Principal component analysis of narrative corpus: A mixed-method design to obliterate the pitfalls of vignette based analysis of personal narratives

Dr. Abhilash Ponnam
Associate Professor, NMIMS Hyderabad
abhilashponnam@gmail.com

:**Dr. Rik Paul**
Associate Professor, BML Munjal University
rik.paul@bmu.edu.in

Abstract

Social scientists engaged in qualitative research rely heavily on the narratives shared by their respondents to attribute meaning to a phenomenon. The authors believe that the prevailing practice of fragmenting long narratives into short vignettes during the analysis of narratives deprives the researchers from capturing the full meaning of narrative, further leading to inferior analysis. The present study describes various pitfalls arising out of vignette based analyses of narratives and suggests an alternative mixed method. The proposed design combines the qualitative rigor of thematic analysis with the graphical explanatory power of nonlinear principal component analysis. The proposed design preserves the respondent stories in their entirety during classifying the codes to themes and while interpreting each theme in detail.

***Keywords*:** Analysis of narratives**;** vignette; principal component analysis; gestalt; mixed method

Introduction

Humans live storied lives. Any experience that a person endures can naturally be cast in a story; generally referred to as a personal narrative (Connelly & Clandinin, 1990). Hence, a story, which is real and a first person narration, can qualify as a fundamental psychological unit of experience (Kohler-Riessman, 2000). A personal narrative is a person's recollection of real life experience surrounding an event or a phenomenon (Lindseth & Norberg, 2004). To this extent, personal narrative can be defined as a logical and temporal production of a gestalt - lived experience that is stated in respondent's own words in the form a story. In qualitative studies involving personal experiences, a researcher's interest in analyzing personal narratives does not lie in the confirming the validity or authenticity of the experience narrated, but to interpret the personal meanings (story) the narrator gives to his or her experience during the process of narration (Huberman & Miles, 2002) .

Narrative analysis is the analysis of singular narratives (Polkinghorne, 1988), while analysis of narratives is the study of personal narratives, recited by several participants about the same phenomenon (Riley & Hawe, 2005). Popular methods for analysis of narratives include content analysis, discourse analysis, grounded theory and phenomenology (Charmaz, 2003; Hsieh & Shannon, 2005; Giorgi, 1985; Wodak & Meyer, 2009). General process outlining all these analytical methods is more or less similar. Stories are generally broken down into short snippets/vignettes and then the meaning of the narrative are derived out of these passages, which most of the time become isolated and out-of-context (Huberman & Miles, 2002). Mishler (1986) has cautioned researchers against this practice of fragmenting narratives for analysis. A story creates meaning not only through its textual content, but also from its background. A reader must be made aware of the related contexts regarding the event(s), such as, the place (socio-cultural factors), the time (past, present and future) and the plot/genre of the story to make sense of the narrative (Connelly & Clandinin, 1990). A major drawback of the vignette-oriented analysis is that the flow of the story is disrupted due to such abrupt puncturing and further analysis of isolated vignettes might lead to framing out-of-context, meanings for the vignette or not being able to fully comprehend the significance of text that forms the part of the vignette. Authors like Spector-Mersel (2011) are of the opinion that a holistic strategy for interpretation should be employed where the narratives are treated as whole units both in content as well as in form. Therefore, the authors of this work believe that any kind of vignette or snippet based analysis of personal narratives is unintuitive, irrational and leads to inferior analysis.

In this paper, we propose a novel mixed-method research design for thematic analysis of personal narratives that does not require puncturing of the story. We call this method as principal component analysis of narrative corpus. In this mixed-method design, we combine the qualitative rigor of thematic analysis with the graphical explanatory power of nonlinear principal component analysis that is used as a data reduction technique in quantitative research.

The rest of the paper is organized as follows: firstly, we demarcate story from other forms of qualitative data. Next, we detail the six stages of (vignette/snippet based) thematic coding, which is popularly used to analyze personal narratives. Subsequently, we discuss in detail the five problems that could ensue because of vignette-based analysis of narrative corpus (a collection of personal narratives about a phenomenon). Later, we introduce a quantitative research method called as nonlinear principal component analysis, which is used for quantitative data reduction of categorical data. We then propose a method, which combines thematic coding scheme with nonlinear principal component analysis to perform analysis of narratives, which obliterates the pitfalls of vignette, based analysis.

Personal narratives must be differentiated from other forms of verbal qualitative data (Cumming, 2007). According to simplified Labov's (1982) typology, personal narrative has an introduction (beginning), a complicated action and an ending. Personal narratives are marked by context, chronological sequence, and plot(s) (Franzosi, 1998). The statements within the narrative, which describes about the location, timing and the manner of occurrence of an event, along with the related entities (persons, objects etc.), form the context of narrative (Labov, 1982). The context of the narrative elicits requisite background information for the reader or listener to fully appreciate the story. A personal narrative is a natural unfolding of events of the past as experienced by the narrator. In a stable narrative, the evaluations of situations remain the same over time, while in a progressive (regressive) narratives, the evaluations become better (worse) over time (Gergen & Gergen, 1984). Hence, structuring the narrative in the order of chronological sequence is important to fully comprehend the latent meanings in the narrative. Plot(s) of a narrative is the organizing theme, which brings coherence to disparate events that compose the narrative (Riley & Hawe, 2005). The plot imparts meaningful logic to various events (which may seem disparate in isolation) and is understood only through full comprehension of the narrative.

General thematic coding system of qualitative data

One generic research skill that all professional qualitative researchers should possess is the ability to see patterned responses across data sets (Holloway & Todres 2003; Johnson, Onwuegbuzie, & Turner, 2007). Different qualitative research traditions exhort seemingly diverse procedures through which these patterns, underlying in the data, can be extracted and analyzed so as to ascertain the meaning of the phenomenon. Braun and Clark (2006) made a meticulous study of various qualitative data analyses, namely interpretive phenomenological analysis, grounded theory, content analysis, discourse analysis and found out common underlying six stage procedure followed in all these methods (also see Ryan & Bernard, 2000 for similar argument). The researchers argue that this general six stage procedure of refining qualitative data should be conceived as a scientific, qualitative method in its own right that is not wedded to any analytic tradition. They named this procedure as thematic analysis.

Since, all popular analytical traditions dealing with analysis of narratives (such as grounded theory, phenomenological analysis, content analysis, discourse analysis, etc.) confine to the broad spectrum of thematic analysis (Braun & Clarke 2006), we describe it in detail. While thematic analysis can be applied to variety of qualitative data sets such as stories, interviews, print media clippings, audiovisual content, we limit our discussion to thematic analysis of a narrative corpus.

In this paper, we define a narrative corpus as a collection of personal narratives, recited by multiple participants with respect to a particular phenomenon. Before conducting thematic analysis, all narratives should be transcribed verbatim. Unit of analysis according to thematic analysis is a single story. In the present context, each personal narrative qualifies as one item (unit of analysis). However, if a respondent elicits two different stories of his personal experience surrounding the same phenomenon, we have two items.

In the first stage, the transcription of stories is read and re-read to identify patterns within a story as well as across all the stories. This process is called as data immersion. Each story is then dissected to form individual vignettes (i.e. codes or meaning units) that describe a particular aspect of the phenomenon in participant’s own words. During this second stage, each vignette is given an appropriate code name that succinctly summarizes the essence of the vignette. Two vignettes from two different stories dealing with the same aspect of the phenomenon will bear the same code name. In the third stage, codes are collated across stories to form fully coherent and internally consistent latent themes such that a code from any story can be mapped to one of the researcher's devised latent themes. After this exercise, researcher no longer interprets a particular participant's story; instead, meanings are derived by interpreting isolated vignettes (Huberman & Miles, 2002) that are assigned to each theme.

In the fourth stage, the researcher examines the codes assigned to each theme for possible omissions, additions and inconsistencies in codes comprising the theme. Themes so generated are reviewed for possible duplication or anomalies and subsequently the researcher may perform some additional exercises, such as, collapsing different themes into one single theme, dividing a single theme into multiple themes or expurgating those themes, which pose as deviation from the proposed research objective. After satisfactorily performing this review, the themes and code names comprising the themes are finalized. In the fifth stage, researcher names every theme in a manner that it captures the essence of all the code names that comprise the theme. In the final stage, the researcher then describes the comprehensive understanding of each latent theme independently. In an attempt to support his/her understanding, the researcher intersperses the interpretive text with disparate and interesting vignettes from multiple stories that cater to the a priori envisaged theme. It is presumed that the practice of embedding vignettes in interpretation of latent theme(s) enhances auditability of the research and helps the reader in appreciating the participants' (subjective) worlds through participants' own words, to a certain extent (Onwuegbuzie & Teddlie, 2003).

Other existing methods of vignette based analysis of narratives

Lincoln and Guba (1985) advocate the cutting and sorting technique to analyze narratives. For example, the index card technique mandates that researchers have to break down narrative passages into small sections and then further deconstruct the sections into shorter snippets or bites. Once the snippets are sufficiently small (say, two to three words), codes are assigned to them and researchers continue towards further analysis. Barkin and Colleagues (1999) advocate theme generation as a two-stage process. At stage one, the researcher identifies themes from theory and in the next stage, independent coders pile sort the themes in the similarity matrix form. Theme based analysis of narratives are standard practice followed by majority of qualitative researchers (Richmond, 2002; Ames, Duke, Moore, & Cunradi, 2009).

However, Gubrium and Holstein (1999) have cautioned researchers that conducting narrative analysis with the help of short snippets may lead to a fragmented analysis of narratives at best, described by researchers as ‘interlocutionary arrangement of external narrative patterns’. Although, Barkin and colleague’s (1999) method incorporates independent coders into the preparation thus reducing the primary researcher bias, one might still argue that the a priori theme selection at stage one itself may lead to the introduction of biases in the analysis.

Pitfalls of vignette based analysis of personal narratives

Mishler (1995) identifies the referential use and order as the degree to which the narrator presents background information about the central characters in his/her discourse. In addition, it reflects the effort put by the narrator to carve out the nature and kind of relationships between the various actors in the narrative. A vignette based analysis of narratives often fail to capture these nuances in the content, which in turn leads to the partial attribution of meaning to a situation/action (Mello, 2002).

In vignette based analysis, even though the researcher attempts to fill the stated context and reference use related voids by corroborating each vignette with a brief description of the context in which the statement is uttered, and referential use and order is mentioned; the evaluation of such statements cannot be fully made from the participants' subjective world. Except for the stable stories, evaluation of the subject and the object of the statement may change temporally in progressive/regressive stories.

Labov (1982) suggests that all narratives have plot, which has temporal and logical sequences attached to it, which, if not counted for, will not give a holistic picture of the story's objective. Therefore, unless the reader is fully aware of the plot of each story, it is difficult to comprehend any event in isolation (Haden & Hoffman, 2013). An event that could be classified as redundant or peripheral when viewed in isolation may as well be a critical plot event that gives overall coherence to a narrative (Denzin, 1989).

Narratives, being a reproduction of gestalt phenomenon, are imbued with contextual information, reference use and order, temporality and plot. They are complex qualitative data sets, which lose their comprehensibility when fragmented. Therefore, the authors of this work believe that narrative data should not be subject to traditional qualitative methods that perform analysis by puncturing the data. Denzin (1989) also suggests that narratives should be interpreted as logical and temporal production of a gestalt phenomenon that which cannot be sliced.

In a recent paper, Haden and Hoffman (2013) have supported the need for preserving contextual information about stories while analyzing narratives in order to keep the temporal and spatial sequences intact in the minds of readers. In their opinion, the context of the story can be best conveyed to readers through details about the place and time where the story is plotted. We opine that the context of the story helps readers understand the latent motivations and thoughts of the story’s characters that later led to their actions. Thus, embedding the story with context enhances explanatory power to the narrative.

Overcoming the pitfalls of vignette based analysis of narratives

Mello (2002) has argued against vignette-based analysis of narratives suggesting the need for analyzing narratives as whole, rather than as the sum of parts since the latter may cause the story to get lost behind the incoherent short pieces of data. As a solution, the author has put forward a collocation analysis as a means of capturing narratives in their entirety in order to create textual and cognitive bridges between the actual story and the interpretation of the same. Collocation analysis is also known as word co-occurrence approach and is developed on the principles of semantic network analysis. The inherent appeal of this approach lies in its ability to produce matrices containing words which are co-occurring simultaneously in multiple responses (Barnett & Danowski, 1992). Since the word co-occurrence matrix is developed from computer software (such as ANTHROPAC), coder bias in the selection of themes gets nullified. However, this method does not incorporate the entire narrative, rather only the recurring words from a narrative; hence, the bias in representation of narrative cannot be ruled out. Such techniques fail to capture the entire fabric of narratives in terms of representing their context, temporal and spatial sequence and the plot inherent in them. Therefore, there is no guarantee of reduction of bias during the analysis of narratives through this method since it is perfectly possible to visualize patterns hidden in the co-occurred words, which may not exist in reality (Ryan & Bernard, 2003).

As a deviation to the norms, Jehn and Doucet (1996) have refrained from cutting and sorting individual narratives in their work. The researchers first identify certain context-specific scenarios and then ask a different pool of respondents to develop a scenario-by-scenario similarity matrix. Following the matrix generated by the participants themselves, the researchers used Multidimensional Scaling (MDS) to analyze the data. This method has its advantage of generating themes from respondents themselves and therefore the issue of investigator bias is countered effectively. However, the problem still persists because the final analysis is done with representative scenarios, not the entire story.

A similar technique known as meta-coding, once again pioneered by Doucet and Jehn (1997), allows for inclusion of a fixed set of short descriptive paragraphs and a fixed set of a-priori themes which are later used for developing a unit -by-theme matrix. The advantage of meta-coding is that the data can be used for further quantitative analyses such as correspondence analysis or factor analysis, which can reveal which a-priori themes are best represented by which type of responses. However, the method is robust only when the texts are short paragraphs and the number of paragraphs is not large in number (Ryan & Bernard, 2003). This is essentially a critical shortcoming of this method, since the analysis of narratives may require interpretation of lengthy verbal accounts.

The life Story approach is one such technique where the researcher immerses so deep in their analysis that it becomes increasingly difficult to distinguish between what constitutes the interviewee’s viewpoint and what is the researcher’s interpretation (Bertaux & Kohli, 1984). This creates the issue of researcher bias. As a remedy to this, researchers have advocated unobtrusive collection of narratives with selective interjections from researchers (Attanucci, 1991; Bell, 1999). Still, the researcher is confronted with the problem of the extent to which he/she should interfere with the respondents to elucidate their comments.

A thorough review of extant literature points out no method in practice, which treats the story/narration as the unit of analysis. It is our belief that by including stories/narration in their entirety and by letting the participants and independent coders generate the latent themes, both the issues of narrative fragmentation and researcher bias can be addressed effectively. In a bid to prove this justification, the following section discusses a novel approach of narrative analysis, which we term as "nonlinear principal component analysis of narrative corpus".

Nonlinear principal component analysis

Linear principal component analysis (PCA) is a data reduction technique that is often used in quantitative research with metric (ratio and interval) data (Abdi & Williams, 2010). Through this technique, a large number of variables are reduced to a smaller set while preserving as much variance as possible in the original data (Wilks, 1938). Each principal component is a weighted combination of all the original variables. Variance accounted by each principal component is called eigen value. Theoretically, the number of principal components that are generated by a PCA solution is equal to a number of variables. The principal components are iteratively derived in a manner such that the variance accounted by any principal component is always less than its preceding component, implying that the first principal component has always the highest eigen value (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Hence, in a standard PCA solution, first few components dominate the solution by incorporating a high proportion of total variance. In practice, a researcher uses only first few principal components instead of the original variables, thus obtaining high parsimony in data representation while sacrificing only a little variance in the process. To perform PCA, desired minimum sample size is generally three times the number of variables.

Nonlinear principal component analysis (NLPCA) is akin to a PCA with few exceptions (Jolliffe, 2002). The first difference is that NLPCA can incorporate non-metric variables such as nominal data and ordinal data (Guttman, 1959). Second, in NLPCA, non-metric data are converted to metric data through optimal scaling procedure (Kruskal & Shepard, 1974). Through this procedure, category level for each non-metric variable is assigned a value, thus rescaling the categorical data to metric data. Consequently, linear PCA is performed on this data. In a nonlinear PCA, researcher has to a priori state the required number of principal components (Ferrari & Barbiero, 2012). Then optimal scaling is performed iteratively, such that the total variance accounted by the solution is maximized for the desired number of principal components. The principal components obtained by NLPCA are orthogonal by default, implying no shared variance between the components. Third, the sample size is not a deterrent in performing NLPCA. A minimum of three data points is sufficient to run NLPCA

Biplot

Biplot succinctly represents the relationships between the variables, objects and variables and the objects in a low dimensional space (Gower & Harding, 1988). In the biplot generated by NLPCA, principal components serve as axes for biplot. Generally, a two dimensional solution is sought in NLPCA as it is conducive for easy graphical analysis. Each variable is represented as a vector. The square of the length the vector is proportional to the variance accounted by the variable in the solution. Variables of relatively small length should generally be discarded as they indicate variable misfit or inadequate explanation of variable in the solution. The cosine of the angle between the vectors represents the extent of correlation between the vectors. This implies that lesser the angular distance between the vectors, the greater is the correlation between the (rescaled) variables. Orthogonal projection of an object point on vector describes the object's evaluation of a particular variable. Higher projection indicates a better evaluation, if the variable is ordinal or metric scaled.

Principal component analysis of narrative corpus

Principal component analysis of narrative corpus is a mixed method research technique that utilizes elements of both thematic analysis and nonlinear principal component analysis for synthesis of narrative data. We now detail this procedure in stepwise fashion using simulated data (see table 1 for a description of the dataset).

The researcher must collect personal narratives surrounding the phenomenon from multiple participants. Depending upon the nature of the phenomenon to be studied, a suitable sampling technique may be used. Participants may be encouraged to tell more than one story surrounding the phenomenon. If new stories do not deal with the previous actors, but are deemed relevant for study context, they can be incorporated into narrative corpus. To this extent, multiple stories recited by single participants serve as independent data units. Researchers need to transcribe the stories stated by the respondents in respondent's own words. Non-verbal gestures like long pauses, false starts, change of tone, and change in facial emotions may be reported alongside transcription if such gestures are anticipated to convey additional insights into the topic. Established transcription guidelines (cf. Powers, 2005) may be used for transcribing the audio / audio-visual data. For the purpose of current analysis, researcher stated words are not transcribed as they were not coded.

After all the stories are transcribed, researcher must read and reread all the story transcripts, both for checking accuracy of coding and getting a comprehensive overview about the phenomenon. This stage is equivalent to the data immersion stage of thematic analysis. In the second stage, the researcher should read every transcript carefully and assign code names to significant lines in the transcript. The exhaustive code name list should be generated after transcribing all the transcripts. Expert opinion may be sought to remove duplications (synonymous code names) and misnomers. Once the code list is finalized, two or more researchers should code the entire narrative corpus using the master coding scheme.

Table 1: Story code array

Participant No.	Story No.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
1	1	2	2	2	2	2	2	2	1	1	2
1	2	1	2	1	2	1	1	1	2	1	1
2	3	1	1	2	2	1	1	1	2	2	1
3	4	1	1	2	1	1	1	2	1	2	2
4	5	1	1	1	2	2	1	2	1	1	1
4	6	1	2	1	1	1	1	2	1	1	2
4	7	1	1	2	2	2	2	2	2	2	1
5	8	2	1	2	2	1	2	2	2	2	1
6	9	2	1	2	2	1	2	2	2	1	1

Notes: Presence or absence of coeds in a story is designated as 1= absent and 2=present. Participant 1 and 4 told multiple stories, but are considered as individual data points for analysis. C1 to C10 is the master code name list. Participant number is not part of story code array. For story 1, we may observe that C1 to C7 are present in the story while C8 to C10 are not present.

When inter-coder reliability is established, they may proceed to the next stage, else the entire coding process is to be reiterated until the master list of codes is agreed upon and high inter-coder reliability is obtained. This procedure is similar to the second stage of naming the codes in thematic analysis.

The third stage deals with generating a story code array. Rarely researcher encounters a story that captures an entire master list. In other cases, every story captures only a few code names from the master coding list. Story code matrix may be generated to find out to which elements of master list each story captures. Story code array is a two dimensional matrix with row as master coding scheme and column as independent story units and elements of the matrix represent whether a particular code name is represented in the story or not. For each story, if the code name is represented in the story, the corresponding element in the matrix should be coded as “2”, otherwise “1”. The entire story code matrix can be generated using this principle. We prefer “2” for presence of code in the story and “1” for absence instead of binary coding scheme (1,0) since NLPCA recognizes “0” as missing data, if implemented through SPSS.

In the fourth stage, NLPCA is performed on story code array. For the purpose of NLPCA, each story serves as an object while every code name serves as an independent variable (vector). All variables are declared as ordinal since the presence of code name is superior to absence of code name [the values are designated accordingly (2 >1)]. A two dimensional biplot is sought. A two dimensional graph is recommended as it is conducive to less strenuous visual interpretation. We use SPSS software with Categories add-on to perform NLPCA as it has easy point and click interface. Open source R software with HOMALS library implementation can also be used to generate NLPCA biplot output using command line interface.

The vectors on the biplot are graphical representations of code names on the biplot map as straight lines emerging from the origin. If two vectors have less angular distance between them, it implies to state that they share similar patterns in the presence and absence of respective code names across all the stories in the narrative corpus. Statistically, this visualization indicates high correlation between the code names, since low angular distances between vectors represent high correlation and vice-versa (mathematically, correlation (r) = cosө; lim (ө 0) cos(ө) =1 lim (ө 0) r = 1). For the purpose of the study, vector cluster is defined as group of vectors which share low angular distance with each other. Each vector cluster serves as a factor (equivalent to that of "theme" in qualitative research), since variables that combine to form a factor should have high correlation amongst themselves. Every vector is assigned to one of existing vector clusters depending upon the criteria of minimum angular distance. However, if the angular distance from the nearest cluster approaches to 90 degrees, it should not be assigned to any cluster since this occurrence indicates low shared variance or near zero correlation with other vector cluster members (r = cos ө; cos 90 = 0 r=0). In such cases, the code name serves as a standalone theme. Every vector is identified as belonging to a vector cluster or designated as stand-alone. Each code name is mapped to only one vector cluster. Akin to thematic analysis, in this stage we categories all code names for possible themes.

<<Insert fig. 1 over here>>

In the fifth stage, each vector cluster and standalone vectors are given a theme name depending upon the code name(s) contributing to it. The theme name should represent the essence of the latent theme described by the code names in the cluster. The names of themes must be cross-validated by fellow researchers for their representation and meaningfulness. Through this process, we have successfully mapped code names to themes (vector clusters on biplot) and name the theme, similar to that of the sixth stage of thematic analysis

In a biplot, stories are represented as individual points. Closer the points on the map, similar are the configurations of the stories with respect to their representation of code names. Every vector cluster is mapped with a story that has the highest orthogonal projection with the largest vector of that cluster. Only the largest vector is considered while mapping stories, since largest vector represents a code name whose variance is most represented in the solution. Highest orthogonal projection is required since, greater the projection, better is the representation of a particular vector (cluster) in that story. We may, at times, evidence a single story sharing highest orthogonal projection with more than one vector cluster; in such cases, the same story could be assigned to more than one vector clusters.

<<Insert fig. 2 over here>>

The sixth stage is the akin to thematic reporting. In this stage, the researcher should describe and interpret the themes obtained from NLPCA biplot analysis. However, instead of interspersing the text with vignettes from multiple stories, the researcher should narrate the story that is assigned to the theme and interpret the entire theme using the story as illustrative evidence/interpretive medium. Story, being a recitation of gestalt experience, can contain information related to more than one theme; therefore to draw attention to the content of the stated theme, significant lines corresponding to the theme should be highlighted in the narrative and they should be interpreted more comprehensively. For lengthier stories, the narrative transcript may be abridged without losing the temporality, reference use and order and plot. However, care should be taken to produce the significant lines corresponding to the theme in its entirety.

<<Insert fig. 3 over here>>

Validating the analysis

Because this approach delivers themes based on statistical output, the need for further corroboration from multiple researchers in establishing validity and reliability of the themes is obviated (if positivist - empirical standpoint is assumed). Researcher's role in establishing qualitative validity and reliability is then limited in extent of incorporating and validating master coding list. However, if the research project is conceived as stage wise to first stage as pilot testing and second stage as confirming the results in different/related contexts, then additional validation step needs to be incorporated as follows. Independent researchers who are not part of the pilot study are asked to replicate the process using a different sample (holdout sample) and the similarity in the themes must be ensued to validate the findings.

Discussion

In this paper, we have attempted a novel method that enables the combined analysis of narrative corpus without fragmenting any narrative. This procedure is unique in many ways. First, without losing qualitative foothold, and at the same time ensuring empirical rigor on the same qualitative data set, this method synergistically combines the prowess of empirical paradigm with the explanatory power of interpretive paradigm. Second, to the best of our knowledge, it is the first of its kind research approach to extract "variance" from stories, and subject it for further analysis. Third, unlike other mixed method designs, same qualitative dataset, which is obtained in the first phase of the study, is quantized during the quantitative phase of analysis obviating the need for successive quantitative data collection. Hence, we believe sampling scheme for this method should be strictly qualitative, since, quantitative data collection is not required for the study.

This method follows a structure similar to that of thematic analysis. As the thematic analysis is not wedded to any paradigm (Braun & Clarke, 2006); accordingly, our method, which is based on thematic analysis, is also postulated as paradigm free. Therefore, once the stories/narratives are mapped to the themes, researcher may advance to interpret them from one of the existing qualitative traditions: interpreting only stated words (semantic interpretation), interpretation that includes assigning meaning to nonverbal gestures (latent interpretation), interpreting a story from a stated theoretical perspective (relativistic interpretation), interpreting the story from the multiple competing theoretical perspectives (realistic interpretation), etc. (Vindrola-Padros & Johnson, 2014).

Direction for future research

Objective quantification of qualitative data is gaining currency only recently (Silverman, 2006). To this extent, we believe that further strides should be made in qualitative research to move from binary coding as depicted in this research to incorporate more complex coding schemes such as -2 to +2 to capture the intensity of the phenomenon in a continuum. For example, in a hospital setting, a patient in the intensive care unit might appreciate frequent monitoring by paramedical staff, while patients in non-critical condition, might construe such frequent monitoring as annoyance or fail to attribute any significance to it. Hence, frequent monitoring as a code when quantized can take multiple values such as +1, -1 and 0 depending upon the interviewee's subjective evaluation of the event. However, if an attempt is also made to capture the degree to which frequent monitoring impacts the patient's evaluation of hospital staff (from very good to very bad), then complex coding system spanning from -2 to +2 must be incorporated. In such situations, the quantized qualitative data are laden with more dispersion and gives more accurate results when subjected to further quantitative analysis involving variance-covariance (correlation) analysis. We anticipate greater synergy between the fields of categorical data analysis, and quantizing qualitative data, in the manner we demonstrated, can lead to more comprehensive insights while interpreting narrators' motivations, ensuing greater clarity in rationalizing their behavior and gaining deeper access to their subjective worlds of beliefs.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*, 433-459.

Ames, G.M., Duke, M.R., Moore, R.S., & Cunradi, C.B. (2009). The impact of occupational culture on drinking behavior of young adults in the U.S. Navy. *Journal of Mixed Methods Research*, *3*, 129-150.

Attanucci, J. (1991). Changing Subjects: Growing Up and Growing Older. *Journal of Moral Education*, *20*,317-328.

Barkin, S., Ryan, G., & Gelberg, L. (1999). What clinicians can do to further youth violence primary prevention: A qualitative study. *Injury Prevention, 5*, 53–58.

Barnett, G., & Danowski, J. (1992). The structure of communication: A network analysis of the international communication association. *Human Communication Resources, 19,*164–285.

Bell, S.E. (1999). Narratives and lives: women's health politics and the diagnosis of cancer for DES daughters. *Narrative Inquiry 9,* 1-43.

Bertaux, D., & Kohli, M. (1984). The life story approach: A continental view. *Annual Review of Sociology*, 215-237.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77-101.

Charmaz, K. (2003). Grounded theory. *Qualitative psychology: A practical guide to research methods*, 81-110.

Connelly, F.M., & Clandinin, D.J. (1990). Stories of experience and narrative inquiry. *Educational Researcher, 19,* 2-14.

Cumming, J. (2007). *MEDIAREVIEW:* Using narrative in social research qualitative and quantitative approaches**.** *Journal of Mixed Methods Research,1*, 200-201.

Denzin, N. K. (1989). *The research act* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Doucet, L., & Jehn, K.A. (1997). Analyzing harsh words in a sensitive setting: American expatriates in communist China. *Journal of Organizational Behavior*, *18*, 559-582

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272.

Ferrari, P. A., & Barbiero, A. (2012). Nonlinear principal component analysis. *Modern Analysisof Customer Surveys: with applications using R*, 333-356.

Franzosi, R. (1998). Narrative analysis-or why (and how) sociologists should be interested in narrative. *Annual Review of Sociology*, 517-554.

Gergen, M. M., & Gergen, K. J. (1984). The social construction of narrative accounts. *Historical Social Psychology*, 173-189.

Giorgi, A. (Ed.). (1985). *Phenomenology and psychological research*. Duquesne University Press.

Gower, J. C., & Harding, S. A. (1988). Nonlinear biplots. *Biometrika, 75*, 445-455.

Guttman, L. (1959). Metricizing rank-ordered or unordered data for a linear factor analysis. *Sankhyā: The Indian Journal of Statistics*, 257-268.

Haden, C.A., & Hoffman, P.C. (2013). Cracking the code: using personal narratives in research. *Journal of Cognition and Development, 14*, 361-375.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*, 1277-1288.

Holstein, J.A., & Gubrium, J. F. (2000). *The Self We Live By: Narrative Identity in a Postmodern World*. New York: Oxford Univ. Press.

Huberman, A. M., & Miles, M. B. (2002). *Narrative analysis. In The qualitative researcher's companion*. Thousand Oaks, CA. pp. 217-270.

Jehn, K. A., & Doucet, L. (1996). Developing categories from interview data: Text analysis and Multi-dimensional Scaling. Part 1. *Cultural Anthropology Methods Journal, 8,* 15–16.

Jolliffe, I. (2002). *Principal component analysis*. John Wiley & Sons, Ltd.

Johnson, R.B., Onwuegbuzie, A.J., & Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*, 112-133.

Kohler-Riessman, C. (2000). Analysis of personal narratives. *Qualitative Research in Social Work*, 168-191.

Kruskal, J. B., & Shepard, R. N. (1974). A nonmetric variety of linear factor analysis. *Psychometrika*, *39*, 123-157.

Labov, W. (1982). Speech Actions and Reactions in Personal Narrative. In *Analyzing Discourse:Text and Talk,* edited by D. Tannen. Washington, D.C.: Georgetown University Press.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Lindseth, A., & Norberg, A. (2004). A phenomenological hermeneutical method for researching lived experience. *Scandinavian journal of caring sciences*, *18*, 145-153.

Mello, R. A. (2002). Collocation analysis: a method for conceptualizing and understanding narrative data. *Qualitative Research, 2,* 231-243.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Mishler, E.G. (1995). Models of narrative analysis: A typology. *Journal of Narrative and Life History*, *5,* 87-123.

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. *Handbook of mixed methods in social and behavioral research*, 351-383.

Polkinghornc, D. E. (1988). *Narrative knowing and the human sciences*. State University of New York Press, Albany

Powers, W. R. (2005) *Transcription techniques for the spoken word.* Lanham, MD: AltaMira.

Richmond, H.J. (2002). Learners' lives: a narrative analysis. *The Qualitative Report*, *7*, 1-11.

Riessman, C. K. (2008). Thematic analysis. *Narrative methods for the human sciences*, 53-76.

Riley, T., & Hawe, P. (2005). Researching practice: the methodological case for narrative inquiry. *Health education research*, *20*, 226-236.

Ryan, G. W., & Weisner, T. (1996). Analyzing words in brief descriptions: Fathers and mothers describe their children. *Cultural Anthropology Methods Journal*, *8,* 13–16.

Ryan, G. W., & Bernard, H. R. (2000). Data management and analysis methods. In *Handbook ofqualitative research*, 2d ed., edited by N. Denzin and Y. Lincoln, 769–802. Thousand Oaks, CA:Sage.

Ryan G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods, 15,* 85-109. Tesch, R. 1990. *Qualitative research: Analysis types and software tools*. New York: Falmer.

Silverman, D. (2006). *Interpreting qualitative data: Methods for analyzing talk, text and interaction*. Sage.

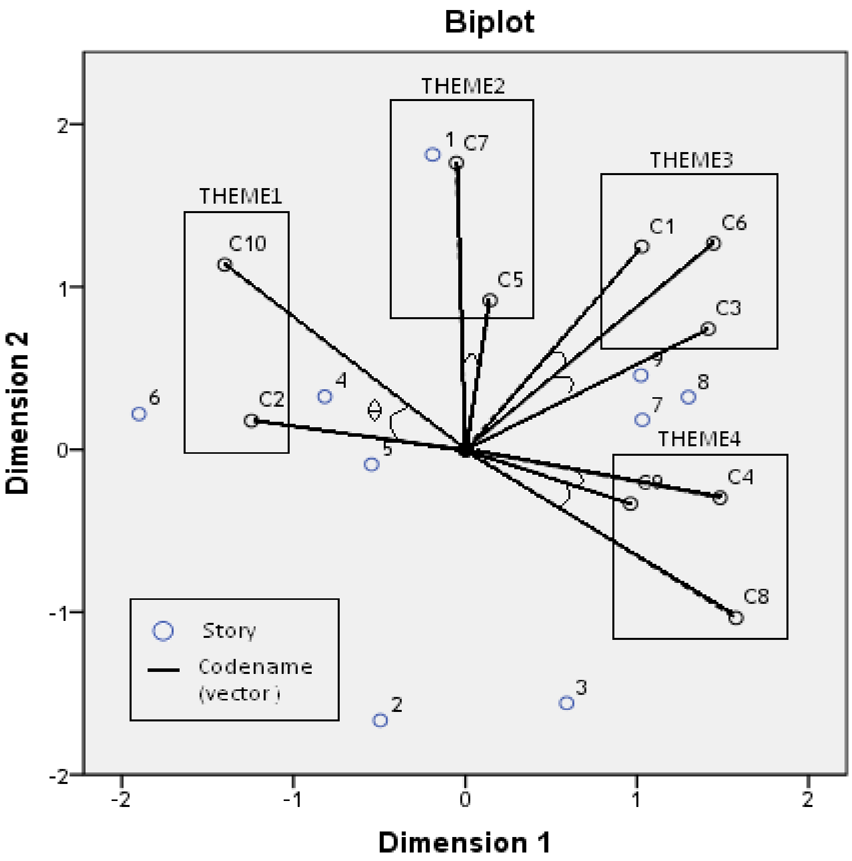
Spector-Mersel, G. (2011). Mechanisms of selection in claiming narrative identities: A model for interpreting narratives. *Qualitative Inquiry*, *17*, 172-185.

Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, *3,* 23-40.

Wodak, R., & Meyer, M. (Eds.). (2009). *Methods for critical discourse analysis*. Sage.

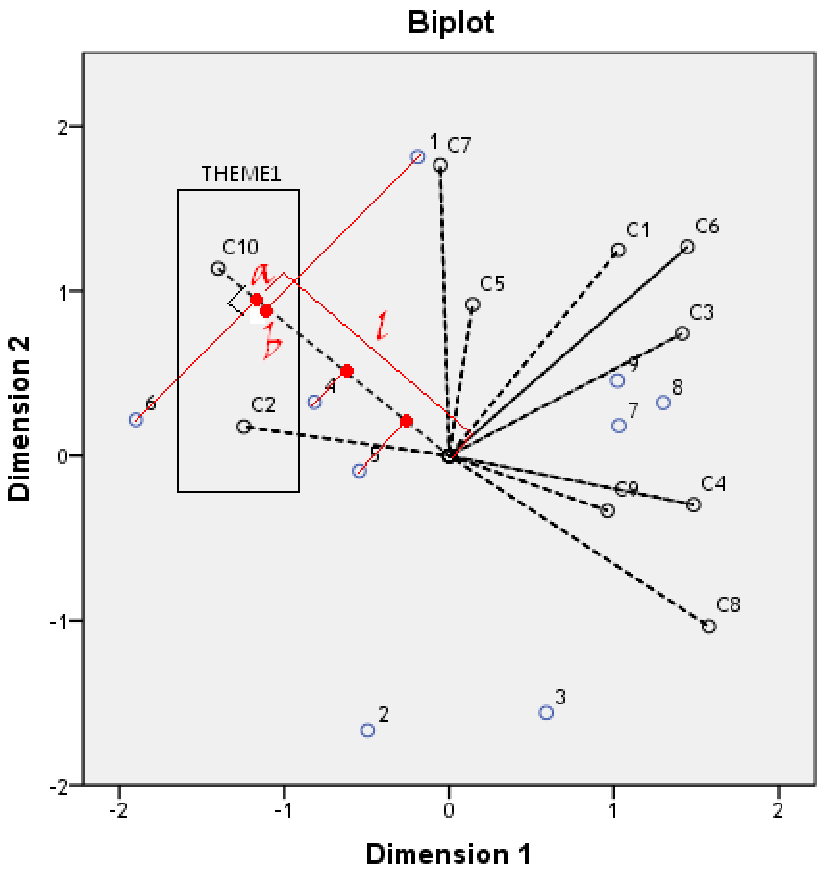
Vindrola-Padros, C. & Johnson, G.A. (2014). The Narrated, non-narrated, and the dis-narrated: Conceptual Tools for Analyzing Narratives in Health Services Research. *Qualitative HealthResearch, 24*, 1603-1611.

Figure 1: Identifying vector clusters aka themes



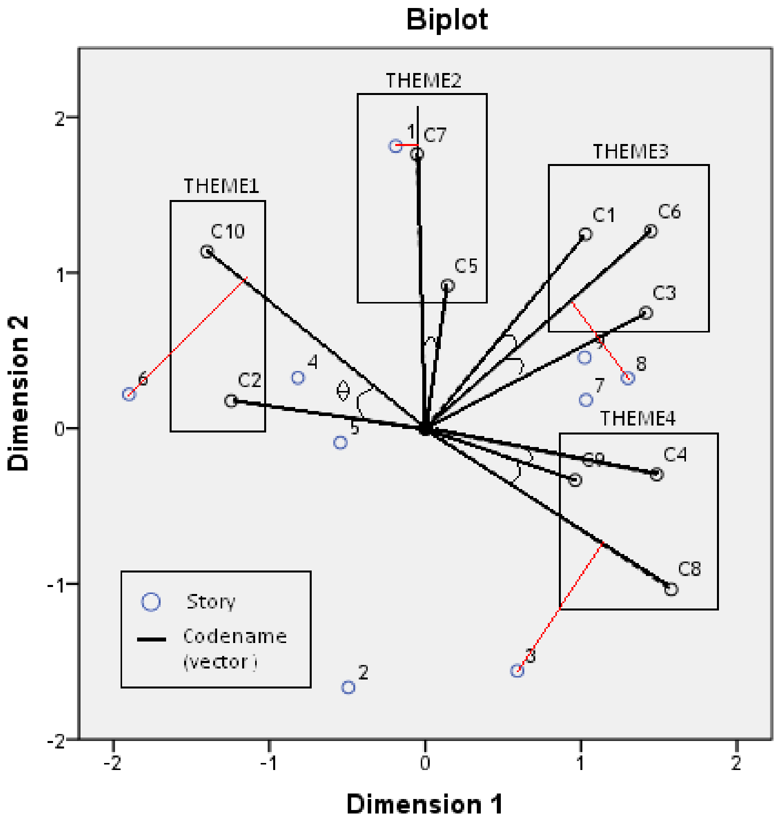
Note: C2 and C10 combine to form a vector cluster (aka theme 1) based on low angular (ө) distance between them

Figure2: Interpretation of orthogonal projections of biplot



Note: Longest vector in theme1 is C10, hence only orthogonal projections on C10 are considered for assigning a story to this theme. Othogonal projection is perpendicular line drawn from the story to the vector. Orthogonal projection of Story 6 intersects C10 at point "a" while story1 intersects origin at point "b", Story 6 is assigned to the theme 1 since length of projection of story 6 (distance of "a" from origin, denoted using "L") is the longer compared to any other projection. .

Figure 3: Complete analysis of biplot



Note: Story 6 is assigned to theme1, story 1 is assigned to theme2, story 8 is assigned to theme 3 and story 3 is assigned to theme 4.