Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 76 (2024) 310–326

Transportation Research Procedia

www.elsevier.com/locate/procedia

12th International Conference on Transport Survey Methods

Evaluating GPS-based Methods of Data Collection from the Users' Perspective

Anton Galich[a]*, Claudia Nobis[a]

[a]*Institute of Transport Research, Rudower Chaussee 7, 12489 Berlin, Germany*

Abstract

This paper evaluates new methods of data collection based on GPS-tracking taking the user's assessment of data quality into account. In contrast to many other studies that assess data quality primarily through a comparison of the results obtained by different methods, we focus on the users' perspective which allows analysing the reasons underlying differing results obtained by diverse methods. In particular, it is highlighted how including the user's assessments can point out potential for improvement in the system of data collection that cannot be detected by looking at the technical output paraments alone.

Keywords: GPS; tracking method evaluation; user perspective

1. Introduction

Data collection based on (relatively) new technical devices has gained increased popularity in transport research in the past two decades. GPS-Loggers have first been applied to track mobility behaviour in the 1990s (Murakami and Wagner 1999, Wagner 1997). Smartphones followed soon after in 2004 in a project carried out in Japan (Itsubo and Hato 2006, Ohmori et al. 2005). Technological development has advanced considerably since these first applications and a myriad of different approaches for data collection based on new technical systems have been put forward in recent years. These include among others smartphone apps relying on Wi-Fi and network location capabilities (Greaves et al. 2015, Sadeghvaziri et al. 2016), smart card data (Chapleau et al. 2018), web-accessible systems merging and processing data from GPS-Loggers, accelerometers, and heart-rate monitors (Kang et al. 2018), smartphone apps

* Corresponding author. Tel.: +49 30 67055-9109; fax: +49 30 67055-283.
E-mail address: anton.galich@dlr.de

10.1016/j.trpro.2023.12.057

using GPS location data (Necula 2015, Richard and Rabaud 2018, Shankari et al. 2018, Sobhani et al. 2019), and mobile phone data based on cell tower networks (Li et al. 2018).

These new approaches have been developed based on the expectations to produce more accurate and more fine-grained data collected over a longer period of time and at a lower cost in comparison to traditional survey methods (Prelipcean and Yamamoto 2018). However, it is not clear yet at the moment, whether the new approaches can already keep their promises. In fact, evaluating their applicability and the quality of the data produced became a big topic in the transport research community (Bonnel and Munizaga 2018, Trépanier and Yamamoto 2015) and diverse methods for quality and applicability assessment have been put forward by different researchers (Stopher et al. 2015). Many of these evaluations focus on comparisons of the data collected by different methods on the basis of specific indicators such as the number of trips per person, average trip length etc. and rely on small and not representative samples.

In contrast, this paper puts forward an evaluation that not only relies on representative samples for each method of data collection but also takes into account the users' own assessments of the collected data and the new technologies. Our data was primarily collected through a new technical system called MovingLab (data collection with GPS tracking and other sensors via smartphones and loggers) and a written questionnaire in autumn 2018 in the city of Berlin (Germany). In addition, the participants who produced data through the technical system answered a questionnaire on their sociodemographic background and mobility routines before the data collection and a questionnaire on their experiences with the technical systems after the data collection. By analysing the combination of sociodemographic information, technology assessments, and actual mobility data from our representative samples, we want to contribute an evaluation of new methods of data collection that focuses more on the users' perspective than most of the existing research in this field.

The remainder of this paper is structured as follows. First, a brief overview on the scholarly literature on the evaluation of new technical systems for data collection in transport research is provided in order to highlight the contribution of this article. Second, the technology and data basis underlying this paper are illustrated. Third, the results of the performance analysis of the MovingLab System with regard to the users' assessments are presented. Finally, the conclusions and insights from this paper are summarized.

2. Literature review

Many of the new technical systems for data collection that have been developed by universities and research organisations in the past two decades are based on GPS-tracking due to its high spatio-temporal precision (Prelipcean and Yamamoto 2018). However, in spite of all the progress made, there remain some key challenges in preparing GPS-based data for further mobility analyses such as automated trip, trip segment and stop detection (Gong et al. 2018, Safi et al. 2016). Consequently, there are diverse studies that attempt to evaluate the quality of GPS-based systems of data collection. Almost all of these studies do this by comparing different methods of data collection on the basis of certain benchmarks such as the number of trips identified.

Kang et al. (2018) and Stopher et al. (2015) for instance evaluate the data collected by equipping people with GPS loggers by comparing it to the information provided in travel diaries by the same people. Kang et al. (2018) found more trips being collected by the GPS Loggers than by the travel diary. By looking at specific examples they assume that this is due to walking trips neglected in the travel diary (trips during work or at night) and to single vehicle trips separated into vehicle and bicycle trips due to low speed by the GPS based system (Kang et al. 2018). However, Stopher et al. (2015) found that more walking trips were reported in the travel diaries than recorded by the GPS Loggers. The authors assume that this is due to the known technical problems of GPS Loggers such as cold start, short duration travel, and travelling in urban canyons (Stopher et al. 2015).

Also other authors tried to evaluate the performance of GPS based systems by comparing it to other methods of data collection by specific benchmarks. Patterson and Fitzsimmons (2016) found that data collection with GPS based smartphone apps results in more trips per day and person in comparison to travel diaries. Allström et al. (2017) concluded that the traditional travel diary tends to miss "return home" trips, whilst the smartphone app tends to miss the "pick-up and/or drop-off" trips. Stopher et al. (2018) found that both data collection via smartphone apps and GPS Loggers resulted in higher daily trip rates than information gathered via travel diaries. Erhardt and Rizzo (2018) compared data collection via GPS Loggers with and without prompted recall options for the users and identified significantly fewer overall trips and among these significantly more car trips in the GPS-only approach. Finally, Neven et al. (2018) highlight that self-reported diaries seem to be more suited for persons with higher disability severities as these more often forget to take their GPS device with them because of several organizational issues related to their mobility limitations (Neven et al. 2018).

All of the mentioned studies have in common that the GPS based systems of data collection are evaluated against certain benchmarks such as the number of trips or mode choice from other sources of data such as travel diaries or GPS approaches with prompted recall options. This allows identifying differences in the data collected but often makes it difficult to elaborate the reasons for these differences. In particular when studies show contradictory findings such as more (Kang et al. 2018) or fewer (Stopher et al. 2015) walking trips detected with GPS based systems in comparison to travel diaries, it would be of great value to elaborate the actual reasons for the different results.

Including the users' perspectives in the evaluation of the technical systems can help in shedding light on the underlying issues for differences in certain benchmarks as Montini et al. (2015) illustrate. In their study the users were equipped with both GPS device and smartphone app and were able to review their collected trip history and to provide manual corrections in the app (Montini et al. 2015). In addition, the users also participated in qualitative interviews and questionnaires in which their experiences and their view on the app were evaluated (Montini et al. 2015).

The statements of the users revealed for instance that not detected trips by the GPS devices rely in part on the technical system not collecting any data although the app was turned on and in part on human behaviour when it was forgotten to turn the app on or when the GPS device was left at home (Montini et al. 2015). In some cases, the app was also deliberately turned off to save battery (Montini et al. 2015). In particular this last result is a valuable insight as it not only highlights one of the reasons for fewer trips being detected by the smartphone app but also illustrates how the app has to be improved in order to solve this issue.

Hence Montini et al. (2015) illustrate the added value of including the users' perspectives in the evaluation of new technical systems for data collection. However, the example of Montini et al. (2015) constitutes the only study found in this literature review that included the users' perspectives. Against this background the paper at hand constitutes another attempt to highlight the benefits of looking at the users' experiences in order to not only evaluate the quality of a new technical system for data collection but also to identify potentials for improvement.

3. Technology and data basis

The following subchapters first provide a brief description of the MovingLab System and the field test in which our data basis was collected. Thereafter, a basic overview of the data collected is presented.

3.1. The MovingLab System

The MovingLab System consists of four basic components: 1) The human participants or users, 2) the different mobile devices used for data collection, 3) a website for trip validation and correction, and 4) a data bank that stores and processes the data collected. All of these four components are connected to each other via the internet as depicted in the following figure:

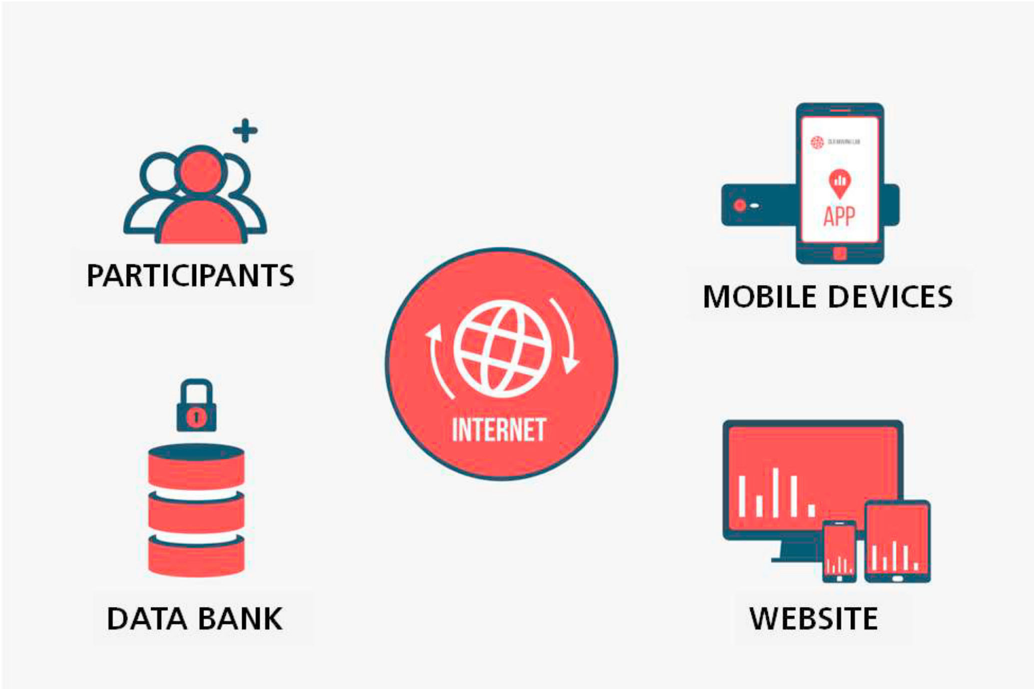

Fig. 1. The MovingLab System.

Three different mobile devices can be used for data collection in the MovingLab System: 1) The participants own smartphones, 2) smartphones provided by the MovingLab project, and 3) GPS-Loggers. Each participant in a project is equipped with one of these devices. The smartphones were equipped with the MovingLab app to collect and send the relevant data to the data bank while the GPS-Loggers directly collect and send the relevant data to the data bank. Both the MovingLab app and the GPS-Loggers provide the opportunity to turn the tracking off if the participants for instance do not want to be tracked for a specific period of time or if they – especially when using the smartphone – want to save battery life.

The data bank functions as the backend system which processes and stores the data received from the smartphone apps and the GPS-Loggers. The information stored include GPS locations and three-dimensional acceleration data. These data are processed in automatic trip, trip segment (i.e. legs), and mode detection algorithms implemented in the backend. Basically, a trip ends if a participant (or rather his/her mobile device) does not move for more than 15 minutes. Trips are split into different segments whenever participants change their mode of transport or when stationary periods (i.e. period of time with no movement up to 15 minutes) occur. Finally, the mode detection process relies on a set of fuzzy rules defined on the basis of test data from previous field tests (Sauerländer-Biebl et al. 2017).

Immediately (meaning a latency of only a few seconds) after the processing of the data in the backend system, the participants can view their trips on their smartphone apps or on the website of the MovingLab. For each participant a private account is created on the website that shows all of his/her trips with dates, times, transport modes, points of departure, destinations, etc. Both on the smartphone app as well as on their website account, participants can correct different parts of the trips detected such as changing transport modes or deleting trip segments. Entire new trips can also be added if the MovingLab System failed to capture a trip (e.g. due to missing GPS signals). The GPS-Loggers themselves do not provide any direct interaction possibilities for the participants and are primarily intended for use cases in which the participants are not familiar with smartphone technologies or in which not human beings but rather specific vehicles are to be tracked. However, users equipped with GPS-Loggers can also use the website to correct or add trips manually.

3.2. The field test

The main objective of the field test carried out in autumn 2018 was to evaluate the performance of the different technical devices applicable for data collection of the MovingLab System. For this purpose, four groups of participants were formed, three of them were equipped with one of the three technical devices provided by MovingLab (own smartphone, provided smartphone and logger), the fourth group was given a written travel diary. In this way the data collected via the three technical devices could not only be compared between each other but also against the traditional default method of data collection in transport research. Furthermore, all participants were asked to fill out two questionnaires on their sociodemographic background and their experiences with the technical system of the MovingLab.

In order to control for other potential influences on mobility behaviour, the participants of each of the four groups should be representative with regard to the population of the city of Berlin in terms of sex, age, car availability, and the area of residence (within/outside of the circular suburban train line and former East-/West-Berlin). The sampling and acquisition of the participants was carried out by the market research company Kantar. Individuals were recruited through door-to-door interviews. To avoid technological bias, individuals were asked to agree to participate in each of the four groups. Assignment to one of the four groups was made later on a random basis. After the formation of the four groups the participants were introduced to the respective devices for data collection that they were to use. For this purpose, information leaflets and video clips were developed and a short test run was performed in which the participants could practise tracking their trips and validating them with the MovingLab technology.

All participants were asked to track all of their trips over an entire week. To control for potential influences through differing weather conditions, the participants of each group were split in two halves, each of which was asked to track their trips in one of the following two weeks:

- Tuesday, 18th September 2018 – Monday, 24th September 2018
- Tuesday, 09th October 2018 – Monday, 15th October 2018

Thus, the field test covered seven days in September and seven days in October 2018. It should also be noted that some participants recorded trips which either both started and ended before the respective period of time or after it. All of these trips were deleted in order to ensure a better comparability between the different groups. Furthermore, trips which started before the respective period of time but ended within it were deleted. However, trips which started within the respective period of time but extended beyond it were kept in the data set.

3.3. Data basis

The 460 participants in the field test produced the following number of trips and trip segments (legs) per each of the data collection devices (hereafter *Split*):

Table 1. Data basis.

Split	Number of participants	Number of trips	Number of trip segments (legs)	Number of trips per person (average)
Own smartphone	98	1.697	5.444	17
Provided smartphone	119	2.271	8.795	19
GPS-Logger	112	2.021	5.842	18
Written travel diary	131	3.183	/ *	24
Sum	460	9.172	20.081	/

* The participants with the written travel diary were not asked to separate their trips into different legs as this would have increased the effort required considerably and in many cases it probably also would not have been practically feasible.

The number of participants ranges from 98 to 131 in the different splits. As can be seen, the number of trips collected varies considerably between the different splits. While the variation between the three technical splits can be explained by the different number of participants as the similar average numbers of trips per person indicate, the written travel diary still stands out with regard to both the total and the average number of trips per person. The potential reasons for this difference will be further elaborated in the results section.

The following table illustrates the sociodemographic distribution of the participants in the different splits:

Table 2. Overview of sociodemographic information.

Split	Sex	Age	Occupation	Driving license	Number of cars in household
Own smartphone	Male: 44%	18-29: 26%	Full-time: 41%	Yes: 80%	0: 34%
	Female: 53%	30-39: 22%	Part-time: 24%	No: 20%	1: 50%
	NA: 3%	40-49: 22%	Homemaker: 1%		2: 15%
		50-59: 12%	VT: 15%		3 or more: 1%
		60-69: 14%	Retired: 12%		
		70+: 5%	Other: 7%		
Provided smartphone	Male: 54%	18-29: 24%	Full-time: 47%	Yes: 79%	0: 40%
	Female: 44%	30-39: 11%	Part-time: 13%	No: 21%	1: 43%
	NA: 1%	40-49: 18%	Homemaker: 0%		2: 14%
		50-59: 20%	VT: 17%		3 or more: 3%
		60-69: 26%	Retired: 16%		
		70+: 1%	Other: 8%		
GPS-Logger	Male: 55%	18-29: 19%	Full-time: 47%	Yes: 78%	0: 38%
	Female: 44%	30-39: 19%	Part-time: 17%	No: 22%	1: 47%
	NA: 1%	40-49: 30%	Homemaker: 0%		2: 14%
		50-59: 12%	VT: 17%		3 or more: 1%
		60-69: 19%	Retired: 13%		
		70+: 2%	Other: 6%		
Written travel diary	Male: 48%	18-29: 19%	Full-time: 50%	Yes: 80%	0: 40%
	Female: 52%	30-39: 18%	Part-time: 16%	No: 20%	1: 48%
	NA: 0%	40-49: 26%	Homemaker: 2%		2: 13%
		50-59: 14%	VT: 8%		3 or more: 0%
		60-69: 18%	Retired: 20%		
		70+: 6%	Other: 4%		

VT: Vocational Training

NA: Not answered

Except for the possession of a driving license where the distribution is almost identical in all of the splits, there are some minor differences in the other categories. While, for instance, male participants constitute the majority in split two and three, they are the minority in the splits one and four. It is also noticeable that there are significantly more participants of the age of 70 years and more in the split of the written travel diary. Correspondingly, also the share of the retired people is higher in this split than in the three other ones. Finally, it is also noticeable that slightly fewer cars seem to be available in the households of the participants with the written travel diaries compared to the participants in the three technical splits. Further statistics of our sample such as the modal split and the average trip length per method of data collection can be found in the tables A.1 and A.2 in the Appendix.

It is noteworthy that for taking part in the field test potential participants had to agree upon using a technical device such as a smartphone or a GPS-Logger to record their trips. Those who agreed on this condition were randomly assigned to one of the three technical splits. People who did not agree on this condition were sorted out and did not take part in the field test. This can have resulted in a certain bias in the sample in different ways. First, it can have resulted in more tech-savvy people in the sample than in the average population in Berlin as only those people participated who feel comfortable in handling smartphones or GPS-Loggers. Second, it is well known that some people have concerns about privacy and data protection when it comes to recording their mobility with technical devices on the basis of GPS signals. Therefore, it is conceivable that the people who took part in the field test are on average less concerned about privacy and data protection than the overall population in Berlin.

While concerns about privacy and data collection should not affect the users' assessments of the technical performance of the data collection system, the potentially above average technical skills of the participants might have had an impact on the results of this study. It is, for instances, conceivable that more tech-savvy people provide better assessments of the functionalities of the different technical systems of data collection than the average population as they find it easier to handle the technical devices. Therefore, this issue should be kept in mind when interpreting the results in the following section and it will be taken up again in the conclusions of this paper.

4. Results

We first provide a descriptive overview of the assessments of the different components of the MovingLab before we analyse whether these assessments correlate with the actual results of the data collection.

4.1. Descriptive analysis

After the field test was completed, the participants with the three technical devices were asked to fill in a questionnaire on the different technical aspects of the MovingLab. Altogether there were more than 40 questions and remarks on issues such as battery consumption, experiences with the app and the website, the quality of mode and trip detection, etc. which could be answered or assessed on a five-level Likert scale. It would go beyond the scope of this paper to provide a complete review of the survey results for all of the questions.

Table 3. General experiences with the technical devices.

Question/Issue	Own smartphone	Provided smartphone	GPS-Logger
Battery consumption was very low.	2.89 (1.32)	2.00 (1.22)	2.03 (1.05)
The download of the app did not cause any problems.	1.58 (1.06)	NA	NA
Mobile data consumption was very high.	3.36 (1.12)	NA	NA
The app crashed frequently.	3.53 (1.39)	3.78 (1.25)	NA
It was easy to remember to turn on the app/ the GPS-Logger before a trip.	3.03 (1.19)	2.39 (1.16)	2.27 (1.20)
The GPS-Logger worked without any problems.	NA	NA	3.18 (1.21)
The LED provided enough information on the status of the GPS-Logger.	NA	NA	2.33 (1.22)
Using the on-/off-button was easy.	NA	NA	2.09 (1.30)
The charging time was too long.	NA	NA	4.01 (1.12)

The table displays the mean values and the standard deviation (in parentheses) of the following Likert scale:
1 – Strongly agree
2 – Agree

3 – Neither agree nor disagree
4 – Disagree
5 – Strongly disagree
NA – Not applicable

Battery consumption seemed to be less of an issue when the participants were provided with a smartphone or a GPS-Logger than when they used their own smartphone. This might be due to the fact that the provided smartphones and GPS-Loggers came with brand-new batteries, while (presumably) some of the participants' own smartphones were already in use for a couple of months or years. Another possible explanation is that the provided smartphone was only used for the actual tracking, while the participants' own smartphones were probably also used for personal purposes in addition to the actual tracking. The download of the app did not cause any issues for most of the participants. However, there appears to be some potential for improvement with regard to mobile data consumption, the stability of the app, and the reliability of the GPS-Logger.

The most interesting aspect of Table 3 is the apparent difference between the own smartphone and the provided items with regard to how easy it was for the participants to remember to turn the tracking on before a trip. The results indicate that it is easier for participants to remember to turn on the tracking if they carry an extra device with them. Possibly they also feel socially more obliged to follow the instructions given to them when they are provided with an additional technical device.

This assumption is in part supported by the fact that 81% of the participants provided with a smartphone answered that they always had the device with them in the reporting week, while further 16% answered that they very often took the device with them. In contrast, only 47% and 28% of the participants equipped with GPS-Loggers answered that they always or very often took the device with them. A bit surprisingly also only 61% and 20% of the participants with their own smartphones answered that they always or very often took their smartphone with them in the reporting week.

The numbers in the following two tables illustrate that the majority of the participants did not encounter any larger problems with the app or the website:

Table 4. Assessment of the app.

Question/Issue	Own smartphone	Provided smartphone	GPS-Logger
The menus and options had distinct and unambiguous names.	2.13 (0.89)	2.09 (0.81)	NA
It took some time to get acquainted with the usage.	3.25 (1.27)	3.21 (1.09)	NA
A proper usage requires remembering a lot of details.	3.88 (1.07)	3.84 (1.04)	NA
The usage is intuitive.	2.65 (0.97)	2.48 (1.05)	NA
The app reacts slowly.	2.96 (1.34)	3.16 (1.37)	NA

The table displays the mean values and the standard deviation (in parentheses) of the following Likert scale:
1 – Strongly agree
2 – Agree
3 – Neither agree nor disagree
4 – Disagree
5 – Strongly disagree
NA – Not applicable

The names of the menus and options, the time to get acquainted with the usage, as well as the usage itself are assessed rather positively with regard to both the app and the website. There appears to be some potential for improvement considering the reaction time of the app. However, the more interesting aspect is that the participants with the GPS-Loggers on average show better assessments of the website than the participants with the smartphones. A potential explanation is that with 71% and 56% the majority of the participants with their own and a provided smartphone used a smartphone for the modification of their trips, while respectively 50% and 34% of the participants with the GPS-Loggers said that they used a laptop or a personal computer for the modification of their trips. Hence the different assessments might rely on different screen sizes and the use of a keyboard and a mouse instead of a touchscreen.

Table 5. Assessment of the website.

Question/Issue	Own smartphone	Provided smartphone	GPS-Logger
The menus and options had distinct and unambiguous names.	2.04 (0.89)	1.94 (0.65)	2.00 (0.81)
It took some time to get acquainted with the usage.	3.38 (1.20)	3.34 (0.97)	3.73 (1.15)
A proper usage requires remembering a lot of details.	3.62 (1.05)	3.80 (1.02)	4.23 (0.88)
The usage is intuitive.	2.55 (0.97)	2.35 (0.84)	2.07 (0.94)
Trips can be modified without any problems.	2.96 (1.06)	2.67 (1.19)	2.16 (1.05)
Trips can easily be added.	2.66 (1.27)	2.45 (1.19)	1.94 (0.99)

The table displays the mean values and the standard deviation (in parentheses) of the following Likert scale:
1 – Strongly agree
2 – Agree
3 – Neither agree nor disagree
4 – Disagree
5 – Strongly disagree
NA – Not applicable

Finally, the table below shows our participants' assessments of the trip and mode detection of the MovingLab system:

Table 6. Assessment of the trip and mode detection.

Question/Issue	Own smartphone	Provided smartphone	GPS-Logger
Trips were not detected although the tracking was turned on.	3.02 (1.40)	2.86 (1.21)	2.46 (1.25)
Trips were split in too many segments (legs).	2.86 (1.41)	2.54 (1.17)	3.02 (1.29)
There were trips with missing segments (legs).	3.50 (1.32)	3.59 (1.19)	2.77 (1.25)
It took a long time before the trips were shown in the app or on the website.	2.87 (1.17)	2.91 (1.28)	3.50 (1.36)
Trip mode was not detected correctly.	2.53 (1.25)	2.28 (1.10)	2.49 (1.16)
The start and/or end time of a trip were not detected correctly.	3.77 (1.29)	3.95 (1.21)	3.79 (1.27)
The start and/or end point of a trip were not detected correctly.	3.66 (1.25)	3.90 (1.19)	3.43 (1.36)
The app or the website showed trips that were not conducted.	3.98 (1.23)	4.01 (1.11)	4.00 (1.21)

The table displays the mean values and the standard deviation (in parentheses) of the following Likert scale:
1 – Very often
2 – Rather often
3 – Rather seldom
4 – Seldom
5 – Never
NA – Not applicable

While the assessments of the participants with their own and with a provided smartphone are often quite close together, the assessments of the participants with GPS-Loggers are significantly more negative or positive with regard to some aspects. While more of the participants with GPS-Loggers claim that trips were not detected although the tracking was turned on and that detected trips had missing segments (legs), less of them complain about too many segments in the detected trips and about the time it took before a trip was shown on the website. The latter might be explained by the fact that most of the participants with GPS-Loggers used a laptop or personal computer to look up their trips, while most of the other participants with technical devices used their smartphones. Thus it could be assumed that more of the participants with GPS-Loggers waited until they were back at home before they reviewed their trips, while the other participants immediately looked up their trips on their smartphones.

In general it can also be seen that there is some potential for improvement with regard to mode, trip and trip segment detection. However, the detection of the start and/or end times and points of the trips seemed to work rather well. In the next section we will explore whether and in how far the different assessments of the participants also relate to actual outputs generated by the MovingLab.

4.2. Performance analysis with regard to user assessments

The table below summarizes the relationships between our participants' assessments of the trip and mode detection of the MovingLab system and the system's actual performance: All variables were recoded so that higher values indicated a higher agreement with an issue which means that positive correlations indicate that a higher agreement goes hand in hand with higher numbers in the different performance indicators.

Table 7. Correlation between user assessments and performance indicators.

Question/Issue	Performance indicator	Own smartphone		Provided smartphone		GPS-Logger	
		S.rho	K.tau	S.rho	K.tau	S.rho	K.tau
It was easy to remember to turn on the app/ the GPS-Logger before a trip.	Number of trips recorded automatically per person.	No cor.	No cor.	No cor.	No cor.	No cor.	No cor.
Frequency of carrying the tracking device in the reporting week.	Number of trips recorded automatically per person.	No cor.	No cor.	0.37 *T	0.30 *T	0.31 *	0.24 *
Trips were not detected although the tracking was turned on.	Number of trips recorded automatically per person.	No cor.	No cor.	-0.34 *T	-0.27 *T	-0.44 *T	-0.34 *T
Battery consumption was very low.	Number of days on which trips were reported per person.	No cor.	No cor.	No cor.	No cor.	No cor.	No cor.
Battery consumption was very low.	Number of trips recorded automatically per person.	No cor.	No cor.	No cor.	No cor.	No cor.	No cor.
Mobile data consumption was very high.	Number of trips recorded automatically per person.	No cor.	No cor.	NA	NA	NA	NA
The app crashed frequently.	Number of trips recorded automatically per person.	-0.29 *T	-0.38 *T	No cor.	No cor.	NA	NA
The GPS-Logger worked without any problems.	Number of trips recorded automatically per person.	NA	NA	NA	NA	0.48 *T	0.38 *T
Trips can easily be added via the website.	Number of trips added manually per person.	No cor.	No cor.	No cor.	No cor.	0.26	0.21
A proper usage of the app requires remembering a lot of details.	Number of trips validated per person.	-0.15	-0.11	-0.18	-0.13	NA	NA
A proper usage of the website requires remembering a lot of details.	Number of trips validated per person.	No cor.	No cor.	-0.23	-0.18	-0.23	-0.19
The usage of the app is intuitive.	Number of trips validated per person.	No cor.	No cor.	0.20	0.15	NA	NA
The usage of the website is intuitive.	Number of trips validated per person.	0.26	0.20	0.26	0.21	No cor.	No cor.

Trips can be modified without any problems.	Number of trips validated per person.	No cor.	No cor.	No cor.	No cor.	No cor.	No cor.
Trips can be modified without any problems.	Number of trips modified per person.	No cor.	No cor.	No cor.	No cor.	No cor.	No cor.
Trip mode was not detected correctly.	Number of trips modified per person.	-0.21	-0.17	-0.27	-0.21	0.22	0.17

S. rho stands for Spearmans rho
K. tau stands for Kendalls tau
No cor. means that the correlation measure showed a value lower than 0,10.
* Significant on a level of 0,05 according to One Way Anova
[T] Significant differences on a level of 0,05 between specific categories according to Tukey's Honest Significant Difference test
NA – Not applicable

One of the strongest and significant correlations exists between the reliability of the trip detection system and the number of trips recorded automatically. Unfortunately, our participants provided with smartphones or GPS-Loggers appear to be correct in their judgements: If in their view it frequently occurred that trips were not detected although the tracking was turned on, then, indeed, less trips were recorded automatically. This result does not differ significantly among various groups of participants in terms of sex, age, or the regular use of different means of transport. With regard to the latter, this is rather positive news as it could have been expected that the lack of a GPS-Signal in underground public transport stations systematically leads to less trips being detected for the public transport mode.

There are two potential explanations for trips not being detected although the tracking was turned on. First, the smartphones or GPS-Loggers might not have received any GPS-Signal during the trips and thus they also did not send any GPS-Coordinates to the backend of the MovingLab System. Second, GPS-coordinates were not processed correctly by the trip detecting algorithm in the backend of the MovingLab System. In any case this illustrates that there is room for improvement in our automatic trip detection system.

This finding is further supported by the significant correlations between the frequency of carrying the device in the reporting week and the number of trips recorded automatically. The assessments of our participants provided with smartphones or GPS-Loggers seem to be valid: More trips were recorded automatically for participants who often had their device with them in the reporting week. In addition, significantly more trips were recorded automatically for participants who agreed with the statement that their GPS-Logger worked without any problems. However, significant differences with regard to sex, age, or the regular use of different means of transport could not be detected for any of these two findings.

This all speaks for the validity of the assessments of our participants. It also makes sense that only for the participants with GPS-Loggers correlations between the statement that trips can easily be added via the website and the actual number of trips added manually per person were found as the website constitutes the only platform for the interaction with our system for them, while the other participants could also use the smartphones to add or alter trips.

It is positive though, that we did not find any of these correlations among our participants who used their own smartphones. However, if participants with their own smartphones often experienced problems with the app, this appears to have significantly reduced the number of trips recorded automatically. This illustrates that further work on the stability of the app is needed up to the point at which the technical functionality does not affect the empirical results anymore. Again, this is an important finding that we were only able to produce through including the user assessments in the interpretation of the actual results of our technical system.

These results also provide an explanation for the difference between the average number of trips recorded per person with the technical devices and the written travel diary outlined in Table 1. The users' assessments in combination with the number of trips automatically recorded indicate that some technical parts of the MovingLab System do not work good enough, yet. For participants who used their own smartphone the instability of the app contributed to significantly less trips being recorded, while for participants with GPS-Loggers or a provided smartphone the MovingLab System too often did not record any trips although the tracking was turned on. Hence it is some technical flaws that explain the lower number of trips recorded per person on average with the technical devices than with the written travel diary.

Besides these findings that illustrate where the MovingLab System still needs improvement, Table 7 also provides some positive news. Battery or mobile data consumption do not appear to be a bigger problem with regard to the number of trips actually recorded or the number of days on which trips were recorded. Hence, even if some users found battery consumption to be high, this did not affect the empirical results. In addition, it appears that the users did not have any problems with modifying their trips or with remembering turning the tracking on before starting a trip. At least, none of this resulted in less trips being recorded, validated, or modified.

Finally, there are indications that users who found the usage of the app or the website intuitive and who did not think that the usage requires remembering a lot of details also validated more of their trips. Yet, some of the correlations found are not very strong and none of them is significant. Also the correlations found between the rate of agreement to the statement that the trip mode was not detected correctly and the number of trips modified are not significant and thus not interpreted any further.

5. Conclusions

Our paper has highlighted the importance of including the users in the assessment and development of GPS-based methods of data collection. By looking at the users' experiences we revealed the potential and need for improvement with regard to various aspects of our technical system that could not have been revealed by relying on the technical measurement data alone.

In particular, our findings have shown that our system indeed recorded fewer trips automatically for those users who complained about the automatic trip detection or the stability of the app or the GPS-Loggers provided. This problem would not have been detected if we had looked on the outputs of our technical system alone. In fact, without a look at the users' assessments, there might even have been misinterpretations of the empirical results and fewer trips recorded for some people due to technical issues might have falsely been ascribed to their sociodemographic features. Therefore, an inclusion of the users in the evaluation and development of GPS-based methods of data collection cannot only reveal the potential for improvement in the technical system but might also be necessary to rule out technical malfunctioning as an explanatory factor in the empirical results.

These results even gain in relevance if it is considered that tech-savvy people might be overrepresented in the sample because agreeing to recording trips with technical devices was a precondition for participating in the field test. First, if tech-savvy people are overrepresented in the sample, then the validity of the users' assessments should be quite high as the participants should be familiar with the functionalities of the technical devices. Second, a potential overrepresentation of tech-savvy people in the sample also means that the assessments of the functionalities of the technical devices might be even worse if made by the overall population of the city of Berlin. All of this confirm the already drawn conclusion that there still is some potential for improvement of the MovingLab system.

In addition, our look at the users' assessments has also shed light on which technical problems might not have a negative impact on the empirical results. Although, for instance, it is well-known that GPS-based methods of data collection can significantly reduce battery life on some smartphones, people who complained about reductions in battery life did not have fewer trips automatically recorded than others users. A potential explanation for this is that people learned to turn on the tracking before they started a trip and turned it off after they finished the trip. In fact, even people who stated that it is difficult for them to remember to turn on the tracking before they start a trip did not have fewer trips automatically recorded than other users.

Appendix A. Presentation of the results in detail

Table A.1. Modal split.

Split	Walking	Bicycle	Car	Public Transport
Own smartphone	25%	9%	36%	30%
Provided smartphone	25%	10%	26%	40%
GPS-Logger	23%	10%	40%	27%
Written travel diary	28%	12%	28%	31%

Table A.2. Average trip length in kilometres.

Split	Walking	Bicycle	Car	Public Transport	All trips
Own smartphone	1.10	3.73	16.22	13.31	10.43
Provided smartphone	0.99	3.73	11.74	9.84	7.41
GPS-Logger	1.03	4.43	11.24	9.08	7.79
Written travel diary	1.60	4.53	15.81	9.99	8.62

Table A.3. Answers to the general experiences with the technical devices in detail (row percentages per split).

Question/Issue	Split	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Do not know
Battery consumption was very low.	Own smartphone	16%	28%	22%	17%	16%	2%
	Provided smartphone	46%	26%	14%	5%	7%	2%
	GPS-Logger	32%	40%	8%	10%	2%	9%
The download of the app did not cause any problems.	Own smartphone	71%	11%	10%	5%	3%	0%
	Provided smartphone	NA	NA	NA	NA	NA	NA
	GPS-Logger	NA	NA	NA	NA	NA	NA
Mobile data consumption was very high.	Own smartphone	5%	10%	23%	23%	12%	27%
	Provided smartphone	NA	NA	NA	NA	NA	NA
	GPS-Logger	NA	NA	NA	NA	NA	NA
The app crashed frequently.	Own smartphone	8%	21%	16%	18%	36%	0%
	Provided smartphone	3%	18%	18%	20%	41%	0%
	GPS-Logger	NA	NA	NA	NA	NA	NA
It was easy to remember to turn on the app/ the GPS-Logger before a trip.	Own smartphone	12%	22%	30%	25%	12%	0%
	Provided smartphone	26%	32%	24%	12%	6%	0%
	GPS-Logger	34%	26%	24%	10%	6%	1%
The GPS-Logger worked without any problems.	Own smartphone	NA	NA	NA	NA	NA	NA
	Provided smartphone	NA	NA	NA	NA	NA	NA
	GPS-Logger	9%	24%	23%	30%	15%	0%
The LED provided enough information on the status of the GPS-Logger.	Own smartphone	NA	NA	NA	NA	NA	NA
	Provided smartphone	NA	NA	NA	NA	NA	NA
	GPS-Logger	32%	29%	20%	14%	6%	0%
Using the on-/off-button was easy.	Own smartphone	NA	NA	NA	NA	NA	NA
	Provided smartphone	NA	NA	NA	NA	NA	NA
	GPS-Logger	48%	20%	14%	11%	7%	0%
The charging time was too long.	Own smartphone	NA	NA	NA	NA	NA	NA
	Provided smartphone	NA	NA	NA	NA	NA	NA
	GPS-Logger	4%	7%	7%	34%	37%	11%

NA – Not applicable

Table A.4. Answers to the assessment of the app in detail (row percentages per split).

Question/Issue	Split	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Do not know
The menus and options had distinct and unambiguous names.	Own smartphone	28%	35%	32%	5%	0%	0%
	Provided smartphone	21%	56%	16%	7%	0%	0%
	GPS-Logger	NA	NA	NA	NA	NA	NA
It took some time to get acquainted with the usage.	Own smartphone	5%	33%	15%	25%	22%	0%
	Provided smartphone	4%	28%	24%	33%	12	0%
	GPS-Logger	NA	NA	NA	NA	NA	NA
A proper usage requires remembering a lot of details.	Own smartphone	2%	12%	16%	37%	33%	0%
	Provided smartphone	2%	10%	19%	39%	39%	0%
	GPS-Logger	NA	NA	NA	NA	NA	NA
The usage is intuitive.	Own smartphone	9%	40%	31%	16%	3%	2%
	Provided smartphone	15%	44%	20%	16%	3%	1%
	GPS-Logger	NA	NA	NA	NA	NA	NA
The app reacts slowly.	Own smartphone	14%	28%	25%	14%	19%	0%
	Provided smartphone	11%	29%	15%	21%	23%	2%
	GPS-Logger	NA	NA	NA	NA	NA	NA

NA – Not applicable

Table A.5. Answers to the assessment of the website in detail (row percentages per split).

Question/Issue	Split	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Do not know
The menus and options had distinct and unambiguous names.	Own smartphone	26%	50%	19%	0%	3%	2%
	Provided smartphone	23%	60%	14%	1%	0%	2%
	GPS-Logger	27%	51%	16%	5%	0%	1%
It took some time to get acquainted with the usage.	Own smartphone	3%	26%%	19%	28%	21%	3%
	Provided smartphone	1%	20%	32%	33%	11%	3%
	GPS-Logger	2%	19%	11%	38%	30%	0%
A proper usage requires remembering a lot of details.	Own smartphone	6%	6%	28%	39%	19%	4%
	Provided smartphone	1%	11%	22%	35%	28%	2%
	GPS-Logger	0%	6%	12%	36%	47%	0%
The usage is intuitive.	Own smartphone	11%	40%	25%	16%	2%	7%
	Provided smartphone	12%	45%	32%	6%	1%	2%
	GPS-Logger	27%	47%	16%	6%	2%	2%
Trips can be modified without any problems.	Own smartphone	7%	28%	26%	28%	5%	5%
	Provided smartphone	17%	31%	21%	21%	6%	4%
	GPS-Logger	28%	44%	15%	10%	3%	0%
Trips can easily be added.	Own smartphone	16%	32%	18%	12%	11%	12%
	Provided smartphone	20%	32%	20%	9%	7%	13%
	GPS-Logger	40%	32%	17%	7%	1%	3%

NA – Not applicable

Table A.6. Answers to the assessment of the trip and mode detection in detail (row percentages per split).

Question/Issue	Split	Very often	Rather often	Rather seldom	Seldom	Never	Do not know
Trips were not detected although the tracking was turned on.	Own smartphone	19%	19%	22%	20%	19%	2%
	Provided smartphone	11%	35%	23%	16%	13%	2%
	GPS-Logger	25%	33%	20%	12%	9%	1%
Trips were split in too many segments (legs).	Own smartphone	19%	25%	26%	7%	21%	2%
	Provided smartphone	18%	37%	21%	15%	7%	2%
	GPS-Logger	9%	33%	19%	17%	18%	3%
There were trips with missing segments (legs).	Own smartphone	7%	21%	16%	24%	29%	3%
	Provided smartphone	7%	10%	26%	30%	26%	1%
	GPS-Logger	16%	29%	20%	19%	10%	7%
It took a long time before the trips were shown in the app or on the website.	Own smartphone	12%	24%	29%	19%	9%%	9%
	Provided smartphone	12%	29%	17%	19%	13%	11%
	GPS-Logger	6%	19%	10%	18%	26%	21%
Trip mode was not detected correctly.	Own smartphone	20%	37%	19%	12%	10%	2%
	Provided smartphone	26%	37%	19%	12%	4%	2%
	GPS-Logger	22%	32%	20%	20%	3%	2%
The start and/or end time of a trip were not detected correctly.	Own smartphone	6%	12%	19%	16%	39%	9%
	Provided smartphone	4%	11%	19%	18%	46%	2%
	GPS-Logger	6%	13%	14%	24%	38%	4%
The start and/or end point of a trip were not detected correctly.	Own smartphone	5%	14%	24%	19%	35%	3%
	Provided smartphone	2%	13%	21%	18%	45%	1%
	GPS-Logger	9%	18%	21%	18%	30%	3%
The app or the website showed trips that were not conducted.	Own smartphone	3%	10%	20%	12%	49%	5%
	Provided smartphone	2%	8%	26%	14%	48%	2%
	GPS-Logger	3%	9%	22%	12%	51%	3%

NA – Not applicable

References

Allström, Andreas, Ida Kristoffersson, and Yusak Susilo. 2017. "Smartphone based travel diary collection: experiences from a field trial in Stockholm." *Transportation Research Procedia* no. 26:32-38. doi: https://doi.org/10.1016/j.trpro.2017.07.006.

Bonnel, Patrick, and Marcela A. Munizaga. 2018. "Transport survey methods - in the era of big data facing new and old challenges." *Transportation Research Procedia* no. 32:1-15. doi: https://doi.org/10.1016/j.trpro.2018.10.001.

Chapleau, Robert, Philippe Gaudette, and Tim Spurr. 2018. "Strict and Deep Comparison of Revealed Transit Trip Structure between Computer-Assisted Telephone Interview Household Travel Survey and Smart Cards." *Transportation Research Record* no. 2672 (42):13-22. doi: 10.1177/0361198118758297.

Erhardt, Gregory D., and Louis Rizzo. 2018. "Evaluating the biases and sample size implications of multi-day GPS-enabled household travel surveys." *Transportation Research Procedia* no. 32:279-290. doi: https://doi.org/10.1016/j.trpro.2018.10.051.

Gong, Lei, Toshiyuki Yamamoto, and Takayuki Morikawa. 2018. "Identification of activity stop locations in GPS trajectories by DBSCAN-TE method combined with support vector machines." *Transportation Research*

Procedia no. 32:146-154. doi: https://doi.org/10.1016/j.trpro.2018.10.028.

Greaves, Stephen, Adrian Ellison, Richard Ellison, Dean Rance, Chris Standen, Chris Rissel, and Melanie Crane. 2015. "A Web-Based Diary and Companion Smartphone app for Travel/Activity Surveys." *Transportation Research Procedia* no. 11:297-310. doi: https://doi.org/10.1016/j.trpro.2015.12.026.

Itsubo, Shinji, and Eiji Hato. 2006. Effectiveness of Household Travel Survey Using GPS-Equipped Cell Phones and Web Diary: Comparative Study with Paper-Based Travel Survey. Paper read at Transportation Research Board 85th Annual Meeting, at Washington DC, United States.

Kang, Mingyu, Anne V. Moudon, Philip M. Hurvitz, and Brian E. Saelens. 2018. "Capturing fine-scale travel behaviors: a comparative analysis between personal activity location measurement system (PALMS) and travel diary." *International Journal of Health Geographics* no. 17 (1):40. doi: 10.1186/s12942-018-0161-9.

Li, Wenxiang, Ye Li, Xuegang Ban, Haopeng Deng, Hanyu Shu, and Dongcan Xie. 2018. "Exploring the Relationships between the Non-Work Trip Frequency and Accessibility Based on Mobile Phone Data." *Transportation Research Record* no. 2672 (42):91-102. doi: 10.1177/0361198118774170.

Montini, Lara, Sebastian Prost, Johann Schrammel, Nadine Rieser-Schüssler, and Kay W. Axhausen. 2015. "Comparison of Travel Diaries Generated from Smartphone Data and Dedicated GPS Devices." *Transportation Research Procedia* no. 11:227-241. doi: https://doi.org/10.1016/j.trpro.2015.12.020.

Murakami, E., and D. P. Wagner. 1999. "Can using global positioning system (GPS) improve trip reporting?" *Transportation Research Part C: Emerging Technologies* no. 7 (2):149-165. doi: https://doi.org/10.1016/S0968-090X(99)00017-0.

Necula, Emilian. 2015. "Analyzing Traffic Patterns on Street Segments Based on GPS Data Using R." *Transportation Research Procedia* no. 10:276-285. doi: https://doi.org/10.1016/j.trpro.2015.09.077.

Neven, An, Ine De Schutter, Geert Wets, Peter Feys, and Davy Janssens. 2018. "Data Quality of Travel Behavior Studies: Factors Influencing the Reporting Rate of Self-Reported and GPS-Recorded Trips in Persons with Disabilities." *Transportation Research Record* no. 2672 (8):662-674. doi: 10.1177/0361198118772952.

Ohmori, Nobuaki, Noboru Harata, and Morimichi Nakazato. 2005. "GPS mobile phone-based activity diary survey." *Proceedings of the Eastern Asia Society for Transportation Studies* no. 5.

Patterson, Zachary, and Kyle Fitzsimmons. 2016. "DataMobile: Smartphone Travel Survey Experiment." *Transportation Research Record* no. 2594 (1):35-43. doi: 10.3141/2594-07.

Prelipcean, Adrian C., and Toshiyuki Yamamoto. 2018. "Workshop Synthesis: New developments in travel diary collection systems based on smartphones and GPS receivers." *Transportation Research Procedia* no. 32:119-125. doi: https://doi.org/10.1016/j.trpro.2018.10.023.

Richard, Olivier, and Mathieu Rabaud. 2018. "French household travel survey: The next generation." *Transportation Research Procedia* no. 32:383-393. doi: https://doi.org/10.1016/j.trpro.2018.10.065.

Sadeghvaziri, Eazaz, Mario B. Rojas, and Xia Jin. 2016. "Exploring the Potential of Mobile Phone Data in Travel Pattern Analysis." *Transportation Research Record* no. 2594 (1):27-34. doi: 10.3141/2594-04.

Safi, Hamid, Behrang Assemi, Mahmoud Mesbah Namini, and Luis Ferreira. 2016. "A trip-detection method for smartphone-assisted travel data collection." In *Proceedings of the Transportation Research Board (TRB) 95th Annual Meeting*, 1-18. United States of America: Transportation Research Board (TRB).

Sauerländer-Biebl, Anke, Elmar Brockfeld, David Suske, and Eric Melde. 2017. "Evaluation of a transport mode detection using fuzzy rules." *Transportation Research Procedia* no. 25:591-602. doi: https://doi.org/10.1016/j.trpro.2017.05.444.

Shankari, K., Mohamed Amine Bouzaghrane, Samuel M. Maurer, Paul Waddell, David E. Culler, and Randy H. Katz. 2018. "e-mission: An Open-Source, Smartphone Platform for Collecting Human Travel Data." *Transportation Research Record* no. 2672 (42):1-12. doi: 10.1177/0361198118770167.

Sobhani, Anae, Hamzeh Alizadeh Aliabadi, and Bilal Farooq. 2019. "Metropolis-Hasting based Expanded Path Size Logit model for cyclists' route choice using GPS data." *International Journal of Transportation Science and Technology* no. 8 (2):161-175. doi: https://doi.org/10.1016/j.ijtst.2018.11.002.

Stopher, Peter R., Vivian Daigler, and Sarah Griffith. 2018. "Smartphone app versus GPS Logger: A comparative study." *Transportation Research Procedia* no. 32:135-145. doi: https://doi.org/10.1016/j.trpro.2018.10.026.

Stopher, Peter R., Li Shen, Wen Liu, and Asif Ahmed. 2015. "The Challenge of Obtaining Ground Truth for GPS Processing." *Transportation Research Procedia* no. 11:206-217. doi: https://doi.org/10.1016/j.trpro.2015.12.018.

Trépanier, Martin, and Toshiyuki Yamamoto. 2015. "Workshop Synthesis: System Based Passive Data Streams

Systems; Smart Cards, Phone Data, GPS." *Transportation Research Procedia* no. 11:340-349. doi: https://doi.org/10.1016/j.trpro.2015.12.029.

Wagner, D. P. 1997. Global Positioning SystemsforPersonal Travel Surveys. Lexington Area Travel Data Collection Test. In *Final Report for OHIM, OTA, and FHWA*: BattelleTransportation Division.