acknowledgements

I am grateful to Jon Lawrence and Jennifer Wilcox for their helpful comments on this paper. Special thanks are due to the members of the 1958 British Birth Cohort Study who have provided detailed information to the study over many years, and to the team at the Centre for Longitudinal Studies who manage the study and document the data.

The Craft of Using NVivo12 to Analyze Open-Ended Questions: An Approach to Mixed-Methods Analysis

Jane Elliott

University of Exeter, United Kingdom

The aim of this paper is to explore and explicate a method for using NVivo12 to analyze the responses to open-ended questions in surveys. It focuses on the features of NVivo12 that facilitate this particular type of mixed-methods research. Open-ended questions are used within surveys across a wide variety of disciplines and provide opportunities for participants to provide their own perspectives on a topic of interest. However, they can be challenging to analyze; if the survey has a large sample size, a very considerable corpus of text will be generated, yet the text associated with any single individual is likely to be modest and lack the context and richness usually associated with qualitative research. The paper builds on a recent contribution by Feng and Behar-Horenstein (2019) and demonstrates how the new features of NVivo12 specifically assist mixed-methods research by adding a new Crosstab query function and making it easier to export and import data directly from SPSS.

Keywords: NVivo, qualitative, surveys, analysis, CAQDAS, future

Introduction

This paper contributes to the methodological literature on how to use NVivo most effectively to analyze the responses to open-ended questions in surveys. It builds on a recent contribution by Feng and Behar-Horenstein (2019), remedies some of the minor errors in their paper, and highlights the updated features in NVivo that assist with mixed-methods research. At a practical level, my aim is to draw attention to the importance of the appropriate use of “stop-words” within NVivo which may otherwise obscure potentially interesting analytic insights. I also highlight the advantage of using the new Crosstab query facility, which is now available in NVivo12. I demonstrate how this is an improvement on the Matrix query advocated by Feng and Behar-Horenstein (using NVivo 11). I conclude by exploring the advantages of using NVivo12 for analysing qualitative data from relatively large samples of individuals in comparison with other approaches to analysis. The paper is mainly aimed at researchers who have some existing familiarity with NVivo or another software package that assists with analysis of qualitative data. My purpose is to demonstrate how these types of software packages can facilitate the analysis of open-ended questions in surveys and how this can be considered a type of mixed-methods research.

My interest in this area of methodology stems from my research work on the 1958 British Birth Cohort study. This is a longitudinal cohort study of all those individuals born in Britain in one week of March 1958. This was originally known as the National Child Development Study, and now as the 1958 British Birth Cohort Study. The original study collected data on a sample of over 18,000 individuals, who have been surveyed periodically from age seven onwards (Power & Elliott, 2006). The majority of the data collected are highly structured and consist of responses to fixed choice questions. However, in recent years there has been more attention paid to the value of including some more qualitative elements to the

study (Elliott, 2008). As a sociologist, I have a specific interest in how people understand their own lives and the influences upon them.

Open-ended questions (OEQs) are routinely used as part of large-scale surveys across a wide variety of disciplines. OEQs may be included in social surveys to capture dimensions not covered by fixed-choice responses or anticipated by the survey designer, to enable respondents to provide an explanation to clarify the option they have selected from a given response set, or even to make respondents feel that the survey has a more human face (Fielding et al., 2012). OEQs can be a very helpful methodological tool within survey research (for a review see Singer & Couper, 2017). In summary, they provide an opportunity for participants to provide their own perspectives on a topic of interest rather than only responding via a pre-determined set of options. In addition to being frequently used in academic research, OEQs are often used in practical settings such as when evaluating services or educational interventions. An insightful example of the analysis of an OEQ to evaluate the National Health Service in the UK is provided by Brookes and Baker (2017), who found that patients' evaluations of staff members working in the health service focussed primarily on their interpersonal skills rather than their technical skills.

The data used as an exemplar in this paper originates from a self-completion survey that forms part of the 1958 British Birth Cohort Study. At age 50, in 2008, 8844 individuals¹ completed a written questionnaire comprised of mainly multiple choice and fixed choice questions (Elliott & Brown, 2011). At the end of the questionnaire, respondents were asked an OEQ, namely, to imagine their lives one decade later: "Imagine that you are now 60 years old ... please write a few lines about the life you are leading (your interests, your home life, your health and wellbeing and any work you may be doing)" (Elliott, 2012). This final OEQ was answered by 7,383 cohort members. The answers were transcribed and anonymized before being archived, and the data is available from the UK Data Archive (Study Number 6978, University of London, Institute of Education, Centre for Longitudinal Studies, 2012).

There are two main advantages of using these data as an exemplar. First, the topic is readily accessible and of potential interest to a wide audience. The study itself is multidisciplinary and asking individuals at mid-life to speculate on an imagined future generates evidence of relevance to scholars from several disciplines including economists, epidemiologists, psychologists, and sociologists (Andrews, 2018; Crow & Andrews, 2019). Second, the data have recently been analyzed using natural language processing techniques (Weber, 2021). This makes it possible to draw some preliminary methodological comparisons between the semi-automated analysis facilitated by NVivo12 and the more fully automated analysis characterised by using two text-analysis methods with the statistical computing software R: relative frequency analysis and topic modelling (Weber, 2021).

For the purposes of this methodological paper, the focus is on the sub-set of responses from individuals whose economic activity status at age 50 was recorded as "permanently sick or disabled." In total, 481 individuals who were part of the age 50 sweep of the 1958 cohort study were classified as permanently sick or disabled (i.e., 4.9%). Of these, 383 returned the self-completion questionnaire, and 287 (135 men and 152 women) completed the OEQ about the future.² The textual data from these OEQs will be used to demonstrate how NVivo12 can be used to assist with analysis.

¹Due to mortality and attrition, the age-50 sweep collected data from 9790 individuals using a face-to-face interview. Prior to the interview, individuals were sent a questionnaire by post, and this was collected by the interviewer.

²As might perhaps be expected, this group is therefore less likely to have completed the OEQ than other groups. One interpretation here is that some individuals within this group had such severe health problems that they felt very pessimistic about the future and preferred not to write about a future imagined life.

Using Computer Assisted Qualitative Data Analysis Software (CAQDAS) for Mixed-Methods Research

There are a number of debates about the value (and dangers) of using CAQDAS for the analysis of qualitative material. Two repeated criticisms of computer assisted approaches have been that first, CAQDAS coding leads to the fragmentation of text whereas researchers should be encouraged to take a wholistic view, and second, CAQDAS inevitably and negatively interferes with the researchers' connection to the data (Jackson et al., 2018). The aim of this methodological paper is to demonstrate how NVivo can be used to aid the analysis of OEQs in a way that addresses these two criticisms.

The version of NVivo used in this paper is NVivo 12. This is a widely used CAQDAS tool that has been developed over many years in collaboration with users (Wolski, 2018). NVivo 12 is very similar to the previous version but has additional features that specifically facilitate mixed-methods research. For example, there is a new crosstabs query which improves on the previous matrix coding query. This will be discussed in more detail below. It is also now more straightforward to import and export directly between NVivo12 and SPSS. This is helpful when analysing OEQs from surveys, particularly if the aim is to integrate the analysis of the quantitative and qualitative material. Other CAQDAS packages could also be used to facilitate analysis of qualitative material alongside more structured quantitative data. For example, MAXQDA (<http://www.maxqda.com>) and QDA Miner (<http://www.provalisresearch.com>) allow users to combine quantitative variable data within the qualitative database and then perform matrix-based analyses of coded text (Bazeley, 2006, 2009). Previous authors have also advocated the use of CAQDAS because it encourages "reflexive moments" during analysis because researchers can use program features to document their processes and decisions, which in turn encourages reflections on interpretive conclusions (Woods et al., 2016).

My aim in writing this paper is to share my experience of using NVivo for the analysis of open-ended questions and to provide other researchers with a practical guide to the process. I have therefore structured my presentation as a series of five practical steps. This updates, corrects, and extends the work of Feng and Bohar-Horenstein (2019), who focus on the use of an earlier version of software (NVivo11) to analyse OEQs.

Step 1: Understanding the Format of Data

The focus of this paper is on analysing responses to OEQs in surveys, and the format of the data is therefore likely to be straightforward, with a row for each unit of analysis (in this example, an individual responding to a self-completion survey at age 50) and a column for each variable or "response." The resulting rectangular data set is relatively easy to interrogate and manipulate in a spreadsheet such as Excel. The techniques described here could be extended to any set of data structured in a similar way. Any large set of short texts, where each text has associated with it some structured data would be amenable to the same analytic approach. For example, a set of tweets or other social media posts could be the focus for analysis, rather than responses to OEQs³.

In the current example, the text responses in the subset of data vary in length from just a single word to a maximum of 235 words and the average (mean) length of response was 47 words (median length 35 words).⁴ An example of a typical response is as follows:

³A recent paper, for example, uses NVivo in tandem with a social media analytics tool, Crimson Hexagon, to analyse a set of tweets about COVID-19 posted by medical professionals (Wahbeh et al., 2020).

⁴When opening exported records from responses to OEQs in text format in Excel, it may not bring all contents to the spreadsheet, or one field may be split across many cells. This is a limitation of Excel. It is important to be aware that in general format, Excel automatically alters the length to just 255 characters per cell, but 32,767

My daughter has been married for nine years and has two children, whom I think the world of. My general health has improved, and I am living life to the full. I do a lot of travelling with my husband and have visited lots of interesting places (N10272K).

As can be seen from panel 1 below, each individual is uniquely identified by a serial number (NCDSID) comprised of letters and numbers. In order to allow for analysis of themes in the responses by some key demographic variables, these variables are also included in the spreadsheet (for example, gender, whether living with a spouse or partner, number of any children in the household, tenure of home, self-reported general health status). The full age 50 survey of the 1958 British Birth Cohort Study includes hundreds of variables; however, NVivo is primarily for the analysis of qualitative data, and it would therefore be very cumbersome to try and import a spreadsheet including the whole dataset. I have found that a good strategy is to select a few key variables likely to be of most interest analytically. Following preliminary analysis, further variables can be imported into NVivo, if necessary, and matched with the initial dataset using the unique case identifier. Alternatively, codes can be exported out of NVivo and analyzed using a quantitative package such as SPSS, Stata, or R. A key improvement in NVivo12 is that this updated version now makes it possible to export codes that have been developed analytically using the qualitative text directly into SPSS, facilitating mixed-methods research.

NCDSID	WORDCNT	IMAGINE	Gender	Partner	Children	Tenure	Health
N10101R	8	I don't imagine I'll ever reach 60 years	2	0	0	4	3
N10108Y	22	My life is very pleasant. Although my he	2	1	2	1	5
N10210V	0		2	0	0	4	3
N10221Y	66	Now that the grandchildren have grown	2	1	0	2	4
N10229G	0		2	1	1	2	5
N10272K	46	My daughter has been married for 9 yea	2	1	1	2	4
N10291N	0		2	1	1	2	5
N10357P	40	If I was about 60yrs I would like to carry	2	1	0	4	5
N10388W	0		2	1	0	2	3
N10393T	32	I live alone im a widow I like living alone	2	0	0	4	4
N10414F	66	IF IN TEN YEARS TIME WHEN I AM 60, A	1	1	1	1	4

Panel 1: Extract from the original data in Excel

Step 2: Data Import and the Data File Within NVivo12

Once the data has been configured within Excel, it is straightforward to import it into NVivo12 using the *Survey Import Wizard*. It is worth noting that the imported data forms a *single* file, with each row of data corresponding to a case for analysis. This can seem somewhat counterintuitive for users of NVivo who are accustomed to projects comprised of multiple files where each file is (usually) a much longer piece of text. For example, I have used NVivo in the past to analyze transcribed interviews with individuals where each individual is represented within the software by a file containing thousands of words. For this type of project, the number of files within the project is equivalent to the number of individuals who have been interviewed. Panel 2 shows the data from Panel 1 imported to NVivo12 and it can be seen that the answers to the OEQ, “Imagine you are 60,” appear in the fourth column.

characters per cell in Text format. This should be plenty for most responses to OEQs. The longest response in the current example was 235 words (i.e., 1321 characters).

ID	NCDSID	WORDCNT	IMAGINE	Gender	Partner	Children	Tenure	Health
1	N10101R	8	I don't imagine I'll ever reach 60 years	2	0	0	4	
2	N10108Y	22	My life is very pleasant. Although my health could be better, my family, home life and friends make my life very happy.	2	1	2	1	
3	N10210V	0		2	0	0	4	
4	N10221Y	66	Now that the grandchildren have grown up I have more time for myself, my husband has slowed down and we have managed to get a retirement plan ready. and we are spending more time together. I have managed to raise more money for cancer research I am coping better with my health maybe medical research has produced medication in which I can manage my pain better.	2	1	0	2	
5	N10229G	0		2	1	1	2	
6	N10272K	46	My daughter has been married for 9 years and has two children, whom I think the world of. My general health has improved and I am living life to the full. I do a lot of travelling with my husband and have visited of interesting places.	2	1	1	2	
7	N10291N	0		2	1	1	2	
8	N10357P	40	If I was about 60yrs I would like to carry on the way I am, to see grandchildren & there mom and dads have a nice home and help with gardening when I could health	2	1	0	4	

Panel 2: The imported data within NVivo12

Step 3: Word Frequency Analysis and the Use of Stop Words

The word count function within NVivo12 is an excellent tool to facilitate preliminary analysis by assisting the researcher to identify key themes within the data. Feng and Behar-Horenstein suggest limiting the word count analysis to words with five or more letters because “words such as ‘him,’ ‘was,’ and ‘I’ do not increase the researcher’s understanding of a phenomenon” (Feng & Behar-Horenstein, 2019, p. 566). However, this limit of five letters is somewhat arbitrary and, depending on the subject of the survey, there may be many four-letter words that are frequently used and help to identify the main topics within the data. For instance, in the current example, words such as “work,” “hope,” and “good” were all of substantive interest. I would therefore advocate initially looking at *all* of the most frequent words and then deciding on an appropriate cut off for word length, or simply using the list of most frequent words as a guide for which topics to pursue for further analysis.

In addition, it is worth noting that NVivo12 (and previous versions) provides default “stop words” for major languages such as Chinese, English (UK), English (US), French, German, Japanese, Portuguese, and Spanish. “Stop words” are words like conjunctions or prepositions. As Feng and Behar-Horenstein note, these are often considered less significant or meaningful for analysis. However, in some contexts, use of these words may be of analytic interest. For example, there is a strand of work that specifically focuses on analysis of the use of pronouns to understand how identities are constructed through discourse (Schriffin 1996, Van de Mieroop, 2005). It is therefore worth viewing the stop words associated with the language you are using, and then customising the default list appropriately.⁵

⁵Customizing stop-words can be done by Clicking the **File** tab and then clicking **Project Properties**. On the **General** tab, then click the **Stop Words** button. The **Stop Words** dialog box opens. Add or remove words from the list. Each word must be separated by a space.

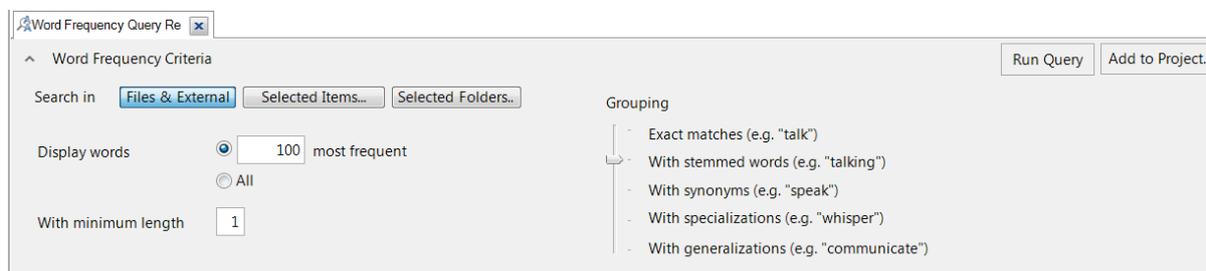
*a about above after again against all am an and any are aren't as at be because been before being below between both but by can can't cannot can't could couldn't did didn't do does doesn't doing don't down during each few for from further had hadn't has hasn't have haven't having he he'd he'll he's he'd he'll her here here's hers herself he's him himself his how how's ii'd i'll i'm i've i'd if i'll i'm in into is isn't it it's its itself I've let's me more most mustn't my myself no nor **not** of off on once only or other ought our ours ourselves out over own said same say says shall shan't she she'd she'll she's she'd she'll she's should shouldn't so some such than that that's the their theirs them themselves then there there's these they they'd they'll they're they've they'd they're they've this those through to too under until up upon us very was wasn't we we'd we'll we're we've were weren't what what's when when's where where's which while who whom who's whose why why's will with won't would wouldn't you you'd you'll you're you've your yours yourself yourselves*

Panel 3a: List of stop words in UK English

As an example of the need to customize the list of stop words, it can be seen from panel 3a that the word “not” is included. However, in many research projects, respondents use of the word “not” might be analytically interesting as it potentially conveys negative sentiment. In the current example, the word “not” was removed from the list of stop words. This meant that it was included in the word frequency query and could also be used as a search term in the *Explore* function. This will be demonstrated and discussed in more detail below.

As can be seen in panel 3b below, NVivo allows the user to look at the occurrence of individual words (“exact matches”), or helpfully allows for “stemmed” words to be grouped together within the word count. In the present example this means that the words “@hope, @hopefully, hope, hopefully, hoping” are appropriately grouped into a single analytic category (the @ sign was used in the transcription to indicate use of a bullet point by a respondent).

Panel 3c shows the most frequent words identified by NVivo in the corpus of responses. It is intuitive to use those words and concepts that appear most frequently (e.g., have a weighted percentage higher than 0.5%) to structure further analysis of the data. However, it is also instructive to look more broadly at a list of the 500 most frequent words to inform further coding. For example, in the “Imagine you are 60” OEQ, individuals were prompted to write about their interests. Responses therefore included terms such as fishing, reading, dancing, gardening, and cooking. None of these individual activities were mentioned very many times; however, as examples of different types of activities, they were of interest in coding.



Panel 3b: The word count facility

Rank	Word	Length	Count	Weighted Percentage (%)	Similar Words	cum %
1	health	6	146	2.27	@health, health	2.27
2	life	4	144	2.24	life, lifes	4.51
3	still	5	138	2.15	@still, still	6.66
4	hoping	6	135	2.10	@hope, @hopefully, hope, hopefully, hoping	8.76
5	working	7	100	1.56	@work, @working, work, worked, working	10.32
6	now	3	83	1.29	now	11.61
7	time	4	71	1.11	time, times	12.72
8	years	5	70	1.09	year, years	13.81
9	living	6	69	1.07	@living, live, lived, liveing, lives, living	14.88
10	enjoying	8	66	1.03	@enjoying, enjoy, enjoyable, enjoyed, enjoying	15.91
11	60	2	64	1.00	60	16.91
12	going	5	61	0.95	@go, @going, go, going	17.86
13	family	6	56	0.87	families, family	18.73
14	good	4	55	0.86	good	19.59
15	like	4	55	0.86	like, liked, likely, likes	20.45
16	able	4	51	0.79	able	21.24
17	get	3	50	0.78	get, getting	22.02
18	homes	5	47	0.73	@home, home, homes	22.75
19	grandchildren	13	46	0.72	grandchildren	23.47
20	happy	5	46	0.72	happiness, happy	24.19
21	love	4	45	0.70	love, loved, loveing, lovely, loves, loving	24.89
22	looks	5	44	0.68	@looking, look, looking, looks	25.57
23	interests	9	42	0.65	@interests, interest, interesting, interests	26.22
24	children	8	42	0.65	children	26.87

Panel 3c: The most frequent words in the corpus

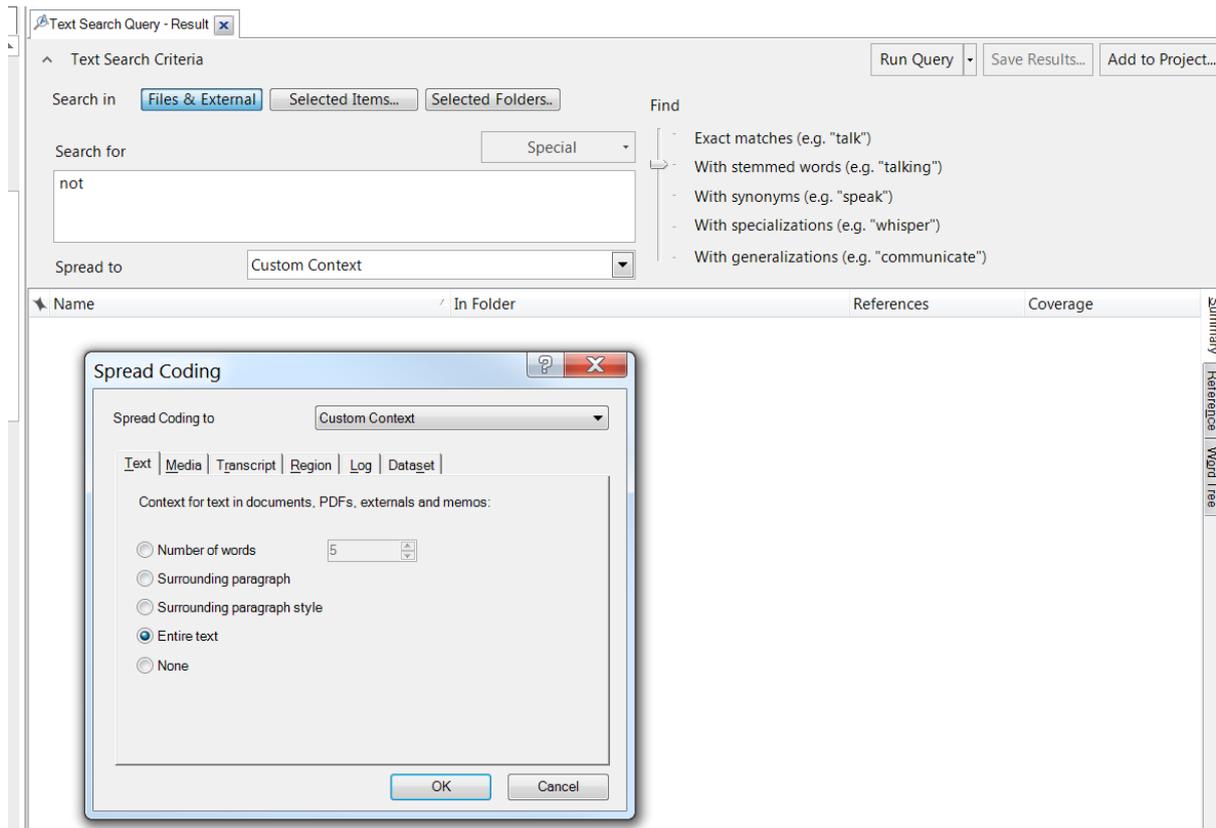
Step 4: Text Coding, Refining the Automatic Coding, and Developing Categories or Meta-Codes

Once a list of frequently used words has been created using NVivo, this can be used to create a new set of nodes by running a series of automatic word searches and saving the results as new nodes. This is accomplished using the “Text search” feature in NVivo. When using this function, it is helpful to be aware that use of the Text search feature with the default settings will only code the word itself, removing the context when you view the results of a crosstab query (see below). I therefore recommend coding the whole text of the response that contains the word of interest; this is shown in Panel 4a. This is specifically applicable to analysis of OEQs, where each text is likely to be short.

Having completed the automatic coding using a text search query, it is then important to clean and refine the contents of the code so that you have an understanding of what it means

in practice. For example, in the analysis of individuals' expectations about life at 60, the word "sport" was used a few times. When the responses that mentioned sport were examined in more detail, it was found that they all referred to *watching* sport rather than *playing* sport. For example:

The life I lead is boring and mundane. @My interests are reading and watching sport on T.V. @My home life is O.K. seeing the grandchildren is a pleasure and am still married after nearly 40 years together. @Health is not bad, but disability is getting worse. I am in a wheelchair permanently. I am not doing any work (N14117R).



Panel 4a: Using the Text Search function in NVivo to create new nodes

The initial individual codes created using the text search query can be thought of as a way of indexing the data (Elliott, 2018). These granular codes can then be refined or aggregated to form parent codes (or categories). Returning to the example of the interests that respondents wrote about when imagining life at age 60, these could be aggregated into two categories: "active interests" and "passive interests" (see Panel 4b), with "active" interests defined as those which required more physical mobility. This process of aggregating individual codes (which simply index or label a portion of text) into more thematic or conceptual categories has been discussed in detail by other authors (Elliott, 2018). What is key here is that while NVivo12 simplifies the practical process of aggregation, the decisions about what constitutes an informative analytic category rest with the researcher (Grodal et al., 2021).

As can be seen from Panel 4b, a further challenge in this example is the heterogeneity of the data, with many different specific interests only mentioned by a very small number of individuals. However, by aggregating these small numbers, more meaningful conceptual categories can be created. This underlines the need for close reading of the text and is a further

reminder that although NVivo can aid analysis, the responsibility for developing concepts is still firmly with the researcher.

Name	Files	References
active interests	1	68
allotment	1	1
Boating	1	1
dancing	1	4
darts	1	2
fishing	1	5
gardening	1	15
golf	1	2
running	1	1
snooker	1	2
swimming	1	8
walking	1	27
passive interests	1	59
Birds	1	3
cooking	1	2
crafting	1	4
crosswords	1	2
football (watching)	1	2
knitting	1	3
music	1	12
painting	1	3
reading	1	20
sport (watching)	1	4
theatre or cinema	1	4

Panel 4b: Aggregating specific codes into broader concepts

If researchers are interested in understanding the “attitude” or sentiment of a writer from their text, this can, to some extent, be deduced from their lexical choices. This is because many words will have inherent positive or negative valences (e.g., happy, healthy, interesting, vs. sad, unhealthy, boring).⁶ The frequency with which such words are used in a text or a response to an OEQ can be used to infer the attitude or mood of the author. However, use of a negative such as “not” can clearly reverse the sentiment expressed in a phrase or sentence. For this reason, there may be words that are of analytic interest but, as discussed above, by default are counted as “stop words” within NVivo 12 (see panel 3a). In the present example, let’s focus on the use of the word “not,” a key example of a “contextual valence shifter” (Polanyi & Zaenen, 2006).

In total, the word “not” was used 112 times within the 287 responses. By creating a node based on a text search query (see panel 4a) we can see that 90 individuals had used the word “not” in their responses (some more than once). This code can then be refined further by examining all of the 90 responses using the word “not” to develop an understanding of how cohort members were using this negative qualifier. As can be seen in Panel 4c below, of the 90 responses which included the word “not,” 54 expressed a negative sentiment, 30 were neutral,

⁶ NVivo12 also provides some automatic sentiment analysis coding based on assigning a sentiment score to each word. It is important to understand that this tool looks at the sentiment of words in isolation—the context is not taken into account. This means that NVivo cannot take account of sarcasm, double negatives, slang, dialect variations, idioms, ambiguity (For further information see: <https://help-nv.qsrinternational.com/12/win/v12.1.110-d3ea61/Content/coding/auto-detect-code-sentiment.htm?Highlight=sentiment%20analysis>).

and 17 expressed a positive sentiment.⁷ For example negative: “My health is **not** very good,” neutral: “Hope I am not in the U.K and living away,” and positive: “My health will have improved or at least not worsened.”

Not text spread to paragraph	1	90
negative sentiment	1	54
not - lack of money	1	5
not alive - existent	1	2
not capability	1	10
not enough time	1	2
not good health	1	13
not other - negative	1	13
not want to reply or think of future	1	9
neutral sentiment	1	30
not different from now	1	3
not other	1	12
not working	1	15
positive sentiment	1	17
health not worsened	1	7
not other positive	1	10

Panel 4c: Refining coding of responses that use the word “not”

The discussion above underlines that NVivo12 can help with analysis first by making it possible to search for specific words or groups of stemmed words. It can then be used to code segments of text automatically based on the occurrence of key words. However considerable interpretation and refinement of coding is still needed by the researcher. Indeed, the use of NVivo could be argued to provide an additional spur to hermeneutic analysis in that its focus on individual words forces close attention to the text and an examination of how individuals have worded their responses.

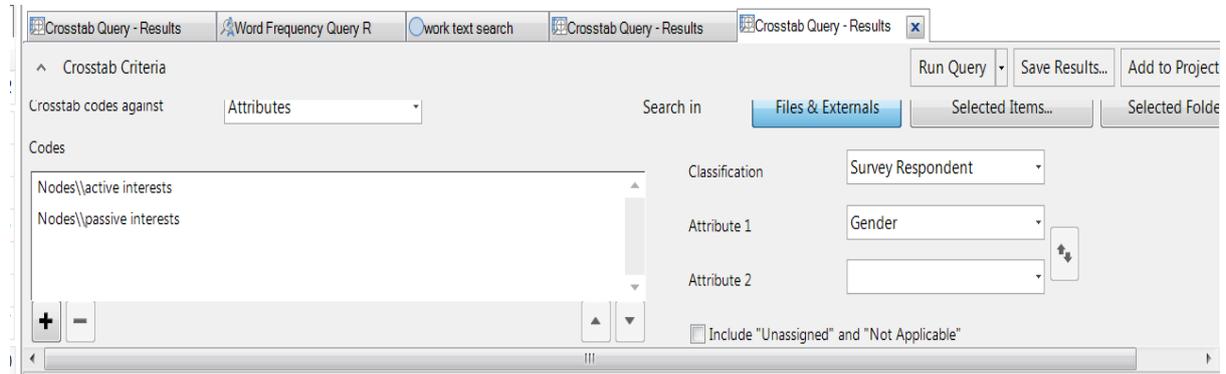
Step Five: Matrix Coding Queries and Crosstabulation Queries

The final stage of the method for analysis of OEQs, described by Feng and Behar-Horenstein (2019) using NVivo11, is to use a matrix coding query to explore the patterning of core themes within OEQs by key attributes of respondents.⁸ However, an improved tool in NVivo12 is the crosstab coding query (Panel 5a), and this is now more appropriate for the analysis of how codes are linked to specific attributes. The crosstab query is accessed via the “Explore” menu in NVivo. As can be seen from Panel 5b, a crosstab has a very similar format to a matrix. The first column shows the nodes, the first row shows the attribute of interest, and each cell of the crosstab represents the number of responses (i.e., the number of cases) which have been coded to the nodes. However, there are three major advantages of the Crosstab tool for the analysis of OEQs, and I therefore wanted to alert readers that Feng and Behar-Horenstein’s description of the use of NVivo is now somewhat outdated. These advantages are

⁷Some responses used the word ‘not’ in more than one way, and so a response could be coded to more than one type of sentiment.

⁸In many surveys that include OEQs, attributes will represent demographic variables such as gender, age, educational level, marital status, employment status etc. However, the method is very flexible and any categorical variables from the survey can be used in this stage of analysis.

that first, the crosstab tool is quicker and more straightforward to use than the matrix coding tool (because attributes are listed on the right-hand side of the criteria box, and values of attributes do not need to be specified one by one), second, the crosstab tool allows for percentages and raw counts to be displayed, and third, two layers of categorisation are possible using the crosstab tool so that patterning of nodes can be examined in the context of two attributes simultaneously.⁹



Panel 5a – The new crosstab tool in NVivo12

To demonstrate the use of the crosstab tool, Panel 5b shows the count (number) of responses from men (Gender=1) and women (Gender=2) which mentioned active and passive interests when imagining life at age 50. For example, 16 of the 240 men mention active interests in their imagined life at age 60 compared with 35 of 241 women, and in total, 51 from the total sample of 481 individuals mentioned active interests. This is a very simple example focussing only on two nodes (namely, active and passive interests), but a very comprehensive list of nodes can be added to the crosstab providing an overview of how the codes within responses are distributed by each attribute.

Survey Respondent	Gender = 1 (240)	Gender = 2 (241)	Total (481)
active interests	16	35	51
passive interests	19	24	43
Total (unique)	35	49	84

Panel 5b: Crosstab of nodes by gender

Panel 5c converts these counts into percentages (an option available in the Crosstab query ribbon) so that we can see that 14.5% of women mention active interests in their response compared with just 6.7% of men. The use of percentages for comparison purposes is particularly valuable when the numbers in the different categories of an attribute are different. NVivo12 does not provide tests of statistical significance, but the data could be exported to a package such as SPSS if this extra level of rigour was required.

⁹The crosstab query tool is discussed here in relation to the analysis of OEQs however it has wider applicability for anyone wishing to look at the relationship between the attributes of texts and the codes attached to those texts.

Survey Respondent	Gender = 1 (240)	Gender = 2 (241)	Total (481)
active interests	6.67%	14.52%	10.6%
passive interests	7.92%	9.96%	8.94%
Total (unique)	14.58%	20.33%	17.46%

Panel 5c: Crosstab of nodes by gender (%)

A Preliminary Comparison Between Using NVivo and More Fully Automated Approaches for Analysis

Where there are thousands of responses to analyze and the focus is primarily on the ostensible content (rather than the form) of the text, a more thoroughgoing automated analysis is likely to be preferred to the use of a CAQDAS package such as NVivo. The lack of precision and possibility of some miss-codes due to the ambiguity of language is likely to be mitigated by the very large sample size. As with all research, however, it is crucial that the research question remains central to any decisions about how the analysis is undertaken. Responses to the open-ended invitation to “Imagine you are 60” in the age 50 sweep of the 1958 British Birth Cohort Study, used as an example above, have also been analyzed using more thoroughly automated techniques (Weber, 2021). Weber has used Natural Language Processing (NLP) software together with relative frequency analysis and topic modelling within R to analyze the complete corpus of over 6,500 responses. Weber’s focus is on whether there are gender differences or social class differences in how individuals answer an OEQ about the future. But the results reported can do little more than sketch out the different topics more frequently mentioned by men and women or by individuals from different social class backgrounds. In contrast, as has been demonstrated above, NVivo12 allows identification and counting of a simple code such as use of the word “not,” but also provides the tools to enable the researcher to look at this code in more detail. A more qualitative or hermeneutic analysis can then be undertaken that seeks to identify the strategies individuals use when responding to a question. Where the sample is in the order of a few hundred responses, it is very manageable for the text to be analyzed using CAQDAS (e.g., NVivo12). An advantage of this approach is that it allows the researcher to remain closer to the data throughout, and manually to check and refine initial automatic coding. This approach also allows for greater attention to be focussed on the form, or style, of response. An advantage of the free text generated by asking OEQs is that it provides access to respondents’ thoughts and ideas in their own words. This means that more subtle analysis can be undertaken that looks for patterns in the precise vocabulary and phraseology used by respondents (Fielding et al., 2012).

Implications and Considerations

As previous authors have highlighted, there is a tendency for those reporting the analysis of qualitative data to mention that they have used a CAQDAS package such as NVivo, without spelling out in detail how they have used the package to aid their explorations and analysis (Paulus et al., 2017). The aim of this paper has therefore been to help remedy this deficit by describing in detail one approach to the analysis of mixed-methods data aided by NVivo12.

One danger of presenting a step-by-step guide to using NVivo for the analysis of OEQs is that it suggests a relatively linear analytic process and can imply that the software is responsible for the analysis. The constraints of the format and length of a journal article mean that the process of analysis is inevitably over-simplified: most qualitative analysis is a messy and iterative process. Hypotheses and ideas are generated through close interaction with

individual extracts of text, which can then be tested by looking at patterns across the whole set of data using the crosstabs query. A key benefit of CAQDAS packages such as NVivo is that they allow for the documentation of this process, for example, by using the memo function to take notes on emerging ideas.

However, it should also be emphasised that the systematic and appropriate use of a CAQDAS package is not in itself a guarantee of high-quality analysis and insightful results. Indeed, it has been argued that OEQs in surveys frequently do not provide a sound basis for rigorous qualitative research (LaDonna et al., 2018). The problem identified by LaDonna et al. is that free-text responses within structured surveys rarely produce data *rich* enough to achieve the sincerity, credibility or resonance that characterises *rigorous* qualitative research (as characterised by Tracy, 2010). For qualitative evidence to be rich, it must have context, personal meaning, and emotional and social nuance (O’Cathain & Thomas, 2004). It will clearly be important for researchers to make their own judgements about the quality of data that has been generated by OEQs within a specific survey, and in turn, judgements about quality will be made in the context of the research topic or research questions being addressed. In the current example, as discussed above, the prospective nature of the longitudinal cohort study which included the OEQ means that a great deal of additional information is available to the researcher providing context to the response to a single OEQ. In summary, however, while NVivo will help with coding and counting elements within the free-text responses generated, it is ultimately the researcher’s interpretation of that data which will be judged according to whether it deepens our understanding of the social phenomenon being explored.

Data

University of London, Institute of Education, Centre for Longitudinal Studies. (2020). National Child Development Study: Age 50, Sweep 8, "Imagine You are 60", 2008-2009. [data collection]. UK Data Service. SN: 6978, DOI: 10.5255/UKDA-SN-6978-1

References

- Andrews, M. (2018). Imagining the ‘baffling geography’ of age. *Journal of Aging Studies*, 47, 90–95. <https://doi.org/10.1016/j.jaging.2018.03.003>
- Bazeley, P. (2006). The contribution of computer software to integrating qualitative and quantitative data and analyses. *Research in the Schools*, 13(1), 64–74.
- Bazeley, P. (2009). Editorial: Integrating data analyses in mixed methods research. *Journal of Mixed Methods Research*, 3(3), 203–207. <https://doi.org/10.1177/1558689809334443>
- Brookes, G., & Baker, P. (2017). What does patient feedback reveal about the NHS? A mixed methods study of comments posted to the NHS Choices online service. *BMJ Open*, 7(4), e013821. <https://doi.org/10.1136/bmjopen-2016-013821>
- Crow, G., & Andrews, M. (2019). Biographical interviews and imagined futures essay writing: Users of two methods in conversation. *International Journal of Social Research Methodology*, 22(6), 557–569. <https://doi.org/10.1080/13645579.2019.1610597>
- Elliott, J. (2008). The narrative potential of the British Birth Cohort Studies. *Qualitative Research*, 8(3), 411–421. <https://doi.org/10.1177/1468794106093637>
- Elliott, J., & Brown, M. (2011). *Data note on response rates and potential response bias relating to the use of the paper self-completion questionnaire used as part of the 2008 sweep of the National Child Development Study*. 10.
- Elliott, J. (2012). *Imagine you are 60*. Centre for Longitudinal Studies. <https://cls.ucl.ac.uk/wp-content/uploads/2017/07/FINAL-CLS-WP-20126.pdf>
- Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *The*

- Qualitative Report*, 23(11), 2850-2861. <https://doi.org/10.46743/2160-3715/2018.3560>
- Feng, X., & Behar-Horenstein, L. (2019). Maximizing NVivo utilities to analyze open-ended responses. *The Qualitative Report*, 24(3), 563–571. <https://doi.org/10.46743/2160-3715/2019.3692>
- Fielding, J., Fielding, N., & Hughes, G. (2012). Opening up open-ended survey data using qualitative software. *Quality and Quantity*, 47, 3261-3276. <https://doi.org/10.1007/s11135-012-9716-1>
- Grodal, S., Anteby, M., & Holm, A. L. (2021). Achieving rigor in qualitative analysis: The role of active categorization in theory building. *Academy of Management Review*, 46(3), 591–612. <https://doi.org/10.5465/amr.2018.0482>
- Jackson, K., Paulus, T., & Woolf, N. (2018). The Walking Dead genealogy: Unsubstantiated criticisms of qualitative data analysis software (QDAS) and the failure to put them to rest. *The Qualitative Report*, 23(13), 74-91. <https://doi.org/10.46743/2160-3715/2018.3096>.
- LaDonna, K. A., Taylor, T., & Lingard, L. (2018). Why open-ended survey questions are unlikely to support rigorous qualitative insights. *Academic Medicine*, 93(3), 347–349. <https://doi.org/10.1097/ACM.0000000000002088>
- O’Cathain, A., & Thomas, K. J. (2004). “Any other comments?” Open questions on questionnaires – a bane or a bonus to research? *BMC Medical Research Methodology*, 4(1), 25. <https://doi.org/10.1186/1471-2288-4-25>
- Paulus, T., Woods, M., Atkins, D. P., & Macklin, R. (2017). The discourse of QDAS: Reporting practices of ATLAS.ti and NVivo users with implications for best practices. *International Journal of Social Research Methodology*, 20(1), 35–47. <https://doi.org/10.1080/13645579.2015.1102454>
- Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. In J. G. Shanahan, Y. Qu, & J. Wiebe (Eds.), *Computing attitude and affect in text: Theory and applications* (Vol. 20, pp. 1–10). Springer-Verlag. https://doi.org/10.1007/1-4020-4102-0_1
- Power, C., & Elliott, J. (2006). Cohort profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology*, 35(1), 34–41. <https://doi.org/10.1093/ije/dyi183>
- Schiffrin, D. (1996) ‘Narrative as Self-Portrait: Sociolinguistic Constructions of Identity’, *Language in Society* 25: 167-203 .
- Singer, E., & Couper, M. P. (2017). Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys. *Methods, Data, Analyses*, 11(2), 115-134. <https://doi.org/10.12758/mda.2017.01>
- Tracy, S. J. (2010). Qualitative quality: Eight “big-tent” criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837–851. <https://doi.org/10.1177/1077800410383121>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2020). National Child Development Study: Age 50, Sweep 8, "Imagine You are 60", 2008-2009:. [data collection]. UK Data Service. SN: 6978, <https://doi.org/10.5255/UKDA-SN-6978-1>
- Van De Mierop, D. (2005). An integrated approach of quantitative and qualitative analysis in the study of identity in speeches. *Discourse & Society* 16(1), 107–130. <https://doi.org/10.1177/0957926505048232>
- Wahbeh, A., Nasralah, T., Al-Ramahi, M., & El-Gayar, O. (2020). Mining physicians’ opinions on social media to obtain insights into COVID-19: Mixed methods analysis. *JMIR Public Health and Surveillance*, 6(2), e19276. <https://doi.org/10.2196/19276>
- Weber, M. (2021). How do 50-year-olds imagine their future: Social class and gender disparities. *SAGE Open*, 11(4), 1-11. <https://doi.org/10.1177/21582440211061567>
- Wolski, U. (2018). The history of the development and propagation of QDA software. *The*

Qualitative Report, 23(13), 6-20. <https://doi.org/10.46743/2160-3715/2018.2984>
Woods, M., Macklin, R., & Lewis, G. K. (2016). Researcher reflexivity: Exploring the impacts of CAQDAS use. *International Journal of Social Research Methodology*, 19(4), 385–403. <https://doi.org/10.1080/13645579.2015.1023964>

Author Note

Jane Elliott is a Professor of Sociology at the University of Exeter, UK. Please direct correspondence to Jane.Elliott@exeter.ac.uk.

Acknowledgements: I am grateful to Jon Lawrence and Jennifer Wilcox for their helpful comments on this paper. Special thanks are due to the members of the 1958 British Birth Cohort Study who have provided detailed information to the study over many years, and to the team at the Centre for Longitudinal Studies who manage the study and document the data.

Copyright 2022: Jane Elliott and Nova Southeastern University.

Article Citation

Elliott, J. (2022). The craft of using NVivo12 to analyze open-ended questions: An approach to mixed-methods analysis. *The Qualitative Report*, 27(6), 1673-1687. <https://doi.org/10.46743/2160-3715/2022.5460>
