Accessible Transportation Technologies Research Initiative (ATTRI) Performance Metrics and Evaluation

Evaluation Plan for the AbleLink Wayfinding Standard to Facilitate Independent Use of Public Transit by Individuals with Cognitive Disabilities

www.its.dot.gov/index.htm

Technical Report—May 2021
FHWA-JPO-21-838

U.S. Department of Transportation

Produced by Cambridge Systematics, Inc. for the
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
Federal Highway Administration
Federal Transit Administration
Intelligent Transportation Systems Joint Programs Office

Notice

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-JPO-21-838		

4. Title and Subtitle	5. Report Date
Accessible Transportation Technologies Research Initiative (ATTRI) Performance Metrics and Evaluation Evaluation Plan for the AbleLink Wayfinding Standard to Facilitate Independent Use of Public Transit by Individuals with Cognitive Disabilities	May 2021
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Anat Caspi, Mark Hallenbeck, Varsha Konda, Dylan Cottrell	

9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)
Cambridge Systematics, Inc. 3 Bethesda Metro Station, Suite 1200 Bethesda, MD 20814	Washington State Transportation Center Taskar Center for Accessible Technology University of Washington 1107 NE 45th St, Suite #535 Seattle, WA 98195-4802	
		11. Contract or Grant No. DTFH61-16-D-00051

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
U.S. Department of Transportation ITS Joint Programs Office—HOIT 1200 New Jersey Avenue, SE Washington, DC 20590	Final
	14. Sponsoring Agency Code HOP

15. Supplementary Notes

Task Order Contracting Officer's Representative: Robert Sheehan

16. Abstract

The Accessible Transportation Technologies Research Initiative (ATTRI) Program provides a venue that funds transit projects and technology development projects, which improves access to mobility and transportation for individuals with disabilities, older people, and veterans with disabilities. For each of the 5 current ATTRI projects, the ATTRI Independent Evaluation Framework [Framework, 2020] structures an analysis of project impacts from performance measures provided by the project partners, as well as an assessment of the business models.

This report constitutes the Independent Evaluation Findings for the ATTRI projects, and includes an independent evaluation for the AbleLink WayFinder 3 mobile application performed per the Independent Evaluation Plan described in [IE Plan, 2020]. It includes the following chapters: Findings Summary, Purpose of the Evaluation, Evaluation Findings, Strengths identified by the Evaluation, Area of Improvement, Conclusion, Recommendation, Appendices.

17. Keywords	18. Distribution Statement
Accessible Transportation Technologies Research Initiative, wayfinding, pre-trip, scenario, logic model, Accessibility Development Projects	No restrictions

19. Security Classification. (of this report)	20. Security Classification. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	96	N/A

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: volumes greater than 1000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	$candela/m^2$	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

Source: Federal Highway Administration.

Table of Contents

List of Figures

List of Tables

Chapter 1. Findings Summary

Introduction to the Accessible Transportation Technologies Research Initiative Project

The U.S. Department of Transportation's (USDOT) Accessible Transportation Technologies Research Initiative (ATTRI) is a joint USDOT initiative, co-led by the Federal Highway Administration (FHWA), Federal Transit Administration (FTA), and Intelligent Transportation Systems Joint Program Office (ITSJPO), with support from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR), and other Federal partners. ATTRI research focuses on removing barriers to transportation for people with visual, hearing, cognitive, and mobility disabilities.

In 2017, ATTRI announced an award of six contracts to Accessibility Development Projects (ADP), to develop applications in the areas of wayfinding and navigation, safe intersection crossing, and pre-trip concierge and virtualization. The contracts follow a two-phase development plan. In 2018, projects made significant progress in their development activities. ATTRI evaluated the applications for their readiness to advance to phase II. Those that met their earlier goals advanced to the demonstration phase. Applications in the demonstration phase: wayfinding and navigation; pre-trip concierge, and virtualization. Independent evaluators assess these demonstrations with a common framework. This work includes an example of an independent evaluation (IE) that follows the proposed Independent Evaluation Plan (IE Plan, 2020) for only one of the ADPs, AbleLink's WayFinder 3 mobile application.

Purpose of the Independent Evaluation for AbleLink's WayFinder3

This report outlines the findings of an evaluation that was conducted by following an independent evaluation plan (Framework, 2020).[1] The proposed plan uses the framework presented in the ATTRI Evaluation Framework Report (IE Plan, 2020) and takes a holistic approach to mobility and transit.[2] The approach focuses on accessibility and usability along the "complete trip," which refers to an individual's ability to plan for and complete a trip from origin to destination without disruptions in the travel chain. This structured approach allows the development of evaluation plans for specific ADP and describes how to

---

[1] Accessible Transportation Technologies Research Initiative (ATTRI) Performance Metrics and Evaluation, Final Evaluation Framework Report, by Anat Caspi, Mark Hallenbeck, Shannon Tyman, July 2020.

[2] Accessible Transportation Technologies Research Initiative (ATTRI) Performance Metrics and Evaluation, Final Evaluation Framework Report Evaluation Plan for AbleLink Project #1, by Anat Caspi, Mark Hallenbeck, Dylan Cottrell, August 2020.

select performance metrics that the ADP uses to conduct IE. Framework 2020 is meant to contribute to the goal of a more equitable and inclusive transportation system.

Prior to discussing the Independent Evaluation findings, it is helpful to contextualize the outcome of this IE, in that it serves as a demonstration evaluation, conducted to gain valuable information about not only the ADP but how the ATTRI Evaluation Framework (Framework, 2020) is operationalized. The evaluation-planning process described in IE Plan 2020 is customer-oriented and sensitive to accessibility issues that arise in ATTRI's populations of interest. The plan focuses evaluators' efforts on refining the ideas described in the Framework 2020. Evaluators will prioritize accessible design aspects and logic models, exposing how the framework operates and how it proposes to assess questions pertaining to the highly diverse, heterogeneous populations who need ATTRI technologies. Although this is outside the scope of this independent evaluation, a next step would be to engage stakeholders (both individuals who are a part of ATTRI as well as the ADP team and its stakeholder population) in a discussion about further learning and questions about "complete trips."

The independent evaluation team formed evaluation questions that clearly identified the information needs and themes that the Framework 2020 proposes, and then matched an evaluation design to those questions to generate valuable data and findings. Specifically, the approach advocates using the framework as a starting point for selecting and applying appropriate performance indicators and measures that are integral to the assessment of an ADP outcome. Three main data types were collected and analyzed to produce the evaluation results herein. We recognize that due to limitations in resources (including staff, volunteer time, and lack of data) and an unusual timing of this evaluation (the evaluation overlapped a global pandemic), some of the evaluation procedures described in IE Plan 2020 had to be modified. Nevertheless, this document provides a valuable demonstration of an exercised evaluation plan, and can be used to benefit other independent evaluations operating a similar program. The independent evaluation team was able to make meaning and sense of the data collected and addressed the evaluation hypotheses. The report conclusion suggests means for the evaluated ADP's technology team to make its program stronger, and means for the ATTRI program to transfer this knowledge to other programs, empowering its funded-teams to reflect new insights. Finally, this document demonstrates how to use both ATTRI's phase I User Needs Assessment (see ATTRI User Needs Assessment: Stakeholder Engagement Report, May 2016) and performance measures in Framework 2020 to understand the ADP technology, but also to better reflect on the ADP performance against the "complete trip" goals and objectives espoused by ATTRI's program.

Brief Introduction to AbleLink's WayFinder 3

The AbleLink #1 ADP aims to create an application for smart phones utilizing native, mobile global positioning system (GPS) technology that provides location-aware (i.e., configurable GPS waypoints that identify the location of the user) visual and auditory prompts. The goal of AbleLink #1 is to provide a mobile-based, trip-concierge application (WayFinder 3) that allows users to select, download, and then follow travel routes independently.

> WayFinder 3 is a specialized application that operates on off-the-shelf, smart phones. It uses GPS and specialized visual, audio, and vibration prompts to allow individuals with cognitive disabilities to use fixed-route transportation independently. With WayFinder 3, multiple travel itineraries or routes, can be programmed into a single device to enhance travel choices and independent transportation. To set up the system, teachers, family members, transportation trainers, or other staff ride the route (either on the bus or in a personal vehicle), and use the system's Route Builder utility to set waypoints and record instructions for those waypoints and optionally take pictures to provide visual cues along the way. Once a route is created with the device, a rider with cognitive disabilities can then select the route and follow the multimedia step-by-step GPS-based prompts to arrive at his or her destination.
>
> AbleLink #1 SSR

Key Partners

Intelligent Transportation Systems Joint Programs Office

ITSJPO is a co-leader of ATTRI. They are the primary funding source for the ATTRI effort. The evaluation is focused on achieving the goals of the ATTRI program.

AbleLink and Merakey

AbleLink Technologies (AbleLink) is the firm that has create the ADP. The main milestones for the AbleLink #1 ADP were captured in the ADP documentation. Merakey is a provider of education and human services to individuals with special needs. Merakey is working with AbleLink to test the Wayfinder 3 ADP. Individuals being supported by Merakey will be testing the ADP.

The AbleLink team will support the evaluation process by responding to queries during evaluation planning and providing technical support during evaluation implementation. AbleLink, in conjunction with Merakey, will support and facilitate initial evaluation deliverables. They provided key information for this evaluation plan. In addition, they will collect and share data that is relevant to this ADP demonstration between May 2019 and September 2019. This data will be shared with the IE team.

Cambridge Systematics and University of Washington

The IE team is working under a contract between the ITSJPO and Cambridge Systematics. The University of Washington, represented by the Taskar Center for Accessible Technology and the Washington State Transportation Center (TRAC), is working with Cambridge Systematics. They will provide the IE documentation, evaluation implementation and reports. The IE team will remain flexible to suggestions and requirements from the evaluation sponsor. The IE team will draft reports, brief the evaluation and project managers and stakeholders on progress and key findings and recommendations. The IE team also will finalize the evaluation, taking into consideration comments and questions on the evaluation plan and report.

Key Findings

Through use of the traveler-centered research and evaluation framework, we identified several findings that may help ADP's manage risk and ensure responsiveness to the target population needs:

- User-centered design and agile product management practices are seen to be the norm in the field, but are coming up against potential challenges in the current ATTRI governance context. The Independent Evaluation team traced back the work performed by the ADP from the mobile application itself, through the traceability matrix back to System Requirements and needs finding. It is possible that barriers posed by a need to comply with documentation structure that presumes a waterfall model of building software, and addresses risk using a narrow definition of risk to the public, made it difficult for the ADP to adopt a style of work that is inseparable from working with the ATTRI target population: namely designing with the target population on an ongoing basis with multiple touchpoints and iterations throughout the design and development cycles. Particularly in addressing the needs of a heterogeneous population, adopting a more holistic evaluation framework allowed the IE to identify some gaps that impacted users, but fell outside the formal ATTRI process.
- Pilot testing: It is worthwhile for the ADP to coordinate a small pilot with the evaluation team before investing the resources and time required to perform field work with a large sample of a population that is difficult to access. When doing user experience and usability studies, it is worthwhile doing a small pilot with approximately 5 people, in order to identify big issues as well as ensure that the study teams (both internal to the ADP as well as the IE) have the data output that was expected. There are two major reasons for this, as we found in the process of performing the evaluation. Firstly, there are usually diminishing returns after as few as 5 participants, and a pilot requires less investment while it exposes the major issues first. As an example, the few experiments the IE did in-lab exposed the main app failures that occurred for multiple MeraKey participants over multiple trips, and resulted in a loss of data for many of the field experiment trips. The difficulty of accessing the target population for many ATTRI technologies highlights the need to do initial pilot field testing (not just lab testing) before advancing the technology for field tests with a larger sample of the target population. Second, a coordinated pilot may expose any miscommunication about the type and quality of data being collected and potentially expose early any data biases. In the MeraKey field test, for example, many trips had missing dashboard data or had data that was not usable for the evaluation, but was already built into the evaluation process. The intent in pilot tests is to leave the higher-value evaluation pieces to the larger field study.
- Detailed, measurable observation of the target population using the system is integral to system evaluation: The most powerful way to understand user behaviors is not only to have users talk about how they use the system but to have them demonstrate how they use the system, and capture detailed recordings of their actions. The intermediaries (caregivers) collecting the Merakey data witnessed important gaps in the Wayfinder app, but there was no formal or detailed mechanism in the field experiment design to capture this information, nor detailed data collection on the mobile platform deployed to enable trace-back for full user activity, to link that behavior to application outcomes or user outcomes. Importantly, this resulted in missing crucial usability data that was only captured through unstructured side comments made by busy caregivers. Our research team learned a lot about system usability, robustness, functionality and failure modes by setting up walkthroughs with a single researcher who self-identifies as part of the target population. By asking this researcher to walk us through his use of the application, both in lab and in the field, we learned how he finds the information

he needs through the WayFinder system and also witnessed some of the application failures that appeared to be similar to those pointed to by the Merakey incidental comments.

- Digital service delivery is a fundamental no ADP can overlook, even though it does not immediately address traveler's apparent needs in completing the trip: three emerging problems appeared with the WayFinder application related to service delivery- suboptimal phone resource usage, low accuracy GPS localization and application crashes. The IE team considers these aspects to be part of digital service delivery because it relates to how the mobile application interacts with or is integrated into the mobile platform, rather than having to do directly with the user experience or interface of the application. Good service delivery can build credibility and trust with the target population, thus enabling the ADP to experiment and innovate further. Bad digital service delivery, or outright failures, damage that trust and credibility in ways that can persist and hold back sustainability and future progress for the ADP. On the flip side, the ATTRI program itself should offer ADP's sufficient funding to invest in the fundamental service delivery aspect of all ADP's, or alternatively, offer ADP's access to government organizations like 18F to help the ADP's focus on user experience and the core technology.
- Needs assessment pays off: The ADP clearly had maintained focus on understanding both the “how” and the "why" of the target population's journey through a complete trip. The ADP clearly understood that data users are looking for is personal and requires customization, and developed the technology to be responsive to that need rather than try to hammer all of the target populations' needs into a single data rubrik. The ADP also understood the manner and workflows through which the target population may be looking to access information, this was demonstrated in the same application containing both the virtual walkthroughs as well as the active on-location wayfinding mode.
- Needs assessment must extend to developing a safety risk model for the target population: The ADP addressed the safety concerns to the target population through providing a baseline assessment to ascertain that the travelers have a required minimum skillset to allow safe use of the mobile application during travel. The Independent Evaluation believe this was a demonstrably important step to verify users and be able to embed baseline assumptions about the safety concerns of a traveler using the application. The IE believes that even with that baseline assessment of primary users, it is important to build in safety features into the application (for example, enables user's call for help in every view of the application) to address any safety concerns that may not have been modeled by the ADP team.
- Needs assessment and proper traceability matrix should extend to any secondary users or supporting users: Many projects concerning people with disabilities involve additional users in the loop (whether they are family members, caregivers, therapists, disability practitioners, or even travel agency personnel). The WayFinder app embeds a reliance on such secondary users by requiring trip routes to be defined in a customized manner for the primary users. While the route-planning workflow was not part of the evaluation process, even in the wayfinding workflow, there are certain touchpoints (like preference selection and configuration) that assumes the involvement of these secondary users, but these users were seemingly invisible in the design and development process (or at least in the documentation provided to the IE). We believe this resulted in a missed opportunity to involve these secondary users in a user experience that was seamless and appropriate for this population.

Chapter 2. Evaluation Background and Data Collection

Accessibility and usability are nonbinary. It is rarely possible to say with certainty that a given technology or program is accessible, or that it is usable by all. This is in part because both accessibility and user experience of a particular technology are defined in relation to the technology's operator(s) and the situation(s) and environment(s) in which the technology is operated. Framework 2020 defines accessibility in relation to its function in a particular situation: is a technology accessible in a certain environment to a certain user population? Similarly, this framework defines usability as a function of a subset of human operators: are particular people able to independently accomplish essential tasks with the technology without becoming frustrated or confused? To define a methodology for performing a useful independent evaluation (IE), we must focus the analysis on both the travel needs that the technology seeks to ameliorate and the people it seeks to help.

This document specifically focuses on exercising and reporting on an evaluation of an Accessibility Development Projects (ADP) technology that already exists. The IE team linked earlier-stage activities of need-finding to other documents (see Accessible Transportation Technologies Research Initiative (ATTRI) User Needs Assessment: Stakeholder Engagement Report, May 2016), and the plan for the evaluation is located in the IE Plan 2020. Additionally, we focus this document on the complete-trip evaluation hypotheses and gap analysis. The IE team understands and practices the process of performing detailed technology validation and audits (both at the architectural and code levels, as well as mobile application performance characteristics). The IE team believes that a comprehensive security audit as well as a mobile performance assessment is essential, but outside the scope of this evaluation.

The purpose of an IE, described in Framework 2020, is to assist the ADP team in creating tools that remove barriers for the target population. Their primary objectives aim to enhance and maintaining access to mobility and transportation. The framework is designed to approach travel comprehensively with special attention to the "complete trip," which allows for the creation of a customized, logic model with the following foci:

- **ADP—**Indicates a particular Accessibility Development Project.
- **ADP Goal Evaluation Hypotheses—**Indicates creating evaluation hypotheses based on each of the project goals for the specific ADP. The project goals clarify the issues addressed by the ADP and what objectives each ADP is trying to achieve. The evaluation hypotheses are derived from project-specific goals.
- **ADP Populations Addressed—**Indicates the step of identifying target populations, their different abilities and needs.
- **ADP Environments Addressed—**Indicates the step of identifying trip activity links (TAL), potential geographic locations, and types of built environments in which the ADP technology might be used in the context of the "complete trip" within each of the 11 TALs.

- **Complete Trip Evaluation Hypotheses—**Indicates the step of identifying evaluation hypotheses for the ADP through the inquiry process that is described by the evaluation contexts named by the Framework 2020 report.
- **Performance Metrics—**Indicates the performance metrics that were used to measure effects of the ADP according to the evaluation hypotheses for the specific ADP.
- **Data Types and Sources—**Indicates each data source that was used in the creation of performance metrics.
- **Method of Evaluation—**Indicates each of the qualitative and quantitative evaluation methods that were used.

Evaluation Workflow Restatement

This report describes an instance of the Framework 2020 process by which ATTRI, IE, and other project partners completed an ADP evaluation in accordance with the IE Plan 2020. The ADP evaluation focuses on equity, accessibility, and inclusion along the complete trip.

To achieve that outcome, the evaluation process contains multiple parts as described in chapter 3 of Framework 2020. In particular, the basic evaluation follows the same workflow. The next two chapters of this document (#REF) contain the last two steps outlined in the following workflow:

1. Set up the IE by identifying the key details required for the evaluation and reviewing the goals and expected performance outcomes of the ADP:
    - Understand the target population for the ADP.
    - Understand the travel activities.
    - Understand the independent evaluation sponsor's priorities.
2. Develop a threat model.
3. Develop an evaluation logic model.
4. Develop logic model hypotheses for changing the performance of travel activities.
5. Develop logic model hypotheses for measuring the ability to mitigate threats.
6. Develop logic model hypotheses for measuring the ADP's ability to address the target population's travel needs.
7. **Perform the evaluation.**
8. **Potential gap analysis**.

This document relies heavily on the independent evaluation plan described in IE Plan 2020.

The Independent Evaluation's Stated Focus

The IE will focus on overall trip completion, user errors during trips, proper user localization, and the effectiveness of user alert notifications. Other aspects of the technology evaluation (such as

comprehensive robustness testing, or the application's ability to communicate information other than warnings effectively) are deemed outside the scope and capabilities of this IE team under the constraints of the evaluation. As the IE team continued the evaluation, mobile application performance arose as an important area of focus. Although application performance ended up directly impacting some of the evaluation hypotheses, in IE Plan 2020, application performance was deemed outside the scope of this IE. This gap will be discussed under the gap analysis presented in chapter 4.

Evaluation Data Collection

Field Testing by Merakey (Field Test Data #1)

The first field test procedure for this IE was completed by Merakey, which allowed the IE team to keep evaluation costs under control by collecting multiple sources of global positioning system (GPS) data simultaneously. Through data collaboration and evaluation data sharing, this IE was able to avoid the costs associated with a completely separate additional deployment and participant recruitment effort. In chapter 3, results of the evaluation will name specific places of the study where a more comprehensive method could have been used but was neglected to remain within reasonable costs and the limitations of the data collected by MeraKey.

The MeraKey data was collected with participants who were either primary users of the application or caregivers who would be secondary users of the WayFinder 3 application. All participants were initially asked background questions to establish the primary user's baseline travel information, which included: age, diagnoses, level of mobility, level of experience with a smartphone, and what modes of travel they used regularly. During the main data collection process, each primary user took from three to twelve trips each. Each trip was along a predesigned, canonical trip route that was programmed into the Wayfinder 3 application. Most participants took one trip to their destination and another trip back from that same destination, so on a day of data collection, they would have taken two trips. Some participants only took one-way trips. There were 106 round-trips and 45 one-way trips. While they took each trip, their phone collected GPS data from the AbleLink application, and whenever possible, it also took GPS data from another application called AWARE. Merakey was not able to collect AWARE data from every phone.

AWARE is an open-source framework that enables users to share mobile data. During the Merakey field tests, AWARE utilized several phone sensors to collect data while participants took their trips. This way there were multiples sources of GPS and battery data for each trip. The IE team initially intended to collect memory data with AWARE, but due to unforeseen errors, memory data was not collected.

During the data collection process with Merakey, a staff member shadowed the participants on each trip. Staff assistance varied from sitting on the bus with the participant to following the bus from a distance or simply meeting them at their destination. After the day of travel, the participant and the staff member completed a field test form that collected their short-answer responses to questions about their trips that day. The staff were asked about their familiarity with WayFinder 3, their relationship to the participant, and whether the trip was completed free of any complication. The participants were asked if they got to their stop using the Wayfinder 3 application, if they went off-course during their trip, if Wayfinder 3 helped them ride the bus, if they enjoyed using Wayfinder 3, if they thought that next time they could take the trip by themselves, and what level of assistance they received from the staff. There were 25 participants total. We received field test forms for 23/25 participants (two of them only provided baseline information), and we have dashboard data for 22/25 of the participants.

The Summary Table contains a conclusive report about any issues with the data. For each trip, it contains the participant identification (ID), date, trip route, start and stop times, trip completion (whether they arrived at their final destination using the Wayfinder 3 application), and if there is dashboard data available. Then the next several columns break down the individual reasons explaining why dashboard data is unavailable. The first of these columns is called “date not provided on field test form”. With no date on the field test form, it is impossible to match the field test form with dashboard data. This issue occurred for 5/267 trips. The next column is called “wrong canonical trip”. This means that we do technically have data for the participant, but they chose the wrong canonical trip in the Wayfinder 3 application. Usually this meant that they chose the trip in the opposite direction, so they started at the end point and proceeded to the start point. Trips with this issue often did not give valid waypoint timestamps but they did give valid battery-usage data. This issue occurred for 17/267 trips.

The next two categories of issues are similar. For both types of issues, we have no dashboard data for the trip and no obvious reason why. The first of these columns is titled “missing data, but nearby data” which stands for “missing data but has unused data within one day of trip”. This means that we have no dashboard data for a trip by that user on that date, but there is unused dashboard data for a trip on one of the surrounding days (one day before or after that trip). It is possible that when the staff member and participant filled out the field test form, they recorded the wrong date. This issue occurred for 16/267 trips. The next column is called “data completely missing”. This means there was no dashboard data for a trip by that user on that date, and there also were no trips by that user on surrounding days. This issue occurred for 52/267 trips.

The final column in table 1 is called “data is confusing”. This means that between the field test forms and the AbleLink dashboard data, the data are unclear, involve large gaps or are impossible to resolve comments on the form. This issue occurred for 5/267 trips.

In total, 94/267 trips were impacted by at least one of the issues outlined above, rendering 68/267 trips completely unidentifiable in the AbleLink dashboard. The trips without corresponding dashboard data will not be used for any of the hypothesis validations. To recover data missing from the AbleLink dashboard, it is necessary to search the AWARE dataset.

Table 1. Summary table for Merakey data collection.

Trip Dates	Trip Completion	Any Dashboard Data Available	Usable/Viable Data	Date not provided	Wrong canonical trip	Missing data but nearby data	Data completely missing	Data is confusing
6/3/2019–8/12/2019	Yes (20)	Yes (20)	7 return, 8 there	0	0	0	0	2
6/4/2019–8/12/2019	Yes (15), No (1)	Yes (12), No (4)	6 there, 7 return	0	0	4	0	0
6/4/2019–8/13/2019	Yes (14)	Yes (11), No (3)	6 there, 5 return	2	3	0	3	No
6/5/2019–7/17/2019	Yes (8)	No (8)	0 there, 0 return	2	0	0	8	0
6/3/2019–8/11/2019	Yes (14), No (2)	Yes (8), No (8)	3 there, 5 return	0	2	2	4	0
6/24/2019–9/13/2019	Yes (13), No (3)	Yes (6), No (10)	3 there, 2 return	0	0	4	6	1
6/4/2019–8/20/2019	Yes (16), No (2)	Yes (14), No (4)	8 there, 6 return	0	0	0	4	0
10/2/2019–10/29/2019	Yes (8)	Yes (5), No (3)	4 there, 1 return	0	0	0	3	0
6/28/2019	Yes (6)	Yes (6), No (2)	6 there	0	0	1	1	0

Table 1. Summary table for Merakey data collection (continuation).

Trip Dates	Trip Completion	Any Dashboard Data Available	Usable/Viable Data	Date not provided	Wrong canonical trip	Missing data but nearby data	Data completely missing	Data is confusing
7/17/2019–9/11/2019	Yes (16)	Yes (16)	8 there, 8 return	0	0	0	0	2
6/13/2019–7/25/2019	Yes (8)	Yes (8)	4 there, 4 return	0	0	0	0	0
4/19/2019–7/12/2019	Yes (6), No (1)	Yes (2), No (5)	1 there, 1 return	0	0	0	5	0
6/5/2019–7/29/2019	Yes (12)	Yes (7), No (5)	1 there, 4 return	0	2	2	3	0
6/13/2019–9/12/2019	Yes (16)	Yes (14), No (2)	7 there, 6 return	0	1	0	2	0
6/4/2019–8/27/2019	Yes (16)	Yes	5 there	0	0	0	10	0
?–7/25/2019	Yes (6)	Yes (4), No (2)	3 there, 1 return	2	0	2	0	0
6/7/2019–8/16/2019	Yes (11)	Yes (9), No (2)	5 there, 4 return	1	0	0	2	0
6/5/2019–8/21/2019	Yes (16)	Yes (16)	9 there, 6 return	0	1	0	0	0

Table 1. Summary table for Merakey data collection (continuation).

Trip Dates	Trip Completion	Any Dashboard Data Available	Usable/Viable Data	Date not provided	Wrong canonical trip	Missing data but nearby data	Data completely missing	Data is confusing
6/28/2019–7/30/2019	Yes (4)	Yes (4)	4 single	0	0	0	0	0
5/31/2019–9/6/2019	Yes (11)	8 single G, 3 single H	7 single	0	4	0	0	0
6/26/2019–8/22/2019	Yes (3)	Yes (1), No (2)	2 single	0	0	1	1	0
5/30/2019–9/19/2019	Yes (3), No (4)	Yes (7)	3 single	0	4	0	0	0
5/30/2019–9/19/2019	Yes (10)	Yes (10)	10 single	0	0	0	0	0

Source: Federal Highway Administration.

Field Testing by Taskar Center (Seattle Field Test Data #2)

The IE team at the Taskar Center performed field tests to fill potential gaps in the Merakey data collection. The field tests took place in the Seattle area and followed three routes in distinct, urban environments: open spaces, dense high-rises, and in the vicinity of waterways. Open spaces and waterways were tested through routes near the University of Washington campus. Areas with high rises were tested through a downtown route. Each of the three routes had at least ten waypoints. Researchers completed each route three times.

The tests were conducted with one professional tester from the IE team and one tester who self identifies as a person with a disability, who typically does travel independently but only on prescribed routes through which this tester has been coached by a parent or caregiver. The professional testers were team members from the Taskar Center who had relevant knowledge about technology development and inclusive design. The tester representing the primary AbleLink target user group who participated in the tests was supposed to perform nine trips; however, the tester was only able to complete three trips, accompanied by the professional tester. The tester attempted two other trips independently, but the tester's phone turned off in the middle of the route due to battery drain by the WayFinder application. Rather than risk their battery running out while they were traveling unassisted, the IE team decided to involve the testers from the primary AbleLink user group only in assisted scenarios, which resulted in three trips for that tester.

The field tests evaluated the accuracy of GPS data in addition to other features support features. During trips, the "contact me" button was turned on so that the IE team could evaluate notifications and assistance features. The IE team also collected information about error messages and the responsiveness of other on-screen displays. To account for differences across operating systems, each trip was taken by the professional testers carrying two devices: one iOS device and one Android device. Data was collected on both devices through the AbleLink dashboard and the accompanying AWARE monitoring application.

Heuristic Evaluation Testing by Taskar Center (Heuristic Evaluation Test Data)

The heuristic evaluation was used to contextualize and analyze the data from the Taskar Center field tests. While the professional tester conducted the field tests, they recorded an audio log to note important information for the heuristic evaluation. The contents of the audio log allowed them to respond "yes", "no", or "N/A" to a series of questions. The questions in the heuristic evaluation were divided into five main categories: robustness; location-based services; error prevention, call for assistance and failure modes; adaptability; and communication across information gaps. Only a few questions for each category will be exemplified in the following paragraph. Appendix A shows the categories of data collected for the complete Heuristic Evaluation).

To evaluate robustness, the heuristic evaluation included questions about whether the WayFinder 3 application functioned properly on both iOS and Android devices, whether the app launched with a display of instructions, and whether the device properly vibrated. These questions evaluated general usability of the application. The next category, location-based services, included questions about whether the WayFinder 3 app launched GPS-based instructions within the appropriate distance, whether it provided signal range, and whether it automatically stopped displaying the waypoint message when the user was

out of range. The third category, “error prevention, call for assistance and failure modes”, included questions about whether the Contact Me button properly sent email notifications, whether the user could request a caregiver to call them from within the WayFinder 3 app, and whether they could send an “I’m okay” message from within the WayFinder 3 app. This category helped to ensure user safety and support. The penultimate category, adaptability, included questions about whether the user could adjust the size of onscreen text, whether they could enable the GPS to automatically optimize for a slower travel speed, and whether they could enable the Exit button to remain visible throughout their trip. These essential options created a customizable experience for users with different levels of mobility. The final category, communication across informational gaps, included questions about whether there are picture prompts for travel instructions, whether the user could enable audible feedback, and whether the Route Label Size enabled users to change to size of on-screen text.

Alongside the data from the AbleLink dashboard, the information from the heuristic evaluation provides a comprehensive picture of the WayFinder 3 application’s performance. The heuristics were designed with consideration of ATTRI’s mission to promote accessibility and inclusivity, Jon Schneiderman’s *Eight Golden Rules for Interface Design*, and Jakob Nielsen and Ralf Molich’s *Ten Usability Heuristics for Interface Design*.

Chapter 3. Evaluation Hypotheses Testing

Traveler-centered Evaluation: Effectiveness, Efficiency, and Equity

The first priority of the Independent Evaluation (IE) Plan 2020 is to understand the technology's impact on travelers' ability to execute the "complete trip." The first hypothesis tested with the WayFinder 3 assesses if users of the technology complete more independent trips after beginning trip assistance with the WayFinder app. This interpretation of Effectiveness is loosely derived from International Standard Organization (ISO)/International Electrotechnical Commission 9126-4 Metrics.

Hypothesis: The WayFinder 3 app is effective in improving primary users' overall independent trip completion.

Hypothesis demonstration: Inter-user, trips using WayFinder show increasingly successful completion of travel, even when users enable multimodality options.

Performance Metrics:

Completion Rate—We derive effectiveness from completion rate, which is a fundamental measurement of usability. It is calculated by measuring trip completion counts over the total attempted trips over a specific period of time.

The IE Plan 2020 describes measuring the effectiveness for a specific participant by segmenting trips attempted and computing the trip completion rate for the participant over that set of trips. Effectiveness for a participant is 1 if participant's completion rate increased over time. Effectiveness for a participant is 0 if the participant's completion rate did not increase over time.

Effectiveness of the application, given a particular study population, is the percent of the study population for whom effectiveness was 1.

For this hypothesis, the IE Plan 2020 outlined a procedure whereby initial trip completion was to be calculated using the prior trip completion rate (as indicated by surveying participants about trip completion rate prior to the Merakey Field test). Unfortunately, the MeraKey background surveys were not instrumental in reliably accessing that baseline information. Therefore, the IE team began with the first four trips completed in the duration of the Merakey Field tests. The IE team evaluated the reported completion rate for every four attempted trips during the Merakey Field test (this should represent about two weeks of trials per participant based on the Field Test Protocol). Then the IE team looked for overall trends in completion rate.

The IE team assessed that the evaluation hypothesis is true if the population mean and two standard deviation of the mean were able to increase or maintain their rate of completion for trips. The IE team did not require 100 percent improvement for the entire participant population due to data collection errors and population variability that is frequent in Field Tests like these. To determine the acceptance threshold for this test, the IE team used the commonly accepted **Chebyshev's rule** (Larose, Daniel T. Discovering statistics. Macmillan Higher Education, 2009) to establish the acceptance rate for this test. The rule states that without making any assumption about the underlying population, at least ¾ of the participants will be within two standard deviations of the mean, that is, in the interval with endpoints $\mu \pm 2\sigma$ for populations. Therefore, the acceptance rate for this test requires overall monotonic improvement in completion rates for at least 75 percent of participants.

Data Elements and Sources:

Trip Completion (source: Merakey Field Test): measured by assigning a binary value of "1" if the test participant completes a trip independently and "0" if he/she does not.

Pre-Study Trip Completion (source: Merakey Field Test): the caregiver reported an estimate of the user's rate of independent trip completion before the AbleLink intervention.

Data Collection Period:

Data collected via the Merakey Field Test period.

Analysis Procedure:

For each Merakey Participant, (i), calculate the following:

1. Collect the Merakey Participant's caregiver-reported trip completion rate per the pre-study survey. Call this the pre-study completion rate (CR) for participant i, CR ip.
2. Assign a completed (1) /not completed (0) status to each of the trips performed (this is a practical application of discretizing data into binary variables).

Calculate the participant's overall in-study trip completion rate as:

$$\frac{Number\ of\ trips\ successfully\ completed}{Total\ number\ of\ trips\ undertaken}\ x\ 100\%$$

Source: IE Field Knowledge.

Figure 1. Equation. Completion rate: calculated as number of tasks completed successfully, divided by total number of tasks undertaken, multiplied by (100 percent).

Segment the participant's trips into four sequential attempts (about two weeks' worth of trips per the Merakey protocol) and calculate the **completion rate as CR_{i1}** as the completion rate for participant i for the first four trips, CR_{i2} for participant i for trips five through eight, etc.

If $\mathbf{CR_{ip} \leq CR_{i1} < CR_{i2}}$, then this participant's completion rate has improved over the course of the study. Let $\mathbf{Eff_i} = 1$ to designate that the app was effective for participant i. Otherwise, set $\mathbf{Eff_i} = 0$. When any of the values are missing (for example, if baseline trip completion for participants is not available), this test will

require that the participants completed at least 10–12 trips to establish 3 CR values to be able to identify a trend.

When possible, graph the CRs for each participant.

Summary effectiveness statistic: Calculate overall effectiveness rate for the AbleLink #1 app, as demonstrated by the MeraKey field study as:

$$Eff = \frac{Sum\ Eff_{i\ \ for\ all\ i \in Participants}}{Total\ number\ of\ participants}\ x\ 100\%$$

Source: IE Field Knowledge.

Figure 2. Equation. Effectiveness of the tested application is the sum of effectiveness for all the participants, divided by the total number of participants in the study.

Hypothesis is accepted if Eff>75. Please note that this means that the app was effective in increasing independent completion of trips for at least 75 percent of participants, not that over 75 percent of attempted trips were actually completed. This diverges from other uses of effectiveness metrics and acknowledges that heterogeneity across the participant population may not yield uniformly-high completion rates.

Analysis Outcome

Data Cleaning

The Merakey pre-study survey contained the following question from which the IE team was intending to estimate each participant's trip completion rates prior to the study:

> *Q5: Current travel methods, frequency, and purpose (2 months of exact data from their records, doctor's appointments, day program, community, and work)*

However, only five of the participants had responses to these questions, and those responses were collected in free form (as opposed to the exact data from record). Therefore, the IE team determined that there was no manner in which this data could be used to deduce precise, pre-study CR for participant i, CR ip, as described by the IE Plan 2020. Consequently, the completion rate trends for each participant are completely reliant on trip groupings of the first, second and third 4-trip chunks executed by the travelers. To calculate at least three completion rates for the trend comparison, the IE team could only use participants who executed at least 12 trips using the AbleLink technology. Specifically, for the data of any participant to be accepted into this analysis, the IE team required the data to calculate the three **completion rates CR_{i1}, CR_{i2}, and CR_{i3}** (the completion rates for participant i for the first four trips, CR_{i2} for trips five through eight, CR_{i3} for trips nine through twelve, etc.

The hypothesis evaluation looking at trends in effectiveness had to be changed from the IE Plan's original (**$CR_{ip} \leq CR_{i1} < CR_{i2}$**) to (**$CR_{i1} \leq CR_{i2} < CR_{i3}$**). The hypothesis evaluation still tracked progression of each participant with the application, but without using the missing pre-study data.

Findings

A total of 11 participants completed at least 12 trips and were included in the analysis for this hypothesis. Participants who attempted fewer than 12 trips were not concluded in this analysis because it was not possible to construct three or more completion rate values (see note above regarding the inability to use the background Merakey Field test survey as a baseline Completion Rate, CR). The 11 participants' trip-completion effectiveness is charted in table 2. Each participant had 3 CR values. A "25" for CR_i1 means that the user completed 1/4 of their first four trips. A "75" for CR_i2 means that the user completed 3/4 of their second four trips.

Table 2. Participant trip completion effectiveness.

Participant ID	CR_i1	CR_i2	CR_i3	Eff_i
AB01	100	100	100	1.00
AB02	100	75	100	0.00
AB03	100	100	100	1.00
AB05	100	100	100	1.00
AB06	25	100	100	1.00
AB07	100	100	50	0.00
AB12	100	100	100	1.00
AB15	100	100	100	1.00
AB16	100	100	100	1.00
AB17	100	100	100	1.00
AB20	100	100	100	1.00

Source: Federal Highway Administration.

Trip completion effectiveness increased or remained constant for 9 out of 11 participants. The overall participant summary effectiveness (Eff) was 0.818 percent, as calculated per figure 2.

Hypothesis Acceptance

Since overall participant effectiveness was over 75 percent, the hypothesis is accepted.

Hypothesis: The WayFinder in-app notifications are effective in reducing primary users' unintended, midtrip errors.

Hypothesis demonstration: Intra-user, over the study period, trips performed using WayFinder show a decline in midtrip, unintended veering off route in response to in-app notifications to the primary users.

This hypothesis examines whether or not specific in-trip errors that can impact the completion of a "Complete Trip" are effectively remediated by using the AbleLink technology. By using these secondary means to assess effectiveness, the IE team analyzes the complete trip holistically, rather than focusing only on end results.

The hypothesis tests whether the AbleLink use of in-app notifications to the primary users are effective in preventing users from making in-trip unintended errors.

Performance Metrics:

We are interested in measurements involving effectiveness of the Accessibility Development Projects (ADP) technology respective to the successful execution of the trip activity links (TAL), which pose difficulties for the population of interest. Two types of midtrip errors that impact completion of TALs are veering off route and off-boarding transit at the wrong stop. Both were assumed to be measurable by the IE Plan 2020 via the WayFinder application. Other important midtrip errors that are not measurable using the MeraKey field studies but are frequent with this primary user population include: interacting with people in socially unexpected ways, misinterpreting traffic during street crossings, slips or omissions that a participant makes while attempting a task (for example, while trying to purchase a ticket at a point of sale). The IE Plan foresaw challenges in combining data from both types of midtrip errors that were assumed to be measurable, due to the requirement for normalization and biased error correction that is outside the scope of this independent evaluation.

There also are additional ethical complexities that are not described in this IE plan. For example, the IE team decided to not put participants in a potentially unsafe situation by having them intentionally veering off a familiar route (to evaluate if the application's methods are effective in redirecting the participant of correcting their mid-trip error). That is, in an ideal, controlled data collection, the collection process would have field testers reporting the number of times the caregivers redirected participants back onto their route after nearly making an error, the number of times the app intervention redirected participants back onto their route, and the number of times neither the caregiver nor the app intervened and the participant either self-directed back to the path or veered off completely. These types of independent tests are difficult to complete and very difficult to control.

The IE Plan suggested using a surrogate, biased aggregate of the number of times the participant veered off the path. The field tests included two defined paths for each participant. One of those two routes was intended to have a walking path as one of its TALs. As the IE team later found out, there was insufficient data for canonical trips that included a walking component in the MeraKey field tests. Therefore, it is ineffective to evaluate mid-trip errors with the Merakey field tests because the participants were on a fixed bus route for the majority of the MeraKey field tests.

The IE Plan 2020 described a procedure by which a performance measure, E^{v}_{ij} could be calculated, estimating the number of attempts by participant (i) in trip attempt (j) to veer off the path. The IE Plan identified such attempts by merging data about in-app notification triggers (which were assumed to be part of the AbleLink dashboard data) with data about when the independent global positioning system (GPS) trace veered more than 10 meters (m) away from the route (the latter assuming that all tester phones will be collecting GPS data both on AbleLink's servers as well as with a separate application named AWARE, described in Ferreira's work.[3]).

[3] Ferreira D, Kostakos V and Dey AK (2015) AWARE: mobile context instrumentation framework. Front. ICT 2:6. doi: 10.3389/fict.2015.00006.

The evaluation of E^{v}_{ij} was to be done by aggregating for participant (i), during trip instance (j), the number of times (t) for which both the Merakey trial dashboard data indicated that veering off the path had triggered and the alternative GPS trace data indicated the participant distance from the intended route was greater than 10m. The calculation rested on several assumptions that did not hold true upon examination of the MeraKey data: (1) routes were assumed to contain some walking path components that were going to be tested, whereas all the participant trips were exclusive to bus travel (2) AbleLink dashboard only recorded in-app message triggers in real-time, and did not record these triggers and associated timestamps as part of the saved archive for that trip (therefore, the IE Team was not able to obtain that information for trips retrospectively) (3) AbleLink GPS trace data saved measurements intermittently in temporal density that would not provide the resolution necessary to identify veering off path behavior. Due to these unsupported assumptions, this metric, which was acknowledged in the IE Plan to be approximate and error prone, was not data that could be collected by the IE team. Instead, the IE team collected the comments relating to in-app triggers and any mid-trip errors (as reported by caregivers along the route) and provided recommendations to the ADP regarding in-app notifications, triggers, and the challenges associated with providing reliable metrics to demonstrate this intervention mechanism's effectiveness.

Data Elements and Sources:

Due to data availability, we will not evaluate trip errors that occurred exclusively during the pedestrian path navigation of a trip as this mode was not tested in the MeraKey Field Tests.

Veering off course is a participant error that is made when the participant navigates on their own more than 10m away from the prescriptive route. If the participant is on the bus, any veering off the prescribed bus route is a decision of the bus driver, rather than the participant.

Collected data:

- Coded caregiver comments fitting one of the following categories:
    1. No triggers.
    2. Delayed triggers.
    3. Messaging post trigger is too fast to be confirmed or noticed.
    4. Triggering end of route, or exit route in inappropriate locations.
    5. Unrecoverable application errors (Application quit unexplainably, error codes coming up).
    6. Triggers helpful.
    7. Route resets on its own.
    8. Staff member intervened to (get back on route, or identify a bus stop).
    9. Staff member said participant went off course.
    10. Misc.

  In the ideal scenario, we would be able to assign a short description along with a severity rating associated with each error and then be able to subclassify them under the respective category.

Data Collection Period:

The data collection period is congruent with the Merakey field study period.

Analysis Procedure:

The IE team conducted a coded caregiver comment evaluation about in-app triggers due to the unavailability of data to support the IE Plan 2020 intended univariate Mann-Kendall Test to assess any trends in the data for each participant.

Analysis Outcomes:

The hypothesis was neither accepted nor rejected on the basis of insufficient data to support a stringent analysis of the application's in-app notifications and their effectiveness.

However, a coded analysis of the comments made by caregivers about errors and observations during the Merakey Field Tests led to a number of findings and recommendations to the ADP team regarding in-app triggers. Caregiver reports were written to be as detailed and specific as possible and included the issue found, but may have left out relevant details such as what the task attempted was, where they encountered the problem, or a screenshot. In our summary of the reported problems, we determined the internal trip identification (ID), the participant whose trip was impacted (to see whether the same trigger issues were related to the same participant and phone or to different participants), the comment made and its coded interpretation.

Comments were separated by the larger comment theme domain or attributes and listed below. It is the IE's assessment that many of these issues are imperative to address before the in-app triggers (or in some cases, the whole product) should obtain wide release.

Findings

Findings are presented by theme:

In-app notifications were helpful where they were invoked.

The results for in-app notifications are promising; however, there is not enough data to support a positive conclusion. Only one user's caregiver (for AB05) had positive experiences with in-app notifications. They reported that "the prompts were extremely helpful, as it brought her attention back to the app."

Recommendation: The IE team believes AbleLink should invest in improving in-app triggers and in collecting data that allows AbleLink to measure in-app notification impact on the user experience and trip completion.

Table 3. Coded Comments for "triggers were helpful..." from MeraKey Field Test.

Code	Comment verbatim	Survey Question	Participant
Triggers were helpful in moving along the route	"The prompts were extremely helpful, as it brought her attention back to the app.	PQ5	AB05

Source: Federal Highway Administration.

Caregivers describe having to intervene (combining coded comment domains: “Staff member intervened to (get back on route, or identify a bus stop)”, and “Staff member said participant went off course”)

In 13/48 trips caregivers reported having to intervene on the participant’s behalf. These trips were associated with 6 of the 25 travelers. Some of these comments explicitly suggested that in-app notifications were not appearing at appropriate times. Regardless, these are instances in which the WayFinder application is not achieving the desired impact of either redirecting the traveler or effectively reminding travelers about their next action. The caregiver who worked with AB06 wrote, "Reminder given to pull the cord," which was a common issue. AB12 also required reminders: “needs verbal prompts on using the app, when to get on/off bus”. Still, other caregivers reported giving assistance with more basic tasks. Overall, our findings indicated that the most common reason a caregiver intervened was to remind the user to disembark the bus. Unfortunately, the unstructured caregiver comments that were coded in these two domains did not always provide the information needed to assess whether the participant was unaware of an in-app notification or whether the notification was never triggered. However, it is indicative that in-trip errors of the type anticipated by the Risk Model described in IE Plan 2020 are frequent in this population and must therefore be addressed.

Table 4. Coded comments for “participant went off course” and “staff member intervened” from MeraKey Field Test.

Code	Comment verbatim	Survey Question	Participant
Staff member commented that participant went off course	"Off course (to the mall) Perfect (to Giant Eagle)"	PQ5	AB15
Staff member intervened to get back on route or id a bus stop	We were able to get where we needed to go with staff assistance.	PQ5	AB03
Staff member intervened to get back on route or id a bus stop	"Reminder given to pull the cord"	PQ5	AB06
Staff member intervened to get back on route or id a bus stop	"Needs verbal prompts on using the app/when to get on/off bus"	PQ5	AB12

Table 4. Coded comments for "participant went off course" and "staff member intervened" from MeraKey Field Test (continuation).

Code	Comment verbatim	Survey Question	Participant
Staff member intervened to get back on route or id a bus stop	"Staff asked several times during the ride if this was the stop participant not sure"	PQ5	AB15
Staff member intervened to get back on route or id a bus stop	"When staff asked if we were at correct stop participants were not sure"	PQ5	AB20
Staff member intervened to get back on route or id a bus stop	"Confused on when to get off and her app while riding bus"	PQ5	AB20
Staff member intervened to get back on route or id a bus stop	"[She] needed verbal prompts to help her stay focused where she was walking [and] on the bus to and from the mall"	PQ5	AB20
Staff member intervened to get back on route or id a bus stop	"Needed verbal prompts on getting to/from the mall. [She] needed verbal prompts on using the app."	PQ5	AB20
Staff member intervened to get back on route or id a bus stop	"[She] needed verbal prompts to help her use the app/when to get off bus"	PQ5	AB20
Staff member intervened to get back on route or id a bus stop	"[She] needed verbal prompts on staying focused while using the app. Also when to get on/off the bus"	PQ5	AB20
Staff member intervened to get back on route or id a bus stop	"[She] needed verbal prompts on the Wayfinder app and how to use it. She had a bit of confusion"	PQ5	AB20

Source: Federal Highway Administration.

Recommendation: The IE team recommends that the AbleLink team build more robust analytics into future research to investigate the triggering and effect of in-app notifications. It would be most prudent to enhance both specific post trip survey information, which would offer better resolution on the relationship between in-trip errors and in app notifications, as well as in-app logging capabilities that improve analytics about when notifications were triggered. The IE team suggests the following questions be added to the post-trip caregiver survey.

For Staff post-trip survey:

If you had to intervene in the trip, please tell us how many times you had to intervene with each type of intervention:

1) NO interventions were needed at all. Participant completed entire trip with app directions. [Mark with an X _____]

2) I had to **tell** the participant or **physically guide** the participant if they made a navigation error too early (for example, turned one corner too early, or got on the wrong bus in anticipation of a bus that would come later, or got off the bus too early) [How many times tell them_____ , how many times **physically guide** them___]

3) I had to **tell** the participant or physically guide the participant if they made a navigation error due to a delay (for example, missed a turn at the corner and kept walking, or missed getting on the bus that came or got off at the next stop) [How many times had to tell them _____ , how many times **physically guide** them____]

4) I had to alert the participant about an up-coming event (for example, bus about to arrive, bus about to stop at your stop) [How many times _____]

5) I had to prevent the participant from making a navigational error (like getting on the wrong transit vehicle) before they made the error [How many times _____]

6) I had to **tell** the participant to pay attention to the app (for example, when I noticed they were not paying attention) [How many times _____]

In-App Triggers not invoking or delayed.

It was previously acknowledged that the WayFinder 3 application must focus on GPS tracking accuracy because the foundational premise of the trip concierge/helper guiding system, for this population of interest in particular, rests on the availability and appropriate use of alerts and in-app notifications to orient or re-direct the user to what is happening, forecast what is about to happen, and provide important instructions on-location. Since the primary user is moving, providing these in-app notifications in a timely, location-based manner is of utmost importance. The comments extracted from the post-trip survey data suggested that further work is required to get triggers to (1) invoke at all (these are considered false negatives) (2) invoke at the correct location (considered true/false positives) (3) play back information at a human-pace, allowing participants time to respond with an acknowledgement or (4) triggering in the wrong sequence (also considered false positives).

Table 5. Coded comments for "no triggers," "delayed triggers" and other trigger issues from MeraKey Field Test

Code	Comment verbatim	Survey Question	Participant
No Triggers	It never told us passing landmarks or when we were arriving	PQ5	AB03
(as above)	"Landmarks don't pop up during the trip"	PQ5	AB07
(as above)	"Landmarks don't pop up during the trip"	PQ5	AB07
Delayed Triggers	"Kept saying (off route) and landmarks came too late (2 blocks past)"	PQ5	AB15
(as above)	"Still slow - landmarks late"	PQ5	AB15
(as above)	"Also, GPS points seemed slightly delayed"	PQ5	AB05
Messaging after trigger doesn't wait for confirmation	However, sometimes the voice prompts did not fully _____ * Cut off at bottom of page *	PQ5	AB05
Triggering end of route or exit route at inappropriate locations	"3x their phone asked if they wanted to exit the route. "	PQ5	AB01

Source: Federal Highway Administration.

Such concerns came up in 9 separate trips, for 6 different users, out of a sample of 48 trips and 24 users that had comments. A general estimate is that one out of five trips that were impacted by an application-related issue had to do with the triggering system. As a note, all of the phones in the MeraKey data collection were iOS-based devices. The IE team experienced similar problems with the Android-based device, but only had one test device. The findings indicate that in-app notifications did not invoke at the right place or the right time with non-negligible frequency.

Recommendation: Although coded comments are not quantitative data and provide a rough estimate, this finding indicates that the AbleLink team should investigate in-app triggering issues further. In particular, since in-app notifications are the primary mechanism used by the WayFinder application to mitigate traveler risk, understanding and correcting triggering problems is of utmost importance. Additionally, the AbleLink team must perform robust testing on multiple platforms to ensure the in-app notifications are triggering properly on both iOS and Android devices.

Hypothesis Neither Accepted Nor Rejected due to lack of formal data. Comment coding resulted in findings and recommendations to improve in-app notification triggering and sequencing.

Hypothesis: The Technology improves or maintains primary users' time efficiency while navigating legs of trips.

Hypothesis demonstration: Intra-user, over the study period, trip legs performed using WayFinder show increased or unchanged relative time efficiency.

The objective of this hypothesis is to test the level of effectiveness achieved using the AbleLink technology compared to the resources expended by the user. Efficiency is generally assessed by the mean time taken to achieve a task. A common measure of efficiency is task time. In the context of trips, the task time is the time (in seconds and/or minutes) the participant takes to successfully complete a task. Since the trips are subdivided into natural waypoints, reaching each waypoint is considered a task in this independent evaluation (IE).

The IE Plan, 2020 described a procedure whereby the time taken to complete a task is calculated by identifying the time differential between two waypoints reached as described below:

*w = Waypoint/Task Time = End Time – Start Time**

**which may be the end time of the previous waypoint if applicable*

Source: Wesson, Janet, and Darelle Van Greunen. "Visualization of usability data: measuring task efficiency." Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology. 2002.

Figure 3. Equation. Time taken to complete task.

Performance Metrics:

The IE Plan, 2020 called for calculating two efficiency metrics:

Per-user time-based efficiency, defined in the field through metricizing (Effectiveness/Task Time):

$$time\ based\ efficiency\ for\ user\ i = u_i = \sum_{w=1}^{T_i} \sum_{k=i}^{N_i} \frac{c_{iwk}}{t_{iwk}}$$

Source: (Bevan, Nigel, and Miles Macleod. "Usability measurement in context." Behavior & information technology 13.1-2 (1994)).

Figure 4. Equation. Per-user time-based efficiency metric.

Per-waypoint relative task efficiency:

The relative task efficiency evaluates the ratio of the time taken by user i to successfully complete a specific waypoint task (through each of the replicated route travel executed during the MeraKey field study), in relation to the total time taken by the same user for all waypoint tasks. The equation is represented as:

$$for\ waypoint\ j\ in\ user\ i^{'}s\ route{:}\ u_{rel\ ij} = \frac{\sum_{w=1}^{Ti} \quad c_{iwj}\ \ t_{iwj}}{\sum_{w=1}^{Ti} \quad \sum_{k=1}^{Ni} \quad t_{iwk}}$$

Source:(Bevan, Nigel, and Miles Macleod. "Usability measurement in context." Behavior & information technology 13.1-2 (1994)).

Figure 5. Equation. Relative task efficiency metric.

Data Elements and Sources:

In executing the evaluation plan, the IE team was able to derive all the data elements from the Merakey Field Testing for participants whose trip data collection included AbleLink Waypoint data. (@dylan- we should identify how many these were)

Let i consistently represent participant i.

Let ui represent user i's time-based efficiency.

Let Ti represent all of user i's trips executed over the study period (in all routes).

Let Ni represent all of user i's waypoints (in all routes).

Let c_{iwk} = the result of waypoint task k during trip w by user i; if the user successfully completes the task, then $c_{iwk} = 1$, if not, then $c_{iwk} = 0$.

Let t_{iwk} = The time spent by user i to complete waypoint k during trip w. If the task is not successfully completed, then time is measured till the moment the user quits the task or requests staff intervention or staff intervenes.

Data Collection Period:

The data collection period is congruent with the Merakey field study period.

Analysis Procedure:

The IE Plan, 2020 identified a mechanism by which to analyze any trends in the waypoint data for any participants whose collections included sufficient data points. With those trends, the IE team conducted a univariate Mann-Kendall Test to identify any trends in the series, even if there is a seasonal component in a data series (Hirsch, Robert M., and James R. Slack. "A nonparametric trend test for seasonal data with serial dependence." *Water Resources Research* 20.6 (1984): 727-732.).

In this case, each participant exhibited a series of relative efficiency for waypoints, going through a waypoint for the replicates of the route. There is a relative efficiency data point for each waypoint in each attempted trip of the particular prescribed trip route in the field tests. In accordance with Merakey protocols, there will be approximately eight replicates per route, per participant who executes that route.

The null hypothesis, H_0, is that there is no trend in the series. The alternative hypothesis, H_a, is that the data follows some trend (may be negative, non-null or positive). The IE team used the implementation distributed as part of the Real Statistics Resource pack for Excel, found in the link below: https://www.real-statistics.com/free-download/real-statistics-resource-pack/.

The IE Plan 2020 set as acceptance criteria AbleLink interventions as effective in increasing traveler efficiency if at least 75 percent of participants trends show a significant upward trend in their relative efficiency for trip waypoints in their routes ($p<0.05$). in the univariate Mann-Kendall Test.

Analysis Outcome:

The application of trend analysis to evaluate the effects of the use of WayFinder 3 depends on the quality of the AbleLink dashboard waypoint timestamp data collection. Data from the AbleLink dashboard was collected by calculating the time difference between the timestamps of consecutive “waypoint hit” demarcations and subtracting any time intervals that were identified between “application paused” time stamps. Give the participant population, the IE team did not expect there to be an abrupt, big change between subsequent trips of the same type.

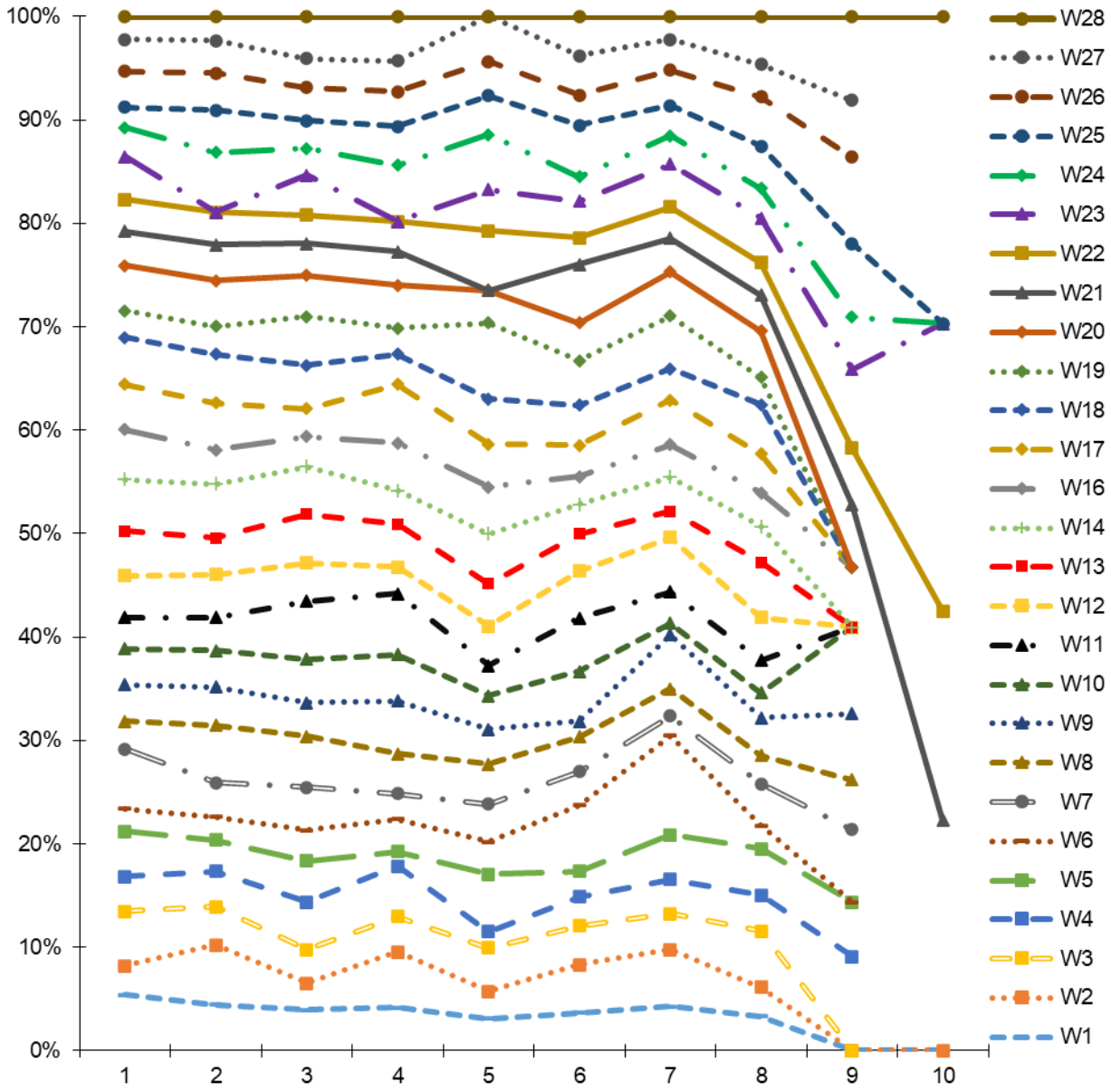
Data Cleaning

The IE team filtered the waypoint timestamp data by participant and then by trip route. This allowed the team to calculate relative increases in time efficiency per user, per waypoint. Next, for each unique trip route that users followed (most participants took two routes), the IE team identified and eliminated data for trip routes that participants traveled fewer than eight times. Finally, the IE team used the univariate Mann-Kendall Test to measure relative time efficiency per user, per waypoint. The results show whether users’ time efficiency increased, decreased, or stayed the same, as they continued taking trips. The inclusion criteria requiring eight trip replicates from each data sequence was motivated by the MeraKey protocol which set eight trips as a goal for participants.

Trend analysis had to account for the variability in quality of data for waypoint timestamps found in the AbleLink data dashboard. The IE team identified many factors that may compromise the quality of data, including:

1. Seasonal cycles in the bus routes.
2. Diurnal changes in traffic of the bus route.
3. Variations in weather conditions impacting driving conditions.
4. GPS measurement error.
5. Participant activities.
6. Actual trends.

The IE team goal was to identify and quantify the actual trend data in a statistically rigorous way. To address the other potential sources of variation, the IE team determined some trips were invalid due to caregiver comments describing the route as having changed (by the bus driver), or the participant not completing the trip. Additionally, the IE team removed any time series data that had fewer than 8 relative waypoint efficiency measurements in order to reduce chances of low frequency errors that would be difficult to characterize in a small data series. Importantly, no transformations were necessary to satisfy assumptions for parametric analysis. Sampling had been collected regularly at every trip interval, and therefore the data can be aggregated to standard trips (e.g., the participant’s first, second, third trip along a particular route, etc.). Therefore, there was no need to adjust data due to changing sampling frequency.



Source: Federal Highway Administration.

Figure 6. Graph. Waypoint relative time efficiency series remaining in the Mann Kendall Analysis.

Findings

The trend data residual (after denoising and filtering) resulted in 29 overall trips, with 4 participants represented in the data (see figure 6) This data is analogous to what is known in the field as fixed-station monitoring. The data included 29 different stations where gradual participant responses such as those that are due to incrementally more efficient trip-taking by participants are of interest. The IE team analyzes monotonic trends that correlated the response variable (i.e., relative task efficiency for a participant) with time and experience using the application.

For this Hypothesis test, the widely used modified Mann-Kendall test was run at 5 percent significance level on the 29 time series data for each of the 29 remaining time series data for the waypoint relative time efficiency collected over the MeraKey Field Test period. The resultant Mann- Kendall test statistic (S) indicates how strong the trend in relative time efficiency is and whether it is increasing or decreasing. For relative waypoint efficiency, 13 waypoints indicated statistically significant increasing trends, 13 waypoints indicated statistically significant decreasing trends and 3 waypoints indicated no significant change. On the contrary, linear trend line plotting indicates mostly no changes in waypoint relative time efficiency for most or all 29 trendlines shown in figure 6.

Table 6. Results of 29 Mann-Kendall analyses for the four participants with waypoint time efficiency data trends. Only one participant's waypoints were trending down in relative efficiency. Three participant's waypoint efficiency trended up or remained the same.

Participant	# Waypoints	# Positively Trending	# No Change	# Negatively Trending	Overall Increasing or Remaining the Same
AB01	10.00	3.00	2.00	-5.00	2.5
AB16	5.00	3.00	0.00	-2.00	2.6
AB20	5.00	1.00	0.00	-4.00	0.2
AI01	9.00	6.00	1.00	-2.00	5.777778

Source: Federal Highway Administration.

We conclude that the AbleLink interventions maintained or slightly increased traveler time efficiency for 75 percent of travelers examined.

Hypothesis Acceptance

The IE Plan 2020 required hypothesis acceptance (i.e., considering the AbleLink interventions as effective in increasing traveler efficiency) if at least 75 percent of waypoint trends show a significant upward trend in their relative efficiency for trip waypoints in their routes ($p<0.05$) in the univariate Mann-Kendall Test. The Hypothesis is accepted because three of the four included participant data did show upward trend ($p<0.05$).

Hypothesis: Participants are satisfied with use of the WayFinder app and their satisfaction rating is independent of their ability to complete tasks or complete trips.

Hypothesis demonstration: Participants' median satisfaction with the app is positive, and two standard deviations below the mean also is positive. Additionally, the satisfaction variable is demonstrably independent of self-reported ability to complete trips or succeed in travel.

The intent of this hypothesis is to measure user satisfaction, and hence it assesses user-perceived mobile application quality. For the purposes of this assessment, the IE team will be using the available MeraKey Field Test Surveys to glean a metric for Overall Assessment (as suggested by Bevan, et al. (1994) when describing the Software Usability Measurement Inventory (SUMI) test). This is a general global assessment of satisfaction, and it is given by a single numerical figure. The global assessment is

useful for setting targets, and for quick comparisons between different products. In this case, Merakey field tests did not ask participants for a numerical rating (as would be typical in an Overall Satisfaction Assessment, as in a Likert Scale). In a typical application, the scores would be between 0 and 100, with a mean of 50 and standard deviation of 10. In our coding, we used the participants' responses to open-ended satisfaction questions ("did you enjoy using the WayFinder app today?") to provide (-1,0,1) sentiment coding for the respondents' answers.

Performance Metrics:

To demonstrate primary-user satisfaction, we examined overall sentiment that primary users express about the technology immediately following use of the app for a trip. While more granular mechanisms of assessment exist, the IE was uncertain what would be reliable to use in this participant population. In addition, given the existing resources and the partnership with Merakey, feasibility and time resources were considered in deciding what questions or survey tools to deploy. The IE Plan 2020 employs a scale that is simple to administer to participants, thus making it ideal for usage with small sample sizes and in populations disinterested in intricate responses. Nevertheless, it is noted that the use of open-ended questions in the population at large and in this population in particular, is considered unreliable and potentially biased.

Data Elements and Sources:

Merakey Field Study User Satisfaction question, and questions about complications during trip.

Satisfaction rating: Positive (+1), neutral (0), negative (-1)—Unfortunately, Merakey respondents were not asked to rate their satisfaction with a Likert scale or other recognized survey tooling, therefore, the IE team coded respondents' answers to the question "**Did you enjoy using the WayFinder app today**?" as "positive," "negative," and "neutral." The coding was done by two independent professional researchers. Any disagreement was disambiguated by a third researcher.

For satisfaction to be measured independently of trip performance, the IE team also must account for trip success or within-trip success. From the Merakey staff survey, the team evaluated the question: "**Did the participant get from their starting location to their destination free of complication**—if complications occurred, please explain from your perspective"; from the participant survey, the IE team evaluated the question: "**Did you get to your stop using WayFinder**?" and "**Did you get off course anytime during your trip? If "yes," what did you do?**" Two independent researchers coded the responses to these questions for a final code for each trip summarizing trip success as positive (+1), neutral (0) or negative (-1). A neutral trip is one in which the trip did not occur incident free, but the participant was not aware of any flaws in the trip (as indicated by the question "**Did you get off course anytime during your trip?**).

Data Collection Period:

The data collection period is congruent with the Merakey field study period.

Analysis Procedure:

Derive the **Satisfaction Rating** assignment from the Merakey post-trip survey for participants. This results in the S variable. This categorical variable collapses to (-1) (0) and (1) values.

Derive the **Trip Success** T categorical variables over (-1, 0, +1, +2). The value assignments for **trip success** are established as follows: three questions will be coded, and appropriately weighted. SQ3: “did the participant get from their starting location to their destination free of complication—if complications occurred, please explain from your perspective.” PQ4: “Did you get to your stop using WayFinder?” and PQ5: “Did you get off course anytime during your trip? If “yes,” what did you do?”

Table 7. Assignment of the categorical variable trip success based on three questions in the post-trip surveys for both participant and caregiver.

SQ3—Did the participant get to their destination free of complications?	PQ4—Did you get to your stop using WayFinder?	PQ5—Did you get off course anytime during your trip?	Negate PQ5	Resulting TRIP SUCCESS	MeraKey Data Set counts
NO (-1)	YES (+1)	NO (-1)	+1	NEUTRAL (0)	29
YES (+1)	YES (+1)	NO (-1)	+1	YES (2)	109
NO (-1)	YES (+1)	YES (+1)	-1	NO (-1)	10
YES (+1)	YES (+1)	YES (+1)	-1	YES (+1)	3
NEUTRAL (0)	YES (+1)	NO (-1)	+1	YES (+1)	2
YES (+1)	NO (-1)	NO (-1)	+1	YES (+1)	2

Source: Federal Highway Administration.

Both categorical variables collapse to several integer values: (-1), (0), (+1) for the S variable, and (-1),(0), (1), (2) for the Trip success values.

Collect all metrics for all participants and assign the appropriate counts for all trips and all combinations. This will be summarized in a 4 x 3 grid representing the (representing all (T,S) possible tuple combinations).

The IE team performed a chi-square analysis testing whether the two variables are independent or if there is sufficient evidence to conclude that they are associated. Hypothesis acceptance was determined to be that the technology itself is satisfying to users if the Chi-squared test shows no association between the satisfaction rating and trip success (significance $p<0.05$) and the mean and median satisfaction rating (computed over all participants) is positive.

Analysis Outcomes

Findings

Our first analysis was to collect from each post-trip survey, the Satisfaction rating (-1,0,+1) and Trip Success (-1,0,+1,2) and place these in a correlation grid, as shown in the table below.

Table 8. Joint count distribution matrix of MeraKey trip ratings. Table of joint counts of user Satisfaction Rating (S) with Trip Success (T).

Trip Success	Satisfaction Rating		
	-1	0	1
-1	2	0	8
0	0	0	28
1	0	0	7
2	1	0	105

Source: Federal Highway Administration.

Since we are looking for a test of independence, and our variables are categorical, the IE must run a chi-square test for independence where the team looked at a relationship between the frequencies in terms of how many trips fit into each category. Since we have no trips that were rated Neutral for Trip Satisfaction, we really only have two levels of satisfaction rating (-1, 1). We had 4 levels of Trip Success (-,0,1,2). We ended up with eight joint tuples altogether (see table 8 for the joint count distribution).

Doing a correlation regression to the observed frequencies in table 8, we calculate the expected counts for joint distribution of trip ratings, as in table 9.

Table 9. Expected joint frequencies table for MeraKey trip ratings.

Trip Success	Satisfaction Rating	
	1	-1
-1	9.801325	0.198675
0	27.44371	0.556291
1	6.860927	0.139073
2	103.894	2.10596

Source: Federal Highway Administration.

Under the null hypothesis, there's a significant relationship between Satisfaction Rating and Trip Success. The IE ran the Chi-Squared test to obtain a p-value for the probability that any relationship between the two is just due to chance (i.e., a larger chi-square results when observed and expected counts are very different). The Chi-Squared p-value was calculated at 0.000447 (using Excel chi-square test). The small p-value indicates low probability that results are due to chance. The acceptance criteria of 5 percent means the p-value is lower than the acceptance threshold, which leads us to accept the null hypothesis that there is a significant non-independent relationship between Satisfaction Rating and Trip Success.

Table 10. Full Chi-squared test calculation.

Trip Success Observed	app satisfaction=TRUE	app satisfaction=FALSE	Number of Observations
trip success =-1	8	2	10
trip success = 0	28	0	28
trip success = 1	7	0	7
trip success = 2	105	1	106
Sum	148	3	151
Trip Success Expected	**app satisfaction=TRUE**	**app satisfaction=FALSE**	**Expected Number of Observations**
trip success =-1	9.801325	0.198675	10
trip success = 0	27.44371	0.556291	28
trip success = 1	6.860927	0.139073	7
trip success = 2	103.894	2.10596	106
Sum:	148	3	151

Source: Federal Highway Administration.

p-value = 0.000447

Rejected Hypothesis

While ratings of app satisfaction were overwhelmingly positive, they were highly interdependent with trip success. The hypothesis of independence was rejected. Participant's opinion was colored by the success of individual trips even if they have used the application before and had trip completion success with it. Two standard deviations below the mean were negative. Generally, the IE team believes that there are some performance issues to address in order to increase traveler satisfaction.

The IE believes that the manner in which data was collected about application satisfaction was not well described to the participants, and was only taken after the trip, therefore creating the interdependence between the variables. The IE suggests that AbleLink can improve data quality by including better-defined questions in the survey and asking about specific features of the application with which the respondents can express dis/satisfaction. Further validation and testing will likely improve data quality in the future.

Ability to Mitigate Threats

Hypothesis: The technology does not adversely impact an individual's ability to utilize other mobile applications.

The multi-dimensional functionality of smartphones requires more power to support people's daily smartphone operation. Battery life limitations are still the bottleneck in smartphone use, compared to processing power, feature-sets and sensors [Corey, G.P.: Nine Ways To Murder Your Battery (These Are Only Some Of The Ways). In: Battcon 2010 (2010)]. This is an important limitation because people, and specifically people with disabilities, who at times use smartphones for multiple assistive device functionality, prioritize the task of managing battery life. Consequently, the power demands of mobile applications, particularly in the assistive technology arena, make a difference and must be carefully monitored to ensure applications can robustly interoperate with other applications in the mobile device environment.

This hypothesis aims to test one dimension of robust interoperability in the mobile device environment. It is the responsibility of the ADP technology team to perform exhaustive technology robustness testing. Specifically, this means the ADP team should be testing the WayFinder app for its functionality, consistency, and usability. This can be done manually or by using automation tools. In such an evaluation, the ADP team should be testing the application against a matrix of operative systems and device models along with functional and nonfunctional testing.

The IE, in the context of addressing concerns of equity for users, is specifically looking to understand whether there are any indications that the primary users of this application are adversely impacted by the use of the ADP intervention. The IE assessed this using the Field Testing strategy, by looking at a small subset of performance metrics used in typical robustness testing. The IE team deployed a separate sensing application onboard the participants' smart phones to assess this subset of performance metrics while the mobile application was in use.

Performance Metrics:

The IE team was initially slated to assess two aspects of performance, Memory usage and Battery consumption. However, neither the AbleLink Dashbaord, nor the monitoring application that was

independently deployed, was tracking memory usage. Consequently, the IE team resorted to battery consumption evaluation:

- **Battery Consumption** while using the application during a WayFinder 3 trip, recording rate every 10 minutes.

Data Elements and Sources:

The data elements available for the IE team via the additional monitoring application used on board the participant smart phones were time-stamped battery usage. Memory usage was initially thought to be collected, but was not.

Data Collection Period:

Merakey Field Testing period.

Analysis Procedure:

Contrary to the IE team's expectations, the additional monitoring application did not include data about memory consumption. To work around this issue, the IE team based the test for this hypothesis on battery consumption data, which was used to calculate battery consumption. The commonly held view in the mobile device field is that while exciting applications place demands on large computing power, memory, and network bandwidth, from the users' perspective, they don't care much about those underlying resources, but do need better performance of their mobile devices which reflects on longer battery life [Thu, Mi Swe Zar and H. Kyi. "Computation Offloading Decision in Mobile Cloud Computing: Enhance Battery Life of Mobile Device." (2020)]. The IE team promotes the idea that even though memory consumption was not, after all, part of this analysis, understanding battery consumption is the one most significant concern towards validating this hypothesis.

Battery Consumption while using the application over a two-hour WayFinder 3 session (there was no mechanism for controlling what participants were running in the background, for this reason we only looked at (same trip, same device) replicates for which MeraKey dataset had at least 8 trips of data. The acceptance criteria was defined as:

- WayFinder 3 will be **conditionally acceptable if it drains less than ⅛ of battery life, but over 1/16 of battery life over a two-hour session** (these thresholds are determined by the expectation that a battery will sustain a user a day-long 16-hour session, and that a heavy-usage app that is still acceptable will only drain as much battery as required to sustain the app running by itself, albeit suboptimal).
- WayFinder 3 will be **fully acceptable if it drains less than 1/16 of battery life** over a two-hour session.
- WayFinder 3 will be **considered unacceptable if it drains more than ⅛ of battery life** over a two-hour.

Analysis Outcomes

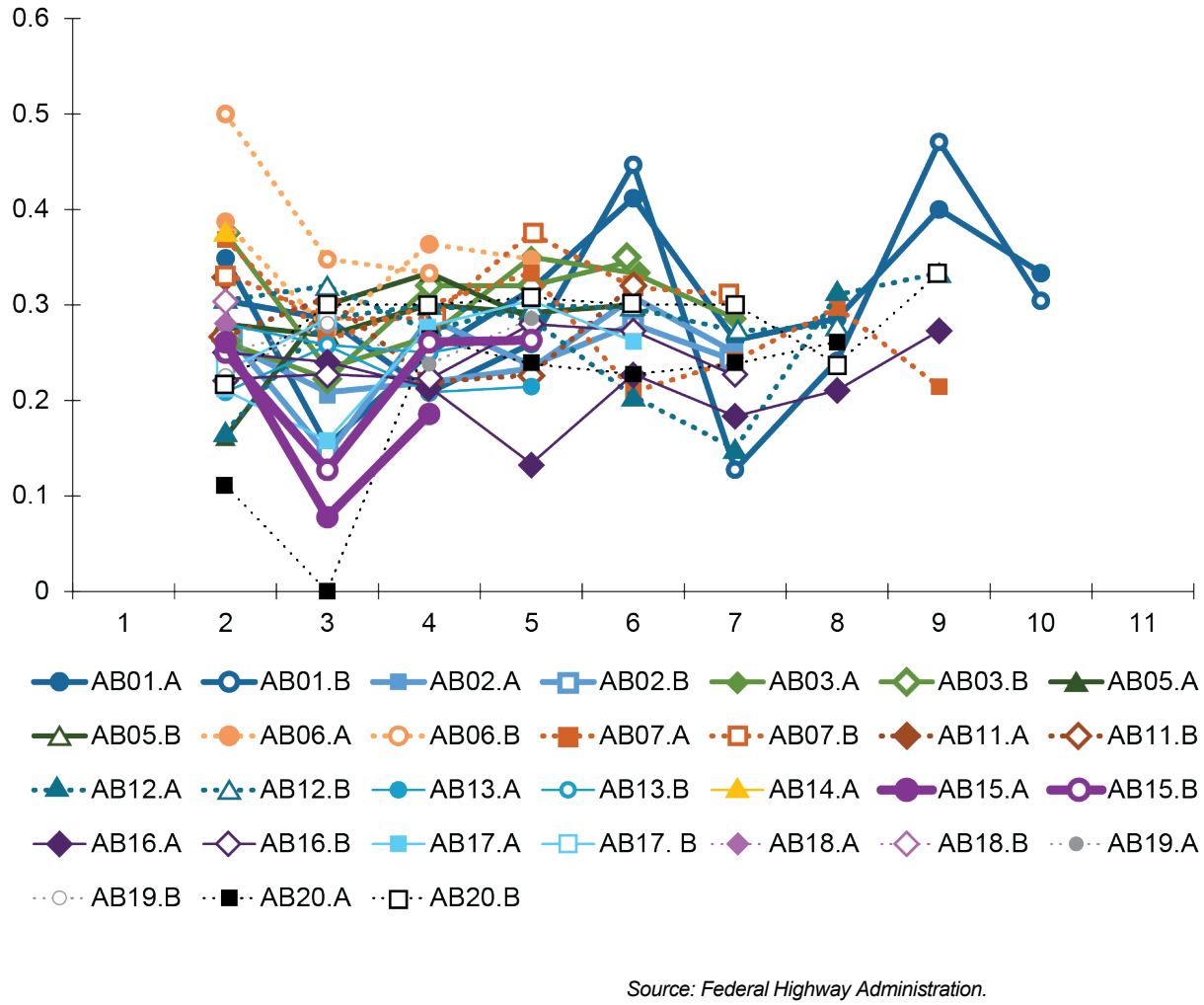
Findings

The analysis of battery data used data points from 22 users and 194 individual trips. To analyze the battery data, the IE team located the earliest and latest relevant timestamp for each trip and noted the battery level at those points. For example, AB01, route A (bus 24 from Kennedy Giant Eagle to Robinson Mall), trip 1, went from 9:32 a.m. to 10:15 a.m. During that time, their battery level fell from 89 percent to 74 percent, so during their 43-minute trip, they lost 15 percent of their battery. Then the IE team calculated the percent change in battery per minute, per trip. For AB01, trip 1, the user's device lost 0.35 percent of its battery per minute, which, at that rate, adds up to 42 percent of their battery over a 2-hour period. Although each user did not actually use the WayFinder 3 app for a full two hours, the IE team used the percent change in battery per minute to extrapolate what the battery level would be after two hours of use.

Next, the IE team calculated the mean percent change in battery level per user, per route. This meant that the IE team took the average percent change in battery per minute for every trip done on a particular route by a particular user. For example, AB01 took 10 trips along route A, so the IE team took the total average of the percent change in battery level per minute of all ten of their trips. Finally, the IE team calculated the standard deviation of the mean percent change in battery per minute, per route.

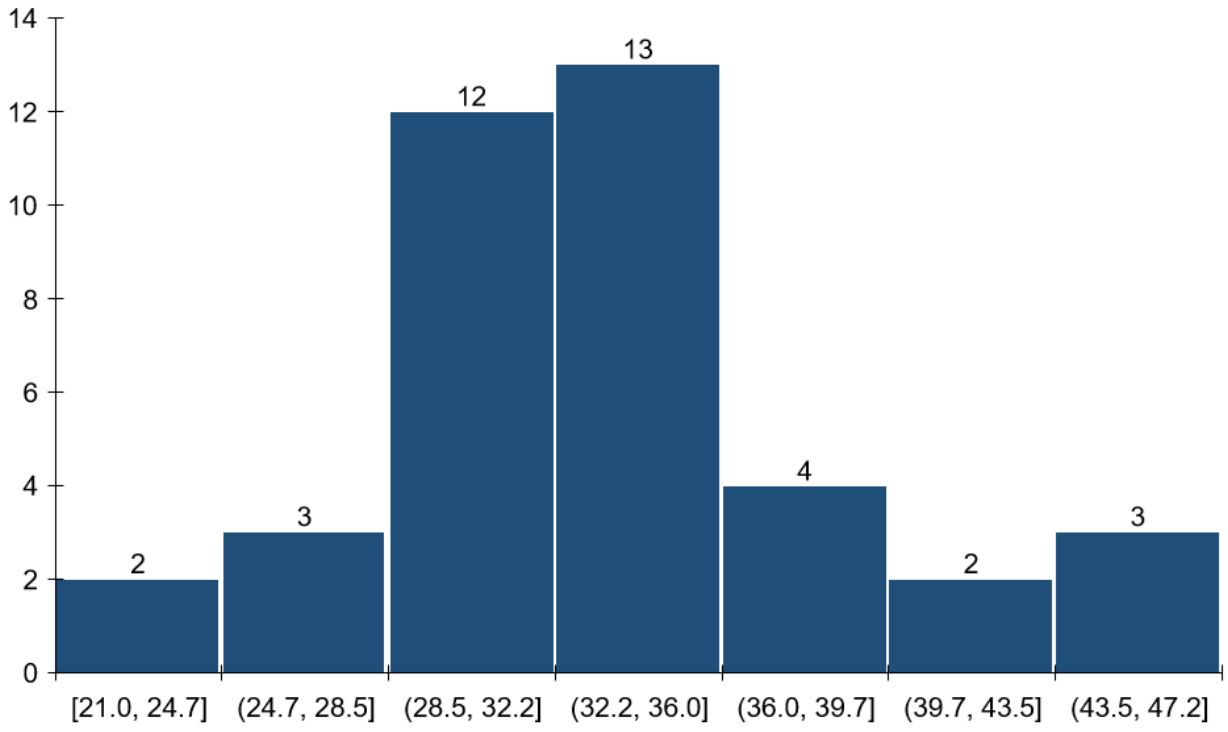
The IE plan 2020 specified that WayFinder 3 must use less than 1/16 of a device's battery life over a two-hour period (6.25 percent of the battery) in order to be acceptable as interoperable in a smart phone environment. The per user, per route, average percent change in battery level per minute, ranged from 0.174787 percent (20.97 percent over two hours) to 0.380952 percent (45.71 percent over two hours). Figure 7 shows the per user, per route, average percent battery drained for the 29 (user/route) pairs.

Only one individual trip (AB15, route A, trip 3), resulted in a conditionally acceptable rate of battery drain per minute (0.077778 percent per minute or 9.33 percent over a 2-hour period). The next closest trip was AB15, route B, trip 3, which drained the battery at a rate of 0.127273 percent per minute or 15.27 percent over a 2-hour period.



Source: Federal Highway Administration.

Figure 7. Graph. Average percent battery drained per user, per trip. The x-axis is the trip replicate number, and the y-axis is the amount of battery drained per minute during that trip.



Source: Federal Highway Administration.

Figure 8. Graph. Extrapolated, percent of battery drained after two hours using WayFinder 3 app. The sections of the x-axis represent ranges of the percent of battery drained over 2 hours, and the y-axis is the number of users whose battery drained by an amount within that range.

Rejected Hypothesis

Based on the analysis of 159 trips, the IE team concludes that the power drain demonstrated by the WayFinder Application during the MeraKey field tests does not pass acceptance criteria. It should be noted that no testing was performed on Android devices towards this hypothesis test. Also, memory usage was not collected as anticipated, and battery usage was used as a surrogate measure for overall resource drain by the application. A more complete test for technology robustness, resource requirements and interoperability with other mobile applications is recommended.

Hypothesis: Over the course of repeated trials to input routes and use routes, the mobile application does not slow down, quit operation or result in unexplainable error.

(surrogate to: “the mobile application works robustly even in computationally constrained environments.”)

It is outside the scope and resources of this IE to provide testing for robust interoperation in the mobile device environment. The ADP technology team must provide both nonfunctional testing by the developers as well as data from real-world testing by users. In the IE Plan 2020, the IE team recommended that the ADP team collect specific, precise performance metrics having to do with Peak response time, Failed

connections, Application Failure or Crash. In the implementation of the independent evaluation, the IE collected anecdotal information that may loosely indicate to AbleLink what performance issues WayFinder users were experiencing in Field tests.

Performance Metrics:

- Peak response time: Measurement of the longest amount of time it takes to fulfill an in-app request or respond to an in-app event.
- Failed Connections: Number of failed connections refused by the client while accessing a route or inputting a new route.
- Application Failure or Crash: Number of application failures during travel attempts while accessing a route or putting in a new route.
- Application Failure or Crash with no error information and/or no mechanism of getting back to route: Number of failures from previous performance metric with no error indicators or mechanism of getting back to route.

The IE could not gather the four attributes as strictly measured performance metrics due to the current data collection capabilities. Instead, the IE team summed findings incidental to field tests, and attempted to connect those with each class of failure. Summary of incidents was collected over the use of the application in the MeraKey Field Tests as well as the Seattle field test, over two hours of variable use. These indices are not sufficient in assessing comprehensive performance of the application in the context of various devices, it should serve as an indicator to the AbleLink team about high priority performance issues that should be addressed.

Data Elements and Sources:

Numerous tools are available on the market for mobile application performance testing (some examples are given here: https://www.testbytes.net/blog/performance-testing-tools-for-mobile-applications/). Data elements will differ based on the tools chosen by the AbleLink team (for the in-depth application testing).

The IE team coded field test comments based on the following comment domains:

Data Collection Period:

August 2020.

Analysis Procedure:

This hypothesis broadly tests two main issues that could significantly impact WayFinder 3's functionality, adoption, and scalability in the field:

- **Speed issues:** Detectable issues having to do with poor user experience due to slow responses and long load times.
- **Application stability issues:** Detectable issues having to do with the application unable to handle the workload, or pushing the device to workload limits. Significant issues will arise in the hands of real users if application failures are not accompanied by error messages and coherent pathways explaining to users how to get back to navigating their route.

Using the above issues as guide, the IE team performed a coarse assessment regarding the status of the application along the following types of issues (https://web.dev/rail/ we use the RAIL model for general ideas about the latency requirements):

- **Application start-up and load time:** Sixty-six percent of iOS travel apps have a launch time of at most 2.0 seconds. The overall average launch time should fall below 2.6 seconds.
- **Average response time:** The average network latency in U.S. in 2016 was 269 milliseconds (ms), which was 49 percent faster than the global average of 525ms iOS 7, iOS 8, iOS 9, iOS 10 Data Reference Period: 2016-08-20 to 2016-09-18. The WayFinder 3 app will be considered acceptable if the average network latency falls within one standard deviation above the mean, resulting in an upper bound of 300ms average response time.
- **Error rate:** Number of application failures resulting in error/no success. Errors should be, if possible, segregated by error output or type. If it is possible to assess the type of failure based on a number of failure categories, that would help the assessment and guidance for the ADP team.
    - **Failed Connections:** Failed connections should be detailed and attached to the mobile application view in which they occurred.
    - **Number of application failures:** Each failure should be explained, and details leading to the failure expressed in a report.
    - **Application Failure or Crash with no error information and/or no mechanism of getting back to route:** Each such failure should be explained and details leading to the failure as well as the application's response should be reported. In particular, failures that produced no legible error messages or helpful mechanisms for users to get back to their route should be noted.
- App should not crash without appropriate error messages.
- If app crashes, an informative error message appears with contact information for the user to call midtrip (it is best if the contact is appropriate for the trip link).

Analysis Outcomes:

Findings

Incidents and comments were gathered from MeraKey Field Test comments and the Seattle Field Tests (inclusive of 9 regional trips). The results below, are listed in the order of issues as they are listed in IE Plan 2020. Acceptance and Severity for each class of issues was listed separately below each title.

Accepted: Application start-up and load time in WayFinder is acceptable

Over the course of 9 recorded trips (and several additional trips that did not go on record), with one iOS and one Android device, **all start-up times** were below 2.0 seconds.

Over the course of 9 recorded trips (and several additional trips that did not go on record), with one iOS and one Android device, **Load times** were variable, but did not exceed 5 seconds.

On both counts, the IE team found the WayFinder Application Start Up and Load Times to be acceptable.

Rejected: Average response time in WayFinder is acceptable

As noted, the IE does not have the mechanism to measure latency in ms. However, a user's expectation of responsiveness changes depending upon the device they are using and the context of use. The IE wishes to raise awareness in consideration of users whose mental model of what is happening with their phone while they're using the WayFinder application does not prepare them to wait for a response, but instead leads them to expect an immediate response. The IE is concerned that users would be tempted to press the screen again or multiple consecutive times, or kill the app and restart it. Latency was particularly bad when waypoints were hit and the screen was about to refresh with the waypoint information.

Over the course of 9 recorded trips on one iOS and one Android device, it became clear that the waypoint notifications experienced delays due to some interplay between GPS tracking issues (covered in a later hypothesis) and the latency possibly created by attempts to compute proximity to waypoints. The **IE team experienced latency issues in (1) calculating device was proximal to waypoint, (2) calculating being off course and (3) calculating and triggering trip endpoints on the application**. These issues were recorded through numerous comments made over the course of the Seattle Field Test:

1. Latency calculating device was proximal to waypoint:

Resulted in either NO waypoint trigger, or DELAYED waypoint trigger (see coded comments in table 11)
Delayed waypoint triggering was more substantial on the Android device. In many instances, the waypoint did not trigger at all. The IE team considers this a critical issue that must be addressed if the application is to be used as a trip concierge application, with waypoint-triggered guidance, for the population of interest.

Table 11. Coded comments for "no waypoint triggers" and "delayed waypoint triggers" from Seattle Field Test.

Code	Comment verbatim	Researcher
Delayed Triggers	Half way up the block from waypoint, there is no signal from the bus and no contact on Android	R1
No Triggers	iOS instance has no triggers since beginning of the trip	R1
Delayed Triggers	Delayed triggers on iOS	R1
Delayed Triggers	Huge reliance on person setting up the route to do it right, especially getting notification to show up at the right time.	R1
Delayed Triggers	This waypoint notification came up after we already have done the walk up the block	R1
Delayed Triggers	Some notifications are often very late. This notification showed up when we already were at the destination.	R2

Source: Federal Highway Administration.

Table 11. Coded comments for "no waypoint triggers" and "delayed waypoint triggers" from Seattle Field Test (continuation).

Code	Comment verbatim	Researcher
No Triggers	The iOS instance never gave a single waypoint notification throughout the entire route. The route was started the Start button was pushed. It went into the wheel. It identified GPS was reading out but never actually brought up any notifications.	R2
Delayed Triggers	iOS showed there was some indication it was moving to the next waypoint. But again, it skipped through both the audio notification and there was about a less than one second indication that something was happening and without any confirmation and moved on back to the idling screen.	R2
No Triggers	We picked up the route in the middle. Both iOS and Android knew where to pick up the route and recognized once we got back on the route. However, iOS was unable to continue the messaging even the short messaging. With Android and also in place to have the messages and received confirmation in order. It then erroneously backtracked us to an earlier part of the route that we had skipped due to being off route. And it played those back. And now waiting to see if it picks up the route where we actually are thereby skipping the 2 messages that already had played.	R2
No Triggers	Android skipped the uVillage waypoint announcements altogether, and instead played the waypoint announcement for the next stop.	R1
No Triggers	The iOS instance is totally silent. No playback.	R1
No Triggers	Android just totally blew by one of the waypoints	R2
No Triggers	iOS was totally off line for this trip. No notifications were replayed	R2
Delayed Triggers	iOS had just a very delayed reaction to a GPS waypoint	R2
Delayed Triggers	iOS would detect the GPS location but didn't play the auditory cue until at least five seconds or six seconds later.	R2
Delayed Triggers	Android played the audio message for the XX stop, despite not having done so on any of the 3 earlier trips, and then proceeds to completely ignore the next waypoint (where in the past 3 trials it had played the audio notification for the following waypoint).	R2
Delayed Triggers	30 feet off the route, "off route" was finally called out by application, but only after having paused and stopped moving at that location. This is a very delayed reaction.	R2

Source: Federal Highway Administration.

Additional delays jointly caused by GPS localization and latency were noted during the application (2) calculating being off course and (3) calculating trip endpoints

This, too, resulted in a number of in-trip delays and confusion, but is less severe than the missed or delayed waypoints described above. See all comments in table 12.

Table 12. Coded comments for "route never finishes" "miscalculating off route" and "triggering end of route or exit route at inappropriate locations" in the Seattle Field Test.

Code	Comment verbatim	Researcher
Route never finishes	When performed, once we arrived and OK'd the previous destination, the route just STOPPED, it didn't say you have reached your destination. This was in the Android.	R3
Miscalculating off route	We picked up the route in the middle. Both the iOS and Android knew where to pick up the route and recognized once we got back on the route. However, iOS was unable to continue the messaging even the short messaging. With Android and also in place to have the messages and received confirmation in order. It then erroneously backtracked us to an earlier part of the route that we had skipped due to being off route. And it played those back. And now waiting to see if it picks up the route where we actually are thereby skipping the 2 messages that already had played.	R16
Route never finishes	iOS totally went on the fritz. Even though the trip had ended, it never played the done message so the trip was not ended.	R23
Route never finishes	iOS enters a weird 'not done' state even after the trip ended. After finishing the route, we happened to pass back near the route, suddenly, iOS woke up to let us know we were off the route.	R24
Miscalculating off route	Android was incorrect when giving one of the off travel route messages.	R30
Triggering end of route or exit route at inappropriate locations	Android, once again, recalls the wrong auditory notification for the specific waypoint (providing the voice note for a previous stop). Perhaps this is a bug in connection with the fact that there was only a text message with the current waypoint as opposed to a text message and an auditory notification.	R31
Miscalculating off route	30 feet off the route, "off route" was finally called out by application, but only after having paused and stopped moving at that location. This is a very delayed reaction.	R35
Triggering end of route or exit route at inappropriate locations	at 15 feet from the final destination, neither iOS nor Android are aware of proximity to trip endpoint.	R36
Route never finishes	Although the application recognized we were back on the path, we were less than 15 feet from the final destination but the final message was not playing.	R37

Source: Federal Highway Administration.

The user experience should be of main focus for WayFinder. While the IE team has no knowledge of the WayFinder backend, below are some concrete recommendations regarding networks calls and computations (such as computing when a user is proximal to a waypoint). Based on testing results,

network calls and computations currently are not performed in a background thread and therefore, the user is impacted through both process delays and interrupts while the application is waiting for a server or compute response. This was much more prominent in the Android instance than the iOS instance, where GPS calculation of the proximity to waypoint was better (in iOS). If this MUST be the case, then the WayFinder app should display some kind of a busy indicator, to inform the user that the application is working. Furthermore, the application should try to load just enough data to draw a screen in the application and allow the user to start observing/comparing to the waypoint while you load the remaining audio data in the background, or anything else.

Overall, the Seattle Field Test results indicate that WayFinder App responsiveness (to user events, to context and location) was below acceptable rates for an application of this type. The IE would highlight this as a major usability problem, and it would be important to fix in order to improve traveler outcomes.

Rejected: Application failures, Crashes and Failure Mitigation in WayFinder is acceptable

As the IE team does not have access to the WayFinder backend, the team is not able to verifiably segregate issue types by the source of the error. The error types that would be important for the AbleLink team to pay attention to are listed below. The tables summarizing the errors identified in both the MeraKey Field Tests and the Seattle Field Tests are segregated by error phenotype rather than its source. Type of failure sources include **Failed Connections**, **Application Failure or Crash**, as described above. The IE team attempted to highlight instances where **no error information and/or no mechanism of getting back to route were provided:**

1. **Findings pertaining to failed connections:**
    - **Over the course of 9 recorded trips** no specific indications of a failed connection were given. It could be possible that some of the application crashes and errors were the result of failure to connect to the server, but no error messages specifically indicated as much.
    - **Loading new routes** in the app was slow. 3 out of 4 attempts never finished loading the new route, but did not offer an error message. Instead, the user interface gave an indication that the route was still loading, but the trip never appeared on the device's route portfolio (3 out 4 trials failed in the iOS device), but this type of operation was outside the scope of this evaluation.
2. **Findings pertaining to Application Failure or Crash:**
    - MeraKey Field Test data included caregiver comments to inform the AbleLink team about any operation interruptions or failures. Comments included information about inexplicable issues with application restarts and shutting down.

Table 13. Coded comments from MeraKey field tests on unexplainable errors and crashes.

Code	Comment verbatim	Participant
General App Malfunction (not specified)	"Reset phone could not use app"	AB07
(as above)	“Phone (App) malfunctioned on the way.”	AB03
(as above)	"A route already was open when first starting Wayfinder"	AB16
Unrecoverable application error (app quits inexplicably or error codes come up)	"Had to restart phone due to an error code coming up multiple times. App worked w/o issue after it was restarted"	AB02
(as above)	"After getting on the bus, the app seemed to reset instead of continuing from "tap and pay" prompt.	AB05

Source: Federal Highway Administration.

Several users reported issues with the WayFinder 3 app suddenly closing, which sometimes resulted in users abandoning their use of the app partway through their route. AB02 reported that they “had to restart phone due to an error code coming up multiple times. App worked w/o issue after it was restarted”. Similarly, AB05 wrote that “after getting on the bus, the app seemed to reset instead of continuing from ‘tap and pay’ prompt.” Finally, another issue that occurred was sometimes the app failed to close properly. AB12 wrote that "A route already was open when first starting Wayfinder", which could easily cause confusion. The commenters were not specifically asked whether an appropriate message was communicated with the user and whether there was a ‘contact’ button that appeared after these issues surfaced. Based on some of the comments, it appears that error codes may have come up, but they were not instrumental in guiding the primary user on what to do or how to recover from the error. More rigorous testing is needed to identify the source of and remedy inexplicable crashes and malfunctions:

- Seattle Field Test data included comments to inform the AbleLink team about any operational interruptions or failures. Issues were identified that were similar to those commented on in the MeraKey Field Tests. Issues are summarized below.

Table 14. Coded comments from Seattle field test indicating unrecoverable application errors and crashes occurred in the field.

Code	Comment verbatim	Researcher
Unrecoverable application error (app quits inexplicably or error codes come up)	iOS crashed 4 times during one trip	R2
Unrecoverable application error (as above)	iOS displayed a runtime error message "attempt to concatenate 'error message' a nil value"	R2
Unrecoverable application error (as above)	Most of the iOS crashes seemed to be associated with the pressing of the contact button. Although two of those crashes were not.	R2
Unrecoverable application error (as above)	iOS crashed once again, with no indication. This time not prompted by any button press.	R2

Source: Federal Highway Administration.

More rigorous testing is needed to identify the source of and remedies for inexplicable crashes and malfunctions. The AbleLink team should in particular note the following:

- Each such failure should be explained and details leading to the failure as well as the application's response should be reported. In particular, failures that produced no legible error messages or helpful mechanisms for users to get back to their route should be noted.
- App should not crash without appropriate error messages.
- If app crashes, an informative error message should appear with contact information for the user to call midtrip (it is best if the contact is appropriate for the trip link).

The errors and failures pointed to above (particularly those experienced by the IE team) did not offer any informative messaging to the user to mitigate the risk involved with the application crashing in the middle of operation. These can be disruptive to an independent trip by user and their frequency can severely impact trusting that the application with provide concierge services for the full duration of the trip. The IE would consider these issues to be the highest priority and highest severity. It is imperative to fix this before the product should be released to the population of interest.

Hypothesis: The technology opportunistically aims to prevent primary user risks as part of strategic and operational planning.

Hypothesis demonstration: Over the course of repeated trials to use routes, the mobile application alerts are used to communicate risk to users.

Given the primary user's risk model, we wish to evaluate whether the ADP took all possible precautions to enable appropriate risk mitigation for the primary users. In practice, in this hypothesis evaluation, we do not assess notification technology; rather, we assess the application's ability to mitigate user risks.

There is a set of notifications to the primary user that such an application might include:

- Whether there are primary user in-app notifications to prevent primary users from risky behavior (examples: reminders not to get distracted, talk to strangers, get off track, etc.).
- Whether the primary user's notifications are responsive to caregiver or secondary user's settings and route markers (relevant and timely activation).
- Whether any automated primary user's notifications occur at relevant and timely points or when user appears unresponsive or not attentive to in-app notifications.

There is a set of notifications to the secondary users that we will evaluate, which involve:

- Alerts launched to secondary user when primary user appears to be at risk (for example, unresponsive or not attentive to in-app notifications).
- Alerts launched to secondary user when the application is down, or out of range, and is unable to provide support for the primary user.

There are a number of triggered events that the app may automatically bring up to alert or remind primary user:

- System launches instructional task within SMART Travel Concierge System in response to relevant triggers (within 10m of the relevant location, at times when no GPS is detected).
- System offers in-app reminders to primary users about information from the pre-trip concierge training that pertain to the leg of the trip.

Performance Metrics:

Performance metrics will be measured through our in-lab tests and heuristic assessment, we will build in triggers for notifications of primary and secondary parties. Each trip will have at least 3 triggers built in. We will repeat the test 3 times under different conditions.

Rows 1 to 6 in the table refer to notifications received by the secondary user. Rows 7 to 9 in the table refer to notification received by the primary user in the app.

The following notification performance metrics should be reported. These notifications are aligned with notifications noted in the AbleLink WayFinder 3 manual:

For notifications to primary and secondary users, we gather the following information:

1. The true distance from the waypoint at the point the notification was sent (acceptance for within 10m accuracy, see previous note about mobile device application GPS sensor accuracy).
2. The latency period of the notification sent (acceptance at average two-second latency).
3. Exactly how far off the path the user is when the notification triggers (acceptance at 10m).
4. Notification completion rate (calculated as total number of this type of notifications arrived within acceptance rates/total number of intentionally triggered notifications of this type).

Table 15. Notification performance metrics table.

No.	Type of notification	True distance of user from the waypoint at the point the notification was sent (acceptance is within 10m)	Time lag between when user reached or passed waypoint and the time notification was sent to secondary user (acceptance at two seconds latency)	Notification completion rate (calculated as total number of this type of notifications arrived within acceptance rates/total number of intentionally triggered notifications of this type)
1	Start/launch of SMART trip.	–	–	–
2	Periodic (every 5 to 60 minutes per user settings).	–	–	–
3	Arrival at destination (completion of trip).	–	–	–
4	Triggered instructional task within SMART Travel Concierge System in response to relevant triggers or waypoints.	–	–	–
5	Notifications triggered by standstill and/or aborting route.	–	–	–
6	Notifications triggered by getting off path.	–	–	–
7	(Primary user) Triggered instructional task within SMART Travel Concierge System in response to relevant triggers or waypoints.	–	–	–
8	(Primary user) Notifications triggered by standstill.	–	–	–
9	(Primary user) Notifications triggered by getting off path.	–	–	–
10	Notifications triggered for “not your stop”.	–	–	–

Note: An endash (–) denotes Not Applicable (N/A).

Source: Federal Highway Administration.

The reports drawn from these intentional triggers are not standardized and not all instances could be triggered in several hours of use. Therefore, this evaluation amounts to a heuristic synthetically manufactured evaluation in an attempt to trigger the events. This evaluation is insufficient in assessing comprehensive notification performance of the application in the context of various devices and operating

systems. Our experiment will utilize the application on two separate smart phones, one iOS and one Android device and will attempt to trigger each of the 10 different types of triggers as described in table 15.

Data Elements and Sources:

- Periodic notifications sent to secondary users.
- Periodic notifications alerting on-device primary users during travel.
- System launch logs (or account from tester of what instructional tasks the app launched).
- Any in-app alert logs.
- Data for table 13 was collected via internal field test.
- Alternative GPS trace capture method that will be deployed on in-lab tester phones.

Data Collection Period:

August 2020.

Analysis Procedure:

The IE Plan 2020 stated that the acceptance criteria required that all the checkboxes in the two leftmost columns of the table above be checked with acceptance for the stated objectives. The rightmost column should indicate at least 75 percent notification completion rate for each type of notification (source: 75 percent is again invoking the Chebyshev rule to cover the mean plus 2 standard deviations from the mean of a population of data points for which we do not know the underlying distribution).

Analysis Outcomes:

Findings

The team found the secondary user's notifications quite difficult to analyze as discretely as required by the table. The IE team performed an in-depth analysis of majority of the trips in the Seattle Field Test and filled out the table below. However, the team found that with the number and frequency of problems that came up with these notifications, reconstructing the narrative of several of the trips to explain how the secondary user's notifications behaved is more telling than the tallies in the table below. See the trip narratives detailed below.

Table 16. Notifications performance in limited field test.

No.	Type of notification	True distance of user from the waypoint at the point the notification was sent (acceptance is within 10m)	Time lag between when user reached or passed waypoint and the time notification was sent to secondary user (acceptance at two seconds latency)	Notification completion rate (calculated as total number of this type of notifications arrived within acceptance rates/total number of intentionally triggered notifications of this type)
1	Start/launch of SMART trip.	NOT ACCEPTED. Values were: • 45.06 m • 772.53 m • 32.68 m • 107.28 m • 49.75 m • 43.50 m 4 values could not be assessed	NOT ACCEPTED: Values were: • 4 notifications were immediate • 3 x 1 min delay • 1 x 2 min • 1 x 7 min delay • 1 MONTH delay	3/10 studied notifications of this type were completely spurious. 6/10 studied notifications of this type had >2sec delays. ALL locations reported to Secondary user were >32m from Primary user's actual location.
2	Location Updates Periodic (every 5 to 60 minutes per user settings).	Could not be assessed because there was no gold truth for the primary user's whereabout. Out of 10 studied notifications, 8 seemed to match AbleLink data to about 20m, 1 was 200m from AbleLink data at the same timestamp and 1 had no dashboard data.	N/A. Unable to assess time lag for this type of notification.	8/10 ACCEPTED.

Table 16. Notifications performance in limited field test (continuation).

No.	Type of notification	True distance of user from the waypoint at the point the notification was sent (acceptance is within 10m)	Time lag between when user reached or passed waypoint and the time notification was sent to secondary user (acceptance at two seconds latency)	Notification completion rate (calculated as total number of this type of notifications arrived within acceptance rates/total number of intentionally triggered notifications of this type)
3	Arrival at destination (completion of trip).	NOT ACCEPTED. Only one value was within acceptable bounds. Values were: • 24.66 m • 21.21 m • 20.21 m • 0 m	4 of 4 were ACCEPTED.	Only 4 of our 15 trips triggered this notification. This notification was triggered 5 times, only 4 of the 5 were correct.
4	Triggered instructional task within SMART.	N/A: functionality does not appear to be implemented.	N/A: functionality does not appear to be implemented.	N/A: functionality does not appear to be implemented.
5	“Route Cancelled” Notifications triggered by standstill and/or aborting route.	NOT ACCEPTED: only 1 of the 5 studied instances had a location that even remotely made sense.	NOT ACCEPTED: 2 of the 5 studied instances were sent to secondary user a MONTH after they triggered. The other three were timely, but only 1 of the three was a true cancelled trip.	NOT ACCEPTED: This notification was triggered 5 times. Only one of the 5 was in fact a cancelled trip. 2 of the 5 studied instances were actually a completed trip that were not cancelled by the primary user, but secondary user was notified of a cancellation. 2 of the 5 studied instances were sent a MONTH after they were initially triggered.

Table 16. Notifications performance in limited field test (continuation).

No.	Type of notification	True distance of user from the waypoint at the point the notification was sent (acceptance is within 10m)	Time lag between when user reached or passed waypoint and the time notification was sent to secondary user (acceptance at two seconds latency)	Notification completion rate (calculated as total number of this type of notifications arrived within acceptance rates/total number of intentionally triggered notifications of this type)
6	Notifications triggered by getting off path.	NOT ACCEPTED: Although the primary user veered off path 4 times, only one such notification triggered and identified the primary user's location appropriately.	NOT ACCEPTED: the one true positive instance triggered 30 m after the primary user left the 20m buffer around the route.	NOT ACCEPTED: 1 out of 4 true off route events actually ended up as a secondary user notification. The trip in which the notification was included was not recorded at all in the AbleLink Dashboard data.
7	(Primary user) Triggered instructional task within SMART Travel Concierge System in response to relevant triggers or waypoints.	N/A: functionality does not appear to be implemented.	N/A: functionality does not appear to be implemented.	N/A: functionality does not appear to be implemented.
8	(Primary user) Notifications triggered by standstill.	N/A: functionality does not appear to be implemented or IE team did not stand still long enough.	N/A: functionality does not appear to be implemented or IE team did not stand still long enough.	N/A: functionality does not appear to be implemented or IE team did not stand still long enough.

Table 16. Notifications performance in limited field test (continuation).

No.	Type of notification	True distance of user from the waypoint at the point the notification was sent (acceptance is within 10m)	Time lag between when user reached or passed waypoint and the time notification was sent to secondary user (acceptance at two seconds latency)	Notification completion rate (calculated as total number of this type of notifications arrived within acceptance rates/total number of intentionally triggered notifications of this type)
9	(Primary user) Notifications triggered by getting off path.	NOT ACCEPTED: Notification did not appear to trigger before primary user was at least 50 meters from path.	Insufficient Data: The delay in triggering may be partially due to latency in computing the primary user's distance to the route.	Insufficient data. There were definitely Off-Route alerts. They appeared further/more delayed than the acceptance criteria suggests. This must be tested further.
10	Notifications triggered for "not your stop".	N/A: functionality does not appear to be implemented, other than if person setting the route creates this notification	N/A: functionality does not appear to be implemented, other than if person setting the route creates this notification.	N/A: functionality does not appear to be implemented, other than if person setting the route creates this notification.

Source: Federal Highway Administration.

While doing the evaluation, the IE team recognized that reducing the notifications to mere performance metrics does not provide an adequate preview of the kind of "complete trip" problems that may be impacted through the inaccuracies, false triggers or delays that the notification performance table exposed. Instead, the IE team sought to better understand the progression of trips by tracing the primary user on the ground (using the AbleLink Dashboard data which was to include GPS tracing of the primary user's device) and the notifications the secondary was getting at the time the primary user was proceeding with the trip. Looking at the data from this standpoint allowed us to understand at a deeper level both the significance to safety and risk mitigation of getting the notifications right and the contextual importance of these notifications.

Reconstructed trip narratives, highlighting the interaction between notifications, primary user risk and secondary user confidence about primary user's safety given limited information in the notifications

iPhone—9:19 a.m. trip start

The secondary user first received a notification at 9:19, which was a "Route-Started" notification which placed the user five miles away from the route. Then, the secondary user received a notification a minute later that the route has completed at the same location five miles away from where the primary user was actually. The AbleLink dashboard shows this trip as a preview, not a real trip.

Android—9:27 a.m. trip start

The AbleLink dashboard had no record of this trip so there was no way to verify GPS location accuracy. At 9:27 the secondary user received a "route-started" notification that placed the primary user 45 meters away from the actual trip start point. At 9:30, the secondary user received a notification that the trip was "canceled" before the route was completed. Then at 9:32, secondary user received another "route-started" notification that placed. the user precisely at the original trip start point. At 9:37 the secondary user appropriately gets a location update placing the user along the route. Again, for this trip, there was no AbleLink dashboard data available to verify GPS location accuracy. At 9:41, the secondary user received an off-route notification that the primary user was away from the route at a seemingly correct location. A minute later, the secondary user received a back-on-route notification that placed the user off-route but closer to the correct route, at a distance 36 meters from the closest point on the route. Three minutes later, the secondary user received a notification that three minutes earlier the primary user completed the route at a location 24 meters from the destination. The AbleLink dashboard had no data to verify the GPS location accuracy.

iPhone—SSI to Frye, 9:52 a.m. trip start

The primary user started the route. The secondary user received a route-started notification 2 minutes later that placed the primary user 107 meters away from actual starting location. Then the secondary user proceeded to receive two location-update notifications that were seven and then five minutes apart. The second location-update notification placed the primary user correctly at the destination. AbleLink incorrectly sent a route-canceled notification to the secondary user; however, the AbleLink dashboard confirmed that the primary user pressed done.

Android—SSI to Frye, 9:52 a.m. trip start

The primary user started the route sometime between 9:45 and 9:52 (based on the other phone as well as AbleLink dashboard annotations). At 9:52, the secondary user received an earlier route-started notification, which placed them 32 meters away from the starting point. This route-started notification was timestamped 9:45 but did not arrive in the secondary user's inbox until seven minutes later, at which point the primary user was ¾ miles from the starting point. Then, the secondary user received a location-update notification at 9:58, which placed the primary user at a location halfway back on the route. This time they were 244 meters away from the starting point. This could have created a false impression in the secondary user that the primary user was ¾ of the way to the destination but then backtracked towards the beginning of the route. The AbleLink dashboard showed that the primary user moved along the route at a typical pace. At 10:13, the secondary user received two notifications at the same time. One was a location update that correctly placed the primary user at a location 26 meters away from the destination. The secondary user also received a route completed notification that correctly placed the primary user at the appropriate destination.

Status Update notification: Android—10:04 a.m., November 26

The secondary user received a location-update notification and a route-canceled notification that both placed the primary user in the correct location.

Android—10:09 a.m. trip start

At 10:09, the secondary user received a route-started notification for the correct canonical trip that placed the primary user at a distance 43 meters away from the starting point. There was no time lag to the notification. Then the secondary user received a location-update notification three minutes later at 10:12. The AbleLink dashboard showed that the user lingered around the start point before they departed. At 10:17 the secondary user received another location-update notification that arrived on time and matched the AbleLink dashboard. At 10:19 the secondary user received a route-completed notification that was dated for 60 seconds earlier and placed the user 21 meters from the destination. During the trip, The primary user did go off-route and triggered the in-app notifications for off-route; however, the secondary user never received off-route or back-on-route notifications.

iPhone—10:09 a.m. trip start

The secondary user received a route-started notification at 10:10 indicating that the primary user had started a route one minute earlier. The notification located the primary user 50 meters away from the starting point. The secondary user received a location-update notification four minutes later at 10:13 that placed the user along the route in the correct position. Meanwhile, the primary user received no waypoint notifications and the AbleLink dashboard showed no waypoint responses. At 10:17 the secondary user received another route-started notification indicating that one minute earlier, the primary user was at a location five miles away (at the primary user's home). The primary user was not at this location for one hour preceding the time of the trip and for two hours following the time of the trip. At 10:18, the secondary user received a location-update notification that placed the primary user back on the route in the correct location. The secondary user received a route-completed notification one minute later that placed the user 20 meters from the destination. During the trip, the primary user did go off-route and triggered the in-app notifications for off-route; however, the secondary user never received off-route or back-on-route notifications.

Issues Summary:

With respect to notifications, the WayFinder backend seemed to have some problems tracking trip state for the primary user, which caused the application to send confusing or poorly sequenced notifications to the secondary user. Throughout the observed Seattle Field Test trips, the IE team encountered several of the notification types, including: “route-started”, “route-canceled,” “location-update,” “off-route,” and “back-on-route.”:

- One common issue occurred at the end of trips. In these cases, when the primary user pressed “done”, the secondary user received a route-canceled notification instead.
- Another issue was caused by incorrect localization of the user. The application sometimes recalled an old GPS location, which caused the application to send incorrect location-update notifications.
- In some cases, there were issues with canonical trip selection. The AbleLink dashboard showed some canonical trips as “dynamic trip” instead of the correct route. There also were instances where a route-started notification failed to trigger, causing the trip to appear as if it started partway along the route.
- Finally, in one case, trip notifications were saved in the phone for a month and were sent out to the secondary user one month later as if they had just occurred.

Rejected Hypothesis

The hypothesis was rejected: The WayFinder application sends notifications to secondary users and alerts to the primary user to mitigate risks to users. The IE team found that notifications were used, but had significant errors and issues which presented barriers to properly using these features towards proactive risk mitigation.

Hypothesis: When either routing primary users or during operational failures, WayFinder provides the primary user with appropriate triggers to enlist assistance or call for help.

Hypothesis demonstration: For every mobile application view or pane, the mobile application provides clear indicators for user on how they can call for help, whether in place or remotely.

Performance Metrics:

In our in-lab tests and heuristic assessment, we will turn application settings “Contact Me” with the intent of providing the primary user the maximal supports and ability to call for help throughout their trip. Evaluations will assess the appropriate indicators to call for assistance. In particular, we will be testing the following three functions enabling primary users to call for help during trips:

1. If the “Contact Me” button is turned on, selecting it sends an email notification and Google Maps link to the appropriate party.
2. If the application crashes or identifies an error, user is provided with information about calling for assistance (we will collect this information serendipitously through in-lab or internal field tests rather than synthetically manufacturing failure events).
3. Every mobile application view or pane will be evaluated for the appearance of a call for help trigger. For each of the application panes, a tabulated report will detail the following aspects:

- Presence of the interaction element to call for support (test: is there a visual interaction element on the screen to call for support? Possible values: yes/no, acceptance: yes).
- Size of on-screen visual element (test: provide element width and height in screen cascading style sheets (CSS) pixels, possible values: 0 … screen width/height; acceptance: 44 by 44 CSS pixels or larger. Source: WCAG 2.5.5 Target Size: "The intent of this success criteria is to ensure that target sizes are large enough for users to easily activate them, even if the user is accessing content on a small handheld touch screen device, has limited dexterity, or has trouble activating small targets for other reasons.").
- Appearance of the trigger in the accessibility tree created by the application (test: using voice over technology, is the visual screen element called out? Possible values: yes/no, acceptance: yes).
- Appropriate naming of the visual element/button to indicate what pressing the button will trigger. (test: using voice over technology, is the visual screen element called out appropriately as "call for help" or other appropriate indication of its function? Possible values: yes/no, acceptance: yes).

Data Elements and Sources:

Data for this hypothesis was collected via internal field test.

Data Collection Period:

August 2020.

Analysis Procedure:

We will consider the AbleLink call for help triggers appropriately placed and featured if all application panes provide affordances to request support (when the appropriate user settings are on). Additionally, upon failure, primary users are provided appropriate messages about accessing help.

Analysis Outcomes:

Findings

In the Seattle Field Tests, the IE team turned on preferences to enable the "CONTACT ME" button onscreen. The AbleLink manual specifies that by pressing this on-screen button, an email notification and Google Maps link is sent to the appropriate secondary user party.

This preference was turned for both of the internal field test phones (one iOS and one Android device) on in tandem with the periodic secondary user notifications (where the IE team elected to have location updates sent to the secondary users every 5 minutes).

In the iOS device, although the "contact me" button was pressed, the secondary user only received the regular "location updates."

In the Android device, the secondary user received both "location updates" and "status updates," However, there were no views in which the "CONTACT ME" button showed up in the Android instance. The IE team is uncertain what distinguishes a "location update" from a "status update" for the secondary user if neither are actively instigated by the primary user in the Android instance.

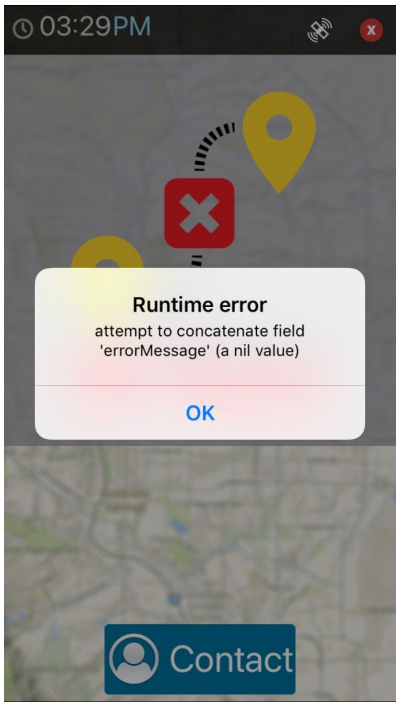
There was no feedback or confirmation provided to the primary user regarding the message being sent to the secondary user. If a user is distressed, such feedback would be very important. This is a usability issue that will be addressed later in the heuristic evaluation.

Overall, for this section, the findings were NOT ACCEPTABLE for the Android instance since the button never appeared, and we are pending acceptance for the iOS device because the AbleLink team expressed some concerns that there was some misstep by the IE team in turning this feature on.

In the Seattle Field Tests, the IE team sought to see whether, when and if the application crashes or identifies an error, user is provided with information about calling for assistance (the IE team collected this information serendipitously through in-lab or internal field tests rather than synthetically manufacturing failure events).

In the iOS environment, the IE team actually found that pressing the “CONTACT ME” button actually caused the iOS application to fail. A comment in table 14 (coded comments from Seattle Field Test) read: “Most of the iOS crashes seemed to be associated with the pressing of the contact button. Although two of those crashes were not.”

Also in the iOS environment, when the application crashed, it either provided no information to the user or it brought up a runtime error that does not guide the user to actionable resolution or the ability to contact someone for assistance, as shown in figure 9.



Source: Federal Highway Administration.

Figure 9. Screenshot. Runtime error after a crash event was instigated by pressing the “contact” button in an iOS instance.

Every mobile application view or pane will be evaluated for the appearance of a call for help trigger. For each of the application panes, a tabulated report will detail the following aspects:

- Presence of the interaction element to call for support (test: is there a visual interaction element on the screen to call for support? Possible values: yes/no, acceptance: yes).
  - The iOS instance had “Contact” button on numerous views, although not all.
  - The Android instance never displayed the “Contact” button.
  - **NOT ACCEPTED.**
- Size of on-screen visual element (test: provide element width and height in screen CSS pixels, possible values: 0 … screen width/height; acceptance: 44 by 44 CSS pixels or larger. Source: WCAG 2.5.5 Target Size: “The intent of this success criteria is to ensure that target sizes are large enough for users to easily activate them, even if the user is accessing content on a small handheld touch screen device, has limited dexterity, or has trouble activating small targets for other reasons.”).
  - The Contact button was sufficiently large, and in the iOS instance, even covered over the “OK” button to confirm the primary user saw the WayPoint notification.
  - There was no Contact button to evaluate in the Android instance.
  - **ACCEPTED.**
- Appearance of the trigger in the accessibility tree created by the application (test: using voice over technology, is the visual screen element called out? Possible values: yes/no, acceptance: yes).
  - The application in iOS instance is not accessible to either Switch Scanning nor Voice Over.
  - **NOT ACCEPTED.**
- Appropriate naming of the visual element/button to indicate what pressing the button will trigger. (test: using voice over technology, is the visual screen element called out appropriately as “call for help” or other appropriate indication of its function? Possible values: yes/no, acceptance: yes).
  - This was not a testable feature since the screen elements were not accessible to Voice Over.
  - **NOT ACCEPTED.**

Rejected Hypothesis

The IE Team found that the option to enlist assistance from a secondary user or caregiver was not utilized to the best extent possible in either the Android nor the iOS instance of the WayFinder Mobile Application.

Ability to Address Target Population’s Travel Needs

Hypothesis: Outdoor Global Positioning System localization and the inferred proximity to a WayPoint is comparable to other leading location-based services and is appropriate for the task.

This hypothesis is testing whether the mobile application’s GPS localization works robustly. The hypothesis is primarily concerned with (1) whether the GPS readings are sufficiently accurate for traveler use: timely notification and alerts; and (2) whether the AbleLink app functionalizes location-based

services to trigger relevant notifications for the user to be able to react within actionable distance. As noted earlier, the system's accurate GPS reading is crucial for providing the targeted population with useful travel guidance at the right time and place. An earlier hypothesis addressed the functional in-app trigger notification and caregiver responses to these features. Here we are evaluating whether the localization internal to the application is accurate. With the understanding that many device-based systems do not have very accurate GPS localization, we set the bar lower, at acceptance requiring that the AbleLink system be on-par with another mobile location monitoring application. In our exploration, we found out that Strava manipulates its GPS readings to conform to a model of the streetscape at that location. While we believe this to be the appropriate way to address uncertainly in on-location services, it gave Stava GPS readouts an unfair advantage against AbleLink. Instead, we utilized the raw GPS readouts from our monitoring application, AWARE, as the point of comparison about standard location based application.

Performance Metrics:

During field tests, the primary user's inferred location is compared to specific WayPoint locations whose precise location is known. We created 3 trips with control points in the Seattle trip dataset. We also had canonical routes for the MeraKey dataset. IE plan 2020 required the use of 30 control points in total, 10 per synthetic trip route that we created for the purpose of the internal field test. The IE trip decided there would be more statistical significance to our results if we added in the MeraKey trip data, since we had the precise control points and the AWARE data available for those trips as well. This data collection enhancement was enabled by the fact that the IE team recorded GPS readings from the AWARE application as the additional monitoring location based service for both MeraKey Field Test and the internal Seattle Field Test collection. We test this hypothesis using the following metrics:

- GPS Location registered through WayFinder 3 app and reported onboard the AbleLink Dashboard.
- GPS Location registered through a common location-based service application (AWARE).
- GPS Location of controlled WayPoints identified through AbleLink Dashboard and designed into the programmed routes for both the MeraKey and the internal Seattle field tests.

Data Elements and Sources:

This data collection involves the integration of WayFinder app-reported GPS location (from the AbleLink dashboard), AWARE-reported GPS location, and known GPS location of selected WayPoints as reported out by the AbleLink Dashboard.

The original IE Plan 2020 indicated the IE team will supply an array of 30 test waypoints (10 per synthetic trip) to function as a gold truth location set to determine if the app has successfully localizing the mobile device and whether the in-app triggers the waypoint. After realizing that by using the AWARE trip data, the IE team can increase the power of our statistical analysis from 9 trips (3 replicates of 3 trips with 10 waypoints each) to 52 trips from 10 contributing users, the team decided this was a worthwhile engagement to appropriately understand the variation in GPS localization in AbleLink, even across two coasts.

Although the IE team enhanced the data considered substantially, the analysis still followed the same statistical techniques: a GPS read out is considered a match if it returns the correct location of search to within 10m (5m if turn-by-turn directions are necessary).

The IE team did not need to supply additional test WayPoints randomly drawn from dense and nondense metropolitan areas to complement the tested WayPoints and fill any gaps (as specified by the IE Plan 2020) because it addressed the concern of sparse data by combining data from both Field Test datasets.

The experiments were therefore conducted in both Seattle, WA and in Pennsylvania (by MeraKey).

Seattle-based experiments were centered around latitude 47° 36' 28.8468" N and longitude 122° 20' 6.6012" W and conducted at altitudes between 0 and 520 feet (170 feet above sea level). The urban area is dominated by low-rise buildings and vegetation primarily comprising of deciduous forests. GPS quality depends on the flora composition, cloud cover conditions, built environment and incline of a particular area, in addition to ambient temperature and precipitation. Hence, three zones are distinguished: University of Washington Campus (with relatively open space and midsized buildings), Seattle Downtown (with low to midrange elevation, dense high rises, and typically cloud covered), and Laurelhurst neighborhood (with land dissected and intercepted by waterways).

Data Collection Period:

August–September 2020.

Analysis Procedure:

Combined Seattle and Merakey Field tests included 12 different phones (one running Android in Seattle, one running iOS in Seattle, all others running iOS in Pennsylvania) with GPS receivers. In the Seattle Data, the Android and iOS mobile devices were used simultaneously throughout the internal field tests. However, the team combined Seattle data with all iOS data from PA, we excluded the Android data from the statistical analysis to ensure we are comparing information from the same types of device. The IE would encourage the AbleLink team to run similar GPS fidelity tests with multiple Android phones and perform the same robust repeated measurements in order to make similar assessments for the Android device platform.

We measured the GPS positional errors by using the WayFinder 3 application on board all devices, while following trip routes in several geographical locations, three Seattle neighborhoods and 6 MeraKey Canonical trips selected by AbleLink in the Pennsylvania region. A trip analysis consists of multiple WayPoints measured in replicates on GPS receivers and recorded by both AbleLink and AWARE applications for each device.

Before starting the field test, we determine the reference coordinates of the WayPoints programmed into the field test routes. The IE team gathered the gold standard waypoint locations from the AbleLink canonical trip definition and for a sampling of the Pennsylvania WayPoints, compared to Google Maps latitude/longitude annotations to confirm that map projections were the same in both systems (this ensures that the tests are comparing and measuring distance in the same coordinate system). During the field tests, the acquired GPS locations in the app were recorded via the AbleLink Dashboard. Observations were conducted as part of the “WayFinder 3 Internal Field Test” and the MeraKey field tests.

The following analytics were calculated from the GPS locations collected and observed:

- Positional Accuracy:

  If a GPS receiver displays position coordinates that are different from the "true coordinates" of the canonical trip position, this registers as a positional error. Since the waypoints we evaluate are predetermined through the design of our field test routes (whether in Seattle or MeraKey tests), we are able directly measure this error. That is, the degree of conformance between the estimated and measured position can be directly assessed.

  We use the Distance Root Mean Square (DRMS) from the known location to a controlled waypoint (x,y) defined as:

$$\sigma_x = \frac{\sqrt{(\sum_{(i=1..n)}(x_i - x)^2}}{(n-1)}, \quad \sigma_y = \frac{\sqrt{(\sum_{(i=1..n)}(y_i - y)^2}}{(n-1)}$$

$$DRMS = \sqrt{(\sigma_x^2 + \sigma_y^2)}$$

Source: https://www.researchgate.net/publication/266150506.

Figure 10. Equation. Distance root mean square error.

  Where σ_x and σ_y denote the standard deviation of the positional error along the x axis and y axis respectively.

- Reliability: at least three replicates will be collected for each waypoint.

  Using these replicates, we can collect the descriptive statistics of the DRMS observed. Specifically, we will calculate the mean DRMS and standard deviation of the DRMS for both the GPS location collected via WayFinder 3 and AWARE.

- Regression: we will perform regression analysis, fitting a generalized linear model to the DRMS figures. This regression will allow us to model how close to the actual GPS location the different receivers registers 50 percent, 65 percent and 95 percent of the time. (this is the tolerance measures at different confidence measures).

- Acceptance: We will perform a Wilcoxon-Sum test pairing the DRMS values for the WayFinder 3 application with the AWARE application. We will accept the null hypothesis that the 2 receivers are comparable if the Wilcoxon-Sum test establishes that the 2 data series are drawn from the same distribution to 95 percent confidence.

Methodology was adapted from https://www.researchgate.net/publication/266150506.

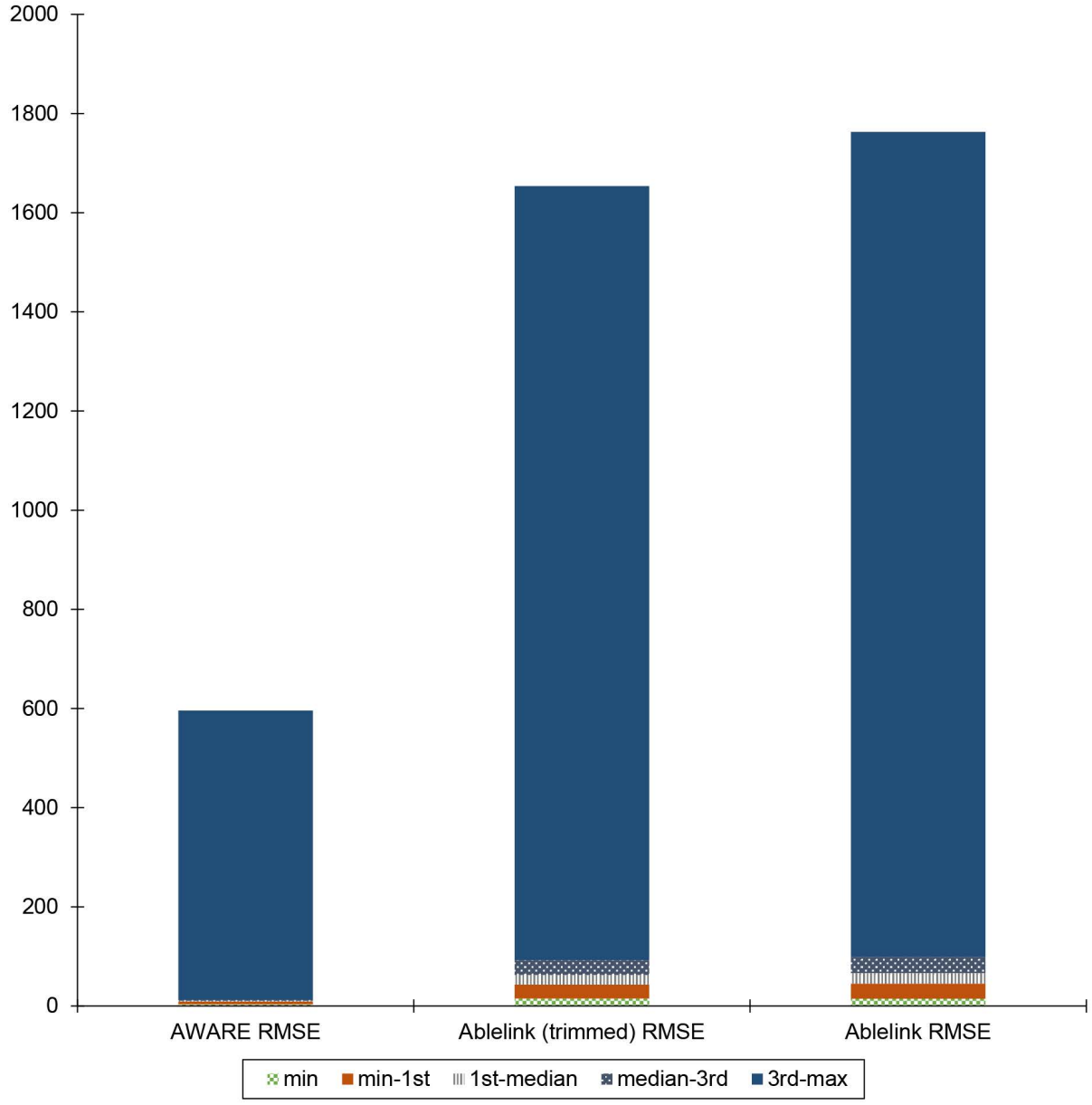
Analysis Outcome:

Findings

Fifty-two trips were analyzed for GPS positional accuracy and the Distance Root Mean Square metric was calculated from the observed location (either via AbleLink or AWARE) to a controlled waypoint.

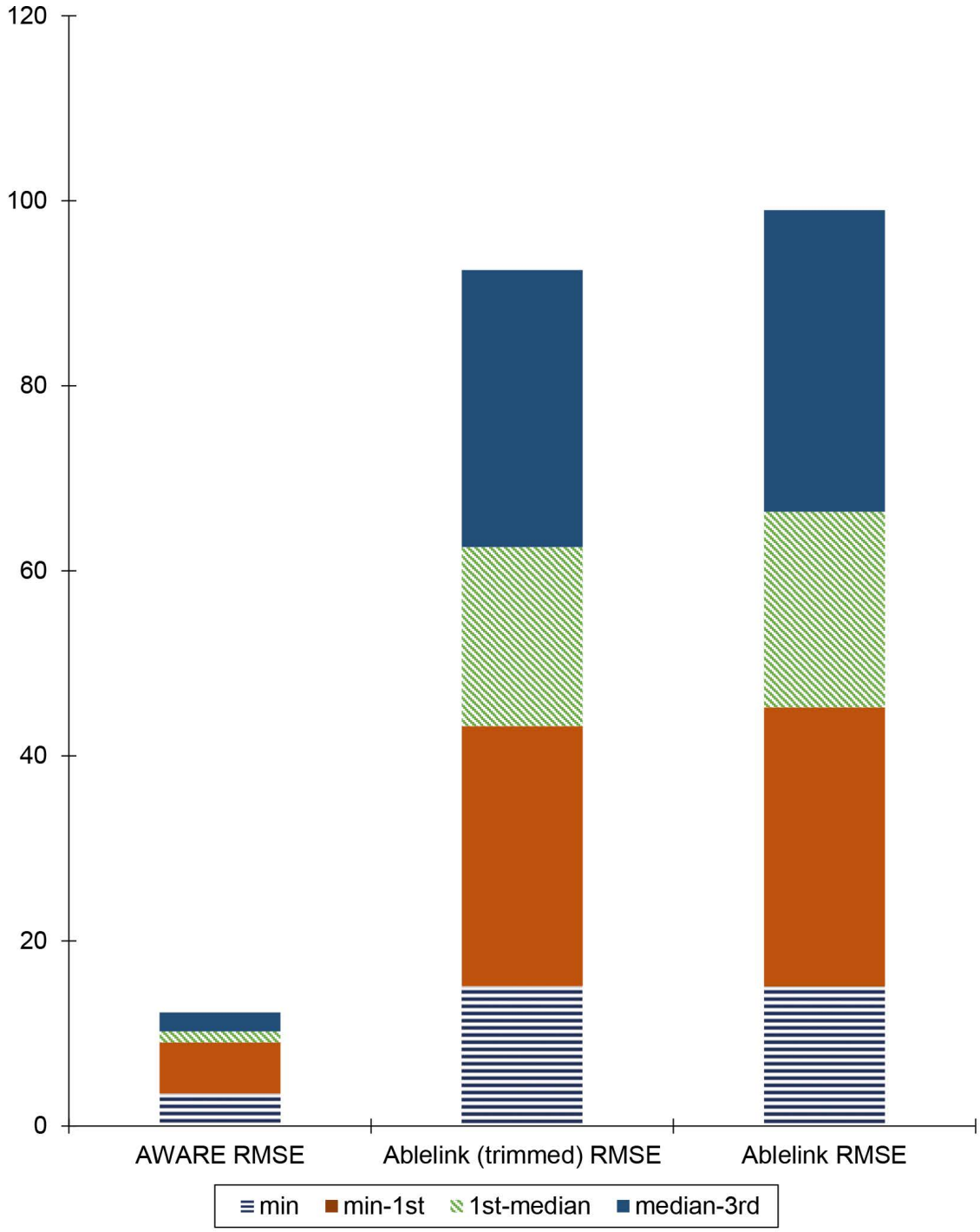
The two graphs show the root-mean-square error (RMSE) values in meters for AWARE GPS location distance from the ground truth and AbleLink GPS location distance to the ground truth. Two data series for AbleLink were offered since the IE team observed that the beginning and ends of GPS trips had

particularly bad error rates compared to the internal GPS locations of the trips. The IE team could not explain this difference, but tried to understand whether trimming the first and last 5 GPS location points of the AbleLink GPS data would improve the RMSE. This is referred to as the AbleLink (trimmed) data in the graphs. figure 11 Shows the full range of RMSE values (in meters). In figure 12, we exclude the 4th quartile in order to show with greater resolution how close the actual GPS location is to the observed location by the two different receivers.



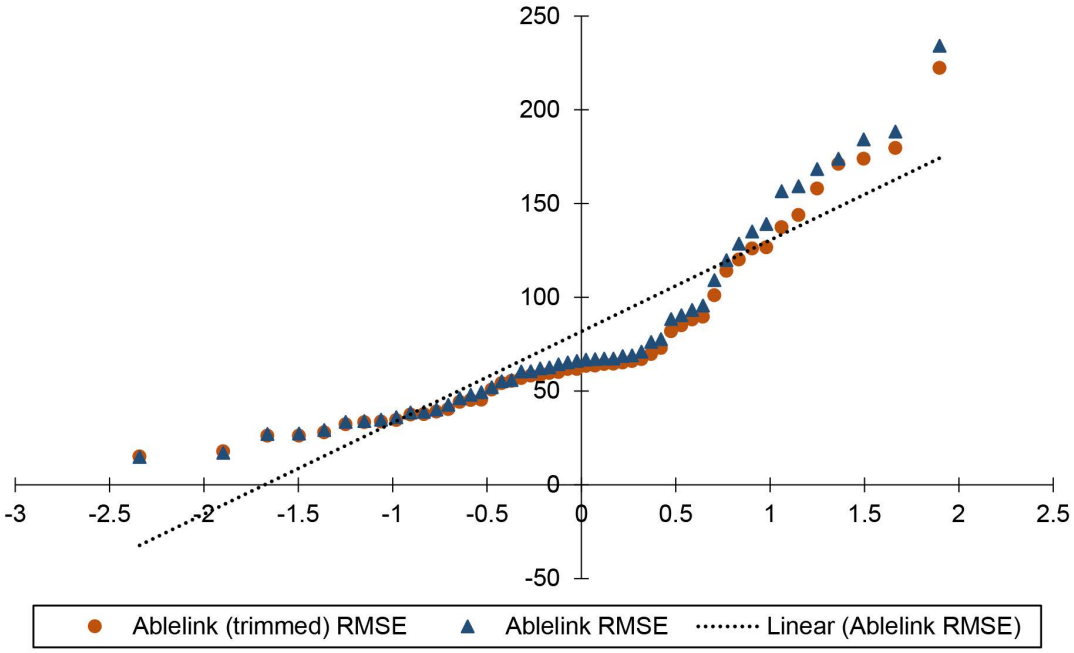
Source: Federal Highway Administration.

Figure 11. Graph. Root-mean square error values in meters for four quartiles of root-mean square error data.



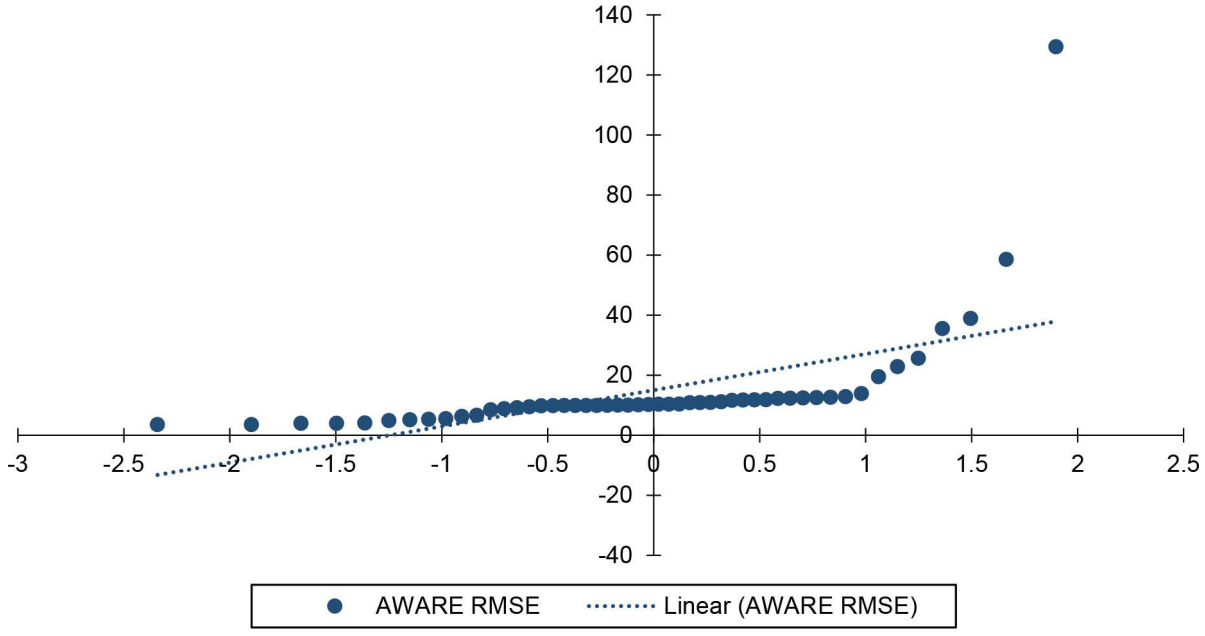
Source: Federal Highway Administration.

Figure 12 Graph. First three quartiles (excluding the last quartile) of data for root-mean square error values in the AWARE and AbleLink global positioning system datasets.



Source: Federal Highway Administration.

Figure 13. Quantile-quantile plot. AbleLink and AbleLink (trimmed) data root-mean square error with comparison to linear 'normal' distribution trendline.



Source: Federal Highway Administration.

Figure 14. Quantile-quantile plot. AWARE data root-mean square error with comparison to linear 'normal' distribution trendline.

Our regression analysis shows that both RMSE datasets are not following a typical normal distribution. The IE team verified this by plotting quantile-quantile (QQ) plots for both datasets. To interpret these plots, if the data values fall along a roughly straight line at a 45-degree angle, then the data is normally distributed. Visual and analytic confirmation shows that the two sensor error distributions are not normally distributed, but seem different from one another. The remaining analysis is to assess whether they indeed are drawn from different distributions. When we look at the QQ plots for the two groups, we see that neither is very normal, but more pertinent to our hypothesis of whether the AbleLink GPS location data is comparable to GPS sensing in other location-based applications, is the quartile plots that show that the AbleLink data is not symmetric (see figure 11). We therefore need to do the Wilcoxon Sign Rank Test (rather than a typical t-test), to assess whether the two datasets are drawn from a different distribution. Indeed, the W statistic for the Matched Pair Test is 103, which is smaller than the critical value for a 52 trip sample, so the finding is significant The critical values for the W test statistic are given in the Wilcoxon Signed-Ranks Table. Here we use $\alpha = .05$ and $n = 52$ (i.e., the 52 trips). From the table we find that $W_{crit} = 241$ (two-tail test, normal approximation). Since $W_{crit} = 241 < 103 = T$, we can't reject the null hypothesis (i.e., $p \geq .05$), and so H1: The median difference is negative and significant at $\alpha=0.05$. What this means in practice is that we expect to see more higher valued distance errors in the AbleLink dataset than the AWARE location data. The test, unfortunately, cannot predict or ascertain the cause for the difference, it is just validating statistically, that the difference we see in the observations is not by chance.

Rejected Hypothesis

Based on the observed location data for the AbleLink and AWARE GPS traces, and by comparison to the ground truth for the canonical trip routes, we cannot conclude that the outdoor Global Positioning System localization and the inferred proximity to a WayPoint is comparable in AbleLink to the AWARE location-based services. Looking at the RMSE values with mean error values around 50m, we believe improvements in the GPS localization in WayFinder would significantly improve traveler outcomes because this will impact all location-based determinations done via the app.

Hypothesis: The usability and design of the WayFinder 3 app interface is accessible and appropriate for the target population both by active field-testing respondents in the target population and by heuristic usability evaluation.

Performance Metrics:

In-lab heuristic evaluation of the technology. We derive all metrics directly from the heuristic evaluation (yes/no/maybe) performed for the first hypothesis evaluated (Effectiveness of Wayfinder in improving independent trip completion.)

Data Elements and Sources:

The In-lab Heuristic Evaluation and Internal Field Test.

Data Collection Period:

August 2020 to October 2020.

Analysis Procedure:

Researchers were given a list of realistic tasks (realistic scenarios) to walk through in a trip scenario. Researchers went through the mobile application flows and respective interfaces independently and analyzed the process and results against the goals and defined heuristics. When coming across an issue or an area for improvement, they recorded it. After following the formal travel tasks, the researchers were encouraged to deviate and explore the system in any way that felt intuitive to them. Appendix A provides a theme-specific heuristic evaluation report that was also furnished to the ADP Technology group.

Analysis Outcomes:

Findings

Here we present some of the incidental findings that came up through the field tests that pointed at some usability concerns with WayFinder. For presentation purposes, we simplify using a combination of heuristics based on Nielsen and Molich's 10 user interface design heuristics and Ben Shneiderman's eight golden rules. We include findings having to do with:

- Visibility of system status.
- Match between system and the real world.
- User control and freedom.
- Consistency and standards.
- Error prevention.
- Recognition rather than recall.
- Flexibility and efficiency of use.
- Help users recognize, diagnose and recover from errors.

Researchers were encouraged to write as much detail and specifics as possible and included the issue found, together with relevant details such as what the task attempted was, where they encountered the problem and why it is a problem. However, since researchers were engaged with actual trip activities, some of the details remained unclear.

In this post-test analysis, the IE team provides:

- Violated heuristic.
- Severity of the issue, following this suggested classification:
    - 0 = Not a usability problem but a suggestion.
    - 1 = Cosmetic problem: does not need to be fixed unless extra time allows.
    - 2 = Minor usability problem: fix is a low priority.
    - 3 = Major usability problem: it is important to fix; high priority.
    - 4 = Usability impediment: imperative to fix.

Example of how issues were coded is shown in Table 17.

Table 17. Examples of How Usability Issues Were Coded.

Severity	Heuristic Violated	Comment verbatim
2	Consistency and Standards	The WayFinder app displays a busy indicator, which is a user interface (UI) device typically used to inform the user that the application is working, in the wrong place. When the user begins going on a route and presses START, a new view with a wheel turning indicator is turned on- it is inconsistently used because at that point, the APPLICATION is waiting on the USER, not the other way around. In fact, even we were confused by why the wheels on the bus were turning, and waited nearly 3 minutes before beginning to walk on the first trip in anticipation of the app about to do something.
0	Consistency and Standards	Inconsistency between iOS messages and Android messages where YES and NO are placed in alternative locations. Important to maintain consistency in UI
0	Visibility of System Status	When just starting, the iOS splash screen was much nicer than the 'nothing' that come up in Android.
3	User Control	iOS, now the ok button is covered up by the contact button and cannot be pressed.
2	Error Prevention/ Visibility of System Status	Pressing the ok button yields nothing and the background goes back to the moving bus the waiting wheels turning.
3	Error Prevention	In general, the waypoints really should play in anticipation of the waypoint. That was way too close to the bus station to play the stop information.
3	Error Prevention	So anytime there's crossover in the routes, Android mistakenly just plays whatever is close to the waypoint regardless less of the ordering. This is not good for trips that require backtracking between different modalities of transportation as it is the case The State Burger Route.
0	Visibility of System Status	Android did finally play the welcome message upon opening of the Android app.
0	Consistency and Standards	General collection issues with secondary user notification emails: for the iPhone, Device Name is always empty. The Subject line specifies what type of phone it is, but sometimes it does not.

Source: Federal Highway Administration.

Chapter 4. Performing Gap Analysis

In any independent evaluation (IE) there are unquestionably remaining gaps. Especially in trying to address heterogeneous populations- someone at some point will be left out. It is important for IE teams to make the Accessibility Development Projects (ADP) teams aware of any issues that appeared in the course of the evaluation that may be important for deployment or travel outcomes. As the project team performed the evaluation tasks described above, we identified some concerns that might limit the deployment of the Wayfinder system. In particular, the IE team identified potential barriers such as:

- The ability of caretakers to gain confidence that specific individuals are capable of taking trips independently. (For example, the issue may not be their ability to follow the Wayfinder directions, but their ability to navigate social interactions in public without support.)
- The accessibility of transit routes to the care facilities.
- The ability of caretakers to adequately construct trip instructions for individuals.

The primary mechanism used for identifying gaps was to work with the caretakers participating in the study to learn about issues and concerns the care staff have that need to be addressed either prior to the system being deployed, or before it could be adopted more widely within the Merakey care system.

Aside from findings incidental to the field tests, the IE team recommends following a specific methodology for the Gap analysis, we encourage IE teams to go back to the User Needs Checklist and the User Threat model to identify those issues that had not been addressed by the IE due to resource or time constraints and identifying priorities and recommendations for the ADP, indicating how these issues may impact future work.

User Needs Checklist Gap Review:

1. Does ADP provide information in a variety of accessible formats? (Yes, medium) The ADP's smartphone application is designed to allow users to input their own visual or audio instructions. The user can customize their profile so that it includes extra instructions, reminders, or assistance information. They choose from specific, preprogrammed routes, and are allowed (but not forced) to add any additional information that they find necessary.
2. Is information from ADP interface accessible in a variety of environments (i.e., amid heavy crowds and noise, underground)? (Yes, high) The ADP's smartphone application is designed to allow users to obtain information both aurally and visually. Both systems are designed to function effectively in street environments, which are both loud and often subject to bad weather (e.g., glare and rain).
3. Does ADP perform a task that improves safety and security or that provides emergency information? (No, high) The ADP technology is not designed to improve user safety; however, in order to mitigate risks to the user, it does allow users to input their own emergency contact information so that the user can easily reach out for additional assistance.

4. Does ADP provide en-route assistance and information? (Yes, high) The current version of the ADP technology does include a “request help” function. The user is responsible inputting the emergency contact information.
5. Does ADP provide connection information (where, who, when)? (No, high) the ADP does not provide transit connection information. Nor does it provide transit schedules or arrival times.
6. Does ADP provide estimated trip length and distance? (No, medium) The ADP does not provide estimated trip length and distance.
7. Does ADP provide comprehensive travel information? (No, low) The ADP does not provide comprehensive travel information. The ADP’s technological function is restricted to guiding the user along a predetermined route and letting them know when they have arrived at their destination. All other information must be inputted by the user.
8. Does ADP require access to equipment (phones, computers, charging, training)? (Yes, high) The ADP technology is smartphone based, so members of the intended user population must have access to a smartphone and the charging infrastructure needed to support that phone.
9. Does ADP allow the user to create a personalized profile? (Yes, high) The ADP’s smartphone application has features that allow users to set choose from a collection of set routes and then they can input any additional information that they need.
10. Does ADP require coordination of information (between agencies, modes)? (No, medium) The ADP only guides the user along predetermined routes that do not require coordination of information between agencies.
11. Does ADP provide real-time transportation information, including 1) real-time vehicle status; or 2) real-time travel condition/obstruction information? (No, medium) The ADP technology does not provide real-time vehicle status updates or travel condition information.
12. Does ADP provide audible mapping/directions? (Yes, medium) The ADP allows the user to input their own audible mapping directions in addition to the visual instructions.
13. Does ADP provide destination information (hours, addresses, entrances, layout)? (No, low) The ADP does not provide destination information.
14. Does ADP provide transit schedule and other transit information (e.g., stop location)? (No, medium) The ADP does not provide transit schedules of arrival information.
15. Does ADP provide information about pathway infrastructure? (No, low) The ADP does not provide information about pathway infrastructure.
16. Does ADP include provision for outside assistance or attendants? (No, low) The ADP technology is not designed so that someone other than the user would access the technology. The technology is controlled exclusively by the user and does not communicate with others.
17. Does ADP provide amenity information (e.g., restroom, shelter, benches, food, drinks)? (No, low) This ADP is not specifically designed to include amenity information associated with buildings along the path.
18. Does ADP provide information about, and interpretation of, signage? (No, high) The ADP is “self contained.” That is, the ADP technology does not interact with external signage and therefore does not provide interpretation of signs. Navigation directions are provided on the basis of

internal map databases and the estimate of the user's current location as identified by the technology (smartphone).

19. Does ADP provide transportation facility information (e.g., maps)? (No, high) The ADP does not provide transportation facility information.
20. Does ADP provide information about weather conditions? (No, high) The ADP technology does not include weather information.
21. Does ADP include information about and/or interpretation of announcements? (No, high) The ADP does not include information about announcements; It is "self contained." That is, the ADP technology does not interact with external announcements and therefore does not provide interpretation of announcements.
22. Does ADP incorporate speech-to-text or text-to-speech that enables the user to communicate more easily? (No, high) The ADP does not provide speech-to-text or text-to-speech functionality; however, it provides visual instructions and it allows the user to input their own aural instructions.
23. Does the ADP provide information in a concise and straightforward manner? (Yes, high) This ADP is designed for people whose primary disability is characterized as cognitive or intellectual.

Threat Model Gap Review

In developing a threat model, the IE team considered the safety threats specific to the population of interest. Traveler safety is essential in any setting where people navigate nonmotorized paths, ride public/demand-responsive transit, or receive transportation services/assistance. Traveler safety is not only contingent on the available infrastructure and services; it also assumes that the system provides proper cues for the traveler so that they can respond to the environment effectively. The sheer complexity of the process means that there are many opportunities for error, especially within the population of concern for the Accessible Transportation Technologies Research Initiative (ATTRI).

There are several areas of concern in which threats to safety ought to be highlighted:

- Traveler is unable to orient in time (examples: keeping to a time schedule, or time of arrival for a bus) (AbleLink will offer timing alerts to leave origin in time, to await the bus, etc.)
- Traveler is unable to orient in space (examples: leaving a store and being unsure whether the bus stop is to the right or left) (AbleLink will offer alerts when the traveler is veering off the route.)
- Traveler must cross the road with skill and attention (AbleLink will remind traveler about appropriate street crossing behaviors, for example, reminding traveler to look both ways, through use of the "pre-trip concierge assessment system.")
- Traveler must know where to stand in a bus stop or platform or pick up location to be seen by transit provider. (AbleLink will remind travelers about staying visible through the "pre-trip concierge assessment system." For instance, to stand by the pole or station on the side of the street, but not in the street, and face the direction from which the bus will be coming).
- Traveler must show intent to board. (AbleLink will remind travelers about showing others their intent through the "pre-trip concierge assessment system.")

- Traveler must be able to negotiate onboarding and payment. (AbleLink will remind travelers about payment systems through the "pre-trip concierge assessment system.")
- Traveler must be able to identify the correct vehicle to board. (AbleLink could be used to show image of bus with route number displayed and if/when real-time alerts are available, alert rider as bus approaches stop/station.)
- Traveler must conform to local appropriate social travel behavior. (AbleLink will remind travelers about appropriate travel behavior through the "pre-trip concierge assessment system," for example, turn taking when boarding a vehicle, or showing a bus pass to the driver)
- Traveler must be able to communicate her/his needs, and be able to identify the right people with whom to communicate those needs (for example, authorities, customer assistance, or driver).
- Traveler must be able to board moving objects and remain secure onboard. (AbleLink will remind travelers about effective travel behavior through the "pre-trip concierge assessment system.")
- Traveler must be able to select a secure position or seat on a vehicle, from which s/he is able to watch the road. (AbleLink will remind travelers about effective travel behavior through the "pre-trip concierge assessment system.")
- Traveler must be able to pay attention to the bus whereabouts and select landmarks in anticipation of stop. (AbleLink alerts such as "get ready for your stop" should be used as additional supports).
- Traveler must be able to identify stop and exit the vehicle at the correct time. (AbleLink will alert users to preempt their stop and off-board the vehicle.)
- Traveler must be able to anticipate stop and request a stop from vehicle operator. (AbleLink will alert users to preempt their stop and off-board the vehicle.)
- Traveler must identify incorrect stop (AbleLink alerts such as "not your stop" should be used as additional supports.)
- Traveler must identify the correct entrance/egress transitioning to/from different travel environments and modes. (AbleLink may be used as a navigation tool where the best transitions through environments are expressed through the caregiver-recorded routes.)
- Traveler must be able to call for help or assistance throughout a trip. (AbleLink call for help should be available from any view in the navigation tool application. It is imperative that these are personalized and available during application failure modes.)
- Travelers must be able to identify emergent or altered travel conditions where emergency or rerouting might be needed. (AbleLink call for help should be personalized and available during all application modes, including after a technical failure of the application.)

In looking at the specific travel safety and security threats for the travel population of interest, it is clear that AbleLink primarily relies on the use of two modes of interaction to mitigate traveler risk. The first is the in-app notification system, and the ability for the app to generate appropriate alerts along the trip. The second is the "pre-trip concierge assessment system," which is used as an educational tool for the traveler. Note: It is difficult to separate the AbleLink #1 (the WayFinder 3 app) and AbleLink #2 (the pre-trip concierge assessment) projects; however, in order to evaluate only the WayFinder app, this document makes certain assertions. Specifically, we treat the pre-trip concierge assessment system (the work product of AbleLink #2) as an existing technology. This is analogous to the user-interaction device known as "in-app notifications," which are a pre-existing technology used by AbleLink #1 in the WayFinder 3

mobile app to produce alerts. It is within the scope of this evaluation to examine whether AbleLink #1 makes proper use and timely invocation of the “in-app notifications” or “pre-trip concierge assessment system.” It is outside the scope of this evaluation to examine whether these external components are effective or efficacious.

In the AbleLink context, the WayFinder 3 application itself does not offer real-time or static information about safety or security. Nor does it provide just-in-time alerts for emergency situations. Instead, the WayFinder mobile app uses two methods: (1) to prepare riders for the unexpected, there is a pre-trip concierge assessment system and (2) in the event the app recognizes that the user has veered off the route, an in-app notification will alert the user. Here we explain how these methods are used by the WayFinder app to mitigate safety concerns for the traveler. The WayFinder app connects to a Web-based, pre-trip concierge assessment system that allows individuals to complete a self-assessment of their public transit experience and skills. The self-assessments evaluate one of the following: transportation skills and experience, street crossing skills, social skills related to traveling in the community, or vehicle identification skills. These interventions address the most frequent human errors that members of the AbleLink population of interest exhibit during travel. The current WayFinder app does not offer a contextually aware decision-support system to improve user practices in real time. It is possible to implement a support system that responds when the system identifies that the user may be making an error or by invoking reminders to the relevant safety intervention (e.g., in-app reminder to the Street Crossing Skills tutorial when the app user is about to cross the street). Through building the foundational infrastructure for a versatile travel concierge tool in WayFinder, the AbleLink team is well poised and positioned to take a leading role in innovative futures for travelers with cognitive disabilities.

Chapter 5. Evaluation Hypothesis Summary

Hypothesis Summary

Accepted: The WayFinder app is effective in improving primary users' overall independent trip completion.

Hypothesis demonstration: Inter-user, trips using WayFinder show increasingly successful completion of travel, even when users enable multimodality options.

No Hypothesis Determination: The WayFinder in-app notifications are effective in reducing primary users' unintended, midtrip errors.

Hypothesis demonstration: Intra-user, over the study period, trips performed using WayFinder show a decline in midtrip, unintended veering off route in response to in-app notifications to the primary users.

Accepted: The Technology improves or maintains primary users' time efficiency while navigating legs of trips.

Hypothesis demonstration: Intra-user, over the study period, trip legs performed using WayFinder show increased or unchanged relative time efficiency.

Rejected: Participants are satisfied with use of the WayFinder app and their satisfaction rating is independent of their ability to complete tasks or complete trips.

Participant's opinion was colored by the success of individual trips even if they have used the application before and had trip completion success with it. 2 standard deviations below the mean were negative. Generally, we believe that there are some performance issues to address in order to increase traveler satisfaction.

Rejected: The technology does not adversely impact an individual's ability to utilize other mobile applications.

Rejected: Over the course of repeated trials to input routes and use routes, the mobile application does not slow down, quit operation or result in unexplainable error.

Rejected: The technology opportunistically aims to prevent primary user risks as part of strategic and operational planning.

Hypothesis demonstration: Over the course of repeated trials to use routes, the mobile application alerts are used to communicate risk to users.

Rejected: When either routing primary users or during operational failures, WayFinder provides the primary user with appropriate triggers to enlist assistance or call for help.

Rejected: Outdoor Global Positioning System localization and the inferred proximity to a WayPoint is comparable to other leading location-based services and is appropriate for the task.

Rejected: The usability and design of the WayFinder 3 app interface is accessible and appropriate for the target population both by active field-testing respondents in the target population and by heuristic usability evaluation.

Appendix A. Theme-Specific Heuristic Evaluation Report

The tables shown in this appendix show the outcomes of the Heuristic Evaluation In-Lab Tests. Test results are grouped by category: 1) robustness, 2) location-based services, 3) error prevention, call for assistance and failure modes, 4) adaptability, and 5) communication across informational gaps.

Table 18. Category I: robustness.

Evaluation Findings	AbleLink element or feature assessed
⊠Yes ⊠No ●N/A	Runs on any Android device running Jelly Bean (Operating System (OS) 4.2) or newer and any Apple device running iPhone Operating System (iOS) 9 or better (Note: Field testing has shown that performance can vary from device to device, depending largely on the quality of the GPS technology in the device. Specifically states GPS is not designed to work indoors, or below ground level, in most cases).
⊠Yes ●No ⊠N/A	GPS localization is sufficiently accurate (5m to 10m distance), performs within acceptable tolerance, is reliable within 5 percent confidence interval and repeatable as observed in triplicates.
⊠Yes ●No ⊠N/A	WayFinder plays Not Your Stop waypoints when appropriate and does not play them if the bus does not stop.
●Yes ⊠No ⊠N/A	Launches app with the ability to select a set of instructions.
⊠Yes ●No ⊠N/A	Launches selected GPS-based location instructions within the set In-Range Distance.
⊠Yes ⊠No ●N/A	Device reliably vibrates during route when turned feature is enabled.
●Yes ⊠No ⊠N/A	If Show Route Exit button is enabled, an Exit button appears that allows user to stop route midtrip.
⊠Yes ⊠No ●N/A	If Show Waypoint Preview is enabled, as soon as one waypoint has been passed WayFinder will immediately show the picture for the next waypoint on the route.

Source: UW Independent Evaluation team.

Table 19. Category II: location-based services.

Evaluation Findings	AbleLink element or feature assessed
⊠Yes ●No ⊠N/A	Launches selected GPS-based location instructions within the set In-Range Distance.
⊠Yes ⊠No ●N/A	When Local Routes Only is enabled, Main Menu only shows routes that start within the established Local Route Distance.
⊠Yes ⊠No ●N/A	Provide signal range and travel speed for walking or vehicle travel.
●Yes ⊠No ⊠N/A	Automatically stop playing the waypoint message when Out-of-Range Distance.
⊠Yes ●No ⊠N/A	WayFinder plays Not Your Stop waypoints when appropriate and does not play if the bus does not stop.
⊠Yes ⊠No ●N/A	If Show Waypoint Preview is enabled, as soon as one waypoint has been passed WayFinder will immediately show the picture for the next waypoint on the route.
●Yes ⊠No ⊠N/A	Accurately represent strength of GPS signal.
⊠Yes ●No ⊠N/A	Accurately represent user location.
⊠Yes ●No ⊠N/A	If enabled, only display routes which start within the user's immediate vicinity.
⊠Yes ⊠No ●N/A	Route status indicator shows the relative position on a travel route during playback.

Source: UW Independent Evaluation team.

Table 20. Category III: error prevention, call for assistance and failure modes.

Evaluation Findings	AbleLink element or feature assessed
●Yes ⊠No ⊠N/A	If feature is enabled, WayFinder automatically sends email notifications (with time, date, Google Map location) when the Start button is tapped after selecting a WayFinder route from the Main Menu, if and when a route that has been started is aborted by the end user, and when a route has been completed as indicated by the user tapping the Done button at the end of a route.
⊠Yes ●No ⊠N/A	If Contact Me button is turned on, selecting it sends an email notification and Google Maps link to appropriate party.
⊠Yes ●No ⊠N/A	If Contact button is enabled and pressed, an email is sent to the address indicated in the Notifications field with a message that the user is okay along with a Google Maps link of the user's current location.
●Yes ⊠No ⊠N/A	If Periodic Notifications button is enabled, email notifications are automatically sent at the indicated frequency in Interval.
●Yes ⊠No ⊠N/A	If enabled, receive periodic email messages with location information to track a traveler's progress when traveling.
⊠Yes ●No ⊠N/A	Request a caregiver to call from within the WayFinder app.
⊠Yes ●No ⊠N/A	Send an "I'm okay" message from within the WayFinder app.
●Yes ⊠No ⊠N/A	If Use Corridor Data is enabled, a notification with a Google Maps location link will be sent when the user goes off the planned route. Notifications also will be sent at the point when the user gets back on route.
●Yes ⊠No ⊠N/A	Automatically stop playing the waypoint message when Out-of-Range Distance.
⊠Yes ●No ⊠N/A	Play "Not Your Stop" when appropriate.
●Yes ⊠No ⊠N/A	If Show Route Exit button is enabled, an Exit button appears that allows user to stop route midtrip.

Source: UW Independent Evaluation team.

Table 21. Category IV: adaptability.

Evaluation Findings	AbleLink element or feature assessed
☒Yes ☒No ●N/A	Route Label Size allows user to adjust size of on-screen text.
●Yes ☒No ☒N/A	If On-Foot Travel is enabled, GPS settings automatically optimize for a slower travel speed.
●Yes ☒No ☒N/A	If Show Route Exit Button is enabled, an Exit button appears that allows user to stop route midtrip.
☒Yes ☒No ●N/A	If Show Waypoint Preview is enabled, as soon as one waypoint has been passed WayFinder will immediately show the picture for the next waypoint on the route.

Source: UW Independent Evaluation team.

Table 22. Category V: communication across informational gaps.

Evaluation Findings	AbleLink element or feature assessed
●Yes ☒No ☒N/A	Audio prompts for travel instructions.
●Yes ☒No ☒N/A	Picture prompts for travel instructions.
●Yes ☒No ☒N/A	Text-based prompts for travel instructions.
●Yes ☒No ☒N/A	Audible feedback offered on WayFinder Settings.
☒Yes ●No ☒N/A	Device vibrates (if feature is present) when GPS waypoints are activated along a route.
☒Yes ☒No ●N/A	Device reliably vibrates during route when turned feature is enabled.
☒Yes ☒No ●N/A	Route Label Size allows user to adjust size of on-screen text.

Source: UW Independent Evaluation.

Appendix B. Analysis Details for Hypothesis Testing

The data and analytical details that support the conclusions drawn and reported for the hypotheses tested in this project have been delivered to USDOT for inclusion in their open data repository. The submission for this project includes a “Read Me” file and eight Excel spreadsheets. Each spreadsheet contains multiple tabs. Those tabs include the data collected and used for this project and the various statistical tests described in chapter 3 of this report. The Read Me file provides a detailed description of, and roadmap for using, those spreadsheets.

U.S. Department of Transportation
ITS Joint Program Office—HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free “Help Line” 866-367-7487

www.its.dot.gov

FHWA-JPO-21-838

U.S. Department of Transportation