System 43 (2014) 74–81

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

To code or not to code: Dilemmas in analysing think-aloud protocols in learning strategies research☆

Yongqi Gu*

Victoria University of Wellington, New Zealand

ARTICLE INFO

Article history:
Received 27 March 2013
Received in revised form 11 December 2013
Accepted 16 December 2013

Keywords:
Coding
Analysis
Think-aloud protocols
Learning strategies

ABSTRACT

Analysis of qualitative data depends on the overall design and the nature of the research questions. Specifically, the approach to coding and analysis should be coherent with the way data is elicited and the epistemological perspective that situates the overall design. This article illustrates these points by using an example from the analysis of think-aloud protocols which revealed patterns of strategy choice and use in completing a listening task among a group of primary school pupils in Singapore. Throughout the discussion, I focus on the major dilemmas that arose during the process of coding and analysis. Two dimensions of these dilemmas were identified, those that pertain to the nature of qualitative data analysis and those that are specifically related to language learning strategies.

1. Introduction

Think-aloud protocols are often used in revealing the strategic decisions learners use in learning and performing tasks in a second language. Two common types of think-aloud procedures have been used, concurrent and retrospective think-aloud (Kuusela & Paul, 2000). The former is elicited while a learning task is being performed. The participant typically either voices aloud thoughts, feelings, and reasoning as the primary learning activity is going on, or stops the primary task every now and then, usually at the prompt of a visual, acoustic or semantic reminder, so that s/he can tell the researcher what has been going on in his/her mind. Retrospective think-aloud happens at the end of a learning task, and is meant to collect the participant's thinking and reasoning processes while they are still in the short-term memory of the learner.

A number of authors have debated the pros and cons in using the think-aloud technique for the discovery of learner's decision-making processes (Cohen, 1984; Ericsson & Simon, 1993; Smagorinsky, 1994). Concerns have mainly centred on the disruptive nature of thinking aloud to the primary process of thinking or learning. Despite this crucial problem, however, it is now widely agreed that various versions of thinking aloud are the most direct and therefore best tools available in examining the on-going processes and intentions as and when learning happens. In addition, recent studies using eye tracking techniques to validate the think-aloud method have also shown very encouraging evidence supporting the usefulness of the method (e.g., Guan, Lee, Cuddihy, & Ramey, 2006). Moreover, it has been found that think-aloud protocols can be elicited not only from adult learners, but also from primary school pupils or younger (Chamot & El-Dinary, 1999).

☆ This paper reports on some of the reflections and thinking during the process of coding and analysis of think-aloud data from a large-scale project supported by the Education Research Fund (EdRF), Ministry of Education, Singapore (EP1/02GYQ) and by the Centre for Research in Pedagogy and Practice (CRPP) Fund, National Institute of Education, Singapore (CRP3/03GYQ). Guangwei Hu and Lawrence Jun Zhang as well as the author constituted the main research team. In this paper, "we" refers to this team of three researchers.

* Tel.: +64 4 4635606.
E-mail address: peter.gu@vuw.ac.nz.

0346-251X/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.system.2013.12.011

Compared to the amount of thinking and debate on thinking aloud as an elicitation tool (e.g., Cohen, 2011; Gu, Hu, & Zhang, 2005), rarely can we find in standard research methodology books as to how think-aloud data can be coded and analysed. Most published research using think-aloud protocols have not presented the nitty-gritty details of coding and analysis, not to mention problems therein.

1.1. An illustrative study

This paper attempts to reveal some of the problems that emerged during the coding, analysis, and interpretation of data in a study of listening strategies employed by primary school pupils in Singapore. In fact, most of these problems in coding and analysis emerged as dilemmas in decision-making. One major dilemma centred on the qualitative nature of the think-aloud data and the resulting tensions in coding and analysis. Another main dilemma lay in the nature of learning strategies and the tensions between strategies as we knew them, strategies that appeared during coding, and how best the tension could be resolved. Revealing these dilemmas and the thinking processes behind the decision-making will provide not only some concrete guides to new researchers, but also insights into the coding and analysis process of qualitative data in general and language learning strategies in particular. In this sense, this article will not present findings systematically, nor will it take a step-by-step description approach. Instead, I will focus directly on the dilemmas and the decisions my colleagues and I in the research team made in order to solve each problem. Some findings will be presented only for illustrative purposes.

2. Coding and analysis of think-aloud data: qualitative or quantitative?

This section discusses bottom-up and top-down approaches to coding think-aloud data, quantitative analysis of qualitative data, and data reduction vs. data richness.

2.1. Bottom-up or top-down: where do codes come from?

One of the first things to do after all the think-aloud transcriptions were done was to explore the transcripts and reduce them to manageable patterns. This process is referred to as coding, by tagging a categorical label (a "node" in NVivo) to a chunk of data. We started from a preliminary coding scheme that was based on three studies on listening strategies which we believed were representative of the field in both research methodology and findings: O'Malley, Chamot, and Kupper (1989), Goh (2002), and Vandergrift (2003). This procedure in itself was a dilemma we had to face right from the outset between our purpose of trawling up as many as possible strategies Singaporean pupils used in learning and using English and our desire to see "meaning" which could fit into our own theoretical background. Notably, approaches to coding have always remained controversial.

Table 1 lists some terms that have described opposing approaches to the coding of qualitative data. On the one hand, we have the idealistic approach of grounded theory (Glaser, 2002; Glaser & Strauss, 1967) which insists on an inductive approach to the "emergence" and discovery of meaning from specific instances, contexts, and individuals. In fact, it was indeed our intention to discover as many as possible naturally occurring strategies our primary school participants in Singapore were using, which might be different from the adult strategies discovered elsewhere. On the other hand, however, there is a reality of the theoretical background and existing research already in our minds which are impossible and unnecessary to ignore. Moreover, our research questions and sampling (comparing primary school pupils representative of top and bottom proficiency groups from three grade levels) were hypothesis-driven in the first place. There was no point pretending that we had no idea about listening strategies before analysis took place. While many researchers insist that coding should start from scratch, others are equally adamant that "a coding scheme should be a theoretically grounded model of the cognitive processes and types of information involved in the activity under study, not a mere list of strategies" (Kasper, 1998, p. 359). With these arguments in mind, we decided to follow the advice of Miles and Huberman (1994) and started from a theoretically and empirically derived list while maintaining flexibility in adding, deleting, and reshuffling the codes based on our data. As a result, our final coding scheme was a result of both top-down and bottom-up approaches to coding. This was, however, a practical solution only, because the tension between what the data told us (bottom-up) and what we wanted to see in the data (top-down) was always present throughout the coding and analysis process. The coding scheme derived from both top-down and bottom-up processes helped us achieve a pragmatic balance only.

Table 1
Approaches to the analysis of verbal protocols.

Bottom-up	Top-down
Inductive	Deductive
Descriptive	Conceptual
Grounded approach	Start-list approach
Emergence of theory from data	Code-and-retrieve
Emic	Etic
Participant-centred	Researcher-centred