On the reliability of identifying design moves in protocol analysis

Gabriela Trindade Perry, Departamento de Design e Expressão Gráfica, Universidade Federal do Rio Grande do Sul, Av. Sarmento Leite, 320, sala 502, CEP 90040-060 Porto Alegre, Rio Grande do Sul, Brazil
Klaus Krippendorff, The Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA 6220, USA

This paper discusses issues and ways of measuring the reliability of segmenting verbal protocols of design activity, a central focus of design research. Reliability is an important issue in distinguishing 'design moves'. In the present study, seven students working for a master in design degree, one graduated designer and two professors segment a 30 min protocol of a product design process into design moves. The intra and inter reliability was calculated for these observers using alpha coefficients. Neither the students', designer's nor professors' segmentation reached the desired cut-off value of 0.8. This negative finding questions the clarity of existing conceptions and urges more concise definitions, better training of analysts, and formulating more decisive instructions.

Keywords: protocol analysis, design research, research methods

According to a bibliometric study (Chai & Xiao, 2011), protocol analyses, using the think-aloud method, detailed by Ericsson and Simon (1993), is one of the most popular design research methods. It involves distinguishing segments within a transcript of verbal accounts of design processes for further analysis. Regarding segmentation, Ericsson and Simon (1993, p. 205) state that 'the appropriate cues are pauses, intonation as well as syntactical markers'. Assuming that these criteria are objectively identifiable, Ericsson and Simon do not foresee reliability issues regarding this step of the analysis (p. 266). They do not mention non-syntactical criteria, which are important in design research and for which their assumption does not apply. Protocol analysis, as introduced in their famous book, aims at analyzing problems whose solving could be modelled – at least to some extent – by tools such as the problem behaviour graph (PBG), which would map the stage the problem solver is at, relative to a problem model. Examples of problems studied by Ericsson and Simon (1993) are: the tower of Hanoi, crypt arithmetic and theorem proving. Regarding design problems, Craig (2001) and Chi (1997) argue that protocol analysis using the think-aloud method might not be the most adequate method to analyze design processes.

Corresponding author:
Gabriela Trindade Perry
gabriela.perry@ufrgs.br

www.elsevier.com/locate/destud
0142-694X $ - see front matter *Design Studies* **34** (2013) 612–635
http://dx.doi.org/10.1016/j.destud.2013.02.001

Chi (1997) suggests modifications on Ericsson and Simon's method, for analyzing how representations change with learning and argues that, regarding ill-structured problems, it is not possible to know in which states of the problem space a problem solver could be. For this reason, she contrast her verbal analysis method with Ericsson and Simon's (1983) protocol analysis, proposing that verbal analysis should aim 'to capture the representation of knowledge that a learner has and how that representation changes with acquisition' (p. 3) — the emphasis on analyzing representations is strong on her paper. Chi states that her proposal differs from protocol analysis regarding (p. 4): 'the instruction, the goal or focus, the analysis, the validation, and the conclusion'. Chi (1997, p. 24) developed detailed instructions on how to segment verbal data, and emphasised the importance of measuring reliability of coded data in every step of the analytical process, i.e. 'during segmentation into units, categorizing or coding of the units, depicting the coded data, seeking pattern(s) in the depicted data, interpreting the pattern(s), and so forth'.

Several design studies used variants of the think-aloud technique and subsequently analyzed the protocol data. Gero and McNeill (1998), for example, developed a broad and wide coding scheme which was applied to 3 protocols of designing electronic devices, using the Delphi method. In their study, coders pause for ten days between each coding, which give the researchers the data needed to examine discrepancies between the two consecutive codings with the aim of locating disagreements and improving the coding scheme. The debate of emerging coding difficulties and revisions of problematic instructions constitute the 2nd and 3rd phases of the Delphi method were 'the group reaches an understanding of the issue' and 'disagreement is explored to bring out underlying reasons for differences and to evaluate them' (p. 34). Employing similar coding instructions as Gero and McNeill (1998), McNeill, Gero, and Warren (1998) investigated two hypotheses about the conceptualization and design of electronic devices. They also combined the Delphi method and a coding protocol, encouraging arbitration and the formation of consensus between coders. The authors argue that 'minimal disagreement between coders is desired but this can be difficult to achieve given the qualitative nature of the coding process' (p. 5). Recognizing the difficulty of achieving high reliability when analyzing textual matter, Krippendorff (2004, p. 3) argues that 'the mathematical complexity of analyzing variably unitized text, while an unquestionable hurdle for replicating research, is no justification for creating the methodological schism between quantitative and qualitative approaches to analyzing textual matter'. Whether the segmentation criteria are syntactical or conceptual, replicability of the coding process by independent coders is essential. Ball and Christensen (2009) also used protocol analysis of verbal data, but in a different way: they segmented the data using syntactical rules, trained one independent coder, and measured the inter-observer reliability [be-