Full Length Research Paper

Qualitative data management and analysis using NVivo: An approach used to examine leadership qualities among student leaders

[1]Noriah Mohd. Ishak and *[2]Abu Yazid Abu Bakar,

[1]Pusat PERMATApintar Negara, Universiti Kebangsaan Malaysia43600 UKM Bangi, Selangor Darul Ehsan, Malaysia
[2]Faculty of Education, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

*Corresponding Author E-mail: yazid3338@ukm.my, Tel: +6019-2246917   Fax: +603-89217525
E-mail: norwmu@hotmail.com, Tel: +603-89217501   Fax: +603-89217525

Abstract

Qualitative researchers, in general, have arrays of methodologies to provide evidence to cope with their inquisitive mind regarding issues pertaining to their interest. In tandem, combining a qualitative research data management and data analysis with technology is frown by some qualitative researchers. However, qualitative researchers now have the option to "*technologized*" their approach in data management and data analysis to ease the complexity of the research process. This is important, due to the fact that qualitative researchers have to handle humongous amount of data and the chunking process of the data is tedious. This paper will descriptively report the researchers' effort in using the qualitative software known as NVivo in managing and analyzing qualitative data that was collected in a study designed to examine patterns of leadership among young student leaders in local universities. Thoroughly, the article discussed about the methodological techniques of using NVivo and briefly presented the results of this 2008 study.

Key words: *Qualitative; data management; data analysis; NVivo; leadership qualities*

INTRODUCTION

Managing and analyzing qualitative data can be a trying experience, particularly for novice researchers who tend to see their data from a single perspective. However, qualitative data is never short of meanings. It is multifaceted, rich in nature, unstructured and at times can be overwhelming where the researcher will be provided with information beyond the scope of the study. Richardson (1998), Richards (2002) and Bazeley (2009, 2007) have all agreed with Miles and Huberman (1994) who suggests that qualitative data can produce meaningful findings if they are managed properly. The humongous data need to be cleaned, chunk into smaller meaningful pieces which then are simultaneously arranged into specific themes before it can be used for any kind of analysis (Miles and Huberman, 1994).

Carrying out qualitative research also means that the researchers will have to triangulate all kind of data to formulate conclusive evidence that could support the existence of the phenomena that they are studying. According to Bogdan and Biklen (1997), Charmaz (2006), and Gleshne and Peshkin (1992), qualitative researchers should not be satisfied with just one type of evidence. Numerous types of data collection are needed to provide stronger evidence to support their hunches. As such, qualitative researchers have to be creative in seeking the best data collection methods for their study. However, more data collection techniques means more measure has to be taken to ensure that the data collected is not wasted and it can actually contribute to a more meaningful findings. This is not an easy task, and can be

challenging for beginning researchers or researchers who are not familiar with mining qualitative data.

In this paper, the authors will describe their experience in managing and analyzing two types of qualitative data commonly used by qualitative researchers: (a) in-depth interview, and (b) focus group interview, using software known as NVivo. At present, there are many software that can be used to manage and analyze qualitative data; for instance, NU*DIST, ATLAS, XSIGHT, Weft QDA and Ethnograph (Bazeley 1999, 2003, 2006). These softwares are available on the market, and can be mastered in a short period of time. The researchers chose to work with NVivo because of the many features that could help them in this study. Examples of such features are: (a) the folders that could organize documents according to type. Subsequently, the researchers were able to arrange all information in the folder, including all literatures read before, and during the study, and integrate the literature with the data for better discussion; (b) the software ability to directly use video or audio data as data source; (c) the many mode of queries that will satisfy qualitative researchers thirst for new information regarding the phenomenon they are studying, and (d) the coder comparison with the Kappa calculation that will enable researchers to determine reliability (and therefore, the stability) of the coding process.

According to Bazeley (2009, 2007) using qualitative software does not lessen the quality of the qualitative research nor does it simplify the whole process of doing qualitative research. What it does is, to ease data management and data analysis processes and make it more manageable as well as rigorous. NVivo or any other qualitative software are just tools that will assist the researcher in undertaking an analysis of qualitative data. In the present research, aside from the benefits gained from the features in NVivo that was mentioned earlier, the researchers also benefited in terms of; (a) cost incurred for services rendered to manage and analyze the data (in which the researchers did the data management and data analysis themselves), and (b) time effective and the research was completed according to the time frame given to them.

The leadership research that was used as an example in this article, is part of a bigger research, conducted in 2008, that examine relationship between emotional intelligence (EQ), emotion-focused solution and leadership qualities among final year students (n=2500) studying at five universities in Malaysia. The research was conducted in two phases. The first phase employed a cross-sectional design and data was collected using three sets of questionnaires. The second phase, a case study design, was used to explore patterns of leadership and emotion-focus solutions among high and low emotionally intelligent students. Concomitantly, this paper will discuss the data management and data analysis based on the second phase of the study, in particular the leadership component. The question that guides this

paper is; how can qualitative data be managed and analyzed effectively to answer research questions?

Using the data from the abovementioned research, this paper will specifically elaborate the method of using NVivo in five essential stages: (a) Developing document for in-depth and focus group interviews, (b) Managing documents using folders, (c) Mapping of variables through nodes, (d) Data Queries using Matrix Coding, and (e) Developing coding reliability using coder comparison and Kappa value. However, steps (a) and (b) will be discussed in the materials and methods section, whereas steps (c), (d), and (e) will be discussed in the results and discussion section. This is done in such a way because both steps (a) and (b) discussed the method used to manage the data, while the other three steps discussed steps to produce the findings (QSR International, 2010, 2008a, 2008b).

Materials and Methods

Developing Documents for In-Depth and Focus Group Interviews

The second phase of the study used two types of data collection techniques: In-depth interview and focus group interview. Nine student leaders (5 from high EQ group, and 4 from low EQ group) from three different universities participated in the in-depth interview, while 21 student leaders from the same three universities participated in the focus-group interview (each group has 7 members and a mixed between high and low EQ student leaders). Their leadership traits were traced using a model developed by the Malaysian Prime Minister's Department which looks at elements of Business Strategy and Leadership Model, Recruitment, Review and Honoring Members, Deployment, Development of Leadership, Engagement and Retention of Members, and HR and line partnership and Capability (JPM, 2007).

Data from both interviews was transcribed and cleaned before it was imported to the NVivo. Although transcription can be directly done using NVivo, the researchers chose to transcribed it using Microsoft Office Word 2007 for a number of reasons; (a) NVivo accept documents saved under .doc, (b) Auto-coding can be done using headings that are available in Microsoft Office, and (c) The researchers share the software, and therefore it is impossible for both researchers to use the software at the same time to transcribe the data.

In developing both documents, the researchers took great care to ensure that the first line of the document encompassed details information of the interviewees such as name, age, gender, program of study, family back ground – number of siblings, socio-economic status, co-curriculum activities, involvement in students' organization, and status or position in organization. Such information will be useful when the researchers want to
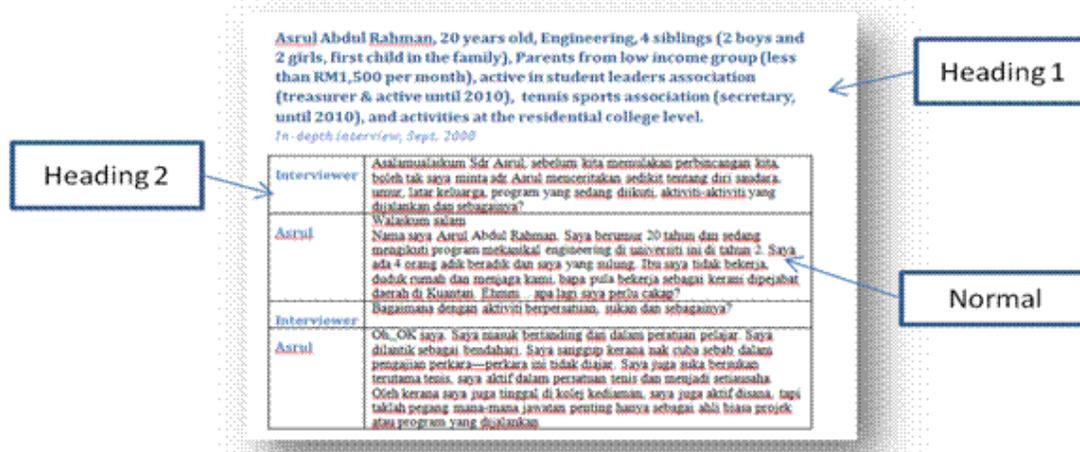
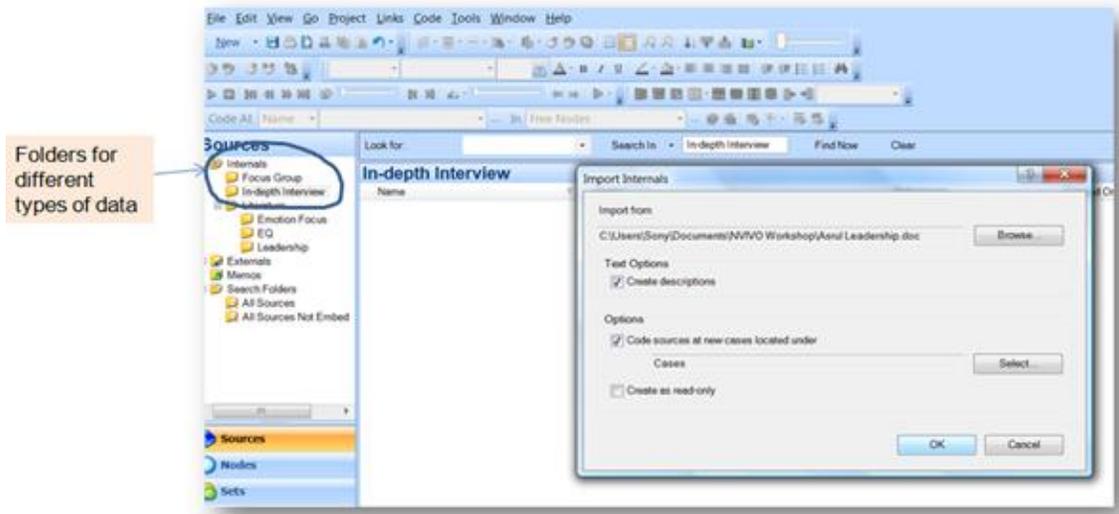**Figure 1.** Document Development using Microsoft Word 2007

Figure 2. Managing Data using Folders

examine the leadership issues from different perspectives based on the demographic information. This information was considered as the first paragraph of the document and was assigned to *Heading 1*. The document was also developed using the following steps: (a) A two by n table was inserted onto the document, and the transcription was done within the table. Column one was used to insert the status of the speaker (either interviewer or interviewee), and the second column was reserve for the transcribed note; (b) The table was merged, and converted to text using paragraph mark; (c) Both column were then given headings and column one was marked as *Heading 2*, while column two as *Normal*; and (d) The document was saved using the interviewees or group name as file name; for example, Asrul.doc, as shown in Figure 1.

The researchers were also very careful when transcribing the focus group interview. In total, three focus groups were conducted, each comprising of 7 members, and the interview process was very active. Each group members eagerly contributed information pertaining to their leadership ability which tantamount to at least 70 pages of transcribed notes for each group. With the help of audio and video tape, the researchers were able to identify group members and the issue he or she put forward during the discussion. Transcription was done directly in the NVivo. A point of challenge to remember is, developing the focus group interview document was more tedious than the document for in-depth interview, in which, the researchers have to ensure "*who said what*" with precision.   Similar format as mentioned earlier followed the rest of   the   document
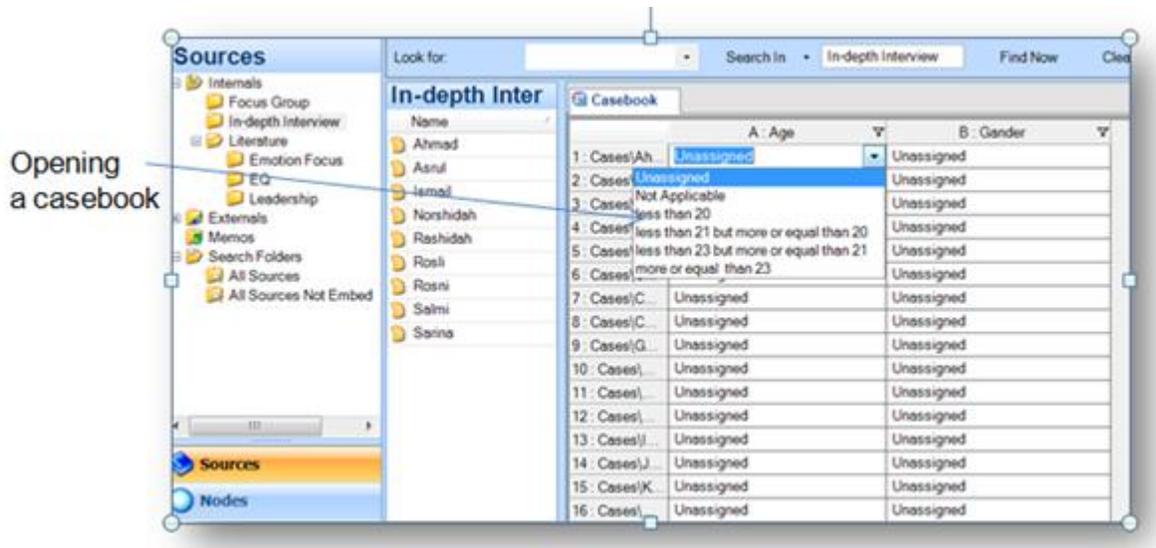
Figure 3. Opening a *Casebook*

Figure 4. Separating Focus Group Members under *Tree Nodes*

development.

Managing Documents using Folders

Management of data was done in NVivo by importing the documents directly into the software. To ease the management of the data, the researchers developed different folders for both in-depth interview and focus group interview under sources in the area called Navigation View of the NVivo software, as indicated in

Figure 2. Data imported was then parked under each folder before it can be opened and analyzed.

Importing in-depth interview documents was a simple process. To ensure that all information was captured, the researchers clicked to "Text option" and "Option". The first command allows the researchers to capture information about the interviewee pertaining to their demographic background. This was basically written in the first paragraph of the document. The second command allows the researcher to use the information to examine the leadership issues among the student
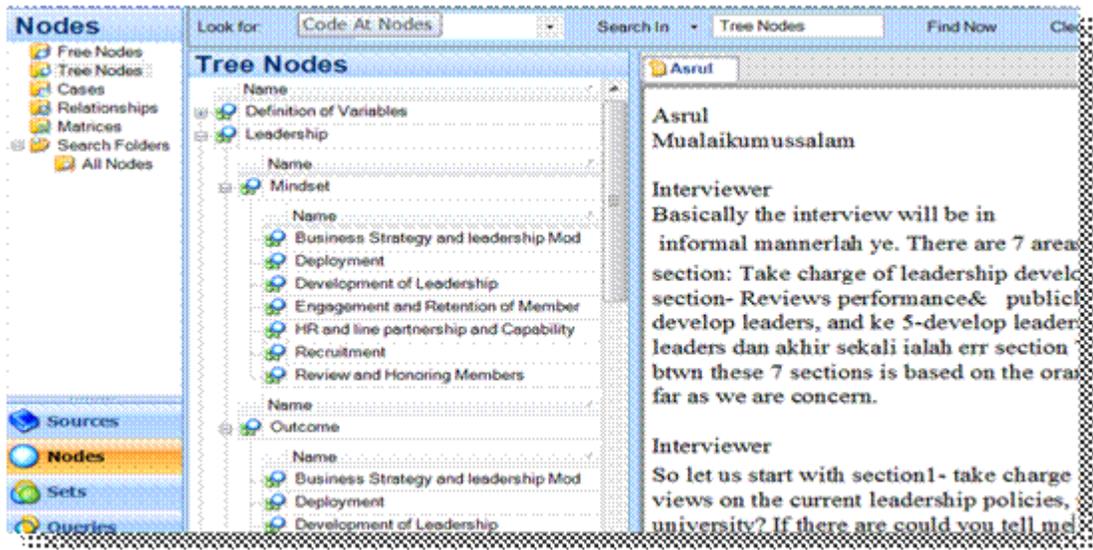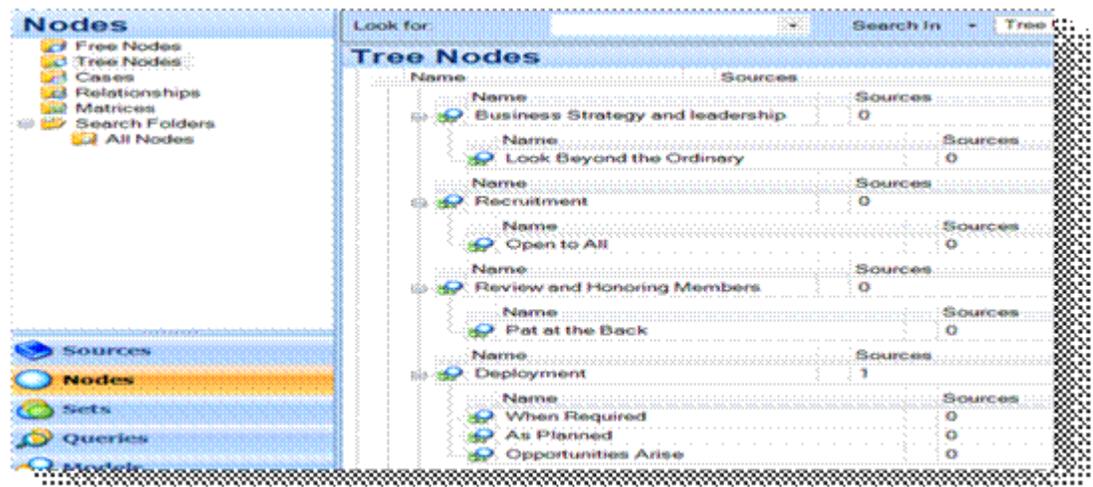
Figure 5. Development of Tree Nodes

Figure 6. Nodes and Sub-Nodes

leaders. The researcher developed *Attributes* for each demographic data, and using the *Casebook*, the attributes was then assigned to each interviewees, as demonstrated in Figure 3.

Focus group documents was imported and placed under the focus group folder, and the interviewees were separated using auto-code process and placed in a section called *Tree Nodes* under the *Nodes*. In order to separate the group members for better analysis and to examine if differences exists among the group members, auto coding using *Heading 2* was done, and the group members name were separated at the tree nodes (refer

Figure 4). The steps allowed the researchers to examine the leadership issue from the groups perspective, and subsequently, similar issues among the group members (this is very important when the researchers were examining differences that exist among group members). Thus, the researchers were able to answer two of the research questions: (a) how do the three groups of student leaders from the different universities differ in terms of looking at leadership development at their respective universities? And (b) how do each members of the group from the same universities differ in terms of looking at leadership development within them?
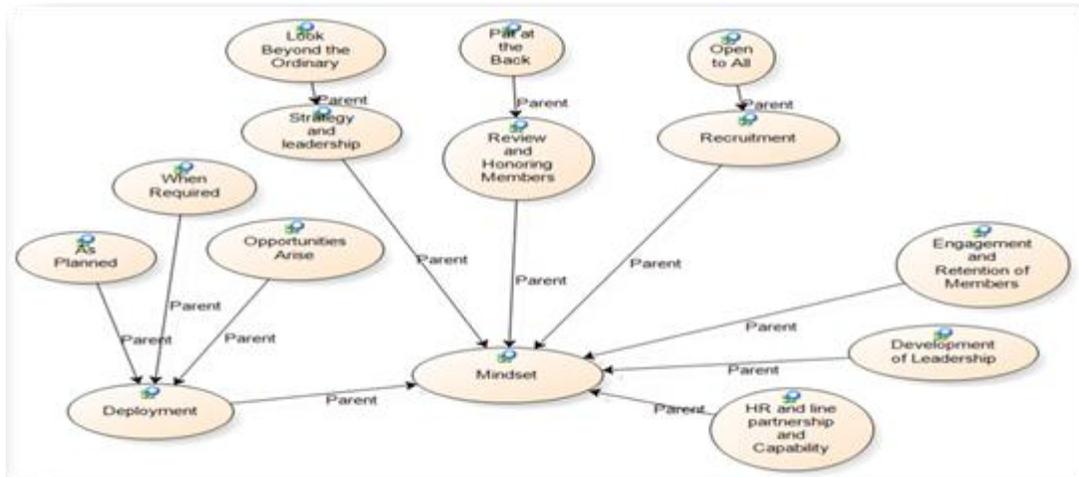
Figure 7. Development of Sub-Nodes

Results

Mapping of Variables through Nodes

Nodes are representation of variables that a researcher is interested in his or her study. NVivo allows flexibility for researcher to identify the nodes which come in different forms of free nodes, tree nodes, case nodes, relationship nodes and matrices. A grounded theorist researcher, for example, may initially want to use free nodes to develop themes of their interest because it allows him or her a free flow of the thematic process (similar to open coding in constant comparative analysis) and the nodes are developed simultaneously as the researcher read, compare, contrast and analyze the data. Tree nodes, on the other hand, are the more concrete nodes and can be developed based on the theoretical framework used by the researcher. This is the approach adopted by many researchers who used case study design.

As mentioned earlier, the present study used a case study approach and one of the objectives is to map the leadership qualities from the leadership model onto the data collected from the stories told by the student leaders. As such, the coding process is guided by the model and the seven leadership elements cited in the model. Thus, in this study, the researchers developed the nodes using Tree Nodes. Since the researchers were trying to examine the seven elements of leadership qualities from three perspective: (a) Policies, process and procedures, (b) Mindset, and (c) Outcomes, the researchers made these three perspectives into the parent nodes. The seven elements, subsequently, were made into children nodes or sub-nodes, as being displayed in Figure 5.

Working through the data, the researchers developed more sub-nodes that explain the patterns of the leadership qualities from the three perspectives. These sub-nodes are illustration of themes that was not stated in the model, but were salient issues mentioned by the student leaders when talking about leadership. Example of such sub-nodes is, when talking about recruitment, the students reiterated that "*we are not limiting our membership to only...*" in which case, the researcher associate that particular statement with a theme or sub-node identified as "*Open to all*". Example of such development and organization of nodes is shown in Figure 6.

The nodes and sub-nodes can also be modelled using the Models in NVivo (located at the Navigation View). For this particular purpose, the researchers selected variables (from the tree nodes) to be added to the model by bringing in the nodes and sub-nodes that interest her into the model. This forms the visual representation of the idea about the quality of leadership that was observed among the student leaders. Figure 7 below is an example of the model developed for the mindset component of the leadership study, which shows the numerous sub-nodes developed for the seven elements of leadership.

Data Queries using Matrix Coding

The researchers were also very interested in exploring the leadership qualities among the student leaders with different genders, and different level of emotional intelligence (research question posed was: "*How do male and female student leaders differ in terms of their leadership qualities?*"). Research conducted by Kerr, Garvin, Heaton and Boyle (2006), Charbonneau and Nicol (2001) and Sheehan and Sheehan (2006) shows a significant difference in leadership qualities between man and woman. Findings from research conducted by Goleman (1996), Mandell (2003) and Hopkins and
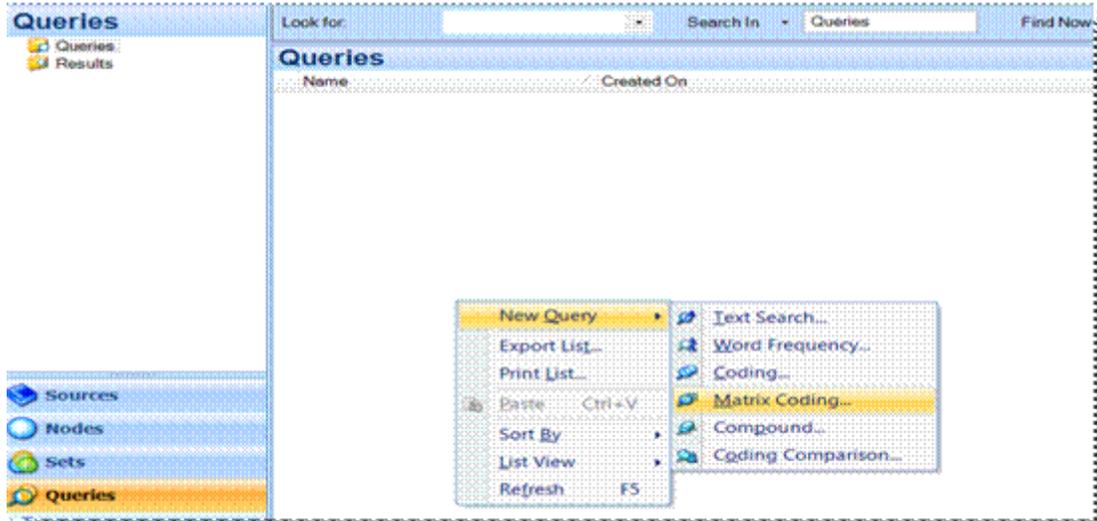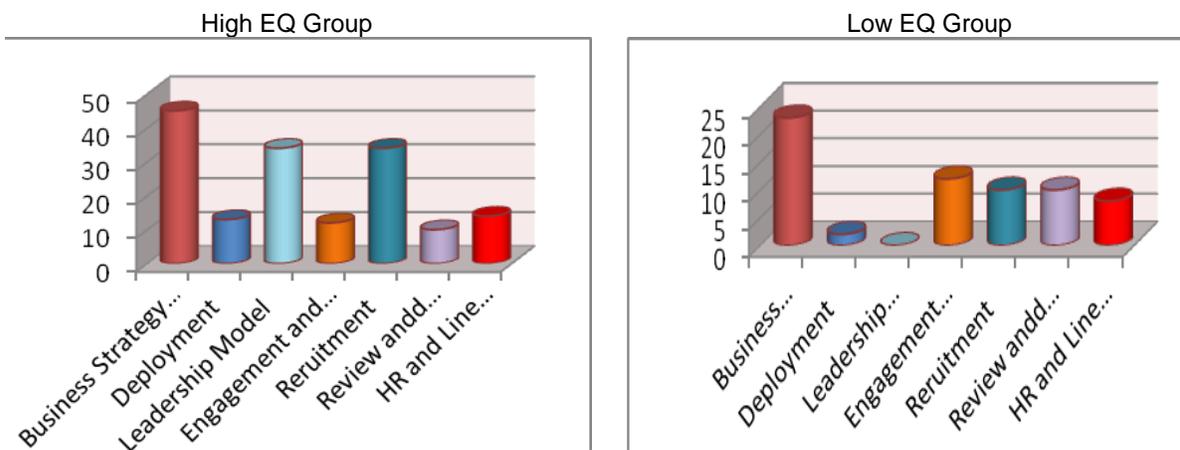
Figure 8. Matrix Coding Query

Figure 9. Chart of Elements of Leadership Qualities against Level of EQ

Bilimoria (2008) shows that executives with high level of emotional intelligence are better performers as leaders when compared with their peers who have lower level of emotional intelligence.

In addressing such research questions, the researcher use Matrix Coding as the mode of data analysis that was found in the Queries section (located at the Navigation View) of the NVivo. The query is an *"n x n"* matrix, and *n* can be any number of rows and columns that make up the matrix (refer Figure 8).

The researchers answered the research question by identifying the themes developed for each of the seven leadership elements and chosen it as rows, while gender and level of emotional intelligence as the columns. All sub-nodes were then added into the query and analysis was conducted. The analysis resulted into a tabulated data which gave the researcher a pattern that suggests similarities or differences among male and female student leaders on all sub-nodes or themes.

Table 1 shows a clear pattern between male and female student leaders in how they perceived the seven elements of the leadership, for example, in terms of building business strategy and leadership model, two student leaders from the high EQ group prefer to look beyond the ordinary strategy used by other student leaders that they have known (e.g. "*I think we have to lead by example. I do not want to be telling my committee what to do all the time like the previous leader. I want creativity in my group, so let them suggest and if agreed by members, do what they have planned. The idea does not have to come from me alone. But if I have ..I think a good idea about certain strategy, say to attract more members, and my members' feel that the strategy is too tedious, then, I try first. But I make sure it is successful,*

Table 1 Example of Matrix Coding Result for Mindset

		High EQ	Low EQ	Female	Male
Business Strategy & Leadership Model	Look beyond the ordinary	4	0	2	0
Recruitment	Open to all	4	0	2	4
Review and Honoring Members	Pat at the back	2	0	4	3
Deployment	When required	2	0	4	0
	as planned	10	3	14	2
	opportunities arise	0	0	2	0
Development of Leadership	Through training	12	6	4	8
	come as a package	4	0	2	2

so they will see and later do what I have accomplished and do the same"). Nonetheless, the same idea was not mentioned by any of the student leaders from the low EQ group.

The table also shows that, for example, for the element "*development of leaders*", both high and low EQ group agreed that leadership can be developed by training (example of the excerpt: "*Very few leaders are born leaders, and I think the best way to develop leadership is to go for training. I know my university conducted such program after the campus election"* and "*You will never know how to act like leader unless you learn. Came on, I don't believe there are such thing as born leaders.*"). However, the themes seem to be more salient to the high EQ group as compared to the low EQ group (the number of the themes being mentioned is more from the high EQ group as compared to the low EQ group.

NVivo also allows the researchers to convert the table to chart (similar to chart in Excel). This gives the visual representation of the data and provides a clearer picture of the leadership qualities among the two different groups (refer Figure 9). For that reason, reporting qualitative research findings is no longer just in the form of narration of the verbatim alone. Content of the verbatim can be interwoven with frequency table, diagram, charts, and supported by excerpt from literature review for triangulation and building chain of evidence. This makes the data presentation more lively and appealing.

NVivo also has an interface with the *Statistical Package for Social Sciences* (SPSS) software which allows for simple non-parametric analysis (frequency, descriptive, Mean, Mode, Median and χ-Square) to be conducted. However, a point to remember is, unless it is meaningful to cross the data to SPSS, it is best for the researcher to keep the analysis within the qualitative approach. It is also unjustified to analyze qualitative data using any of the inferential statistics for a number of reasons, which include: (a) the research sample is normally very small,

and (b) the sample is not randomized.

In this writing, the researchers only discussed the use of matrix coding queries. However, NVivo can perform other types of data analysis, from simple to advance queries. It allows the researchers to do text search, coding and node search, compound search as well as "coder comparison" which is useful to examine the consistency of the coding process conducted by different coder within the same project.

Developing Coding Reliability using Coder Comparison and Kappa Value

One of the most important procedures in ensuring reliable findings is by calculating the Cronbach Alpha Coefficient for instrument used in a quantitative study or the Kappa Coefficient for coding consistency in a qualitative study. For this study, the Kappa value was calculated using the Coder Comparison Queries in the Navigation View of the NVivo. An expert in leadership field was consulted and asked to code two documents (in-depth and focus group transcribed documents). Prior to the expert coding, the researcher have to add the expert as user of the project. This is done by putting the expert name and initial in the field provided in the Application Options at the Option menu.

Coder Comparison Analysis provides information about: Node Type, Node, Source Type, Source, Source Folder. Source Size, Kappa Value, Agreement (%), and the possible agreement patterns [A and B (%), Not A and Not B (%), Disagreement (%), A and Not B (%) and B and Not A (%)], in which case, A is defined by one of the selected node coded by the researcher and B is the selected node coded by the expert. The Kappa value increases as more nodes and text are used for comparison. In NVivo, the agreement is calculated on

Table 2 Kappa Coefficient Values and Interpretation

Kappa Value	Interpretation
Below 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

both uncoded and coded text, which tends to push the agreement levels higher (and distort the impressions given that if neither coder uses a code, the level is 100% agreement!), but may not impact so much on the Kappa value – the latter is calculated on a character basis (rather than words). In producing the Kappa value, the expected agreement is calculated by multiplying together the units of measure coded by each user and dividing this by the total number of units of measure in the source document. This value is then subtracted from the number of units of total agreement between the two coders which is then divided by the total units of measure, or categories, minus the expected frequency value calculated above. The result of this calculation gives a value f for Kappa. To calculate the Kappa, NVivo takes into account the respective length, and duration or size of each source being considered for agreement.

The Kappa value for this study was found to be 0.89, which is an almost perfect reliability value (Landis and Koch, 1977). Although there are no absolute cut-offs for Kappa coefficients, according to Fleiss (1981), values exceeding .75 suggest strong agreement above chance, values in the range of .40 to .75 indicate fair levels of agreement above chance, and values below .40 are indicative of poor agreement above chance levels. In addition, Gardner (1995) recommends that the Kappa to exceed .70 before a researcher should proceed with additional data analyses. Table 2 displays the Kappa interpretation scale to interpret degree of agreement, as being suggested by Landis and Koch (1977):

Discussion

In essence, managing and analyzing qualitative data using software does not lessen the quality of the research findings or even the study itself or quantifying the qualitative research data. It can actually improves the research process (by cutting the amount of time needed, and of course, the amount of boredom and frustration you get from reading too many text), by systematically managing the humungous amount of data, and make it

more manageable for interpretation (Spencer 2007; Bazeley 1999, 2003, 2006; Crowley, Harré, and Tagg 2002). NVivo for instance, is just another set of tools that will assist a researcher in undertaking an analysis of qualitative data. The software does not prescribe a method, but rather it supports a wide range of methodological approaches. It manages and organizes the qualitative data, provides a system to answer the research question and develops image of the whole study for better understanding.

However, regardless of the type of software being used, the researcher has to dutifully make sense of all the data him or herself, without damaging the context of the phenomenon being studied. Unlike quantitative data which is numeric in nature and the meaning of the numbers is not attached to the numbers but the patterns arising from analyzing the numbers statistically, the meaning of each qualitative data is multifaceted. Researcher has to read and re-read each line of text to understand the meaning that underlies each line before coding at the appropriate three nodes or deciding the themes suitable for the text. Similar process is required for all types of qualitative data (video, audio, etc.). Inevitably, the software cannot replace the wisdom that the researcher brings into the research because at the back of every researcher's mind lies his or her life history that will influence the way he or she sees and interpret the world.

Conclusion

In a nutshell, the assessment of leadership qualities among student leaders in this study had broadened the understanding of analyzing qualitative data using the NVivo software among local researchers. Indeed, the procurement of qualitative data had supported the quantitative inputs of this study; hence, strengthened the overall findings of the study. However, the increase of future research works using mixed method design will promote the usage of qualitative analysis software like NVivo in local scene.

References

Bazeley P (2009). Analysing qualitative data: More than 'identifying themes'. *Malaysian J. Qualitative Res.*, 2: 6-22.

Bazeley P (2007). *Qualitative data analysis with NVivo*. London: SAGE Publications.

Bazeley P (2006). The contribution of computer software to intergrating qualitative and quantitative data and analyses. *Research in the Schools*. 13(1): 63-73.

Bazeley, P. (2003). Computerized data analysis for mixed methods research. In A. Tashakkori and C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, CA: SAGE Publications. 385-422.

Bazeley P (1999). The *bricoleur* with a computer: Piecing together qualitative and quantitative data. *Qualitative Health Research*. 9(2): 279-287.

Bogdan RC, Biklen SK (1997). *Qualitative Research for Education: An Introduction to Theory and Methods*. Los Angeles: Allyn and Bacon.

Charmaz K (2006). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: SAGE Publications.

Charbonneau D, Nicol AAM (2001). Emotional intelligence and leadership in adolescents. *Personality and Individual Differences*. 33: 1101-1113.

Crowley C, Harré R, Tagg C (2002). Qualitative research and computing: Methodological issues and practices in using QSR NVivo and NUD*IST. Int. J. Soc. Res. Methodol., 5(3): 193-197.

Fleiss JL (1981). *Statistical Methods for Rates and Proportions.* 2nd ed. New York: John Wiley and Sons.

Gardner W (1995). On the reliability of sequential data: Measurement, meaning, and correction. In J.M. Gottman (Ed.). *The Analysis of Change.* New Jersey.: Erlbaum.

Gleshne C, Peshkin A (1992). *Becoming Qualitative Researchers: An Introduction*. White Plains, NY: Longman.

Goleman D (1996). *Emotional Intelligence*. New York: Bantam Dell Publishing Group.

Jabatan Perdana Menteri. (2007). *The Orange Book: Leadership Development*. Kuala Lumpur: Khazanah Sdn. Bhd.

Kerr R, Garvin J, Heaton N, Boyle E (2006). Emotional intelligence and leadership effectiveness. *Leadership and Organization Development Journal*. 27(4): 265-279.

Hopkins MM, Bilimoria D (2008). Social and emotional competencies predicting success for male and female executives. *Journal of Management Development*. 27(1): 13-35.

Landis J, Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics*. 33: 159-174.

Mandell B (2003). Relationship Between Emotional Intelligence And Transformational Leadership Style A Gender Comparison. *J. Bus. Psychol.,* 17(3): 387-404.

Miles MB, Huberman M (1994). *Qualitative Data Analysis: An Expanded* Sourcebook. 2nd ed. Sydney: SAGE Publication.

Richards T (2002). An intellectual history of NUD*IST and NVivo. Int. J. Soc. Res. Methodol., 5(3): 199-214.

Richardson JTE (1998). Introduction. In J.T.E. Richardson (Ed.), *Handbook of Qualitative Research Methods for Psychology and the Social Sciences*. Oxford, UK: Blackwell Publishers. pp.3-10.

Sheehan JJ, Sheehan OT (2006). The American Presidency: Categorizing and Assessing Leadership Qualities. J. Soc. Stud. Res., 30: 9-14.

Spencer C (2007). Qualitative data analysis with NVivo. *Journal of Emergency Primary Health Care*. 5(3): 1-2.

QSR International. (2010). *NVivo 9: Getting started*. Victoria, Australia: Author.

QSR International. (2008a). *NVivo 8 fundamentals: Starting to work with your material*. Victoria, Australia: Author.

QSR International. (2008b). *Moving on in NVivo 8: Exploring, visualizing and sharing your material*. Victoria, Australia: Author.