

ON THE ISSUE OF COLLECTING SUBJECTIVE DATA: EXPERIENCES FROM A SMALL-SCALE FIELD OPERATIONAL TEST

Karlsson, I.C.M. and Englebretsson, P.
SAFER - Vehicle and Traffic Safety Center and Division Design & Human Factors,
Department of Product and Production Development
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
mak@chalmers.se, pontus.engelbretsson@chalmers.se

ABSTRACT: The paper presents a subset of the experiences made regarding the collection of subjective data made in a small scale field operational test (FOT), in particular the design, distribution, and overall management of questionnaires. One conclusion drawn is that the standardised instruments for collecting data on driving behaviour need further development in order to address drivers of (private) cars and professional drivers. A second conclusion is that many of the instruments used to collect information on personality traits etc. have aged and require updating in order to become valid again. Overall, standard questionnaires/instruments need to be “contextualised” in terms of time, national culture, and community. Furthermore, a web-based survey solution will not necessarily solve general problems regarding loss of data or the need for careful monitoring and reminders. Finally, the importance of carrying out complementary and in-depth interviews is stressed if the objective of the study is not only to identify “that”, but also “why”.

1 INTRODUCTION

1.1 *The SeMiFOT project*

SeMiFOT (Sweden-Michigan Naturalistic Field Operational Test) is an example of a smallscale FOT project. Altogether 39 participants (22 drivers of private cars and 17 professional lorry drivers) have over approximately six months tested different combinations of driver support systems including Adaptive Cruise Control (ACC), Blind Spot Information (BLIS), Electronic Stability Control (ESC), Forward Collision Warning (FCW), and Lane Departure Warning (LDW). Over these six months, data has been collected including objective data (e.g. speed, acceleration, road characteristics, steering wheel angle, and brake pressure, etc.), and subjective data (e.g. on the participants’ demographics, personality traits, and on user acceptance).

1.2 *Objective*

Most often the objectives associated with a field operational test (FOT) are to investigate the effects of different technical systems, for instance advanced driver assistance systems (ADAS) on driver behaviour and, as a consequence, on traffic safety. However, the overall purpose of the SeMiFOT project has been to develop further knowledge on the methodology related to FOTs. The overall purpose of this paper is to present a subset of the experiences made in the project and to discuss some methodological implications with focus on the

collection of subjective data, in particular the design, distribution, and overall management of questionnaires.

2 PROJECT METHODOLOGY

2.1 Collection of subjective data

Subjective data plays an important role in a test such as SeMiFOT. Not only is the subjective data most often necessary to explain the outcome of the objective data collected, but it is also vital for testing certain hypotheses, for instance regarding individuals' acceptance (or rejection) of a particular technical function and/or device.

In the SeMiFOT project, subjective data on the participating drivers and their acceptance of the functions evaluated were collected by the means of a set of different questionnaires and by means of a personal interview (with a sub-set of the participants). The questionnaires included: (1) Background Questionnaire (car and truck version); (2) Driver Behaviour Questionnaire, DBQ (Reason et al. 1990), car and lorry version; (3) Decision Making Questionnaire, DMQ (French et al. 1993); (4) Traffic Locus of Control Questionnaire, T-LoC (Özkan et al. 2005); (5) User Uptake or Acceptance Questionnaire, including the van der Laan Acceptance Scale (van der Laan et al. 1997), one per function tested, repeated at three occasions; and (6) Evaluation Questionnaire, one per function tested. Initially was also the Sensation Seeking Scale (SSS) (e.g. Zuckerman, 1979; 1994) considered but its inclusion was cancelled before the trials began.

The relatively large number of questionnaires to be administered was in itself a methodological issue. The decision was however to avoid developing new questionnaires specifically for the project unless considered necessary, but rather to use questionnaires and/or instruments commonly used in studies of drivers and driving. The questionnaires distributed were therefore a combination of custom made questionnaires (number 1; part of numbers 5 and 6) and established questionnaires and instruments (numbers 2; 3; 4 and part of 5 and 6), often used in field tests of, for instance, safety systems. Unless originally designed or already available in Swedish, the questionnaires were translated from English into Swedish.

2.2 Rationale and procedure

The Background Questionnaire distributed before the test began contained questions on the participants' demographic data (e.g. age, sex, and height); questions on their driving experience and driving patterns (e.g. type of roads used), self assessment of driving skill, as well as questions on their knowledge and/or use of the different functions that were to be tested in SeMiFOT. It was assumed that individuals with good knowledge and experience of a certain function would react differently than would individuals who had little or no knowledge and/or experience of the function in question.

Also the Driving Behaviour Questionnaire (DBQ) was distributed before the test. This questionnaire was included in order to collect information on drivers' behaviour. In SeMiFOT, the specific questionnaire included 31 statements to which the participants had to respond according to a six-point Likert scale. For

the professional drivers, who were to describe their driving when driving trucks, some questions were omitted as they were interpreted to refer to the drivers' behaviour when driving their private car.

The Decision Making Questionnaire distributed to the drivers at the beginning of the trials was based on the questionnaire developed by French et al. (1993). The intention was to collect information on the individual's decision-making style as this may also reflect in his/her driving, and in different risks of having a road traffic accident. The project participants had to indicate how often they acted according to 21 different statements on a six-point Likert scale.

Also personality traits are considered to influence driving behaviour. Differences in personality should therefore be carefully mapped in a study such as SeMiFOT. Part of the drivers in SeMiFOT were asked to fill in an adapted Locus of Control Questionnaire, the Traffic Locus of Control Questionnaire, originally developed by Özkan and Lajunen (2005).

They had to agree or disagree to 17 statements on possible causes for traffic accidents. The reason for including this questionnaire was that in relation to different safety systems, it can be anticipated that drivers with an internal locus of control will rely more on their own ability no matter what safety systems are available, whereas externally-oriented individuals may be more likely to give up their control to a device (see e.g. Brown & Noy, 2004). This may affect the effect of the systems on driving, as well as the degree to which the systems are accepted.

Furthermore, three different User Acceptance Questionnaires were designed and used. One was distributed to the participants before the trials started, one after a short trial period (approx. one week into the trial), and one after a longer period of time (more than a month into the trial). The participants had to assess the different functions according to Likert type scales. Included as part of the questionnaire was the van der Laan Acceptance Scale (van der Laan et al. 1997). This scale consists of nine semantic differentials that are used to calculate two values, one indicating how usable a person finds the function (or product, device) and one indicating how satisfactory the function is perceived to be. As the participants had to answer one questionnaire per function, and since they had more than one function installed in their vehicle, there were a substantial number of User Acceptance Questionnaires to be filled in across the trial period.

After the trials, a final Evaluation Questionnaire, one per function, was distributed to the participants. This questionnaire was similar to the User Acceptance Questionnaires but contained additional questions on the participants' thoughts on the design of the specific function, their perception of possible benefits (or the opposite) of using the functions, and whether they considered keeping the function or not, etc.

The main part of the questionnaires were distributed as web-based questionnaires, using a commercial web-survey service. However, some of the participants (all of them truck drivers) received the questionnaires on paper as it was considered uncertain whether they had access to computers and the Internet or not. The questionnaires were in this case filled in by the respondent

and the responses were later added to the web-survey database by the research staff.

Complementary personal or group interviews were to be carried out with all participants. At the end of the test period, group interviews were conducted with a sub-set of the participating drivers (truck drivers). The interviews were semi-structured to their character, supported by an interview guide. The questions or themes covered included, for instance the participants' knowledge of the respective functions, their assessments of the functions and if/how it had changed during the test, the design of the user interfaces, as well as questions on how they had experienced participating in the field test as such.

3 RESULTS, EXPERIENCES AND IMPLICATIONS

The following section reports on some of the results and the experiences made related to the collection of subjective data by means of questionnaires and complementary focus group interviews.

3.1 *Design of questionnaires and instruments*

3.1.1 On the issue of private vs professional drivers

The participants were all experienced drivers. The car drivers drove on average 26.500 km/year, the truck drivers considerably more, on average 126.000 km/year with truck plus 19.000 km/year with their private car. Typically, the truck drivers had driven more than 5 million kilometers in total.

The car drivers were in general "positive" towards driving. They rated their driving style as "somewhat offensive" rather than "defensive". The truck drivers were "very positive" towards driving and especially positive towards driving trucks. They rated their truck driving style as "slightly defensive" while the considered their car driving style as "slightly offensive".

The fact that SeMiFOT addressed both private car drivers and professional drivers caused some problems regarding, for instance, the Driver Behaviour Questionnaire (DBQ). It could be argued that the DBQ should reflect the individual's driving behaviour in general, and therefore that it could be used independent of the participants being private or professional drivers and/or drivers of cars or trucks (or buses). However, in SeMiFOT, the participating truck drivers' self-assessment of their driving styles indicated a difference between their driving style in the trucks compared to in the private cars. Hence it can also be argued that if, for instance, the driver assistance systems to be tested in a FOT are to be introduced into a professional driver's truck (or bus), it is more appropriate to consistently focus on the driver and his/her driving style in this particular vehicle.

A particular predicament regarding the DBQ concerned the items included in the instrument. One obvious difficulty was, for instance, to ask the professional truck drivers how often they had felt uncertain about where they had parked their vehicle (i.e. the truck) on a large parking lot, but there were also other items formulated in a way that could be interpreted as questioning the individual in his/her professional driver role.

Given this observation, the questionnaires used in SeMiFOT should have been

modified and items reformulated in order to better fit the behaviours of professional drivers. The time available, between the collection of background information and the distribution of the DBQ did not, however, allow for such a redesign. The decision was instead to erase problematic items, a decision that could be questioned as the number of items measuring driver behaviour issues, such as “slips” or “violations”, as a consequence was reduced and hence also the basis for the categorisation of drivers.

3.1.2 On the issue of time

Most of the questionnaires used in SeMiFOT were established, validated questionnaires /instruments. According to the original study design, the Sensation Seeking Scale (SSS) was to be one of the instruments used, more specifically a Swedish version of the SSS-V. This instrument contained 40 forced choice items and respondents were to indicate which of two statements they perceived to best describe them as persons. However, the pilot studies carried out indicated that whereas some items were unproblematic, such as “I like to listen to loud music” (Swe: “Jag tycker om musik på hög volym nivå”), other items were awkward.

One reason was the terms used were clearly outdated, for instance “hippies” used in the statement “I would like to make friends in far-out groups like artists and hippies” (Swe: Jag skulle vilja ha lite udda vänner - som artister och “hippies”). Another reason, related to changes in societal attitudes over time, was the relevance of items such as: “I stay away from anyone I suspect of being homosexual” (Swe: ”Jag håller mig undan personer som jag tror kan vara homosexuella”). Together, these issues lead to the decision not to include the SSS in the SemiFOT study.

The problem is, evidently, not a new one – even though the literature contains less input than could be expected. The same problem has been identified, specifically regarding the SSS, for instance by Arnett (1994) and later also by Gray and Wilson (2007). The latter conclude, as did the SeMiFOT project, that the items in the SSS are “... dated in tone or reflect[ing] attitudes that are perhaps less useful indicators of sensation seeking at the start of the 21st century”. In response to the critique, Zuckerman has agreed that “... some of the item content was worded in the vernacular or slang of the time. This is a mistake test constructors should be warned about because if the test survives beyond a generation the vernacular of an older generation may become incomprehensible to the following generations.” The instrument has therefore been revised several times (see Zuckerman, 1979; Zuckerman, 1994), including replacement of terms (e.g. the term “hippies” has in later version been replaced by “punks”).

It is possible that a later version of the SSS, or other versions of the same instrument, would have passed the pilots carried out in the SeMiFOT project. Nevertheless, the experiences made emphasize the need to carefully pilot the questionnaires/instruments chosen and assess whether they will, indeed, reflect for instance the terms and values of the time and the target community.

3.1.3 On the issue of language and cultural differences

All questionnaires/instruments administered in the SeMiFOT project were in Swedish. However, most of the questionnaires/instruments used were originally designed in a language other than Swedish, most often English. Some of instruments had been translated in earlier projects, and a few of the translations have also been validated. Even so, translations and translating questionnaires/instruments were still an issue in the SeMiFOT project.

An option could have been to distribute the questionnaires without translations. However, this was never regarded as an alternative. The participants' knowledge in English (as a second language) may have been good but not necessarily good enough not to result in fundamental and serious errors in the interpretation.

One problem in translating a questionnaire from one language to another is evidently to find the exact, correct wording for each question/item. Common approaches to overcome possible errors include translation and back-translation, and multiple forward-translations, i.e. when two or more translators translate the instrument from the original language to the new language, and the versions of the instrument in the new language are then compared.

However, several studies show that one cannot assume that the translated items are valid simply because they have been translated. The cultural context of the second language must also be taken into consideration. Cultural contexts differ and as a consequence the meaning and intention of a word, sentence, or reference may be understood differently (compare e.g. Griffiee, 2001) Due to the context specificity of validity, a questionnaire written in one language and translated into another cannot be considered as an equivalent survey instrument. For instance, what is considered as an item measuring a "sensation seeking" personality in the UK or the Netherlands may not reflect a sensation seeking behaviour in Sweden, and/or may the item not be understood by the specific respondent as the he/she may not be able to relate to the references used. Thus, given that questionnaires/instruments to be used in a FOT need to be translated, substantial work has to be allocated the translation, as well as to the further verification of the translations and the validation of the new questionnaire before the actual data collection can begin. Any questionnaire translated into another language must be subjected to further analysis and pilot studies to confirm its validity. Based on the experiences from SeMiFOT, this phase in the project should have been allocated more time and resources.

The SeMiFOT project involved translations of instruments, from English to Swedish, but as all participants received the same translation, the results can at least be compared across individuals. The identified problems are, however, of another dimension when the (supposedly) same questionnaire/instrument is to be used in a project encompassing studies in several countries (as is the case for instance in ongoing FOTs, such as EuroFOT and TeleFOT) and where cross-national analyses are part of the scope.

3.2 Choosing data collection method

3.2.1 On the issue of web-based versus paper-based questionnaires

The main part of the questionnaires in SeMiFOT was distributed as web-based questionnaires, using a commercial web-survey service. The assumed advantages of using the Internet include cost savings associated with eliminating the printing and mailing of survey instruments, as well as time and cost savings of having returned survey data received already in an electronic format (cf. Cobanoglu et al. 2001). Another issue is the possible

increase in response rates proposed for instance by Saltzman (1993) as well as by Cheyne and Ritter (2001). The decision can evidently also be questioned. Despite the growth of the Internet there is still a large number of people who do not have access and/or choose not to use the Internet. This was (at least assumed to be) the case in the SeMiFOT project why part of the participants instead filled in traditional paper-and-pen based versions of the questionnaires.

Nevertheless, a major part of the participants had access to the Internet but even though they were highly motivated test drivers of different OEMs, and the OEMs had allocated persons responsible for monitoring the data collection and reminding the participants to fill in the questionnaires, there is data missing. In particular this concerns the questionnaires administered towards the end of the test. Thus, although no strict comparison has been made, the web-based survey solution at least does not necessarily ensure a higher response rate or fewer drop-outs over time than do the traditional pen-and-paper based solutions administered by mail. This same conclusion has been reached in other studies, for instance by Kaplowitz et al. (2004).

The need for reminders and a dedicated person with specific responsibility for monitoring that participants receive the information that they should receive and that they also fill in the questionnaires appears absolutely necessary also when the data is collected by means of web-based questionnaires (cf. Kaplowitz et al., 2001). At the same time the questionnaires that were answered were also completed which could imply that the web-based format had a positive effect, as proposed for instance by (Griffis et al. 2003) – but probably more so that those participants motivated enough to fill in the questionnaires were also motivated to complete them.

Another concern is evidently if there are differences in the nature of the data gathered by web-based and paper-based questionnaires, in particular if the study uses a mix of both. In SeMiFOT the number of participants is too small to conduct any comparisons but other studies, carried out with the specific purpose of comparing the outcome of web-based and paper-based responses, have found little evidence of any such mode effect (e.g. Ekman et al., 2006; Meckel et al. 2005). Nevertheless, these studies concerned other topics and in designing a study, such as SeMiFOT, the attitude towards computers and the Internet (as technical artefacts), positive or negative, might project onto the attitude towards and hence the assessment of the technical functions tested (compare Denscombe, 2006). Such an impact needs to be considered in studies of a FOT character.

3.2.2 On the issue of questionnaires vs interviews

Methods for collecting subjective data include, for instance, questionnaires and personal interviews. In SeMiFOT questionnaire data was used in order to investigate the participants' acceptance of the functions tested and any changes in acceptance over time. According to the results most functions were, according to the van der Laan Acceptance Scale, considered moderately usable and satisfactory, with FCW, ACC, and ESC receiving the highest scores and BLIS the lowest. The results also showed that FCW and ACC received higher ratings the longer the participants used the system. On the other hand, the participants' satisfaction with the LDW function decreased over time whereas perceived usefulness remained on the same level. This could be interpreted as though the participants considered LDW to be 'a good function' as such, but that the performance of the specific systems installed were less satisfactory.

However, the fact that the participants assessments changed over time is one issue, the reasons why is another. One way to find out more about the motives behind assessments and ratings would be to provide close-ended questions and answer alternatives that offer the sensitivity needed for the interpretation. Another possibility is the introduction of open-ended questions, which was the decision in the SeMiFOT project. A relevant critique is that openended questions often result in data missing. In SeMiFOT most participants but far from all wrote comments. Furthermore, the comments were most often short, ranging from a single word to a single sentence, for example "More false alarms than I expected" (referring to BLIS) and "If you consider the consequences in the long term run, I can see the benefits of LDW". In contrast to findings by, for instance, Denscombe (2009) the web-based format did not in this case appear to motivate the participants to elaborate their answers to any higher degree than do the traditional paper-and-pen version.

The introduction of open-ended questions can be debated also from the perspective of resources. The analysis requires considerable effort and resources to be allocated preprocessing, encoding and (in some cases) also translation of the verbatims. Actually, the analysis of the relatively few comments provided by those participating in the SeMiFOT project took a substantial amount of time to compile and analyze. At the same time the short comments provided at least some further insight into the reasons behind the assessments.

Studies with the purpose of investigating the relationship between, for example, personality traits and acceptance, or driving style and acceptance, can most probably rely on data collected by means of the type of questionnaires used in SeMiFOT. Also for studies with the purpose of investigating the overall reasons for acceptance or rejection of a new technical function, such questionnaire data may suffice. However, if the purpose is to more in-depth understand the underlying reasons for indicating, for instance, that a function is not useful or is difficult to use – in particular when the reasons are sought in the design of the function – the participants must be given the opportunity to explain and elaborate on the theme. The participants may even need to be supported in reflecting on the issues, something that may require a skilled interviewer enabling the interviewee through different probing techniques or by providing

the possibility for the participant to demonstrate his/her interaction with the device/system evaluated. In SeMiFOT the initial intention was to conduct complementary individual interviews, or group interviews, with all participants. Interviews were carried out with the truck drivers involved in the test whereas the car drivers have not been interviewed. Compared with the single word or single sentence collected by means of the open-ended questions in the questionnaires, the richness of the data is evident. A question regarding to what degree the drivers trusted the systems resulted for instance in the following elaboration from one of the participants: "No, you don't trust it - or yes - I trust it when I'm on the motorway and it keeps the distance. it reacts when, for instance, the car in front brakes in panic. Then it beeps. And then you can rely on that it will stop the car without me braking too. But I cannot ... that is I still have to observe things. So trust is a difficult question."

Based on the experiences made in SeMiFOT, as well as in a number of other studies, there are strong arguments for carrying out personal, in-depth interviews with at least a subset of FOT participants.

4 ACKNOWLEDGEMENTS

Acknowledgements are due to the participants in the SeMiFOT project and to the project partners, in particular Emma Johansson at Volvo Technology, Katja Kircher at the Swedish Road Administration, Martin Lönn at Saab Automobile, and Josefin Nilsson at Scania.

5 REFERENCES

- Arnett, J. (1994): Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, 16, pp. 289-296
- Brown, C.M. & Noy, Y.I. (2004): Chapter 3. Behavioural adaptation to in-vehicle safety systems. In: *Traffic & transport psychology. Theory and application*. Rothengatter, T. and Huguenin, R.D. (eds). Elsevier, Amsterdam, pp. 25-46
- Cheyne, T. L. & Ritter, F.E. (2001): Targeting audiences on the Internet. *Communications of the Association for Computing Machinery (ACM)*, Vol. 44, No. 4, pp. 94-98
- Cobanoglu, C., Warde, B. & Moreo, P.J. (2001): A comparison of mail, fax, and web-based survey methods. *International Journal of Market Research*, Vol. 43, No. 4, pp. 441-452 (2001)
- Ekman, A., Dickman, P.W., Klint, Å., Weiderpass, E., & Litton J-E. (2006): Feasibility of using web-based questionnaires in large population-based epidemiological studies. *European Journal of Epidemiology*, 21(2), pp. 103-111
- Denscombe, M. (2006): Web-based questionnaires and the mode effect. *Social Science Computer Review*, Vol. 24, No.2, pp. 246-254
- Denscombe, M. (2009): Item non-response rate: a comparison of online and paper questionnaires. *International Journal of Social Research Methodology*, Vol. 12, Issue 4, pp.281-291

French D.J., West R.J., Elander J., & Wilding J.M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents, *Ergonomics*, Jun; 36(6), pp. 627-644

Gray J.M. & M.A. Wilson (2007): A detailed analysis of the reliability and validity of the sensation seeking scale in a UK sample, *Personality and Individual Differences* 42, pp. 641–651

Griffiee, D.T. (2001): Questionnaire translation and questionnaire validation: Are they the same? Paper presented at the Annual Meeting of the American Association for Applied Linguistics. St. Louis, MO, February 24-27

Griffis, S.E., Golsdby, T.J. & Cooper, M. (2003): Web-based response and mail surveys: a comparison of response, data and cost', *Journal of Business Logistics*, Vol. 24, No. 2, pp. 237–258

Kaplowitz, M.D., Hadlock, T.D. & Levine, R. (2004): A comparison of web and mail survey response rates. *Public Opinion Quarterly*, Vol. 68, No.1., pp. 94-101

van der Laan, J.D., Heino, A., & De Waard, D. (1997): A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research – Part C: Emerging Technologies*, 5, pp. 1-10

Meckel, M., Walters, D. & Baugh, P. (2005): Mixed-mode surveys using mail and web questionnaires. *The Electronic Journal of Business Research Methodology*. Vol.3, Issue 1, pp. 69-80

Reason, J., Manstead, A., Stradling, S., Baxter, J., & Campbell, K. L. (1990): Errors and violations on the roads: A real distinction? *Ergonomics*, 33, 10/11, pp. 1315-1332

Saltzman, A. (1993): Improving Response Rates in Disk-by-Mail Surveys. *Marketing Research*, Vol. 5, No. 3, pp. 32-39

Zuckerman, M. (1979): *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, NJ: Erlbaum

Zuckerman, M. (1994): *Behavioral expressions and biosocial bases of sensation seeking*. New York: Cambridge University Press

Özkan, T., Lajunen, T., & Kaistinen, J. (2005): Traffic locus of control, driving skills, and attitudes towards in-vehicle technologies (ISA and ACC). 18th International Co-operation on Theories and Concepts in Traffic Safety (ICTCT), Helsinki, Finland (<http://www.ictct.org/>)

Özkan, T., & Lajunen, T. (2005): Multidimensional Traffic locus of control scale (T-LOC): factor structure and relationship to risky driving. *Personality and Individual Differences*, 38(3), pp. 533-545