Evaluation Methods and Results for Intermodal Mobility Applications in Public Transport

Ulrike Stopka(✉), Katrin Fischer, and René Pessier

Technische Universität Dresden, Dresden, Germany
ulrike.stopka@tu-dresden.de

Abstract. Against the background of changing social values towards resource efficiency and environmental benefits today's mobility is lived less as an alternative choice between different modes of transport, but rather in their intermodal combination during a trip. The increasing diversity of mobility services is combined by the users under very pragmatic aspects in order to get from A to B as fast, reliable, convenient and cost effective as possible.

Within the scope of the research project "Dynamic Seamless Mobility Information" (DYNAMO) conducted by a consortium of different public transportation, service developing and IT companies as well as scientific institutions such as the Technische Universität Dresden a comprehensive prototype of an intermodal mobility application was developed between 2013 and 2015. The scientific monitoring and evaluation of this process will be presented in the following article.

Keywords: Mobility information · Mobile application · Intermodal services · Public transport · Evaluation methods · User requirements · Operating concepts

1 Project Presentation

The target of the research project DYNAMO was the development and prototypical implementation of dynamic information services to support the travelers before and during their journey. The user is in the focus of the project and provides the basis for the "Design-for-all" concept. The scientific evaluation of the project was divided into three phases, which were conducted at the beginning, during and at the end of the application development (Fig. 1):

Within the scope of the project, the following basic services or core areas had to be designed and integrated into the application:

- dynamic trip guidance (real time information and identification of route deviations)
- indoor positioning (enables navigation in buildings)
- social media (integration of modern interaction channels for networking with passengers)
- routing and navigation (reliable, user-friendly and continuous routing)
- accessibility without barriers (barrier-free routing)

M. Kurosu (Ed.): HCI 2016, Part III, LNCS 9733, pp. 343–354, 2016.
DOI: 10.1007/978-3-319-39513-5_32

Fig. 1. Scientific evaluation steps in the DYNAMO research project

- intermodal crosslinking (intermodal information including car-/bike sharing, ride sharing and ride selling, taxi)

The development of the intermodal application and the surveys took place in two test regions: in the area of the Rhine-Main Transport Association and the area of the Munich Transport Association.

2 Goals and Methods of the Different Evaluation Phases

2.1 Analysis of User Requirements

The aim of this upstream stage was to elicit requirements for mobility applications with regard to the above mentioned basic services and core topics by using scientific evaluation methods. After an intensive literature review and the development of general requirement catalogs, focus group and individual interviews for the qualitative determination of user requirements for the DYNAMO prototype were conducted. In total 14 guideline-based group or individual interviews with focus on dynamic trip guidance (16 subjects), routing and navigation (13 subjects), intermodal crosslinking (20 subjects) and social media (15 subjects) were performed. The results of the focus group discussions served as a basis for the following standardized online survey. With the help of this survey the insights gained by the qualitative methods (focus group and individual interviews) were supposed to be validated quantitatively.

The questionnaire was distributed electronically via the student and staff mailing center of the Technische Universität Dresden. With 1,985 completed questionnaires an utilizable response rate of 5.3 % was achieved. The features and services found out in the previous interviews were divided and grouped in a reasonable manner and the subjects had to evaluate them by means of a rating scale (5-point scale: very needed … not required). Thereby different scenarios (sections of the travel chain) have been predefined. The sample was made up as follows: 51 % of respondents were male and 49 % female. 73 % of respondents were aged between 16 and 25 years and a quarter between of 26 and 35. 87 % of the subjects stated that they currently pass a training, apprenticeship or study. 6 % were fully employed and 4 % partially employed. The sample cannot be regarded as representative. However, it allows to identify trends in general mobility behavior.

2.2 Operating Concept Evaluation

The aim of the second assessment phase was the evaluation of eight operating concepts with regard to their acceptance by potential users. Operating concepts comprise the operation and interaction features of software users [3]. This second step was very helpful for the iterative approach in the research project, because the results of the survey were sent as feedback to the developers of the application prototype.

1,884 oral interviews in November and December 2014 both at public transport stops and in vehicles of the Rhine-Main Transport Association provided the data base for this evaluation. To participate, the subjects had to use at least one mobility application on their smartphone. The standardized questionnaire was made up of two parts: a first section on mobility behavior, attitudes and personal information and a second section for the actual evaluation of the operating concepts. Following operating concepts had to been evaluated: Menu design, route options overview, route search, route details, social media, modal choice, indoor navigation and map display. The subjects were due to assess their acceptance by evaluating three different screenshots for each operating concept shown on a tablet. The range of the rating scale went from 1 (poor operating features) to 7 (very helpful operating features). Beyond that more in-depth questions were asked especially interesting from a developer's point of view.

2.3 Overall Evaluation

The overall evaluation of the DYNAMO prototype application has the following investigation targets:

1. evaluation of the usability on the functional and application level
2. evaluation of the user acceptance by:
   (a) the intensity of use and their changes
   (b) the willingness to recommend the application
3. elicitation of the potential for improvement of the application
4. assessment of the impacts with regard to mobility behavior by using the application and further mobility service supply options

The usability of an application is a hypothetical feature. After DIN ISO 9241-11 it is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". For this analysis, the above mentioned three criteria effectiveness, efficiency and satisfaction should be regarded both on the functional and application level. The definitions of these three terms can be found in Table 1.

Table 1. Usability criteria [1]

Criterion	Definition
Effectiveness	Accuracy and completeness with which selected users can achieve defined goals in specific environments
Efficiency	The used resources in relation to the accuracy and completeness of goals achieved
Satisfaction	The comfort and acceptance of the work system to its users and for other people who are affected by the use of it

In order to evaluate software, a variety of methods can be used. Figure 2 provides an overview and also clarifies the trade-off between the greatest possible objectivity and strong user involvement.

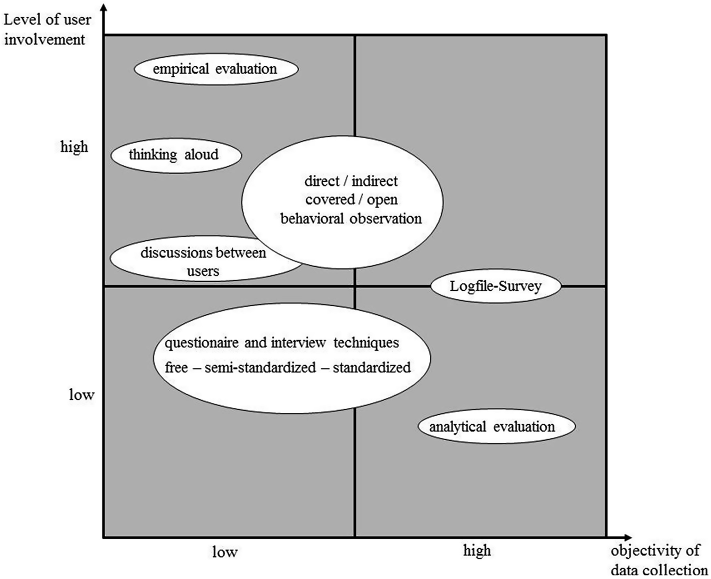

Fig. 2. Evaluation method overview [1]

Table 2. Objectives of the survey and methods used

Objective of the survey/content	Indicator	Used method of collecting data	Used evaluation method
Usability	Effectiveness	• Standardized questionnaire • Semi-standardized (explorative) interview	• Quantitative analysis • Content analysis
	Efficiency		
	Satisfaction		
Acceptance	Intensity of use	• Logfile-survey • Standardized questionnaire	• Quantitative analysis
	Change of use	• Logfile-survey • Standardized questionnaire	• Quantitative analysis
	Willingness to recommend	• Standardized questionnaire	• Quantitative analysis
Potential for further development and improvement of the application		• Standardized questionnaire • Semi-standardized (explorative) interview	• Quantitative analysis • Content analysis
Impacts with regard to mobility behavior and further service supply options		• Standardized questionnaire	• Quantitative analysis

To minimize this trade-off, a combination of quantitative and qualitative approaches (see Table 2) was used for the present study.

The operationalization of these approaches can be seen in Table 3.

Table 3. Details on surveys in the test region Munich and test region Frankfurt

Type of survey	Details of data collection	Test region Munich transport association	Test region Rhine-Main transport association
Semi-standardized (explorative) interview	Goal questions: • Use problems • Particularly positive/negative implementations • Suggestion for improvement	20 individual interviews, each about 45 min	16 individual interviews, each about 45 min
Standardized questioning (online)	Once, as far as possible identical repetition of the survey for the same sample (panel)	Wave 1: n = 388 (response rate: 40,7 %) wave 2 (+ 3 weeks): n = 148 (response rate: 76,4 %)	Wave 1: n = 207 wave 2 (+ 3 weeks): n = 116

The semi-structured exploratory interviews were characterized by great flexibility in response to statements and behavior of subjects [4]. For the interviews goal questions (see Table 3) were developed as an interview guideline. During the interview the subjects first were confronted with a hypothetical scenario (e.g. trip to work, travel to acquaintances), which corresponded with the information on mobility behavior the subjects had provided before.

The standardized questionnaire (see Table 3) was sent to two times online. Participation criteria for the subjects were the use of a smartphone and an Android operating system so that the prototype could be installed. Furthermore, at least an occasional use of a mobility application was a prerequisite for participation. The questionnaires of the two waves have been very similar. Only questions about travel behavior and personal features have been left in the second survey.

The questionnaire was composed of four parts according to the survey targets. The questions primarily comprised closed response formats:

The question blocks for usability were developed according to the GQM model ("Goal Question Metric") of Hussain and Kutar [2]. This model includes dimensions (usability criteria: effectiveness, efficiency and satisfaction), associated questions and metrics that will finally lead to an overall usability assessment of the tested software. Table 4 summarizes the questions used for the presented study. The response format was a endpoint named 5-point scale with numbered values.

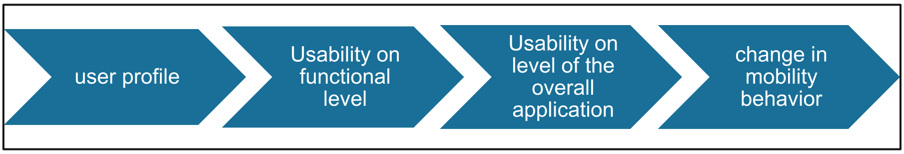

Fig. 3. Questionnaire structure

Table 4. Questions per usability factor of the evaluation questionnaire [2]

	Effectiveness	Efficiency	Satisfaction
Questions	• ease of use • speed of learning • average response time • successful operations from the 1st try • successful solution of the given task from the 1st try • virtual keyboard available	• needed time to solve the task • time until error-free handling the application • suitability of the solutions provided by the app	• sense of security • satisfaction with the user interface • overall satisfaction with the app

3 Results

3.1 Analysis of User Requirements

The following results refer to the quantitative survey described in Sect. 2.1. 60 % of the respondents indicated to use public transport at least once or twice a week for the way to work or training. 53 % of the respondents use public transit regularly (once or twice a week) for leisure trips and 45 % for shopping and other purposes. 70 % of respondents own a smartphone, 60 % use the mobility application once per leisure trip and 33 % once on shopping trips and other routes. After all, nearly one-third shows the same behavior patterns on trips for shopping and other purposes. On trips to job and training the majority (45 %) is using the application only occasionally or rarely.

Within the requirement analysis, the users' needs were investigated with regard to different route sections. Table 5 shows the three most important needs in each section. The most desired user requirements per basic service can be found in Table 8 below.

3.2 Evaluation of Operating Concepts

In this phase of the project should be provided support for the application development by examining eight operating concepts each with the help of three different screenshots shown to the subjects. In the following the results for the operating concept "modal choice" (see Fig. 4) and "maps" (see Fig. 5) belonging to the basic services routing & navigation and intermodal crosslinking are shown.

Table 5. User requirements per route section (priority 1 to 3)

#	Trip preparation	Way to stop/station	Orientation at stop/station	Travel	Transfer	Way to destination
1	Route or stop/station favorites	Alternative route in case of disturbance				Spontaneous change of destination
2	Automatic information in case of delays or disturbances					/
3	Information about special features of the route (replacement services, stations etc. …)	Spontaneous change of destination	Departure overview	Spontaneous change of destination	Departure overview	/

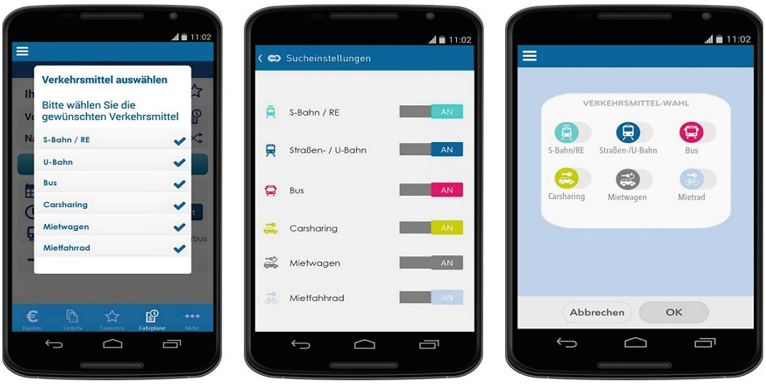

Fig. 4. Screenshots of the operating concept “Modal Choice”

The left screen shows a tabular list of different means of transport which can be selected by setting or cancelling a check mark. This procedure is taken over in the second (middle) screen by a slider. The right screen use pictograms as a slider for the selection of means of transport (Table 6).

Table 6. Evaluation results of the operating concept “Modal Choice”

	Screen 1 (left)	Screen 2 (middle)	Screen 3 (right)
Mean/standard deviation	5,25/1,65	5,32/1,45	4,99/1,73

The screen in the middle represents the top ranked operating concept. Sliders tend to be assessed with higher values by the subjects than other notations. However, these

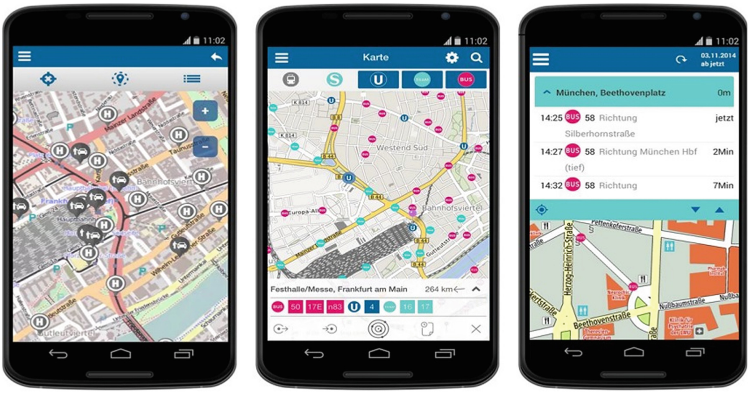

Fig. 5. Screenshots of the operating concept "Map"

sliders should have a clear description and should not contain pictograms as in the right screen.

Additionally imposed evaluation criteria include the age of the subjects, the frequency of public transport use and possible disabilities in personal mobility. With regard to these items, it was found that subjects older than 65 years and disabled people rated the left screen as the best to almost 60 %. Also rare public transit users had a clear preference for the left screen.

The screenshots for evaluating the map design show that the left screen has no additional settings by the user, as they can be found in the headline and footer of the middle screen. A deeper detailing and dichotomy of the screen in a graphical representation and a departure monitor can be seen in the right screen.

On average, 475 subjects evaluated the right screen at its highest. Information to be seen on this screen includes stops, departure boards, information about connections from the selected station as well as the possibility to set start and end points. The use of interactive icons on maps is recommended. With regard to the age and the intensity of public transit use was determined that both younger and middle-aged subjects and rare public transit users prefer a map display with larger icons (see left screen). However, elderly subjects and frequent public transport users prefer to split screen composed of a map and the departure monitor (see right screen). That is the same result as the mean rating already showed (see Table 7).

Table 7. Evaluation results of the operating concept "Map"

	Screen 1 (left)	Screen 2 (middle)	Screen 3 (right)
Mean/standard deviation	4,36/1,87	5,04/1,39	5,38/1,58

In the prototype of the intermodal mobility app and its basic services the operating concepts preferred by the majority of the subjects were implemented.

Table 8. Requirements and qualitative evaluation in the test region Frankfurt

Requirements per basic service	Qualitative evaluation by subjects
Intermodal crosslinking:	
Intermodal inclusion of alternative means of transport	• Intermodal combinations perceived partly as interesting • However only reasonable combinations and • Relevant alternatives should be displayed • Clear identification of means of transport with sufficient information about their options are required
Setting options regarding alternative means of transport	• Selection of the relevant means of transport must be very easy to understand and to implement
Routing und navigation:	
Possibility of multimodal comparison	• The decision for transportation does not change due to the presented alternatives (especially in case of individual car traffic alternatives), because with the use of the app the decision for public transport has already been made • If needed, the comparison of different means of transport is well resolved and clear
Specify settings/preferences for route search	• Options should be presented in an easily understandable manner, possibly in the start screen as directly apparent option list • Sidebar as a place for options is very useful • Selection of walking speed is well accepted • Partly great skepticism towards many settings in the options menu (e.g. modal choice, maximum walking distance)
Display all stops/stations nearby	• Very practical for unknown routes • Request for every stop/station to display the serving lines
Detailed information to stops and stations	• No clear opinion because the depth of details on the one hand is perceived as very advantageous and on the other hand as information overload
Indoor-navigation:	
Overview of the whole building	• Up to the moment of writing the paper the available interviews did not focus this aspect
Navigation with the help of pictures and important points in the surroundings	
Barrier-free accessibility	
Barrier-free routing	• Voice response and list display of the route is important for people with disabilities

(*Continued*)

Table 8. (*Continued*)

Requirements per basic service	Qualitative evaluation by subjects
Social media:	
Combination of calendar and mobility application	• Very useful when the trip can be stored in the calendar
Alarm in case of disturbance on the planned route	• Alarm is an important, useful function
Opportunity to provide feedback about the trip, vehicle, staff, etc. to other users and the transport company	• With the possibility to input the feedback very fast in the application, it is realistic that the feedback function is used intensely • The feedbacks' trustworthiness given by other passengers is rated less than a feedback given by the transport company • The reason of a disturbance is interesting to know in any case, even if it has been given by other passengers and not by the transport company • Subjects requested that customer feedback should be verified in any case by the transport company • It must be clear who is the sender of the feedback
Dynamic trip guidance:	
Real-time information about departure, arrival (push messages)	• Very useful because new departure and arrival times are calculated automatically
Reliable information about the reason and duration of disturbance	• Users need longer time to build confidence and to accept the information as "reliable"
In case of delays automatic display of alternative connections	• Explicit desire of subjects • Only alternatives really relevant for the user should be • Displayed
Storage of routes and corresponding documentations	• The term "travel folder"[a] as a possibility to store route information is difficult to understand • "Favorites" function, however, is far better known and understood, but is confused with the term "travel folder" →Aggregation in a "favorites" function may be reasonable

[a]Virtual folder of planned trips with stored time data; favorites are distinguished from travel folder and store frequent start and destination points.

3.3 Overall Evaluation

At the time of finishing this paper the qualitative and quantitative surveys of the overall evaluation have already taken place in the two testing regions Munich and Frankfurt.

So far, the evaluation results are only partially completed. The following statements therefore comprise only parts of the objectives formulated in the survey.

The qualitative surveys in the Frankfurt region revealed that the multimodal information is very useful in specific situations. The subjects' expectations regarding a public transport application rarely include the offer of alternative modes of transport. The trade-off between the greatest possible depth of information combined with low "overloading" of the application appeared in the interviews repeatedly. Different requirements in different situations require both "slim" solutions as well as highly detailed information. The sender of the information is important for to the subjects. Confidence in reports of other passengers is significantly lower than the confidence in information of transport companies. Moreover, the quality of the data plays an important role. The subjects assess real-time information to be very positive and valuable.

The quantitative results of the overall evaluation show a nearly balanced picture (see Table 9). The effectiveness, that is the accuracy and completeness with which users reached the objectives, is evaluated a little bit higher than the degree of satisfaction and efficiency, but only slightly.

Table 9. Results of usability measurement for the application prototype in total

	Effectiveness	Efficiency	Satisfaction	Usability
Mean of the ratings (Scale: 1 (poor) – 5 (very good), n = 152)	3,78	3,58	3,62	3,66

With regard to the objectives, to gain insights about changes of the mobility behavior due to the use of mobility applications (see Fig. 3), the following statements can be derived from the survey results (n = 154 subjects):

- the likelihood that public transport is used more often with such an application is significantly higher than in the case of other means of transport
- 54 % of the subjects reported that the chance to use public transport more often is at least 80 %, but
- with increasing age and with declining frequency of public transit use this chance is reducing.

4 Conclusions

Within the scope of the DYNAMO research project an intermodal door-to-door application was developed and at the same time this process was accompanied scientifically. The article dealt with the three-stage approach of this scientific evaluation. At the end of the project the great benefit of a broad-based methodology in which all the quantitative and qualitative approaches are combined is visible. In particular, in the development process of a prototype it is very helpful to integrate qualitative user research as often as possible, because it delivers valuable insights for software engineers during the development process. The involvement of the software development teams in these surveys enhances this effect.

All three survey levels showed that during the development of intermodal applications a trade-off between the complex information needs of users and the applications' ease of use always exists. In the future, this trade-off may be handled with the help of specialized applications for individual types of users or groups. Probably, such solutions, however, can be fully implemented on the basis of self-learning systems and artificial intelligence.

References

1. Hegner, M.: Methoden zur Evaluation von Software. IZ-Arbeitsbericht Nr. 29., Informationszentrum Sozialwissenschaften der Arbeitsgemeinschaft sozialwissenschaftlicher Institute e. V., Bonn (2003)
2. Hussain, A., Kutar, M.: Usability metric framework for mobile phone application. In: The 10th Annual PostGraduate Symposium on The Convergence of Telecommunications, Networking and Broadcasting, Liverpool (2009)
3. Rhein-Main-Verkehrsverbund Servicegesellschaft mbH. Projektphasen. http://www.nahtlosmobil.eu/arbeitsplan.html (2016). Accessed 29 Jan 2016
4. Honer, A.: Das explorative Interview: zur Rekonstruktion der Relevanzen von Expertinnen und anderen Leuten. Schweizerische Zeitschrift für Soziologie **20**, 3 (1994)