

Usability Evaluation of a Paratransit Mobile Application

Elizabeth Mintmire, Chen Ling, and Randa Shehab
The University of Oklahoma, School of Industrial & Systems Engineering

TransitMobile is a mobile application that provides logistic information to drivers in Oklahoma community-based paratransit operations. Passenger reservation information is transferred from the home agency to in-vehicle mobile devices to notify the drivers as reservations are made in real-time. In this study, user testing and a questionnaire were used to evaluate the usability of the application. Task completion times and subjective measures from the questionnaire pointed to areas of the system design that caused errors. The effects of practice revealed the importance of training for drivers in the field. While TransitMobile is a promising application, the system in which it functions is complex, and users must receive adequate training in order to use the application to its fullest potential.

INTRODUCTION

Paratransit is defined by the Transportation Research Board's Committee on Paratransit as "alongside transit," or any mass transportation mode that supplements individual transportation (Lave and Mathias, 2000). Paratransit refers to transportation between customer-specified pick-up and drop-off locations at a customer-specified time. It is often characterized as demand-responsive, which means that it reacts to instantaneous demand from potential customers. With ever-increasing demands on public paratransit systems, it is vital to develop effective tools to assist agencies in managing transit data. Some approaches to this issue involve advanced public transportation systems technologies and computer-aided dispatching systems (Schweiger, 1996; Chira-Chavala, Gosling, & Venter, 1997).

TransitMobile is an Android-based mobile application used by drivers to manage paratransit information. Information is transmitted to the application using the parent program, TransitAssistant. The TransitAssistant program is an information management system developed by members of the School of Computer Science at the University of Oklahoma. While TransitAssistant is currently used by many Oklahoma paratransit agencies, use of TransitMobile is not yet as widespread. It is important to understand design factors influencing the system's usability to create a tool that intuitively and effectively satisfies user requirements.

In the field, schedulers and dispatchers at community-based paratransit agencies receive calls from customers who wish to make reservations. The employee records the relevant information in TransitAssistant (e.g. pick-up location and time, drop-off location and time, passenger name, and special requirements). Once the scheduler records the reservation, he or she then assigns the trip to a manifest—a list of passengers served by a single vehicle. The scheduler then transmits the manifest to the appropriate driver's mobile device. The driver can access the trip information for all passengers using TransitMobile. As the driver proceeds along the route, he or she enters the odometer readings associated with each passenger pick-up and drop-off event. Drivers may also create new trips if a potential customer arrives at a pick-up stop without a reservation. At the end of the shift, the driver transmits the information back to his or her agency then reports the data to state-level organizations. Agencies that do

not use TransitMobile rely on paper-based methods for trip planning and data reporting, but such methods can be more time-consuming for drivers and schedulers. An electronic management system allows agencies to streamline communications and reporting methods in a way that allows them to satisfy demand loads more efficiently.

This study sought to evaluate the usability of the TransitMobile application through user testing and an adapted Questionnaire for User Interface Satisfaction (Chin, Diehl, & Norman, 1988). Participants completed a series of tasks similar to those that a driver would perform.

METHODOLOGY

Participants and Demographics. To evaluate the TransitMobile system, this study used a convenience sample of 18 participants from the University of Oklahoma. All participants but one were students in the School of Industrial & Systems Engineering. The average age for reporting participants was 22.9 years, with a breakdown of 10 males, 5 females, and 3 non-respondents.

In comparison to TransitMobile users, participants were similar in gender distribution, but were younger on average than agency counterparts. In addition, study participants may have had a higher level of education and technology exposure than the average TransitMobile user. While these differences between sample and user population existed, previous usability studies have shown that subjects with a background in human-computer interaction tend to identify interface usability issues, whereas end-users tend to identify task performance issues (Yen & Bakken, 2006). Despite certain differences between subjects and participants with respect to age and background experience, the researchers chose to use students as a sample due to their experience with human factors and ergonomics.

Equipment. The evaluation ran the TransitMobile application on a Motorola Droid X Android touchscreen phone with a 4.3-inch screen height. The evaluators used the TransitAssistant program to input passenger data for each participant. The TransitAssistant program ran on a Dell Optiplex 745 with a dual-monitor setup and a Windows XP Professional 32-bit operating system.

To create the questionnaire, the researchers adapted the Questionnaire for User Interface Satisfaction (QUIS) (Chin,

Diehl, & Norman, 1988) to suit the requirements of the current usability evaluation. The questionnaire addressed screen design, terminology, the user's learning experience, system capabilities, and multimedia. Though QUIS is not a standardized instrument, it was chosen over standardized tools because it addressed areas of user interface interaction that the researchers wished to analyze.

Tasks. As each participant arrived, a researcher explained the purpose of the evaluation and gave a brief overview of the TransitAssistant and TransitMobile system. Login information was pre-programmed into the phone, and the participant only had to tap the "login" button to begin completing the tasks. Each participant received a list of 10 passenger names with corresponding pick-up and drop-off location odometer readings. Passenger manifests were created prior to the evaluation by the researchers and uploaded to TransitMobile using the desktop TransitAssistant program.

To complete the data entry tasks, each participant had to click the login button on the opening screen. After logging in, the participants tapped the "begin shift" button, which took them to a screen with the name of one manifest on it. They selected the manifest by touching the name, which then brought up a preview of all the passengers in the manifest, represented in Figure 1. In the full list screen, the participants then decided if they wanted that listing and confirmed their choice by pressing the "start manifest" button.

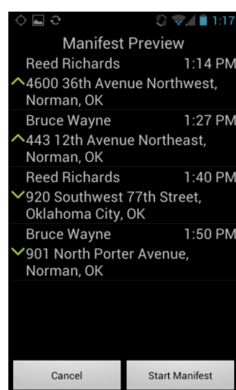


Figure 1. Manifest Preview Screen

After pressing "start," a screen appeared that asked the participants to enter the beginning odometer reading. Participants pressed the "enter mileage" button and then used an on-screen number pad to do so (see Figure 2). After entering the data, the participants exited the number pad by pressing "save," which returned them to the screen of origin. They moved into the actual evaluation data entry tasks by yet another "start shift" button.

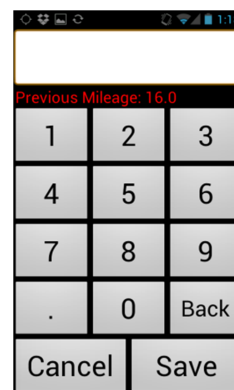


Figure 2. Number Pad Screen

The following screen listed all passengers twice—once for pick-up and once for drop-off. Pick-ups were indicated with a green upwards-facing arrow and drop-offs were indicated with a green downwards-facing arrow. The participants selected the first passenger's name, which took them to a screen with an option to enter the pick-up mileage. After pressing the "pick-up" button, another screen showed a "enter mileage" button which then took participants to a number pad screen where they entered the mileage and pressed "save" or "cancel." After the participant saved the mileage, the passenger data screen returned but this time with a "drop-off" button. Participants selected that to drop off passengers and the procedure for entering the drop-off mileage was identical to the pick-up procedure. Once a passenger was picked up and dropped off, both entries disappeared from the manifest. The participants repeated this process for all 10 passengers.

After entering each passenger's data into the system, the participants ended their shift, entered the ending mileage and logged out of the application. Finally, they completed the QUIS questionnaire to judge their experiences with the system.

RESULTS

User Testing Results. Before starting the application, the researchers gave the subjects a basic overview of the tasks and mileage data, and then let the participants learn the system through trial and error. Though participants reported having some difficulty with getting started, several commented that after practicing data entry with one or two passengers, the process was simple to understand. However, one subject mentioned that his or her learning curve was steep because she was "unclear of the program's purpose and features at the beginning."

The researchers observed that participants with a first language other than English appeared to have more trouble at the beginning of the trial than native English speakers. However, because language background was not collected during the study, the researchers could not measure any definite differences. It may be worthwhile when working with agencies to determine if any drivers have a first language other than English, as this may affect their comfort with using the system.

Regarding physical system attributes, participants mentioned that it the phone was “large enough and had fairly good touch screen sensitivity,” so the phone used in the experiment was an appropriate tool on which to display the information. Participants appeared to have mixed feelings about the screen layout; some found the layout “easy and intuitive” while others thought that between-screen changes in button position and size was confusing.

During each evaluation, a researcher video recorded the participant interacting with the system. Using the 10 clearest videos, the researchers measured average times for task completion. Over the course of the evaluation, tasks involving the data entry steps reflected the learning curve, as shown in table 1. A t-test for equal sample sizes, unequal variances was conducted to compare performances between the first passenger’s and final passenger’s data entry ($\alpha = 0.05$). The p-values from the t-test reveal that participants experienced significant gains over time for simple button selection tasks.

Table 1. Mean time to complete data entry task for first and final passengers, n=10 (in min:sec.milliseconds)

Task	First Passenger		Tenth Passenger		t-Test
	Avg.	Stdev	Avg.	Stdev	p-val
Choose passenger name from manifest screen	00:16.01	00:22.73	00:00.80	00:00.14	0.046
Press "pick-up"	00:06.83	00:05.58	00:02.21	00:00.57	0.033
Press "enter mileage"	00:10.66	00:11.37	00:01.58	00:00.61	0.025
Enter odometer reading & press "save"	00:08.06	00:04.86	00:03.75	00:02.42	0.024
Press "okay"	00:07.64	00:05.76	00:02.20	00:01.13	0.032
Press "drop-off"	00:05.18	00:04.39	00:01.89	00:00.65	0.072
Press "enter mileage"	00:02.56	00:00.82	00:01.91	00:00.87	0.132
Enter odometer reading, press "save"	00:03.77	00:02.02	00:03.60	00:01.34	0.843
Press "okay"	00:02.37	00:01.06	00:01.41	00:00.36	0.039
Press "done" to return to manifest	00:04.06	00:02.25	00:01.80	00:00.42	0.025

Questionnaire Results. Surprisingly, participants responded that learning to operate the system was easy and the time it took was reasonable. Participants thought that remembering commands and operation rules was easy, as well. Participants could usually perform tasks in a straightforward manner, and they could complete tasks in a logical sequence of steps. They reported that entering mileages did not require an excessive number of steps and that the system provided somewhat clear feedback at the completion of steps.

On the other hand, participants reported mild difficulty with discovery of new features, which may conflict with the positive response to the learning curve. To some degree, this is attributable to the nature of the evaluation; the procedure

limited the participants to using the same commands for each passenger.

With respect to screen design, responses were very positive. The phone screen was easy to read and the fonts used were very legible. Characters and images were sharp, and the amount of information that the screen displayed was adequate for the tasks. Participants found that the arrangement of the information was in a logical order. Screen layouts generally helped users understand what to do.

When following the steps to complete the data entry tasks, participants followed a sequence of screens. Some participants found the sequence of screens clear, but others commented that they experienced confusion. This may be due to misunderstanding the purpose of the TransitMobile program. Participants believed that the screen sequencing was somewhat predictable and that the direction for progress through work-related tasks was somewhat clearly marked. Additionally, participants found that returning to a previous screen was also fairly simple.

In general, participants felt positively about the color choices for the system, though some commented about a dislike of displays with dark backgrounds and light text. On average, participants rated the colors in the display as somewhat natural and that the colors available were somewhat adequate.

Use of terminology throughout the system was consistent, and on-screen message content and positions were consistent. Participants found on-screen messages and instructions for completing commands fairly clear, as well. Error messages caused slight confusion for users; however, not every participant encountered errors, so not all were exposed to error messages. This may have influenced the response to related questions. For the participants who did experience errors, participants found that instructions for correcting them were only somewhat clear and the message content was only somewhat helpful.

Performing an operation almost always led to a predictable result. Participants reacted very positively to the appropriateness of the length of delay between operations. Participants felt less positive about the level that the system kept the user informed about its progress on a scale of never/always keeping the user informed.

On the whole, the system capabilities were very strong. The system occasionally warned the user about potential problems. Participants reported that correcting typos and mistakes was usually easy. However, the average for correcting mistakes may be an inflated estimate; not all participants completed the associated QUIS question. Participants also reported that the system adequately allowed them to undo operations. Participants felt that it was relatively easy to accomplish the tasks with few commands. They generally agreed that ease of operation depended on their level of experience, which may go back to user comments regarding the system’s learning curve. However, some users commented that “it was not hard to figure out how to go back when [they] missed something.”

Table 2. QUIS Results (9-point Likert scale)

	Question	Mean	Stdev
Overall reactions	Terrible/wonderful	6.11	1.60
	Frustrating/satisfying	5.87	2.03
	Dull/stimulating	4.93	1.53
	Difficult/easy	6.67	1.68
	Inadequate power/adequate power	7.67	1.45
	Rigid/flexible	4.60	1.88
Screen	Characteristics on the computer screen, hard to read/easy to read	8.17	1.10
	Image of characters, fuzzy/sharp	7.89	0.96
	Character shapes (fonts), barely legible/very legible	8.33	0.91
	Screen layouts were helpful, never/always	6.56	1.34
	Amount of information that can be displayed on screen, inadequate/adequate	7.44	1.25
	Arrangement of information on screen, illogical/logical	7.17	1.42
	Sequence of screens, confusing/clear	5.83	2.01
	Next screen in a sequence, unpredictable/predictable	6.67	1.88
	Going back to the previous screen, impossible/easy	7.11	1.75
	Progression of work related tasks, confusing/clearly marked	6.72	1.87
Terminology & System Information	Use of terminology throughout system, inconsistent/consistent	7.89	1.23
	Work related terminology, inconsistent/consistent	7.82	1.29
	Computer terminology, inconsistent/consistent	7.82	1.13
	Messages which appear on-screen, inconsistent/consistent	7.83	1.15
	Position of instructions on the screen, inconsistent/consistent	7.71	1.10
	Messages which appear on-screen, confusing/clear	7.28	1.45
	Instructions for commands or functions, confusing/clear	6.61	1.58
	Instructions for correcting errors, confusing/clear	6.54	2.11
	Computer keeps you informed about what it is doing, never/always	5.67	1.91
	Animated cursors keep you informed	6.21	1.25
	Performing an operation leads to a predictable result, never/always	7.56	1.34
	Controlling amount of feedback, impossible/easy	6.93	1.27
	Length of delay between operations, unacceptable/acceptable	8.22	0.88
	Error messages, unhelpful/helpful	6.75	1.96
	Error messages clarify the problem, never/always	6.50	2.35
	Phrasing of error messages, unpleasant/pleasant	6.75	1.66
Learning	Learning to operate the system, difficult/easy	6.17	1.92
	Getting started, difficult/easy	5.11	2.22
	Learning advanced features, difficult/easy	5.60	2.10
	Time to learn to use the system, slow/fast	6.33	1.85
	Exploration of features by trial and error, discouraging/encouraging	6.47	2.07
	Exploration of features, risky/safe	6.82	1.81
	Discovering new features, difficult/easy	5.82	1.81
	Remembering names and use of commands, difficult/easy	7.33	1.28
	Remembering specific rules about	7.39	1.38

System Capabilities & Multimedia	entering commands, difficult/easy		
	Tasks can be performed in a straightforward manner, never/always	7.11	1.45
	Number of steps per task, too many/just right	6.63	2.47
	Steps to complete a task follow a logical sequence, never/always	7.56	1.62
	Feedback on the completion of the steps, unclear/clear	6.89	1.88
	System warns you about potential problems	6.83	1.59
	Correcting your mistakes, difficult/easy	7.46	1.61
	Correcting typos, complex/simple	7.91	0.83
	Ability to undo operations, inadequate/adequate	7.57	1.16
	Ease of operation depends on your level of experience, never/always	7.50	1.15
	You can accomplish tasks knowing only a few commands, with difficulty/easily	7.44	1.50
	You can use features/shortcuts, with difficulty/easily	6.06	2.02
	Colors used are, unnatural/natural:	6.82	1.74
	Amount of colors available, inadequate/adequate:	6.65	1.50

DISCUSSION

Participants commented that the program was an “easy-to-use app” and that “with proper training, [it seemed] easily navigable.” However, some participants had trouble understanding the concepts “drop-off” and “pick-up,” and others initially did not understand how to interpret the mileage table.

The TransitMobile application’s interface lists each passenger’s pick-up and drop-off information in separate lines on the display. However, when completing the data entry tasks, dropping off a passenger occurred by touching a button on the same screen in which the participant entered the pick-up data. This led to confusion among the participants. This confusion could be addressed by populating the manifest with one data line per passenger, which would lead to a shorter, easier to read list. However, some of the difficulty users had may be attributable to the evaluation’s procedure. In real applications, drivers may need to pick up many other passengers before dropping off a single passenger; in such a case, it may be more helpful to see two entries for the passenger in order to tell if he or she had been picked up already. Another evaluation approach would be to change the procedure by asking participants to pick up several passengers, and then drop off one passenger, instead of using a pick-up/drop-off/pick-up/drop-off approach.

During each participant’s learning curve, several errors commonly occurred. The first step of the procedure required participants to choose a manifest, which led to a screen in which previewed the list of trips (see Figure 1). Several participants expected the preview to be interactive, and tried to select passenger names from the list. This may have led to confusion later when the participants interacted with the “show trips” screen. The “show trips” screen was nearly identical to the preview screen, with the addition of allowing the participants to press the names and view associated trip

information. This may have led to some of the hesitation revealed in the task completion time measurements.

Another common error occurred in the number pad screen. Some participants went to the screen, but did not know what to do and attempted to return to the previous screen by pressing "OK." This caused an error message to appear notifying the participant that they needed to enter a number. A correct response after not entering a number would have been to press cancel instead.

One participant mentioned that the "error message helped [him] to understand problems but not how to correct [them]." This was a common reaction for other users who experienced errors. Several participants experienced errors inputting starting mileage, but this appears to have been more of a system operational error than a usability problem. One participant made a suggestion that the system could benefit from the use of graphics on the display instead of text.

The task completion times measured during the user testing appear to reflect participant comments about problems more than the QUIS results. Only two of the measures on the questionnaire, "1=rigid, 9=flexible" and "1=dull, 9=stimulating," fell below 5 points on the 9-point Likert scale. These positive scores indicate that TransitMobile's initial design presents data in an adequate way, though there is surely room for improvement, such as in the user's learning phase. The task completion time measurements did reflect significant improvements over time during each evaluation. The higher-than-expected questionnaire results may be attributable to the fact that while participants struggled to understand the program's intent at the beginning of the evaluation, once they became familiar with the layout and functions, the process was easily repeatable.

The improved task performance over time indicates that training and practice is a requirement for users to have early success when interacting with TransitMobile. Tasks with noticeably longer completion times for the first passenger's information primarily involved button selections. Based on user testing observations, these delays were likely caused by the participants' unfamiliarity with the screen layouts. An adequate training for TransitMobile should familiarize users with the screen layouts, screen progressions, and system terminology.

CONCLUSION

Participants rated the TransitMobile's system capabilities and text design very positively; it reacted to their demands and performed tasks quickly. Using a mobile phone with a touch screen is also feasible—participants felt that the screen size was appropriate to representing all the information they needed in a single screen. Some participants had difficulty understanding the screen progression through tasks, and most experienced difficulty getting started through trial and error. However, many participants found that after one or two data entries, the tasks were easy to understand.

Participants who ran into errors found some easier to fix than others. If caught immediately, typos when entering mileage were simple to fix, but if left unrecognized after finalizing a passenger's trip information, the erroneous

mileage could lead to a nearly unresolvable conflict with another passenger's data. The application clearly informed them that there had been an error, but did not clearly tell them how to fix it. Improved error messages or even a help section may be useful to include in the program in future models. Additionally, though participants rated the app as somewhat rigid and dull, it appears that this initial design satisfies the drivers' needs adequately.

Understanding the needs of TransitMobile users—primarily drivers—is critical in ensuring a high-quality product that effectively supports paratransit operations. In many cases, technology has improved operations in transit organizations; with intentional effort to understand user requirements, TransitMobile has great potential to positively affect the productivity for Oklahoma paratransit agencies.

ACKNOWLEDGEMENTS

This work was supported through funding provided by the Department of Transportation Federal Transit Administration (DOT-FTA) Grant # OK 03-7008. The work reported in this paper reflects the ideas and opinions of the authors and not the DOT-FTA. The authors would also like to acknowledge Aaron Albus for contributions to this work.

REFERENCES

- Chin, J. P., Diehl, V. A. and Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of SIGCHI '88*, (pp. 213-218), New York: ACM/SIGCHI.
- Chira-Chavala, T., Gosling, G., & Venter, C. (1997). Automation of Paratransit Reservation, Routing, and Scheduling. *Journal of Advanced Transportation*, 34(2), 191-211.
- Lave, R., & Mathias, R. (2000). State of the art of paratransit. In *Transportation in the New Millennium: State of the Art and Future Directions, Perspectives from Transportation Research Board Standing Committees*. Wash., DC: Transportation Research Board.
- Schweiger, C. L. (1996). Deployment of Technology for Paratransit: Examples of Innovation. In *Proceedings for the Vehicle Navigation and Information Systems Conference, VNIS '96*, vol. 7.
- Yen, P.-Y. & Bakken, S. (2009). A Comparison of Usability Evaluation Methods: Heuristic Evaluation versus End-User Think-Aloud Protocol – An Example from a Web-based Communication Tool for Nurse Scheduling. *AMIA Annual Symposium Proceedings 2009*, 714-718.