The Qualitative Report

Volume 23 | Number 7 How To Article 10

7-15-2018

Visual Inspection of Sequential Data: A Research Instrument for Qualitative Data Analysis

Alaa A. AlDahdouh
University of Minho, AlaaAlDahdouh@gmail.com

Visual Inspection of Sequential Data: A Research Instrument for Qualitative Data Analysis

Abstract

Sequence analysis has been widely used to investigate the patterns of similarities and differences of sequential data in biology and sociology. However, the debate on the usage of sequence analysis in social sciences has not been settled yet. Among a long list, sequence analysis methods have been criticized for ignoring the qualitative information behind the sequences. This paper presents a new instrument for inspecting sequential data visually in qualitative studies. The method includes building a hierarchical tree of relations among the categories which is then used to recode the categories systematically. The recoding process is meant to give meaning to the differences among categories and, therefore, increases our ability to see the differences. The instrument is a fruit of a qualitative study carried out to explore student's learning patterns. The focus in this paper will be on the algorithm of recoding the categories and how the emergent codes can be plotted to generate insights for further qualitative investigation.

Keywords

: Sequence Analysis, Qualitative Studies, Learning Patterns, Data Analysis, Categorization, Sequences

Acknowledgements

I am grateful to Professor António J. Osório and Professor Susana Caires for their valuable advice and encouragement throughout the qualitative study in which this instrument has been developed.

The Qualitative Report 2018 Volume 23, Number 7, How To Article 2, 1631-1649

Visual Inspection of Sequential Data: A Research Instrument for Qualitative Data Analysis

Alaa A. Aldahdouh
University of Minho, Braga, Portugal

Sequence analysis has been widely used to investigate the patterns of similarities and differences of sequential data in biology and sociology. However, the debate on the usage of sequence analysis in social sciences has not been settled yet. Among a long list, sequence analysis methods have been criticized for ignoring the qualitative information behind the sequences. This paper presents a new instrument for inspecting sequential data visually in qualitative studies. The method includes building a hierarchical tree of relations among the categories which is then used to recode the categories systematically. The recoding process is meant to give meaning to the differences among categories and, therefore, increases our ability to see the differences. The instrument is a fruit of a qualitative study carried out to explore student's learning patterns. The focus in this paper will be on the algorithm of recoding the categories and how the emergent codes can be plotted to generate insights for further qualitative investigation. Keywords: Sequence Analysis, Qualitative Studies, Learning Patterns, Data Analysis, Categorization, Sequences

Sequence Analysis (SA) refers to a wide set of (qualitative and quantitative) methods to analyze the evolution of a phenomenon over time (Abbott, 1995; King, 2013; Macindoe & Abbott, 2004). Obviously, a sequence is the main data type utilized for conducting SA and it refers simply to an ordered list of states or elements (Macindoe & Abbott, 2004). The sequence is sometime seen as a trajectory of individual evolution through predefined states over a period of time (King, 2013). For example, if the state space includes a set of jobs, then the sequence refers to the ordered list of jobs an individual has passed through over a period of time. SA is conceptualized here as the process of understanding and categorizing the patterns of similarities and differences in, and among, sequences.

In the history of SA development, some quantitative SA methods have been considered as a key factor for a coming *revolution* in social sciences (Abbott, 1995; Abbott & Hrycak, 1990) but the fact of the matter is that there has been little agreement on using quantitative SA to date (Arosio, 2004; King, 2013; Lalanda Nico, 2016). As King (2013) has pointed out, a longitudinal analysis of *ordinal* data has enjoyed a satisfactory acceptance among researchers, but it fails to have such acceptance for *categorical* data. Among a long list of criticisms, there is a concern on how to justify representing categorical data on an ordinal scale and how to justify representing the differences between categories in numbers. In Optimal Matching Analysis (OMA), for instance, researchers have often made subjective decisions by building what they called *a replacement cost matrix* and *indel cost* (Abbott, 1995; Abbott & Hrycak, 1990; Macindoe & Abbott, 2004). Bakeman and Gottman (1997) has added that "Developing a coding scheme is very much a theoretical act,… and the coding scheme itself represents a hypothesis, even if it is rarely treated as such" (p. 15). This subjective intervention has been criticized (King, 2013) while advocates of quantitative SA (Abbott, 1995; Abbott & Hrycak, 1990; Bakeman & Gottman, 1997; Bakeman, Quera, & Gnisci, 2009; Macindoe & Abbott, 2004) argue that the subjective decisions in social science are justified by the nature of the subject being measured. After all, and as Johnson and Onwuegbuzie (2004) said, "the conduct

of fully objective and value-free research is a myth, even though the regulatory ideal of objectivity can be a useful one" (p. 16).

The last two decades have seen a growing trend towards mixed methods research which rejects the dichotomous distinction between quantitative and qualitative methods (Johnson & Onwuegbuzie, 2004). The idea is to make use of the strengths and minimize the weaknesses of each single research approach and, hence, increases the possibility to understand a phenomenon. Lalanda Nico (2016) contends that the discussion of quantitative SA has dominated the field for many years while far too little attention has been paid to the role of the qualitative SA. Omitting or dismissing the meanings and narratives of individuals behind the sequences has resulted in wasting the opportunity to gain more insights from the data. Debate continues but the researchers in the field have reached a conclusion that each technique of analyzing the sequences has its own value and contribution, where probing the same problem using different methods may be a better choice (Arosio, 2004; King, 2013).

This paper proposes a new instrument which could be in hand of researchers who are interested in qualitative SA studies. The instrument was designed as part of a qualitative study to investigate learning patterns of higher education students in Palestine. As it is the case in the most of qualitative studies, the major characteristics of the study were induction, discovery, exploration, and theory generation (Johnson & Onwuegbuzie, 2004). In this sense, the proposed instrument was of benefit to generate insights for further investigation. It is important to note that the proposed instrument is particularly useful in provoking thought more than judging and drawing conclusions from data. This means that the visual inspection of the sequence should be accompanied by a qualitative inquiry of the underlying relations of the emerging patterns. In the sections that follow, the reader may have a sense that the instrument is mostly quantitative tool which involves a lot of calculations. I don't argue that this instrument has nothing to do with the quantitative analysis. It most probably does. But I argue that it may be useful for the qualitative researchers who track phenomena that grow over time. And such phenomena are certainly prevalent in social sciences. A researcher who is interested in learning about the way the students experience their school day, for instance, may also need to track the sequence of classes the students have attended throughout the entire day. There may be a consistent pattern of experiences among the students in a given class or may be a consistent pattern of experiences across the classes for a given student. The proposed instrument helps qualitative researchers to explore such questions.

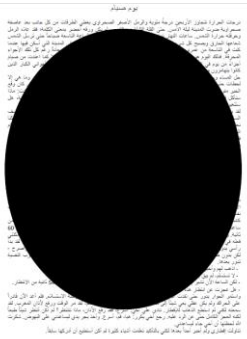
Data of Analysis

The data used in the analysis examples came from a qualitative study in which nine participants were asked to disclose their learning activities while trying to solve 10 experimental tasks. The aim was to investigate the applicability of connectivism learning theory and its ability to interpret the learning activities of students in higher education institutions (Aldahdouh, 2017; Aldahdouh & Osório, 2016; Aldahdouh, Osório, & Caires, 2015; Downes, 2006, 2008; Siemens, 2005, 2006; Siemens & Tittenberger, 2009). The participants were allowed to use all resources they wish without being restricted by any constraint, including the time required to complete the task. After solving each task, they were asked to participate in an aided retrospective think-aloud protocol (Hofer, 2004; Kuusela & Paul, 2000; Van den Haak, De Jong, & Schellens, 2003, 2004). Participants in the aided retrospective think-aloud protocol, also known as "prompted retrospective protocol" (Kuusela & Paul, 2000) or "retrospective verbal protocol" (Ericsson & Simon, 1980) or "actual retrospective protocol" (Van den Haak et al., 2003), are usually asked to complete each task of the experiment silently and then verbalize their thoughts in a follow-up session while watching a video recording of their activities. In this study, however, a slight modification on the method

was adopted. Namely, the participants were not monitored *while* solving the tasks and this was to conform with the connectivism's principles which insist on giving the students the freedom to do whatever they want as well as to conform with the recent findings in the think-aloud literature where the presence of the researcher was reported as less functional and more harmful to the participant's performance (Peute, de Keizer, & Jaspers, 2015; Van den Haak et al., 2003, 2004). At the beginning of each session, the participants were asked to provide a sequence of their steps in an ordered list. Since the experiment involved nine participants and 10 tasks each, 90 sequences were generated in total. The average number of steps *per* sequence was 4.82 while the minimum and the maximum number of steps *per* sequence were one and 16 steps respectively. The tasks were organized into 10 different categories as shown in Table 1 below.

Table 1. Categories and the questions provided for the participants

Code	Category	Description	Example
Q01	Information Search	Simple task that requires gathering information which is available but scattered over the internet. The participant should find information from different resources.	Prepare a comprehensive list of toxins and antibiotics.
Q02	Investigation of Person	Search for a Palestinian character and create a complete profile of him/her. Plenty of information about the character already exist on internet. The participant should be able to orient him/herself and decide which information to read and which one to skip. The participant should also be able to provide an opinion about the character.	Gather info about a Palestinian character, Imad Farajin. Prepare a complete profile of his life, history, and achievements. Include your opinion about him.
Q03	Question in a Field of Study	Each participant has a unique question related to his/her field of study at the university. The task should induce the participant to search for an evidence-based information in his/her major specific area. The participant should search in trustworthy resources.	Search for the medicine named "Gaviscon." Specify for what purposes it has been used and what its components are. Prepare a list of all equivalent medicines and their prices.
Q04	Self-Motivation Question	The participant is given a chance to pick the topic they wish to search for. The question is designed to allow the participants to exercise their volition and pursue the topic they really interested in.	There might be a question that you had encountered and didn't find time to search for or to read about in the past. Take your time to remember and to search for it.
Q05	Info Validation	This task induces a participant to search for a topic which is controversial or uncertain on the internet. The participant should be able to make his/her decision based on uncertain information.	Check the validity of the information that eating fish with milk (or any dairy products) is unhealthy? Support your answer with details and evidences.

Code	Category	Description	Example
Q06	Compound Task	This task involves sub-tasks. The participant is asked to search for a named scholarship and to apply for it. The task involves writing essays and gathering information for a scholarship.	Search for a "Hani Qaddumi Scholarship Foundation." Prepare your files to apply for a scholarship to cover your tuition fees in the next year.
Q07	Essay Writing	The participant is asked to write a scientific essay about a topic related to his/her field of study. The participant should gather information from trusted resources, remix and repurpose them in a blend.	Write an essay about citrus fruits. The essay should include the gossips about their harmful effects on human health along with the reality and misconceptions of their benefits.
Q08	Design Question	A participant is asked to provide a new sketch or design for something he/she is familiar with. The task targets the ability to imagine. The participant should use his/her imagination to see old things differently.	Design a sketch for your mobile's home-screen as you wish it to be.
Q09	Creativity	This High-level task requires the participant to be greatly creative. In this question, the participant is also invited to imagine, but his/her imagination should be guided and framed by a given set of clues and constraints. The participant should make use of the clues and the connections between them while using his/her imagination in the rest. The task should not exist on the internet at all.	The sheet, you have, is a short story with a hole appearing on in the middle. The hole covers a considerable part of the script. Do your best to recover the missing part. Note, you should make use of all parts shown so the whole story becomes consistent. 
Q10	Technical Question	This task is out of the participant's field of study. It is for an expert in the field of Information Technology. It is a call for action which is very difficult for anyone out of the field. The task is meant to monitor how the participant establish a connection to one of the experts in a field.	You have an Excel file that you work with every day. You need to backup this file every day at a certain time. You should name the backup file as the same as the file name followed by the date, and to keep those backup files up to one week. This backup mechanism should be done automatically.

For purposes of illustration and for the sake of simplicity, a simple example of two participants and two tasks was selected from the experiment data to exemplify the method in

detail. Table 2 shows the description provided by participants A and B of their steps on tasks 1 and 2.

Table 2. The steps of participant A and B on tasks 1 and 2

	Task 1		Task 2	
	#	Step description	#	Step description
Participant A	1	Ask her friend (chat on Facebook)	1	Internet search (laptop, general exploration)
	2	Refer to a book (Principles of Medical Pharmacology)	2	Internet search (laptop, copy and paste some information)
	3	Refer to a study notes (printed PowerPoint slides on pharmaceutical chemistry)	3	Asked her brother (Face-to-Face or F2F, to give her a link)
	4	Refer to a book (Pharmacology 4th Edition Lippincott's)	4	Thinking (connecting and rewriting)
	5	Refer to a book (Clinical Medicine)	5	Cooperation with sister (F2F)
	6	Internet search (laptop)		
Participant B	1	Hand writing on her notebook	1	Search internet (laptop, found another person of the same name)
	2	Search internet (mobile, to see the Table of Content of the school textbook)	2	Search internet (laptop)
	3	Search internet (laptop, for possible solutions)	3	Send a message to the character directly (chat on Facebook)
	4	Think on her own (to create new games)	4	Search internet (laptop)

Categorization

The term *categorization* can broadly be defined as the process of demarcating the state space in order to serve the purpose of the study (King, 2013). According to Bakeman and Gottman (1997), categorization process is the lens through which researchers see the world and it is the stage where mistakes are irrevocable. Deciding whether to include, exclude, or merge elements should be carefully taken. For example, in Table 2, participant A referred to some *books* in addition to a *study note* in task 1. The decision here is whether one should add *books* and *study note* as two distinct states in state space or simply combine them into one state called hard copy material, for example. The state space refers here to a collection of all states, steps, or stages the participants have passed through while in the study. Some researchers tend to detail the state space by adding all elements and then count on the statistical analysis to figure out which elements seem to make differences (Bakeman & Gottman, 1997; King, 2013). However, the researcher did not pick this suggestion because it harms the analysis and makes the visual inspection of the sequences harder.

Of course, the categorization is a theoretical process par excellence and is closely related to the aim of the study. Since the aim of the qualitative study was to examine the applicability of connectivism learning theory, the starting point was to see how connectivism defines learning and how it distinguishes between different learning resources. In connectivism, learning is defined as "a continuous process of network exploration and patterns finding; it is a process of patterns' recognition" (Aldahdouh et al., 2015, p. 14). According to Siemens and Tittenberger (2009), there are three main levels of learning networks: neural, conceptual, and external. Therefore, our attention should be on those resources which magnify the differences among the three broad levels. Accordingly, a distinct resource list was generated as shown in Table 3.

Table 3. Distinct resources in tasks 1 and 2

Step
Ask people (online)
Ask people (F2F)
Refer to hard copy material
Internet Search (laptop)
Internet Search (mobile)
Write
Think

Hierarchal Tree of Relations

Apparently, the "distinct" resources in Table 3 still have relative similarities which can be used to build a hierarchy of relations (Bernauer, Lichtman, Jacobs, & Robinson, 2013; Ford, Oberski, & Higgins, 2000; Schulz, 2012). The hierarchy is usually built inductively with careful consideration of the aim of the study too (Bakeman & Gottman, 1997; Ford et al., 2000). Connectivism theory does not provide a detailed description on how the categories should be linked together to form a hierarchical tree (Braun & Clarke, 2006). The links between the categories – and not the categories themselves – are the ones which need to be built inductively. Even though, the theory does provide a broad categorization which includes the internal and external nodes. And this broad categorization guides the process of linking the categories in one way or another. It is also important to know that connectivism theory is still developing (Aldahdouh et al., 2015) and, thus, the broad categorization itself may not, and should not, be complete. Elo and Kyngäs (2008) contended that when the experiment data present categories outside the boundary of the theory, one may adopt *unconstrained categorization* which starts with, but is not limited to, the categories presented in the theory. The categories which appear in data but do not appear in the theory suggest that the theory needs to be updated. Finding a pattern of relations between categories is a common practice in qualitative analysis (Bernauer et al., 2013; Elo & Kyngäs, 2008; Ford et al., 2000; Johnson & Onwuegbuzie, 2004; Mayring, 2000). Since we are interested in determining when participants consulted external nodes (people and things) and when they consulted internal nodes (think and write), the resources were relinked to amplify the differences between internal and external nodes as shown in Figure 1.

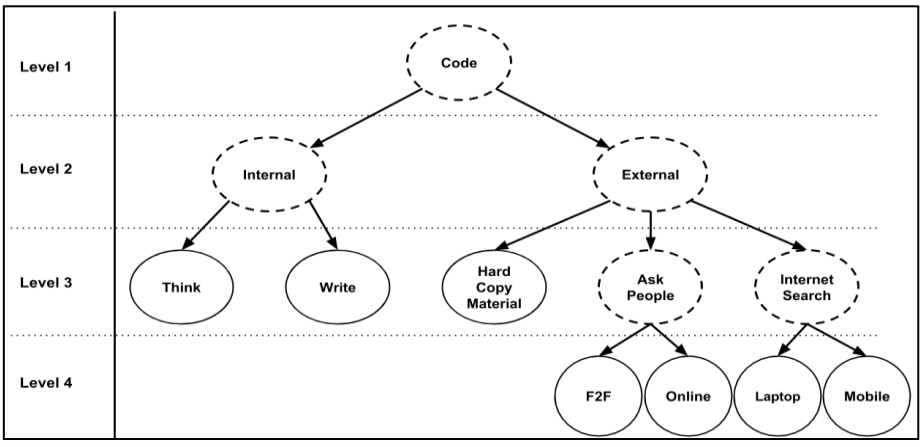


Figure 1. Simple hierarchical tree of relations

While building a hierarchy, one should note that the nodes with no children (leaf nodes) are the same as those distinct resources identified previously in Table 3. Nodes with dashed lines are theoretical nodes and they are used to link the leaf nodes up to the top node in the tree (root node). The root node resides in the highest level of the tree (level 1) while other theoretical

nodes reside in the next levels (level 2 and 3). It's worth noting that the deepest level in the tree (level 4) should always be occupied by leaf nodes but not all leaf nodes must reside in the deepest level.

Recording the Categories

Recoding the categories refers to the process of giving numerical value for each category (leaf node) in the hierarchical tree. The numerical values will be used later to represent the categories in the graphical representation of the sequences. The recoding process is meant to give a meaning to the differences between categories. In other words, the numerical differences between the nodes should reflect the theoretical differences between the categories in the hierarchical tree depicted in Figure 1. It is quite important to recognize that we should agree in this stage on the merit of the hierarchical tree to portray our theoretical understanding. The recording process will make use of the hierarchical tree to infer about the numerical differences between the nodes and, certainly, the failure of the hierarchical tree to depict our thoughts implicitly means that the subsequent analysis becomes useless.

Theoretically, it seems reasonable in our example if we can depict the data so that it appears clearer (with greater distance) when the participants move from writing on their own (Write node) to referring to book (Hard-Copy Material node) than when they move from asking people online (Online node) to face-to-face (F2F node). After all, both asking people online and face-to-face indicate the participant's preference to communicate with other people. Thus, both nodes deserve to appear closer to each other in the sequence illustration. The question now is, how can we infer this kind of theoretical differences from the hierarchical tree solely? It is by recognizing that, in the hierarchical tree, both asking people online and face-to-face share the same parent while Write and Hard-Copy Material nodes do not have such property. And thus, identifying whether the nodes share the same parent in the hierarchical tree has helped us to infer about the difference. The next section will present the criteria used to infer from the hierarchical tree to the theoretical differences.

But before delving into details, why don't we think to keep it simple? The simplest way to recode the categories is to give each leaf node a number sequentially. Therefore, the codes start with 1 for Think node and end with 7 for Mobile node. If recoding is done sequentially as such, the difference between Write node (2) and Hard-Copy Material (3) will be one, which is exactly the same as the difference between F2F node (4) and Online (5). Of course, this does not reflect the theoretical differences in hierarchical tree and we must find an alternative way.

In the pages that follow, the researcher proposes an algorithm to recode the categories systematically. A software program was developed (see Appendix) to facilitate the recoding process. While the discussion of the algorithm may seem daunting, the benefit to the reader in understanding of the operation is fundamental. The algorithm will be explained below with some sort of detail and will be supported by illustrations to make it easier to follow.

The Criteria and the Technique

Based on the hierarchical tree created in the previous step, I refined four criteria the algorithm should adhere to and for good reasons. First, the algorithm must generate a *distinct* code number for each leaf node in the hierarchical tree to guarantee that each leaf node has a chance to appear in the illustrations. Suppose otherwise that the algorithm generates the same code number for two distinct nodes. What this really mean is that the difference between those distinct nodes is completely vanished and it becomes impossible to notice as to when the participants move from one node to another. Second, the difference between sibling nodes (nodes of the same parent) must be smaller than the difference between cousin nodes (nodes in

the same level but of different parents). This criterion mirrors the fact that, in any hierarchical tree, the sibling nodes have more in common than the cousin nodes. In Figure 1, take the nodes in level 3 as an example. Think and Write nodes are sibling nodes since they share the same parent (Internal node) while Write and Hard-Copy Material nodes are cousin nodes because they lie on the same level, yet they do not share the same parent. Assuming that the hierarchical tree depicts our understanding correctly, we are most probably interested in seeing greater difference between cousin nodes than between sibling nodes. Being close to each other in the hierarchical tree should be then reflected in the differences between the nodes, as we initially proposed. Third, the deeper the level, the lower the difference between the sibling nodes. In Figure 1, this means that the difference between sibling nodes in level 4 must be smaller than the difference between sibling nodes in level 3; which in turns must be smaller than the difference between sibling nodes in level 2 and so on. The criterion takes into account that the sibling nodes in higher level are quite distinct and the difference between them should be seen more clearly in the illustrations. Fourth, the algorithm must be flexible, so it allows researchers to stretch out the differences among all nodes. The point is to empower the researchers to make salient differences in the graphical representations but without forfeiting the comparative value of each node as listed in the previous three criteria.

In order to satisfy the criteria, the algorithm calculates three types of variables: weight, gap, and code number. The *weight* is a number given to each node to emphasize its weight in comparison to other nodes. The weight will be used to calculate the code number so that the larger the weight, the larger the code number. The *gap* refers to the distance between sibling nodes in each level. Each level has one gap value for its sibling nodes. The *code number* is the output of the algorithm and it refers to the value which is going to represent each category in the illustrations. The algorithm traverses the hierarchical tree twice. In the first round, the weights and gaps are computed while the second round is dedicated to compute the code numbers.

Recoding Algorithm

- Assigning the weights for nodes in the deepest level.

Recoding process begins by assigning weights for the nodes in level 4, the deepest level in the current example. The weight of the sibling nodes starts with 1 and is separated by a *separation factor* (SF).

$$weight(node) = \begin{cases} 1, & rank(node) = 1 \\ weight(node_{-1}) + SF, & rank(node) \neq 1 \end{cases}$$

Weight(node) and *weight(node$_{-1}$)* denote the weight of the current node and the previous sibling node respectively. The position of the current node in regard to its siblings is referred to as *rank(node)*. *SF* is a positive integer number selected by researchers. The larger the separation factor, the greater the distance between nodes. For this example, SF will be assigned a value of 1. Each group of sibling nodes should start counting again (see Figure 2). Therefore, the weights of the deepest level's nodes are as the following: F2F (1), Online (2), Laptop (1) and Mobile (2).

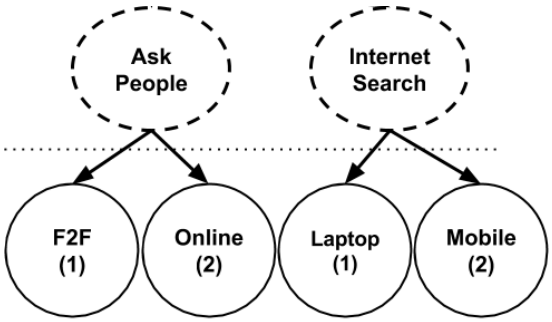


Figure 2. Assigning weights in the deepest level

- Finding the gap between the parents in level 3

To compute the gap between the parent nodes, one should recall that the distance between the parents should be greater than the distance between the children nodes. In this case, the number of children of each parent matters. Suppose one parent has five children then the gap between this parent and other parents should take into consideration the number of the children of this parent (5) along with the number of children of the other parents. Recall also that the gap between sibling parents in each level has one value. To maintain all constraints, I compute the gap between the sibling parents in each level based on the two parents who have the maximum number of children, which will be referred to as $Parent^{max}$ and $Parent^{max\text{-}1}$. In specific, the gap is calculated based on the weight of the last child of $Parent^{max}$ and $Parent^{max\text{-}1}$ as the following:

$$gap = 2 \times \mathrm{SF} \times WLC(Parent^{max-1}) + WLC(Parent^{max})$$

Where *WLC* denotes the weight of the last child of a given parent.

Accordingly, the gap between sibling nodes in level 3 is calculated as the following (see also Figure 3): The gap = 2 × SF(1) × weight of Mobile node (2) + weight of Online node (2) = 6.

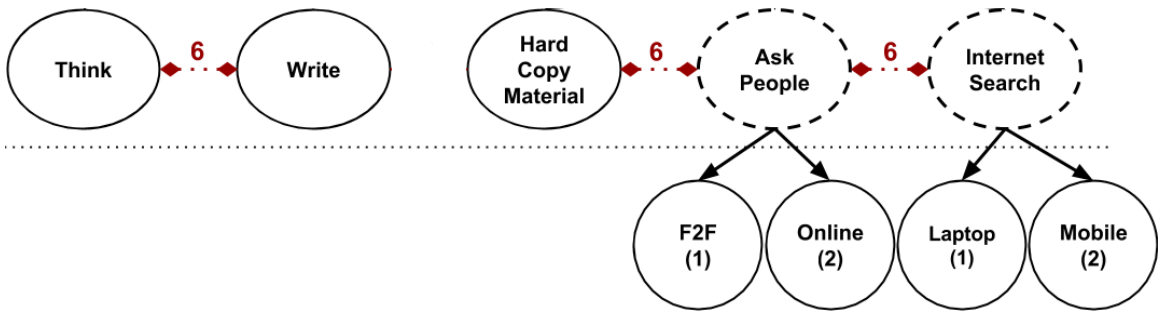


Figure 3. Finding the gap in level 3

- Assigning the weights for nodes in level 3

Finding the weights in level 3 and the successive levels is similar to the process of finding the weights in level 4 (the deepest level) except that it depends on the calculated gap instead of SF. The calculated gap implicitly takes the value of SF into consideration as shown in the previous step. The weight of the sibling nodes in level 3 starts with 1 and is separated by the gap.

$$weight(node) = \begin{cases} 1, & rank(node) = 1 \\ weight(node_{-1}) + gap, & rank(node) \neq 1 \end{cases}$$

The weights are calculated as follows: Think (1), Write (1+6), Hard Copy Material (1), Ask People (1+6), and Internet Search (7+6). Figure 4 shows the gap and the weight of each node in level 3.

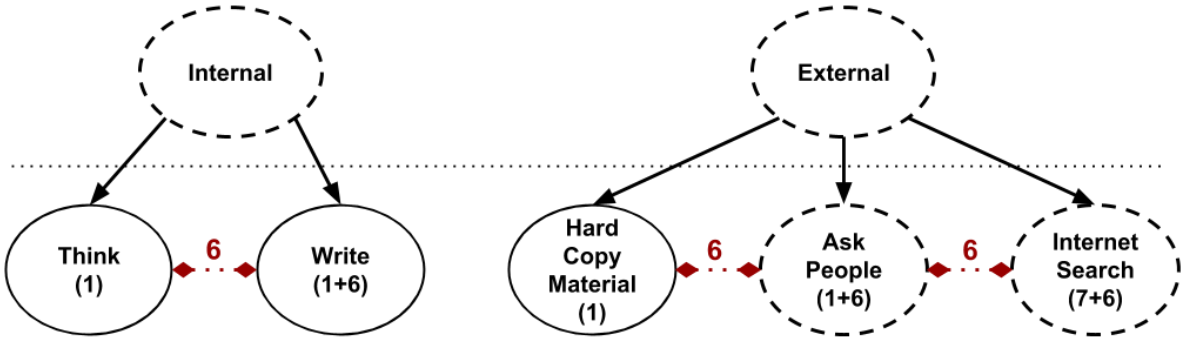


Figure 4. Assigning weights in level 3

- Finding the gap and assigning the weights of nodes in Level 2

The previous two steps are used to calculate the gap between the sibling nodes in level 2 along with their weights. The gap is 27. Therefore, the weights of Internal and External nodes are (1) and (1+27) respectively. The whole tree will be as shown in Figure 5 below:

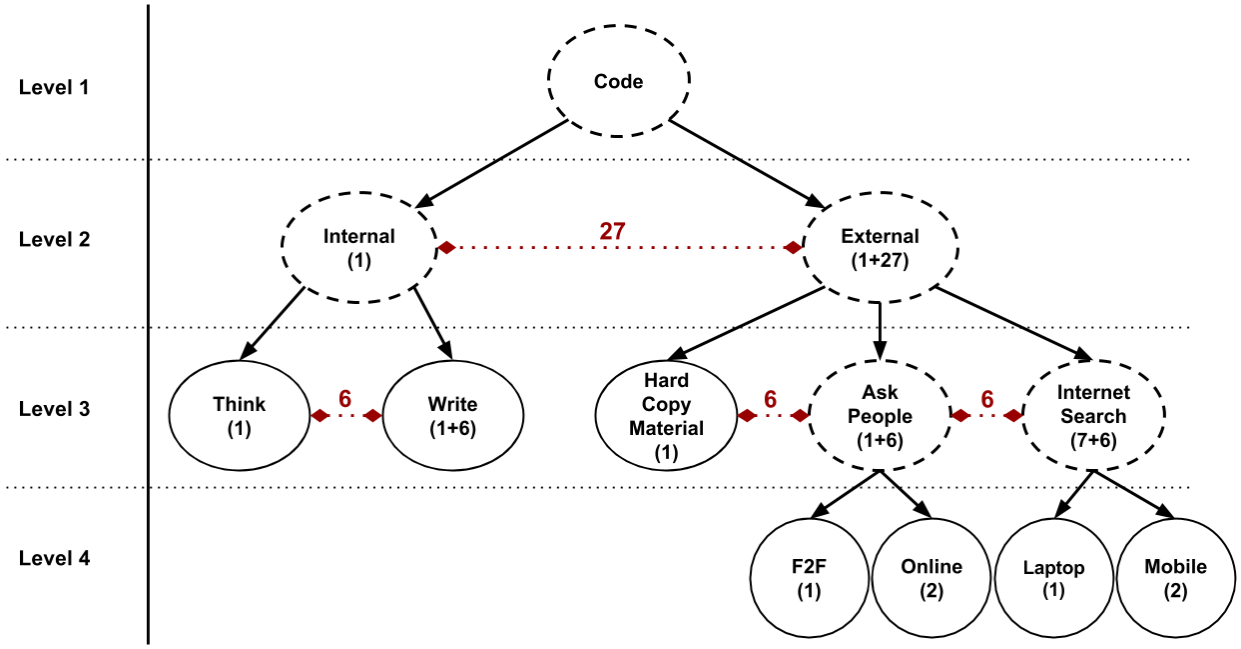


Figure 5. Finding the gap and assigning the weights in level 2

- Finding the code numbers.

The code number of any node is the sum of all weights up to the root node. For example, Mobile code is calculated as the following: Mobile (2) + Internal Search (7+6) + External (1+27) = 43. The code numbers of all leaf nodes are shown in Figure 6 below:

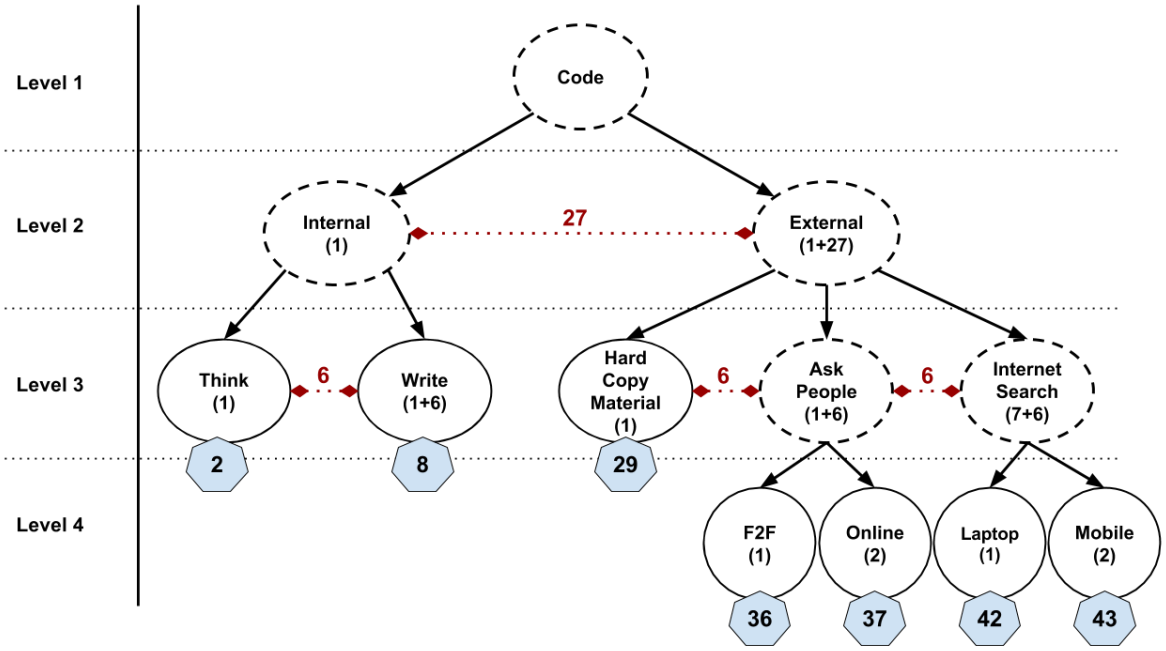


Figure 6. The code numbers of all nodes

The algorithm generates the hierarchy tree in Figure 6 which satisfies the four criteria crafted earlier. For example, the algorithm generates distinct code number for each leaf node. The biggest difference between two adjusting leaf nodes are between Write (8) and Hard Copy Material (29). This reflects the fact that those two nodes are belonging to completely different top parent nodes (internal and external) and this will make inspecting the jump between these nodes easily. Moreover, the difference between sibling nodes in upper levels is bigger than the difference between sibling nodes in lower levels. In short, using the suggested algorithm, one generates differences which have meaning, and which can be used in the visual inspection of the sequences.

Consequently, the code number of all participants' steps are generated in Table 4.

Table 4. The code number of participants' steps on tasks 1 and 2

	Task 1		Task 2	
	#	Step code	#	Step code
Participant A	1	37	1	42
	2	29	2	42
	3	29	3	36
	4	29	4	2
	5	29	5	36
	6	42		
Participant B	1	8	1	42
	2	43	2	42
	3	42	3	37
	4	2	4	42

Analysis and Presentation

The presentation is a means through which the results are communicated with the community. John W. Tukey (1977), a statistician who developed one of the most used methods to calculate the significant differences between a set of means, argued that the best way to communicate and to get useful insight from numbers is not by submitting them to the statistical tests; it is by plotting, charting, and graphing them. This paper suggests visualizing participants' steps and recognizing the patterns of similarity and differences among tasks and participants using charts. As King (2013) argues, figures are not only complementing the text, they are telling the story behind the sequences.

There are two dimensions which can be analyzed using the code numbers: (1) comparing tasks – the question can be formulated as: Does the pattern of participants' steps in task 1 differ from their patterns in task 2? (2) comparing participants – the question can be formulated as: Does the pattern of participant A's steps differ from participant B's steps in all tasks? Data should be rearranged according to the desired dimension. Table 5 below rearranges data to compare tasks.

Table 5. Rearranging participants' steps to compare between tasks

	Task 1					
Steps	1	2	3	4	5	6
Participant A	37	29	29	29	29	42
Participant B	8	43	42	2		
	Task 2					
Participant A	42	42	36	2	36	
Participant B	42	42	37	42		

Any charting software can be used to plot the data. For this example, I used Microsoft Office Excel to draw two charts (one for each task) as shown in Figure 7.

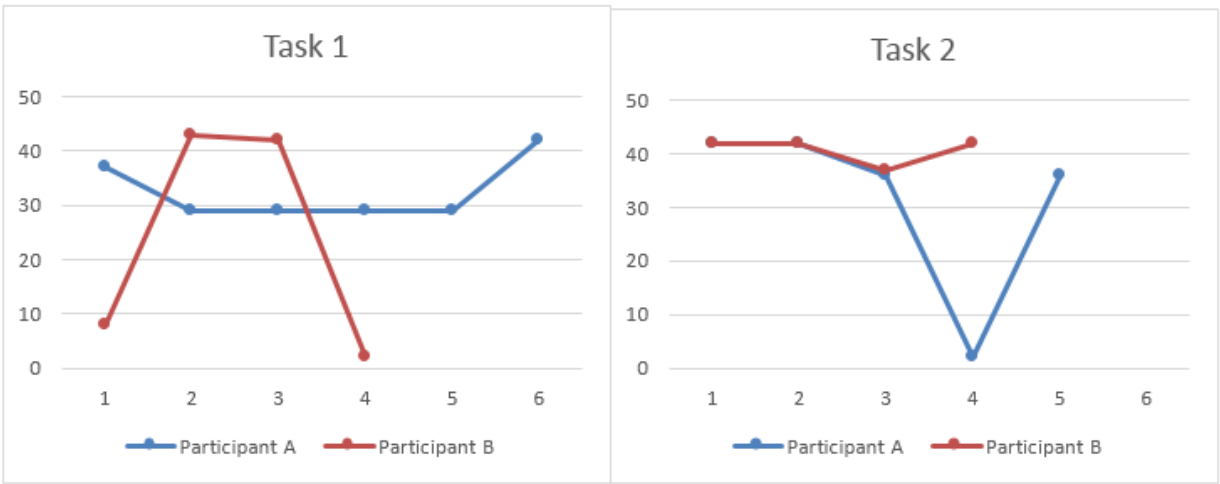


Figure 7. Charts of participants' (A and B) steps on tasks (1 and 2)

The horizontal axis represents the steps while the vertical axis represents the nodes (categories). The lower values in the vertical axis (2 to 8) are those nodes belonging to Internal category while the upper values in the vertical axis (29 to 43) correspond to External category. Each chart line (red or blue) represents one participant. Attention should be paid to the following points while searching for patterns of similarities and differences between tasks:

- How many steps, in average, did the task take for the participants to complete?
- Where do most participants spend their steps, in the upper (External) or in the lower (Internal) side?
- Is there any notable consistency between participants in their starting and ending steps?
- When do participants seem to "jump" from one side (upper) to another (lower)?
- How does the "shape" of the behavior's pattern look like (e.g., an animal with four legs, braid, or slide)?

The charts should be used to stimulate thinking on how each task may be compared to other tasks or on how each participant may be compared to other participants.

Visual Inspection in the Qualitative Study

The sequences in Figure 7 show two participants' steps in two tasks. Of course, the visual inspection becomes harder – and more useful – when the number of sequences increases. Figure 8 provides examples of how the illustrations were used in the qualitative study to generate insights from the data. Four questions (Q04, Q07, Q09, and Q10) are selected for this purpose from the list of the questions in Table 1.

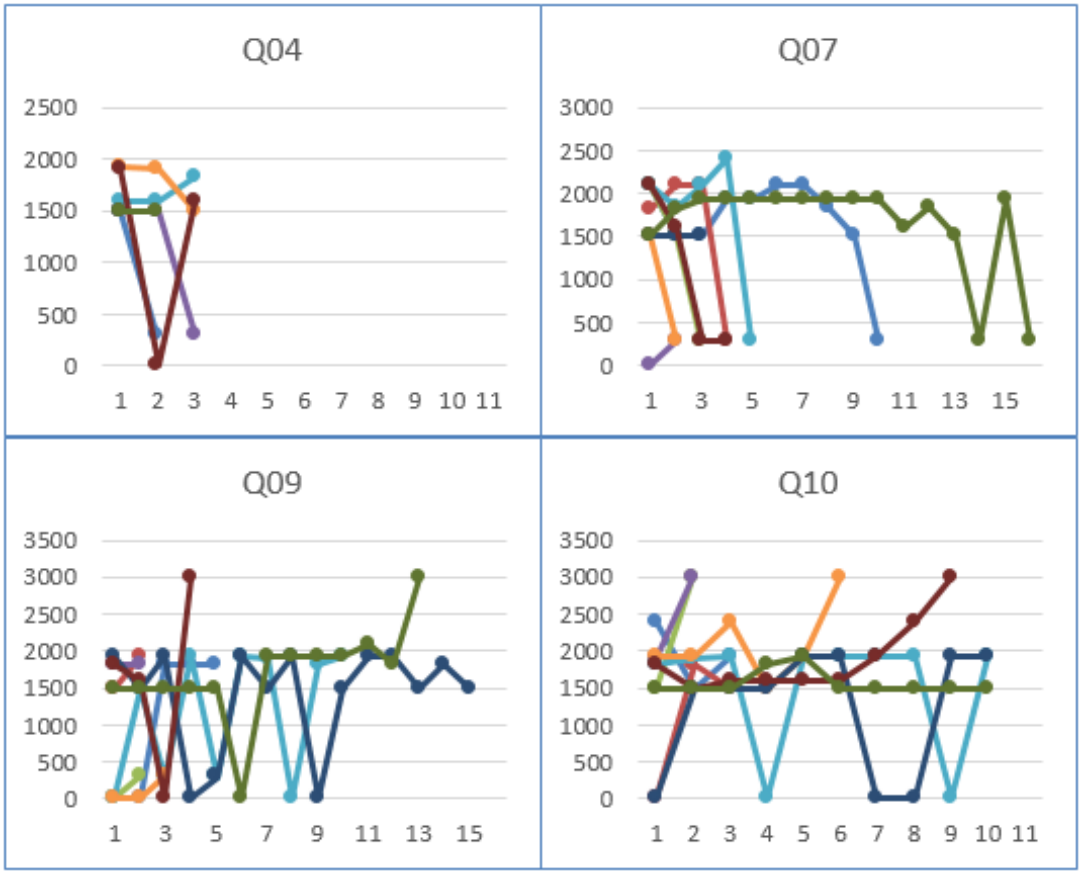


Figure 8. Sequences of Q04, Q07, Q09 and Q10

The lower values in the vertical axis (0 to 300) are those nodes belonging to Internal category while the upper values in the vertical axis (1500 to 2400) correspond to External category. The value 3000 represents Give Up node. Each colored line links the selected nodes of one participant out of nine in the experiment.

Comparing between tasks in Figure 8 and by looking at the first step in each question, it is notable that all questions, except Q09, were started mainly by consulting external nodes. In Q04, for example, none of the participants started by thinking or writing while in Q09 four participants started by thinking and writing. In addition, there is an oscillation between external and external nodes in Q09 which can't be observed in other questions. Taking this as a hint, a qualitative investigation followed the reason of why participants tended to think and write more in Q09. It turns out that most participants experienced a kind of confusion (Pekrun, Frenzel, Goetz, & Perry, 2007; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011; Pekrun, Goetz, Titz, & Perry, 2002). On one hand, they believed that the task is hard to be solved by thinking or writing. On the other hand, they had no idea how to search the internet for such a task. Therefore, they fluctuated between thinking and searching.

By exploring the average number of steps per question (see also Table 6), surprisingly, Q04 appeared to be the shortest. The question was designed to let the participants be autonomous and exercise their volition to choose the topic they want to search for. The participants were also given a time to think of the topic, in order to avoid the effect of the anxiety and the pressure of finding the topic immediately. These results suggested that self-motivation questions do not induce learners to select topics far beyond their current knowledge. Tracking back the participants' selected questions clarifies that all of them, except one, formulated their questions around "What is something?" In other words, they searched for facts. They mainly answered their questions by a simple search on the internet. In digital literacy studies (Coiro, Castek, & Guzniczak, 2011; Kammerer, Bråten, Gerjets, & Strømsø, 2012; Kiili, Laurinen, & Marttunen, 2008, 2009; Kiili, Laurinen, Marttunen, & Leu, 2012; Leu et al., 2013; Mason, Boldrin, & Ariasi, 2010), fact-retrieval question (or factual question which has only one answer) was identified as trivial question which should be avoided for a learner who has access to the internet. Of course, one can argue that the qualitative analysis alone is conducive to the same conclusion without painstakingly recoding the categories and then inspecting the sequences visually. And this could be right but consider the following case. A researcher conducts a study similar to the current study in which he has 10 participants solving 10 tasks. While running the experiment and while qualitatively coding the transcripts, he handles the participants in sequence, one after another. By doing so, he limits his chances to compare between tasks unless he re-read, or maybe re-analyzed, the data task by task. The visual inspection would give us the holistic picture we lose in those cases where we dive in the data from one direction. Nevertheless, the visual inspection is not proposed as a counterpart of the qualitative coding; they should be partners.

Table 6. Questions' first step and average number of steps

Q#	Average number of steps	Nodes in the first step	
		Internal	External
Q04	2.11	0	9
Q07	5.44	1	8
Q09	6.22	4	5
Q10	6.11	2	7

In contrast to Q04, Q09 and Q10 took very long for the participants to complete. Those questions seemed to be difficult for the participants, considering the average number of steps

as a sign of complexity. Another sign of complexity is the number of participants who gave up (3000 on vertical axis). Q10 seemed to be the trickiest question (with four participants giving up) followed by Q09 (with two participants giving up). The result is consistent with our theoretical classification of the tasks, since the questions were sorted by their expected difficulty where Q01 is the easiest and Q10 is the hardest question. However, one should consider these signs as indicators for further qualitative analysis since we usually seek to know how and why the participants experienced this difficulty. Mainly, the combination of three perceptions were identified as the reasons for giving up and experiencing difficulties: (1) the ease of finding information on the internet, and (2) the authority of knowledgeable people, together with (3) the participants' perceptions about themselves (low Self-Efficacy). The following excerpt of one of the participants would clarify the reasons:

> *When I first saw the story, I found out that most of the text is hidden. It may be – or certainly – impossible for me to recover the whole text. I said to myself "I will find it as it is on the internet."*

It is worth noting that the concentration in this paper is not on the results of the qualitative study. It is on how the visual inspection has succeeded to drag the attention to the parts of the study which need to be investigated thoroughly.

In Q07, it is notable that there is a consistent pattern between participants since most of them consulted the external nodes all the way up to the end of the sequences where they turned to internal nodes (namely writing). The purpose of question Q07 was to examine the pattern of behavior when students are asked to write a scientific essay. Tracking back participant's behaviors reveals that they tended to gather, copy, and paste information from different resources (especially the internet and their friends on Facebook) and then remixed the information near the end of the task. Therefore, it seems that writing tasks are equivalent to the process of remixing information on the internet from the participants' perspective. The result casts doubt on the effectiveness of these type of questions to encourage students to write on their own.

Conclusion

Gaining an accurate and deep understanding of qualitative data is not always immediate and intuitive. The process is daunting and researchers in the field may be interested in probing the qualitative data from different sides. This paper proposes a new instrument which may be well in hand to inspect sequential data visually with little computational needs. The method generates codes systematically out of a hierarchical tree of relations among the categories. The generated codes amplify the differences between nodes in higher levels and minimize the differences in the lower levels and this facilitates the process of tracking the transition of the participants between the most important nodes of the study. In addition, the codes allow sibling nodes to appear closer and cousin nodes to appear distinct. The output of the algorithm mimics the characteristics of the hierarchical tree to a certain degree. All in all, the visual inspection of the sequences is particularly beneficial to qualitative researchers, and in at least two ways. First, if the primary objective of a study is to track the evolution of the phenomenon over time, then the purpose of the proposed instrument completely matches that of the study. For instance, it may be our interest to track the development of the participants' emotions over the course of the study, or maybe their engagement with the learning activities. Second, the sequential data are often occurred as the result of the participants' activities, as it was the case of the example used in this study. The development of the events over time may not be our main goal but we just analyze the sequences of the nodes as a means to an end. Similarly, a researcher who aims

at studying the reasons of the high dropout rate in schools may utilize the visual inspection of the schools' sequences that the students have passed through as a means to investigate the reasons behind the high dropout rate.

Unlike other methods of analysis, the visual inspection of the illustrations serves as a starting point for wondering about the data and as a base for further investigation. Several examples have been presented of the interplay between the visual inspection and the qualitative inquiry and how that has led to better understanding of the story behind the observed patterns. Although the recoding algorithm has found to be useful for qualitative SA studies, it may be of no less value to the quantitative SA studies as well. In specific, one may find it useful to use the automatically calculated gaps in the hierarchical tree to feed in the replacement cost matrix in OMA.

The most important limitation of the proposed instrument lies in the fact that it does not address the time spent in each state. Ignoring the length of the experience in each state is identified as a key limitation for other SA methods such as OMA and cumulative event tables (Abbott, 1995; Arosio, 2004; Macindoe & Abbott, 2004). The suggested mitigation technique of this issue is to unify the time unit of each transition (Abbott & Hrycak, 1990; Macindoe & Abbott, 2004). Alternatively, and since the sequences are inspected visually, one may enlarge the circle of each state in the illustration according to the time spent in the state, where a larger circle refers to longer experience. Secondly, the suggested instrument shows a limited contribution to detect frequent transition from one state to another. Other methods such as cumulative event tables (Arosio, 2004) and code sequence analysis (Derobertmasure & Robertson, 2014) may complement the proposed method in detecting a frequent transition.

Notwithstanding these limitations, the proposed instrument offers valuable tool to explore sequential data in qualitative studies. It is also recommended that further research be undertaken to examine the usefulness of the proposed method in quantitative SA. There is still room for many improvements on the proposed method to overcome the pre-mentioned deficiencies.

References

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology*, *21*, 93-113.

Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *American Journal of Sociology*, *96*(1), 144-185.

Aldahdouh, A. A. (2017). Does artificial neural network support connectivism's assumptions? *International Journal of Instructional Technology and Distance Learning*, *14*(3), 3-26. http://doi.org/10.5281/zenodo.545695

Aldahdouh, A. A., & Osório, A. J. (2016). Planning to design MOOC? Think first! *The Online Journal of Distance Education and E-Learning*, *4*(2), 47-57. http://doi.org/10.5281/zenodo.48804

Aldahdouh, A. A., Osório, A. J., & Caires, S. (2015). Understanding knowledge network, learning and connectivism. *International Journal of Instructional Technology and Distance Learning*, *12*(10), 3-21. http://doi.org/10.5281/zenodo.46186

Arosio, L. (2004). Occupational careers and longitudinal data: Tools and perspectives of research. *Quality and Quantity*, *38*(4), 435-456. http://doi.org/10.1023/B:QUQU.0000043117.38242.a3

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge, UK: Cambridge University Press.

Bakeman, R., Quera, V., & Gnisci, A. (2009). Observer agreement for timed-event sequential

data: A comparison of time-based and event-based algorithms. *Behavior Research Methods*, *41*(1), 137-147. http://doi.org/doi:10.3758/BRM.41.1.137

Bernauer, J. A., Lichtman, M., Jacobs, C., & Robinson, S. (2013). Blending the old and the new: Qualitative data analysis as critical thinking and using NVivo with a generic approach. *The Qualitative Report*, *18*(31), 1-10. Retrieved from http://nsuworks.nova.edu/ tqr/vol18/iss31/3

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. http://doi.org/10.1191/1478088706qp063oa

Coiro, J., Castek, J., & Guzniczak, L. (2011). Uncovering online reading comprehension processes: Two adolescents reading independently and collaboratively on the internet. *60th Yearbook of the Literacy Research Association*, 354–369.

Derobertmasure, A., & Robertson, J. E. (2014). Data analysis in the context of teacher training: Code sequence analysis using QDA Miner. *Quality and Quantity*, *48*(4), 2255-2276. http://doi.org/10.1007/s11135-013-9890-9

Downes, S. (2006). Learning networks and connective knowledge. In H. H. Yang & S. C.-Y. Yuen (Eds.), *Collective intelligence and elearning* (pp. 1-26). New York, NY: Information Science Reference.

Downes, S. (2008). An introduction to connective knowledge. In T. Hug (Ed.), *Media, knowledge & education: Exploring new spaces, relations and dynamics in digital media ecologies* (pp. 77-102). Innsbruck, Austria: Innsbruck University Press.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107-115. http://doi.org/10.1111/j.1365-2648.2007.04569.x

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *78*(3), 215-250.

Ford, K., Oberski, I., & Higgins, S. (2000). Computer-aided qualitative analysis of interview data: some recommendations for collaborative working. *The Qualitative Report*, *4*(3), 1-14. Retrieved from https://nsuworks.nova.edu/tqr/vol4/iss3/1

Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist*, *39*(1), 43-55. http://doi.org/10.1207/s15326985ep3901_5

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26.

Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2012). The role of internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. *Computers in Human Behavior*, *29*(3), 1193-1203. http://doi.org/10.1016/j.chb.2012.10.012

Kiili, C., Laurinen, L., & Marttunen, M. (2008). Students evaluating internet sources: From versatile evaluators to uncritical readers. *Journal of Educational Computing Research*, *39*(1), 75-95. http://doi.org/10.2190/EC.39.1.e

Kiili, C., Laurinen, L., & Marttunen, M. (2009). Skillful internet reader is metacognitively competent. *Hanbook of Research on New Media Literacy at the K-12 Level: Issues and Challenges*, *II*(2005), 654-668. Retrieved from http://www.igi-global.com/chapter/handbook-research-new-media-literacy/35943

Kiili, C., Laurinen, L., Marttunen, M., & Leu, D. J. (2012). Working on understanding during collaborative online reading. *Journal of Literacy Research*, *44*(4), 448-483. http://doi.org/10.1177/1086296X12457166

King, T. (2013). A framework for analysing social sequences. *Quality and Quantity*, *47*(1), 167-191. http://doi.org/10.1007/s11135-011-9510-5

Kuusela, H., & Paul, P. (2000). A comparison of concurrent verbal protocol analysis retrospective. *American Journal of Psychology*, *113*(3), 387-404.

http://doi.org/10.2307/1423365

Lalanda Nico, M. (2016). Bringing life "back into life course research": Using the life grid as a research instrument for qualitative data collection and analysis. *Quality & Quantity*, *50*(5), 2107-2120. http://doi.org/10.1007/s11135-015-0253-6

Leu, D. J., Forzani, E., Burlingame, C., Kulikowich, J., Sedransk, N., Coiro, J., & Kennedy, C. (2013). The new literacies of online research and comprehension: Assessing and preparing students for the 21st century with Common Core State Standards. In S. B. Neuman & L. B. Gambrell (Eds.), *Quality reading instruction in the age of common core standards* (pp. 219-236). Newark, DE: International Reading Association.

Macindoe, H., & Abbott, A. (2004). Sequence analysis and optimal matching techniques for social science data. In M. Hardy & A. Bryman (Eds.), *Handbook of data analysis* (pp. 387-406). London, UK: Sage. http://doi.org/10.4135/9781848608184

Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the web to learn about a controversial topic: Are students epistemically active? *Instructional Science*, *38*(6), 607-633. http://doi.org/10.1007/s11251-008-9089-y

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *1*(2), 1-7.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions. An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Ed.), *Emotion in education* (pp. 13-36). Amsterdam, The Netherlands: Academic Press. http://doi.org/10.1016/B978-012372545-5/50003-4

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36-48. http://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91-105.

Peute, L. W. P., de Keizer, N. F., & Jaspers, M. W. M. (2015). The value of retrospective and concurrent think aloud in formative usability testing of a physician data query tool. *Journal of Biomedical Informatics*, *55*, 1-10. http://doi.org/10.1016/j.jbi.2015.02.006

Schulz, J. (2012, August 14). *Analysing your interviews*. Retrieved from https://www.youtube.com/watch?v=59GsjhPolPs

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, *2*(1), 3-10. Retrieved from http://www.itdl.org/Journal/Jan_05/article01.htm

Siemens, G. (2006). *Knowing knowledge*. Retrieved from http://www.elearnspace.org/KnowingKnowledge_LowRes.pdf

Siemens, G., & Tittenberger, P. (2009). *Handbook of emerging technologies for learning*. Manitoba, Canada: University of Manitoba. Retrieved from http://elearnspace.org/Articles/HETL.pdf

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley Publishing Company.

Van den Haak, M., De Jong, M., & Schellens, P. J. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, *22*(5), 339-351. http://doi.org/10.1080/0044929031000

Van den Haak, M. J., De Jong, M. D. T., & Schellens, P. J. (2004). Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: A methodological comparison. *Interacting with Computers*, *16*(6), 1153-1170. http://doi.org/10.1016/j.intcom.2004.07.007

Appendix

I developed a software to facilitate calculating the code numbers of the sequential data using Oracle Database 11g Express Edition Release 2. The code does not have a user-friendly interface and may require extra effort to understand and run. Interested researchers may contact the author on his email to cooperate in using the code or to provide a newer version of the software, if it is available. The code can be found on the following link: https://adahdouh.wordpress.com/2017/09/13/visual-inspection-of-sequential-data-recoding-algorithm/.

Author Note

Alaa Aldahdouh is a PhD candidate at University of Minho, Portugal. He has a computer engineering degree with more than 10 years of experience in software development. His major area of interest resides in investigating online learning, digital literacy, online experience, learning theories, connectivism, and MOOC. Correspondence regarding this article can be addressed directly to: AlaaAlDahdouh@gmail.com.

I am grateful to Professor António J. Osório and Professor Susana Caires for their valuable advice and encouragement throughout the qualitative study in which this instrument has been developed.

Article Citation

Aldahdouh, A. A. (2018). Visual inspection of sequential data: A research instrument for qualitative data analysis. *The Qualitative Report, 23*(7), 1631-1649. Retrieved from https://nsuworks.nova.edu/tqr/vol23/iss7/10