INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN
ICED 99 MUNICH, AUGUST 24-26, 1999

TECHNIQUES FOR RESEARCH METHOD VALIDATION IN PROTOCOL STUDIES

Srinivas Nidamarthi, Amaresh Chakrabarti and Thomas P. Bligh

Keywords: Research Methods, Research Method Validation

1 Introduction

Protocol studies in engineering design employ audio and video recording of designers' actions and spoken statements, and researchers use this recorded data to observe, analyse and describe the design process. Several researchers applied these studies to individual designers [1, 2, 3], as described in [7], as well as to design teams [5], as described in [8]. In both the cases, researchers first segmented the protocol data into punctuated, meaningful sentences, which we call *events*, and then *coded* the events into several *categories*. The *coding* scheme is essentially labelling an event with one of the *categories* the researcher intended to observe. Because the coding method is essentially a researcher's interpretation of the design process in terms of the categories s/he chooses, the method and its end results are subjective to the categories [5] and the interpretation. Therefore, we argue that research using these methods should try to assess and minimise the degree of subjectivity. In this paper, we discuss our efforts and observations to validate the method by identifying and evaluating its subjectivity.

2 Research Context

This paper is part of the first author's doctoral study on understanding the design process in terms of "how designers satisfy requirements". We used protocol studies on individual designers as well as teams. In these studies, designers were given a design problem, which included a set of requirements to be satisfied, and were asked to produce a detailed concept. This concept was evaluated by three independent designers with respect to the given requirements in the design problem. In each case study, we have used these evaluation reports and the captured design process to understand the designers' process of satisfying the requirements. Initial research results were published in [6]. As discussed in the previous section, our research method, too, has two forms of subjectivity. These are 1) the subjectivity of the categories to the research aim, and 2) the subjectivity of the coding method because of the researcher's interpretation. The following sections discuss methods to assess the above forms of subjectivity, and efforts to minimise the same.

3 Validation of the categories

Minneman [5] discusses the deleterious effect of using protocol studies with research aims of proving pre-selected research hypotheses (often, in terms of pre-determined categories). However, in order to initiate the research, a researcher needs to have, at least, a broad and abstract research aim. For example, Minneman's research aim was "to understand social interaction in design teams" [5]. Therefore, in order to minimise the subjectivity of our research aim, we avoided any pre-selected hypotheses, and let our categories evolve freely during our protocol data analysis. After the categories have been evolved, we validated them, as discussed in the following sections, in terms of

- bench-marking them with the categories of other researchers who had related, but different, research aims, and
- their *completeness* in describing the design process.

3.1 Evolution of categories

In order to understand "how designers satisfy requirements", we initially analysed one case study in terms of the requirements given in the design problem. This initial analysis helped us to assess the protocol data in terms of its explicit and implicit details of the design process (Stauffer et. al. [7] discuss this issue further, using the terms "Declarative Knowledge" and "Procedural Knowledge"). During this initial study, we realised that the designers' *activities* seem to constitute a major part of their attempts to satisfy requirements. Moreover, the protocol data captured the activities more explicitly than any other form of detail of the design process. Therefore, with a renewed research focus—to describe requirements satisfaction in terms of the design activities, we searched existing literature for an *initial set* of activities. The literature search helped us in two ways. First, it helped us to understand the existing view of the design process. Second, it provided us with generic definitions, free of bias towards our research aim, for the activities. Blessing [1] and Hubka [4] are some of the works that provided us with initial descriptions of design activities.

During the protocol analysis we refined the activities set, upgraded it by adding new activities, and detailed it further. Our categories, the design activities, thus evolved as shown in Figure 1. During this evolution, deciding to add a new category is a critical step. Whenever a new category was felt to be necessary, we reviewed the existing set of categories to see if a refined definition of a category describes the event. If it was not an option, the event is reviewed for possible segmentation, and coding each segment with the existing activities. After failing in this option too, a tentative definition of the new category is formed and evaluated against the existing categories and the *research scope*.

The first evaluation is essentially determining relationships between the new activity and the existing activities. We carried out this evaluation in terms of a structured activities set, and structured the categories into two design stages—*Problem Understanding* and *Problem Solving*, and in each stage, two levels of *abstraction—primary* and *secondary*. Table 1 shows the activities and their hierarchy. Discussion of these activities is beyond the scope of this paper.

For example, "question a solution" (in "Evaluate" in "Problem Solving", Table 1) was an added category during the evolution. In the team design process, we observed designers questioning each other as a way to assess performance of solutions. We realised that this activity belongs to "evaluation of solutions", and so far such events were not described by the current set of categories.

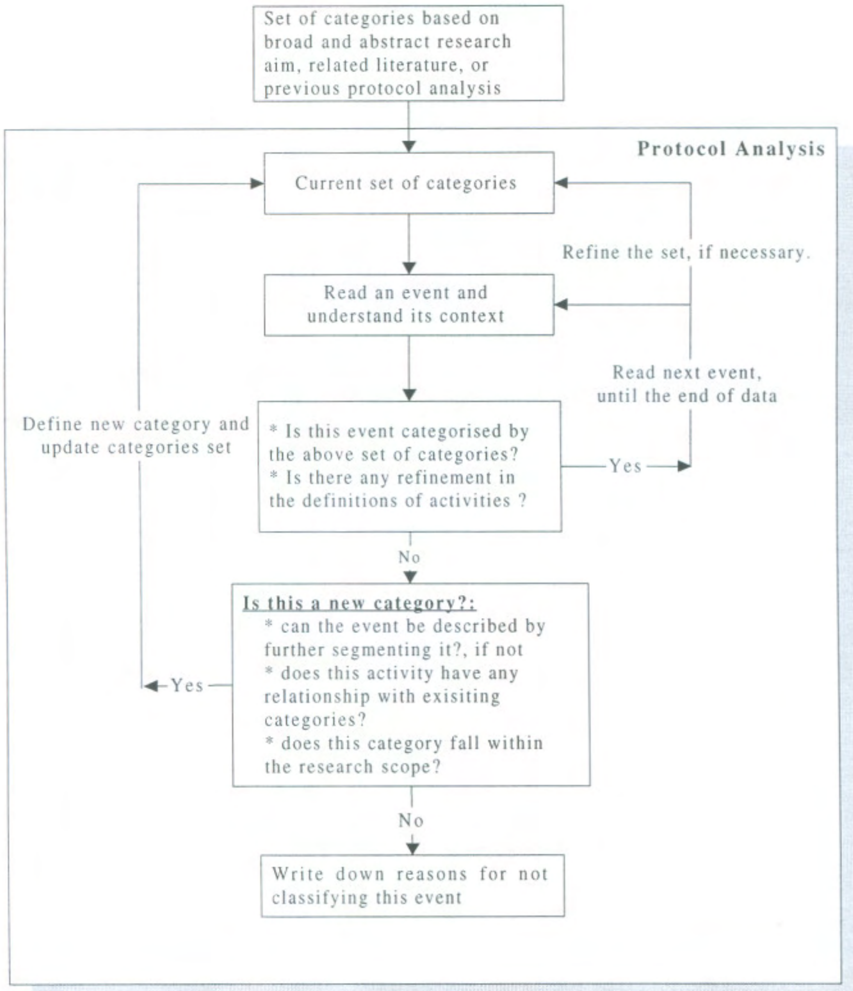

Figure 1: Evolution of categories during protocol analysis.

The evaluation of a new category with respect to the research scope, is to check if the category serves our research aim. For example, the activity "question a solution" is considered because it reflected the designers' efforts to "assess solutions' performance" which seem to serve our aim. Moreover, the research scope helped us to set limits for the expansion of categories as per our resources. Section 3.3 discusses the effect of limiting the research scope.

Table 1: The designers' activities—the "problem understanding" activities are the activities on requirements, and the "problem solving" activities are the ones on solutions.

Stage	Problem Understanding			Problem Solving		
Primary Level Activities	Identify	Analyse	Choose	Generate	Evaluate	Select
Secondary Level Activities	Perceive, Infer, Modify	Question, Relate, Weigh, Verify, Visualise	Decide	Create, Modify, Detail	Identify characteristic, Question, Relate, Verify	Identify things to do, Compare, Decide

In each case study the protocol data was analysed, at least, two times. We ended the analysis of a case study only when there were no significant changes to already coded events in the latest iteration. The very first case study was analysed more than four times as changes in the category set in later analyses triggered re-analysis of earlier protocols. Such repeated analysis helped us arrive at unambiguous definitions for the categories. At the end of analyses of three case studies, the categories seemed to have completely evolved, and there were no changes to the category set during the analysis of the fourth case study.

3.2 Bench-marking of categories

The bench-marking of our categories with the work of others was carried out to

- evaluate our categories with respect to the categories of others whose evolved with different research aims, and
- evaluate, and thereby defend or critique, our end research results and that of the others (this discussion is beyond the scope of this paper).

Table 2 summarises the similarities in our research categories and those of Blessing [1], Dwarakanath [2] and Gero & McNeil [3].

Table 2: Summary of comparison of categories in our research method and with those of [1, 2 and 3].

Authors	Research Aim (briefly)	Similar Definitions
Blessing [1]	To support designers activities using computers (emphasis is prescriptive).	Her "Generate" and "Evaluate" categories, and our primary activities in "Problem Solving".
Dwarakanath [2]	To support designers "decision making" activity (emphasis is descriptive).	His "Criteria" and "Decisions", and our "Requirements" and "Solutions".
Gero and McNeil [3]	To describe the design process using "design strategies" (sequence of micro actions, emphasis is descriptive).	Their "Proposing Solution" and "Analysing solution ", and our "Generate" and "Evaluate".

While Blessing's categories are comparable with our categories at the primary abstraction level, Gero and McNeil's "micro" activities are at the secondary abstraction level. Most categories differed with ours in their details and hierarchy of definitions. However, the overall definitions of their categories do not conflict with our findings.

3.3 Completeness evaluation for categories

Completeness of categories conveys the amount of protocol data that was described by the categories. Because our research scope is limited to design activities, the categories set was evaluated to find out how much of the design process is reflected in the coded protocol data. The protocol data consisted of three parts—1) data coded into categories, 2) data discarded (unrelated statements to the design process, such as "it is tea time"), and 3) data unclassified but related to the research aim. We analysed completeness of the coded data in terms of the unclassified data.

We marked each unclassified event with a reason, and found that three reasons were very frequent—the designer's expressing "strategy" and "confidence", and in teams, "co-ordination". The rest of the data included designers statements that seem to influence their design process but cannot be related the design context of that particular moment. Examples for such unknown reasons include statements like "Ah! that is cunning…" (no valid suggestion to why that particular solution is "cunning"). Table 3 presents percent of time spent by the designers on various reasons within the unclassified data.

Table 3: Percent of the designers' time on events that were not classified by our categories, but that might relate to the research aim.

	Individual-A	Individual-B	Team-A	Team-B
Strategy	0.8 %	0.4 %	2.1 %	2.7 %
Confidence	0.1 %	0 %	0.1 %	0.4 %
Co-ordination	--	--	0.9 %	0.7 %
Unknown	3.2 %	1.4 %	2.4 %	5.0 %
Total Unclassified	**4.1 %**	**1.8 %**	**5.5 %**	**8.8 %**

Table 3 suggests the possibility of about 10 % of the protocol data that we fail to consider in addressing our research aim. Instead of expanding the category set to minimise the percent of unclassified data, we tried to assess its influence on our end results. For example, if designers follow their strategy (plan of activities), the activity set may capture it, and therefore should not have a significant affect on the end result. Discounting such insignificant influences, we seem to have the category set that was unable to account for about 5 % of the protocol data. Completeness evaluation also provided us a decisive factor as to whether or not re-analysis of the data was required. For example, if the amount of unclassified data is more than 10 % it is probably worth re-analysing the data one more time to see if more data can be coded.

4 Validation of the coding method

So far the paper has discussed the interpretation of the protocol data by one researcher. A major criticism for such interpretation is *reproducibility* of the coded data irrespective of human subjectivity [7]. In order to assess such subjectivity, we used a different coder to analyse the same research data, using the same categories set. During this study we also wanted to observe if the other coder could segment an event, similar to the researcher's process (Figure 1), so as to describe finer details of the design process. Therefore, the event lengths in the cases of the researcher and the coder differ. Two different end results were obtained after using two coders analysing the same data. We compared these two sets of coded categories to find 1) differences in coded categories, 2) why these differences occurred, and 3) how these differences can affect our research results.

The total number of differences found in the two coders' categories was 21%. By analysing the differences in the coding, we found that one third of the differences are due to mistakes by the other coder which seem to be primarily due to his misunderstanding of the categories, and not segmenting the events. The misunderstanding was evident in his coding of "Problem Understanding" activities (activities on requirements) as "Problem Solving activities" (activities on solutions). We verified this misunderstanding by checking these gross differences with the coder. Interestingly, even after the coder was informed of how to segment an event for its finer description, he coded the event as it was. This comparison also helped us to refine the definitions of activities, as we found one tenth of the differences are due to the researcher's interpretation.

Based on the structure of design activities (Table 1), we found that about one seventh of the differences do not affect our end results in terms of our primary level activities, because the differences occurred in the secondary level activities but not in the primary level. Therefore, by discounting mistakes by the other coder, and those differences that will not affect our theory, we get about 14% net difference that can influence our research results. We found that these differences are mainly due to the interpretations, both of which seem to be valid. For example, we found designers "evaluating" solutions at the same time as "generating" them. We could not further segment such overlapping events in the protocol data, nor refine the definitions of our categories. This analysis conveys a complex mix of design activities that need to be considered in describing the design process or supporting the designers.

5 Conclusions

This paper presented a method to evolve categories free of specific research hypotheses in protocol studies. We found that the overall definitions of the evolved categories similar to those of others. The measure of completeness of the categories describing the protocol data, revealed the unclassified data of about 5 % that can affect our end results. The validation in terms of the reproducibility of the coding method provided us with an estimate of human interpretation that can influence our research results.

Another form of reproducibility of the coding method, repeating similar classification for similar events in two different case studies, was not analysed in this research. Such analysis requires two coders analysing more than one case study data, and comparative evaluation of similarly classified events by the same coder in each case study. However, efficiency of such rigourous analysis, in terms of human effort and time, is considered to be very low [7].

References

[1] Blessing, L. T. M. "A Process-Based Approach to Computer-Supported Engineering Design", PhD Thesis, University of Twente, the Netherlands, 1994.

[2] Dwarakanth, S. "Decision Support in Design", PhD Thesis, University of Cambridge, UK, 1996.

[3] Gero, J. S., and McNeill, T. "An approach to the analysis of design protocols", Design Studies, Vol.: 19, No.: 1, 1998, pp.: 21–61.

[4] Hubka, V. "Principles of Engineering Design", Butterworths, 1984.

[5] Minneman, S. L. "The Social Construction of a Technical Reality: Empirical Studies of Group Engineering Design Practice", PhD Thesis, Stanford University, USA, 1991.

[6] Nidamarthi, S., Chakrabarti, A. and Bligh, T. P. "The significance of co-evolving requirements and solutions in the design process", Proceedings of ICED, Tampere, Finland, Ed. A. Riitahuhta, Vol.: 1, 1997, pp.: 227–230.

[7] Stauffer, L. A., Diteman, M. and Hyde, R. "Eliciting and analysing subjective data about engineering design", Journal of Engineering Design, Vol.: 2, No.: 4, 1991, pp.: 351–366.

[8] Tang, J. C. and Leifer, L. J. "An observational methodology for studying group design activity", Research in Engineering Design, 1991, Vol. 2, pp.: 209–219.

Srinivas Nidamarthi
Engineering Design Centre
University of Cambridge
Cambridge CB2 1PZ
UK
Phone: +44-1223-332790 Fax: +44-1223-332662
E-mail: sn210@eng.cam.ac.uk