

**WHAT YOU SEE IS WHAT YOU GET? CORRESPONDENCE OF VIDEO AND  
INTERVIEW DATA ON SECONDARY TASK ENGAGEMENT WHILE DRIVING-  
A NATURALISTIC DRIVING STUDY**

Maria Kreusslein\*, Katja Schleinitz\*<sup>+</sup>, Markus Schumacher<sup>#</sup>  
\*TU Chemnitz, Germany, <sup>+</sup> TÜV | DEKRA arge tp 21, Germany  
<sup>#</sup>Federal Highway Research Institut (BAST), Germany  
Email: maria.kreusslein@psychologie.tu-chemnitz.de

**Summary:** Numerous studies use questionnaires or interviews to investigate the prevalence of secondary task engagement while driving. This data may be subject to memory distortion. This study aims at investigating the extent to which interviews are valid instruments to assess secondary tasks. Therefore, we evaluated interviews and video data allowing the observation of secondary task engagement from a Naturalistic Driving Study. We equipped the vehicles of 94 subjects with cameras filming the driver's vehicle cabin. Video and interview data were collected twice within the study period of 3 days. We then determined hit rate, misses, false alarms, correct rejections, sensitivity, as well as specificity for 15 secondary tasks. We found 594 secondary tasks in the videos. In 10% of all comparisons ( $N_{all}=2.187$  for 15 tasks) the interview correctly identified task engagement (hit). In 17% of the comparisons drivers missed to report a task. In 9% of the comparisons there was a false alarm and in 64% we found correct rejections. More conscious and long-lasting tasks (hands-free phoning, smoking) were remembered best. The interview method seems to be a valuable and valid tool to assess rather consciously conducted and legally prohibited secondary tasks while driving.

## INTRODUCTION

Distraction caused by performing secondary tasks while driving is assumed to be one of the main causes for traffic crashes (e.g., Dingus et al. 2016, Huemer & Vollrath, 2012). To determine crash risk accurately, research needs to provide information about the secondary task engagement frequency while driving (Kidd & McCartt, 2015) and the riskiness of the task execution. Subjective as well as objective methods have been applied to assess the prevalence.

## Subjective methods

Until now, a large quantity of studies providing information on secondary task prevalence is based on information provided by the drivers themselves (e.g. questionnaires, interviews). Amongst others, the main advantage of these methods is that large samples can be assessed within a short time frame with minimal costs. However, these methods have a fundamental problem regarding the validity of the data (Petzoldt & Utesch, 2016), as they rely on the respondents giving complete and honest details about secondary task performance while driving. Some of the tasks may be prohibited by law in certain countries and therefore might be withheld in interviews due to social desirability biases (Philips & Clancy, 1972). Furthermore, survey data is subject to the naive understanding of secondary activities and memory distortion (Petzoldt & Utesch, 2016). The question arises whether all relevant activities that could affect the validity of the data have been assessed. It is supposed, that remembering a secondary task depends on the type and nature of the task and the situation in which it is carried out. Strenuous, long-lasting

secondary tasks like phoning (hand-held, hands-free) might be remembered better. Tasks that occur more frequently and are performed less consciously (e.g. changing radio stations) might not be remembered. Objective methods should solve the problem of underestimation.

### **Objective methods**

There is a variety of methods that can be used to objectively assess secondary task engagement while driving. Beside more technical methods (digital image processing) there are numerous observational studies on secondary task engagement (e.g. Huemer, Schumacher, Mennecke, & Vollrath, 2018). Direct observations are characterized by the fact that an experimenter sitting in the vehicle or standing at the roadside (stationary e.g. at intersections) registers all visible secondary activities. Indirect observations are conducted when secondary activities are identified via videos recorded during driving and coded subsequently. An increasingly used approach is the Naturalistic Driving Study (NDS). Hereby, drivers are observed while driving by unobtrusively built-in data acquisition systems (e.g. cameras in the vehicle; Klauer, Dingus, Neale, Sudweeks & Ramsey, 2006, Eenink, Barnard, Baumann, Augros & Utesch, 2014). Here, drivers are expected to demonstrate everyday driving behaviors. Further, driver's attention is not focused on omitting secondary tasks. Conducting NDS's is, however more cost- and time consuming than questionnaire studies. Nevertheless, it is suitable for the validation of questionnaire data.

### **Objectives**

The question arises whether interviews are a valid tool to assess secondary tasks. Therefore, the objective of this study was to investigate the validity of the survey methodology using a NDS combined with interviews. We refined the secondary task engagement semi structured interview guide by Huemer and Vollrath (2011).

## **METHOD**

### **Sample**

We acquired a total of 94 subjects, which met our inclusion criteria: a valid drivers' license and a weekly mileage of at least 100 km, which guaranteed that subjects were driving for 30 minutes at least once within the study period. All in all, we sighted 67 hours of video footage. Due to technical problems the final sample consists of 86 drivers (Men=57, Women=29), with a mean age of 38 years ( $SD= 12.2$ ,  $Min= 19.0$ ,  $Max=74.0$ ). Drivers had been holding their license for 19 years on average ( $SD= 11.7$ ,  $Min= 2.0$ ,  $Max=54.0$ ) and showed a weekly mileage of 382.15 km ( $SD=310.2$ ,  $Min=60.0$ ,  $Max=1900.0$ ).

### **Methodology**

We combined a naturalistic driving study (NDS) with interviews. Electronic objective observations were realized with the NDS approach. We refined the semistructured interview guide by Huemer and Vollrath (2011) to assess secondary task engagement (17 tasks) objectively. Subsequently, the validity of this interview data was tested with the video data of the NDS. In order to increase the sample size, two interviews were conducted with each driver for trips occurring within the study period of three days. We conducted the first interview via telephone after camera installation and a drive lasting approximately 30 minutes. The second interview occurred during the deinstallation of the cameras of the driver's vehicle.

## Material

*Data acquisition system (DAS).* For the NDS, we used dashcams (Blackvue DR650S-2CH IR), consisting of two cameras: One Full-HD front camera (920X1080 pixels, with 30 frames per second) and one HD rear camera (1280 x 720, 30 frames per second) including infrared LED (Model RC100-IR). The infrared extension also ensures the visibility of activities during trips in the dark. The viewing angle of both cameras was 129°. In addition, the system featured a 3-axis G-sensor for logging the G-forces during vehicle movement. This sensor made it possible for the drivers to record the videos without intervening in the system. It registers any movement of the car, which starts the video recording. In addition, a GPS sensor made it possible to identify the distance travelled. Camera position in the car is depicted in Figure 1.



Figure 1. Camera position in vehicles

*Interview guide.* . In the interview, we applied a two-stage process when asking for task engagement. First, the drivers were asked an open question. As a second step, the experimenter asked about the execution of all secondary tasks from a list of 17 that had not been mentioned in the first answer, by means of a closed format. We only included tasks executed within the final 30 minutes of the last trip to minimize memory distortion bias. The following tasks were assessed (extended; following Huemer & Vollrath, 2011):

- Phoning hand-held
- Phoning hands-free
- Writing of text messages
- Reading of text messages
- Internet usage
- Usage of the navigation system
- Vehicle- related adjustments (relevant to driving, e.g. adjusting seat, mirror)
- Inputs to internal devices (e.g. radio, AC)
- Inputs to external devices (e.g. mp3-player)
- Interacting with another passenger
- Self-initiated tasks (e.g. self-talk, singing, intense reflection)
- Hygiene actions (cleaning nose, combing, make-up)
- Changing clothes (also putting on glasses, to button clothes)
- Eating & drinking
- Smoking
- Searching, picking up or placing something; cleaning
- Distraction from the outside

## Procedure

Subjects were approached via a participant database of the department, as well as the university webpage and the local newspaper in spring 2017. After giving informed consent, subjects were deceived about the contents of the study by means of a cover story, to secure that the drivers did not focus their attention on the distracting activities or consciously refrain from secondary tasks. Participants were told that the study's aim was to compare different camera systems. After the camera installation, subjects drove three days with the DAS installed in their vehicles. The first interview was conducted via telephone after a trip lasting at least 30 minutes. Interview duration was between 20 and 30 minutes and included questions about camera systems (to keep up the

cover story), as well as questions on secondary task engagement. We conducted the second interview (face-to-face) on the third day, while the DAS was deinstalled. Subjects were subsequently debriefed and informed about the actual study content and were asked to once again give informed consent. Finally, they received a remuneration of 40€. Prior to the study, the procedure gained ethical approval by the faculty's ethics commission.

### Data analysis

Due to technical issues, we could only include one of two trips (videos) of some participants. Therefore, the analysis included 154 videos ( $N_1=76$ ,  $N_2=78$ ). In the beginning, we identified relevant video sequences (the final 30 minutes of the respective trip that were also addressed in the interview) and coded them with regard to secondary task engagement. We then compared and analyzed interview and observational data regarding accordance of secondary task engagement. The comparison included only the occurrence of a task in the video (executed vs. not executed) and not the frequency of secondary tasks in the video. Based on the Signal-Detection Theory (Green & Swets, 1966), we created a four-field cross-table (Table 1). For the evaluation of the interview method it is of particular relevance to evaluate whether the drivers reported all secondary activities in the semistructured interview that were annotated in the videos (a). In addition, it is important to know the quantity of activities that is not reported in the interview but was seen in the video (c) as well as those that were reported but not seen at all in the video (b). In case of a correct rejection no secondary task occurred, neither in the interview nor the video.

**Table 1. Four-field cross-table based on signal-detection theory**

Number of activities		Video	
		Executed	Not executed
Questionnaire	Mentioned	Hit (a)	False alarm (b)
	Not mentioned	Miss (c)	Correct rejection (d)

We calculated sensitivity and specificity of the answers in the interview guide. Sensitivity describes the probability that the interview correctly identifies activities seen in the video. Specificity describes the probability that the interview correctly identifies activities not performed in the video. The frequency of the activities performed in trip 1 and trip 2 was calculated using the odds ratio (OR) to exclude social desirability biases. OR is a measure of association between an exposure and an outcome. OR of 1 means no association, while  $OR > 1$  indicates the presence of an association. *P*-values were calculated according to Sheskin (2004).

$$Sensitivity = \frac{a}{a + c}$$

$$Specificity = \frac{d}{b + d}$$

$$OR = \frac{a/b}{c/d}$$

Internet usage was only mentioned by one driver. Due to positioning of the camera it was not possible to clearly identify this kind of mobile phone use in the video. The same holds true for drivers who positioned external devices outside the camera range. Therefore, internet usage and handling of external devices besides mobile phones were not considered in the analyses (15 tasks remaining).

## RESULTS

### Frequency of secondary task engagement

In order to maximize the data base, we included video data from both interviews (telephone and face-to-face). The majority of secondary tasks did not appear significantly less in the second drive (Table 2; Columns: 3 [OR] and 4 [*p*]). Overall, videos revealed 594 secondary tasks. The most frequently performed tasks were activities for personal hygiene (*n*=147), followed by inputs to internal devices like a radio or AC (*n*=75), changing clothes (*n*=67), and self-initiated tasks (*n*=67). Almost all drivers interacted with passengers when there were any in the car with them (26 of 31). Mobile phones use (hand-held/free phoning, reading/writing messages) was identified in 7% of the videos.

### Correspondence of video and interview data

We found agreement in approximately 10% of all video-interview comparisons (*N*=2.187 for 15 tasks). In 17% of the comparisons the interview missed the task, in 9% drivers mentioned a secondary task not seen in the videos, and in 64% the analysis revealed correct rejections. Descriptive results (Table 2) showed the highest number of hits was found for inputs to internal devices, but also a relatively high number of false alarms and misses. Overall sensitivity (Table 2) of the interview method was 50%. However, we found large differences between the tasks. It was highest for smoking, interacting with passengers, and input to internal devices and lowest for actions such as changing clothes, searching, picking or placing items inside the car, and personal hygiene. Overall specificity (Table 2) was 84%. For a few tasks, it was rather low e.g. inputs to internal devices and interacting with passengers.

**Table 2. Frequency of secondary task engagement from video data in first and second drive; correspondence categories, sensitivity and specificity of interview.**

	Drive				Correspondence categories				Sensitivity in %	Specificity in %
	1	2	OR	<i>p</i>	Hit	False Alarm	Miss	Correct Rejection		
Phoning hand-held	2	10	0.18	.032	3	1	9	141	25.00	99.29
Phoning hands-free	7	5	1.48	.519	9	6	3	136	75.00	95.77
Writing text messages	5	5	1.02	.096	5	5	5	139	50.00	96.52
Reading text messages	5	7	0.71	.580	8	16	4	126	66.66	88.73
Navigation system	1	1	1.02	.985	1	12	1	140	50.00	92.10
Vehicle-related adjustments	26	21	1.28	.475	10	17	37	90	21.27	84.11
Inputs to internal devices (Radio, AC)	40	35	1.36	.335	58	50	17	29	77.33	36.70
Interacting with passengers	17	9	2.20	.077	22	3	4	2	84.61	40.00
Self-initiated tasks	36	31	1.36	.340	48	24	19	63	71.64	72.41
Activities for personal hygiene	72	75	0.72	.674	19	0	128	7	12.92	100.00
Changing clothes	34	33	1.10	.761	6	7	61	80	8.95	91.65
Eating & drinking	9	14	0.61	.291	13	5	10	126	56.21	96.18
Smoking	2	2	1.02	.979	4	0	0	150	100.00	100.00
Searching, picking up, placing things	30	24	1.46	.258	5	8	49	92	9.25	92.00

Distraction from the outside	26	11	3.16	.004	18	42	19	75	48.64	64.00
Sum of activities	312	283	-	-	229	196	366	1396	-	-

## DISCUSSION

The goal of the study was to investigate the validity of interview method to assess secondary task engagement while driving. Therefore, we combined a NDS with semistructured interviews including a questionnaire about secondary task engagement.

In general, the results indicate large differences between the secondary tasks with regard to a proper assessment via a semistructured interview. Our results show that the sensitivity of interviews to detect secondary tasks is particularly high for tasks like smoking, interacting with passengers, and inputs to internal devices (>75%). However, for an interview to be considered valid, the specificity has to be high too. This was true for hands-free phoning and smoking. These are rather conscious and long-lasting secondary tasks which might be remembered very well and thus stated correctly when executed. Although the probability to detect input to internal devices and interactions with passengers is relatively high (>75%), the interview method falsely reports these tasks even though we could not see them in the videos (<50%). Presumably, drivers are not able to clearly assign these activities to the relevant trip period. Interview answers are always dependent on the respondent's memory performance. Probably drivers performed these tasks prior to the period of relevance. Drivers are not able to determine the reference period due to the driving situation complexity, which supposed Petzoldt and Utesch (2016) as well.

We found medium sensitivity (50-75%) and high specificity values (>75%) for the tasks such as texting and reading text messages, as well eating and drinking, and usage of the navigation system. This supports the assumption that subjective data underestimates the frequency of such activities to a large extent. The interview misses to report up to 50% of these executed tasks. It is possible that drivers did not mention these tasks due to social desirability biases, which is supported through the high specificity. Even if specificity was high, we can see some false alarms with the interviews. Both, sensitivity and specificity were medium for self-initiated tasks such as talking or singing while driving and distraction from the outside. For these activities, we determined the highest number of secondary tasks drivers reported but could not find these in the videos. The tasks are evaluated rather “subjectively” and might not be visible to external observers, e.g. distractions from the outside of the vehicle and self-initiated task like intense thinking or mind-wandering, but also inputs to internal devices.

We found the lowest sensitivity values for (<50%) secondary tasks that include very detailed actions e.g., activities of hygiene, changing clothes, searching items, and vehicle related adjustments. An explanation might be that these tasks are performed rather unconsciously (e.g., activities of hygiene like nose cleaning, biting nails) and cannot be remembered. This assumption is supported by the high specificity (>75%). Moreover, sensitivity for hand-held phoning is very low. This category includes actions such as searching, picking up as well as holding the mobile phone. Most likely drivers ascribe this behavior to the category texting or reading messages. Therefore, these categories should be included to future interviews.

Secondary task recognition from the videos material was limited due to the limited range and/or resolution of the cameras, i.e., inputs that drivers performed via the steering wheel could not be

seen very well by the coders. An even larger data set should be generated for checking the correspondence for activities that occur less frequently in daily driving because they are prohibited legally (e.g. phoning, reading and writing text messages) or the basic rate is rather low (smoking). However, the results demonstrate that interviews are a valid and efficient tool to assess these secondary tasks. On the downside, valid results for distracting tasks which are not considered as those by the drivers, since they are probably performed rather unconsciously (e.g. activities for personal hygiene) are missing. Depending on the kind of distraction, questionnaires, NDS or observations from the roadside are more suitable. The methods can also be combined to exploit the respective advantages and compensate for the shortcomings.

## ACKNOWLEDGMENTS

This report is based on parts of the research project carried out at the request of the Federal Ministry of Transport and Digital Infrastructure, represented by the Federal Highway Research Institute, under project No. 82.0642/2015. The author is solely responsible for the content.

## REFERENCES

- Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, 113(10), 2636-2641. <https://doi.org/10.1073/pnas.1513271113>
- Eenink, R., Barnard, Y., Baumann, M., Augros, X., & Utesch, F. (2014). UDRIVE: the European naturalistic driving study. *In Proceedings of Transport Research Arena*. IFSTTAR.
- Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.
- Green, D., & Swets, J. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Huemer, A., & Vollrath, M. (2012). *Ablenkung durch fahrfremde Tätigkeiten – Machbarkeitsstudie*. Berichte der Bundesanstalt für Straßenwesen, Mensch und Sicherheit M225. Bremerhaven: Wirtschaftsverlag NW.
- Huemer, A. K., Schumacher, M., Mennecke, M., & Vollrath, M. (2018). Systematic review of observational studies on secondary task engagement while driving. *Accident Analysis & Prevention*, 119, 225–236. <https://doi.org/10.1016/j.aap.2018.07.017>
- Kidd, D. G., & McCartt, A. T. (2015). The relevance of crash type and severity when estimating crash risk using the SHRP2 naturalistic driving data. *4th International Conference on Driver Distraction and Inattention 2015*. Sydney, New South Wales, Australia.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data*. Virginia Tech Transportation Institute.
- Petzoldt, T., & Utesch, F. (2016). Trying to validate subjective reports with a naturalistic driving data-a case against questionnaires and surveys to quantify driver distraction. *European Conference on Human Centered Design for Intelligent Transport Systems, 30. June - 01. July 2016*. Loughborough, UK.
- Phillips, D. L., & Clancy, K. J. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology*, 77(5), 921-940.
- Sheskin, D.J. (2004) *Handbook of parametric and nonparametric statistical procedures*. 3rd ed. Boca Raton: Chapman & Hall /CRC.