Taking the lab on the road and bringing the road to the lab: On using mixed-methods and virtual reality to study a location-based task

by

Kofi Whitney

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Human Computer Interaction

Program of Study Committee:
Les Miller, Co-major Professor
Stephen Gilbert, Co-major Professor
Sunghyun Kang
Jon Kelly
Johnny Wong

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

DEDICATION

For my son Thomas, who is my greatest inspiration. For my mother and father, who gave me my roots and my wings. For the Whitney family, the Knox family, the Knapp family, and for my closest friends and loved ones—they have been there for me through the thick and thin of it… always encouraging me… always believing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

The terms and abbreviations used in this dissertation are as follows:

CAVE	CAVE Automatic Virtual Environment
EFS	Ethnographic Field Study
FE1 / FE2	Field Experiment 1 / Field Experiment 2
HighVZ	High Spatial Visualization Participant Group
IVE	Immersive Virtual Environment
LowVZ	Low Spatial Visualization Participant Group
PT	Perspective Taking Ability
PS	Perceptual Speed Ability
RLE	Rusch's Lab Experiment
RW	Real World
RWT	Real-world Treatment (or Experimental Group)
RWT+HighVZ	High Spatial Viz. Participants from Real-world Treatment
RWT+LowVZ	Low Spatial Viz. Participants from Real-world Treatment
VE	Virtual Environment
VM	Visual Memory Ability
VR	Virtual Reality
VRFE	VR/Field Experiment
VRT	Virtual Reality Treatment (or Experimental Group)
VRT+HighVZ	High Spatial Viz. Participants from VR Treatment
VRT+LowVZ	Low Spatial Viz. Participants from VR Treatment
VZ	Spatial Visualization Ability based on VZ-2 test; Ekstrom, et al, (1976)

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Les Miller, for his guidance and devotion—words cannot express my gratitude for the impact that he has had on my growth as a researcher and as an individual. A big thanks to my co-major professor, Dr. Stephen Gilbert, for his tremendous support throughout the research and to my committee members—Dr. Jon Kelly, Dr. Johnny Wong, and Professor Sunghyun Kang—for their commitment to the work and for sharing their knowledge and insights.

Thanks to Dr. Sarah Nusser for her engagement and valuable contributions. Thanks to Dr. Georgi Batinov for his partnership from day-one and for his collaboration and ceaseless optimism. Thanks to the members of my research group, Dr. Thitivatr (Randy) Patanasakpinyo and Adel Sulaiman, who hit the ground running. Thanks to Dr. Al Taylor for his encouragement.

This work would not have been possible without the partnership and support from the faculty, staff, and students of the Virtual Reality Applications Center (VRAC), the HCI Graduate Program, and the Computer Science Department—their dedication to research and their responsiveness to the needs of the students cannot be overstated; the same can be said for the support provided by the GMAP Program as well as The National GEM Consortium. I would also like to thank the U.S. Census Bureau for their contributions and funding.

I am grateful to my friends, colleagues, and professors for making my time at Iowa State University a wonderful experience. I want to also offer my appreciation to those who were willing to participate in my research, without whom, this dissertation would not have been possible.

ABSTRACT

In earlier work, a lab and field study were employed to evaluate participants who used a digital map to complete a surveying task. The lab incorporated photos to simulate various scenarios within the task environment. It had a high degree of experimental control, strengths in quantitative data collection, and it could be easily replicated. Whereas the field study took place in the task's natural setting (a neighborhood); it afforded participants to navigate the environment on foot, which allowed for more ecologically valid task outcomes and rich qualitative data collection. The strengths of the lab method were desired, but the rich context and the ecological validity of the field study proved to be critical to outcomes.

In this research, three follow-up experiments were conducted. The first two field experiments explored differences in task outcomes between fieldworkers with high spatial visualization ability and low spatial visualization ability. Participants completed a series of surveying tasks using paper maps while navigating a neighborhood. The results indicated that task performance outcomes and behavior could be linked to participants' spatial visualization ability, their map usage patterns, and environmental factors.

In the third experiment, a VR lab was used to replicate a field experiment on the task as it was performed by digital map users. An approach is highlighted to recreate the task environment—a neighborhood that was large in extent—using an immersive virtual environment (IVE). Outcomes from the field are compared to those of the VR lab, which enabled participants to move through the simulated neighborhood using a hands-free interface. Using this approach, strengths of the lab method (i.e., its experimental control) are combined with the ecological validity afforded to natural setting research. The results

indicate that real-world behaviors observed in the field—and some of the expected task performance outcomes—were also evident in the VR lab. Many of the findings corroborate those of the two prior field experiments. Comparisons made across experiments show that task outcomes were linked to participants' spatial visualization ability, their workflows, the street layout of the neighborhood, as well as the type of map used (i.e., paper vs. digital) and the styles of map use.

This methodology can be applied—in the field and in lab settings—to evaluate location-based tasks that involve pedestrian navigation and map use; it can also be used to assess and validate VR labs that are designed to replicate mobile HCI field studies by simulating real-world task environments.

CHAPTER 1 - INTRODUCTION

Researchers in mobile human-computer interaction can find themselves in a "catch-22" situation when deciding between a field study and a lab study. The challenge is that the environmental context of a mobile computing task (i.e., its ecological validity) is oftentimes critically important to task outcomes—for which a field study is well-suited. However, researchers have gravitated toward lab studies—which are less ecologically valid—because they afford greater experimental control, are more familiar to practitioners, and they are easier to conduct and manage (Kjeldskov & Graham, 2003). The next section presents two studies that the author contributed to; they exemplify the trade-offs made when choosing either a lab or a field study for evaluating a mobile computing task; they also highlight the benefits and complementary nature of these evaluation methods.

1.1 Earlier Studies

A lab experiment was conducted by members of our research group to examine how digital map users carried out a series of location-based tasks. A subsequent, one-day ethnographic field study on a related task was also conducted in coordination with the U.S. Census Bureau. This dissertation refers to the lab experiment as *Rusch's Lab Experiment (RLE)*—named after Michelle Rusch, the principal investigator (Rusch, Nusser, Miller, Batinov, & Whitney, 2012; Rusch, 2008); the field study is referred to as the *Ethnographic Field Study (EFS)*.

1.1.1 Rusch's Lab Experiment (RLE)

When planning user research for a location-based task, the environmental context of the activity is an important consideration. If the lab method is to be used, there is the challenge of recreating the task's context of use in the lab. In an initial lab experiment on the location-based task that is central to this dissertation (address verification), Rusch et al. (2012) incorporated views of the task environment into an office-based lab by using two PC monitors to show participants vantage points (i.e., photos) of a real-world neighborhood. While seated at a desk, participants referenced these vantage points and used a tablet device to verify the correctness of residential addresses shown on a digital map; ten different address verification scenarios were "simulated" in this format (see *Figure 1.1*). In the real world, however, fieldworkers would complete the task in a standing position while holding the mobile device, they would be able to walk and freely explore the neighborhood, and they would be responsible for all task-related activities—not just the verification of residential homes, but also route planning and navigation; the absence of these elements in the experiment of Rusch et al. (2012) underscores the challenges of ensuring that such a method leads to ecologically valid outcomes.

Figure 1.1 *Photo of the lab experiment used by Rusch (2008).*

The experiment of Rusch et al. (2012) also made evident the strengths of using a lab to simulate and evaluate a location-based task. For example, unfavorable weather, untimely construction zones, or other potential disruptions that we encountered in the field were no longer a concern in the lab; safety was no longer a concern; data collection was streamlined and precise; and the setup, protocol and logistics of the experiment were relatively simple and easy to replicate. Furthermore, in the lab, participants were given tests on several cognitive abilities prior to completing the address verification scenarios. Measures of task performance and logs of participants' digital map usage were collected and analyzed. Rusch et al. found that participants' spatial visualization and perspective-taking abilities were correlated to task performance and digital map usage outcomes; these results exemplified the lab method's strengths in in terms of quantitative analyses and they served as a motivation for our group to continue to explore relationships between task outcomes and participants' spatial visualization ability.

1.1.2 The Ethnographic Field Study (EFS)

Following the lab experiment of Rusch et al. (2012), the author was among four researchers that used ethnographic methods to observe Census Bureau fieldworkers for a day. The fieldworkers had recently completed training on a task similar to address verification. Our group shadowed Census Bureau fieldworkers from their office, where planning activities began (see *Figure 1.2*), to various neighborhood areas so that we could observe fieldworkers as they used a handheld device and map-based software to capture information on residential homes. Despite the fact that we shadowed just a few participants over the course of a workday, we learned a great deal from observing authentic users execute a series of location-based tasks in "the wild." The lab experiment of Rusch et al. focused on a fraction of the overall task, whereas the field study allowed researchers to examine the entire process. During the field study, we recorded qualitative observations of participant behavior and workflows that were grounded in real-world scenarios. We also recognized that, with the right planning and equipment, we could augment the quantitative measures used by Rusch et al. (e.g., task time, distance traveled, errors made) and complement them with the qualitative observations that were of focus in the *Ethnographic Field Study*.

Figure 1.2 *Photo taken during one-day ethnography of Census Bureau fieldworkers*.

Table 1.1 provides an overview of the methods and approach that were adopted for the earlier research activities.

Table 1.1 *Overview of the Earlier Studies on Location-Based Tasks*.

	Research Method					**Data Method**		**Environmental Context**			**Spatial Viz.**		**Map Type**	
	Setting		**Type**											
Study or Experiment	Field	Lab	Ethnography	Experiment	Observ. Study	Qualitative	Quantitative	Natural setting	Photo-based	Immersive VE	Tested on it.	Screened on it.	Digital	Paper
Rusch's Lab Experiment[1]		X		X			X		X		X		X	
Ethnographic Field Study[2]	X		X			X		X					X	

[1] A prior lab experiment used for comparison (Rusch et al., 2012; Rusch, 2008); related, but not an experiment covered in this dissertation.
[2] A prior field study used for comparison; related, but not a study covered in this dissertation.

For example, *Table 1.1* indicates that for *Rusch's Lab Experiment*:

- The research method is a lab experiment (per the "Research Method" column).
- Only quantitative data was collected (per the "Data Method" column).
- The environmental context was provided (i.e., the task was simulated) using a photo-based technique (per the "Environmental Context" column).
- Participants were tested on spatial visualization ability (per the "Spatial Viz." column).
- Lastly, the participants used a digital map to complete their tasks (per the "Map Type" column).

In addition to providing an overview of the earlier studies, this format will be used to introduce and compare the three experiments that are central to the present work; they are discussed in the next section.

1.2 Overview of the Experiments

The three experiments detailed in this dissertation—*Field Experiment 1 (FE1)*, *Field Experiment 2 (FE2)*, and the *VR/Field Experiment (VRFE)*— all are centered on the location-based task that we refer to as "address verification". An overview of each experiment is shown in *Table 1.2.* A mixed-methods approach was used for each experiment. Quantitative data consisted of participants' task performance metrics and measures of their cognitive abilities. Qualitative data consisted of observations recorded by attending researchers in conjunction with participants' use of the think aloud method and their completion of an exit questionnaire at the end of each experiment. All participants were shadowed by at least one observer per experiment. *FE1* and *FE2* were conducted in a natural setting (i.e. a real-world neighborhood), whereas the *VRFE* compares a field study that is conducted in a neighborhood to one that is conducted in an immersive virtual environment that is a replica of the real-world neighborhood. Other important distinctions between the experiments are discussed in the *Objectives Section*.

Table 1.2 *Overview of Experiments Covered in this Dissertation*.

	Research Method					**Data Method**		**Environmental Context**			**Spatial Viz.**		**Map Type**	
	Setting		**Type**											
Study or Experiment	Field	Lab	Ethnography	Experiment	Observ. Study	Qualitative	Quantitative	Natural setting	Photo-based	Immersive VE	Tested on it.	Screened on it.	Digital	Paper
Rusch's Lab Experiment[1]		X		X			X		X		X		X	
Ethnographic Field Study[2]	X		X			X		X					X	
Field Experiment 1 (oblique streets)	X			X	X	X	X	X			X	X		X
Field Experiment 2 (oblique + grid)	X			X	X	X	X	X			X	X		X
VR/Field Experiment (oblique streets)	X	X		X	X	X	X	X		X	X	X	X	

1 A prior lab experiment used for comparison (Rusch et al., 2012; Rusch, 2008); related, but not an experiment covered in this dissertation.
2 A prior field study used for comparison; related, but not a study covered in this dissertation.

1.3 Objectives

In addition to gaining valuable insights about users' approaches to the address verification task, this research also addresses several other key objectives; they are described in this section.

1.3.1 Improving Methods Used to Evaluate a Location-based Task

In the preceding work, there were clear strengths to the office-based lab that was used by Rusch et al. (2012): It had a high degree of experimental control, a relatively simple implementation, strengths in quantitative analysis, and it could be easily replicated. In comparison, we appreciated that the ethnographic field study was situated in an authentic task environment (a neighborhood), thereby lending itself to rich, qualitative data collection as well as more ecologically valid task outcomes. The complementary nature of these methods motivated our group to improve upon them both, hence, iterations of the field study method are used in *Field Experiment 1*, *Field Experiment 2*, and the *VR/Field Experiment (VRFE)*. The *VRFE* is unique in that it examines the feasibility of replicating a field study by using a virtual reality lab to simulate the task environment, thus, both the lab and field study methods are employed as experimental treatments and compared in the *VRFE*. A major objective of the *VRFE* is to exemplify the VR lab's ability to reproduce field study outcomes for a location-based task (i.e., demonstrating its ecological validity), while retaining the high degree of experimental control that was afforded to the office lab used in the earlier work (see *Table 1.3*).

Table 1.3 *Characteristics of the Field and Lab Evaluation Settings*.

	Ecological Validity	**Experimental Control**
Field Setting[1]	Highest	Lowest
Office Lab[2]	Lowest	Highest
Proposed VR Lab[3]	High	High

[1] Field Setting: Represents the real-world neighborhoods that were used in the field studies.
[2] Office Lab: Represents the photo-based lab that was used by Rusch et al. (2012) to simulate the task in an office.
[3] Proposed VR Lab: Represents the proposed VR lab tasked with replicating a field study using an immersive virtual environment.

1.3.2 Understanding the Impact of Spatial Viz. Ability on Task Outcomes

The experiment of Rusch et al. (2012) evidenced connections between participants' task performance and their spatial visualization ability. There was further interest from our research group to examine groups of individuals who possessed either high or low spatial visualization ability relative to their peers; these individuals were of interest due to the spatial reasoning skills known to be involved in the task—for example, the task requires digital map use, navigation, and comparisons of residences seen in the field to their depictions on a map. Thus, in *FE1*, *FE2*, and the *VRFE*, spatial visualization ability was used as a primary factor in the experimental designs. Participants were screened into either high or low spatial visualization ability groups prior to the experiments. By screening participants based on their spatial visualization ability, which has a well-established connection to task performance in computing applications (Campbell, 2011; Egan & Gomez, 1985; Kozhevnikov, Hayes, & Kozhevnikov, 2013; Vicente, Hayes, & Williges, 1987; Zhang & Salvendy, 2001), we expect to gain insights into behaviors, strategies, and workflows that can be taken advantage of (or mitigated) to improve mobile computing solutions tailored for location-based tasks such as address verification.

1.3.3 Exploring Other Factors that Might Influence Task Outcomes

Field Experiment 1 and *Field Experiment 2* are unique because participants used paper maps to complete the address verification task rather than the mobile device and digital map combination used in the preceding work and the *VR/Field Experiment*. The motivation for using paper maps was to free participants from the constraints of pre-existing software, which could influence participants' task approach and behavior. We intend to compare map usage outcomes across the studies to identify whether the map type has a noticeable effect on task outcomes.

Printed and written materials were used by participants for *FE1* and *FE2*. This may seem counterintuitive to research centered on a mobile computing task, but there is a rationale to this aspect of the study designs. We sought to observe the task as it has been performed in the past, running up to the time of the study, which is when the Census Bureau began incorporating handheld computers for such tasks. This would enable us to gain an understanding of the task as it has been performed historically and prior to the organization's mobile computing transition. The needs and approaches of users could be identified in a free-form fashion.

Field Experiment 1 and *Field Experiment 2* also differ in their task environments. The *FE1* study area was situated in a neighborhood that consisted of a complex network of streets, which ran and intersected at odd, oblique angles. Half of the *FE1* study area was retained for *FE2*, however, *FE2* incorporated a new portion of the neighborhood whereby the streets were arranged in a grid-like, orthogonal pattern that was more typical of neighborhood block layouts. The address verification scenarios of *FE2* were divided between the two neighborhood halves. These differences in the study locations between *FE1* and *FE2* allowed us to examine whether or not the street layouts had a noticeable impact on study outcomes.

1.4 Contributions

1. An approach is presented to assess and validate a VR lab designed with the purpose of evaluating location-based tasks. This method involved the comparison of a field study to its VR lab equivalent. In the VR lab, the field study procedure was replicated within a simulated field environment. Mixed-methods were used to compare and contrast the outcomes of the two evaluations. Using this technique, connections were made between quantitative data (i.e., task performance metrics, map usage statistics, counts of participants' behaviors and actions) and qualitative data obtained from researchers' observations of participants who follow a think aloud protocol; participants also completed exit questionnaires for additional insights. This approach was used in the field and in the VR lab to effectively evaluate location-based tasks that involve map use and navigation.
2. A VR lab was built that can be used to evaluate a location-based task in a controlled setting, while preserving some of the ecological validity afforded to field studies. The technique used involved the creation of an immersive virtual

environment (IVE) that is an accurate, geo-located replica of the field. The IVE is used in conjunction with a CAVE to simulate task environments during evaluations. This implementation allows for: Hands-free, simulated locomotion in the IVE; use of external devices and tools (e.g., a paper map or a mobile device and digital map); automated data collection; and the ability for observers and participants to interact in a shared, virtual environment.

1.5 Organization

The rest of this dissertation is organized as follows: In *Chapter 2*, literature is presented on the methods and approaches used in this research. *Chapter 3* is divided into two sections that cover *Field Experiment 1* and *Field Experiment 2*. In *Chapter 4*, the *VR/ Field Experiment* and its results are presented and discussed. Lastly, *Chapter 5* summarizes the entire body of work, key findings are discussed, and this dissertation is concluded.

CHAPTER 2 - REVIEW OF LITERATURE

2.1 Introduction

In this dissertation, an exploratory approach is taken to study a mobile computing task that is referred to as “address verification,” which is a location-based task that involves pedestrian navigation and the use of a map for the purpose of surveying homes. Experiments are conducted on the task using a combination of lab and field evaluation techniques. We examine relationships between task outcomes and factors such as: Users’ spatial visualization ability; differences in the task environment; differences in map type (i.e., paper or digital), map preferences and strategies; as well as task-related behaviors such as route planning and navigation. This literature review will focus on the factors that have a bearing on task outcomes as well as the research methods that are utilized to examine the task.

2.2 Evolving Evaluation Methods in Mobile HCI

Until the late 1970s, HCI practices were mostly limited to researchers and IT professionals (Carroll, 2009; Myers, 1998). However, the subsequent emergence of personal computing meant that computers would become tools used by the masses. Caroll (2009) notes that personal computing “vividly highlighted the deficiencies of computers with respect to usability.” As computers extended their reach into people’s everyday lives, a more human-centered approach to HCI became necessary. A focus on user-centered design (UCD) and usability during evaluations helped to reign in the problems that were associated with early computer designs.

Whereas the personal computing revolution empowered users to access information “anytime,” the mobile computing revolution empowered computer users to access information “anywhere.” The evaluation techniques that were once reliable in fixed settings were oftentimes inappropriate for mobile computing scenarios (Coursaris & Kim, 2011; Johnson, 1998; Kjeldskov & Graham, 2003; Poupyrev, Maruyama, & Rekimoto, 2002). Grudin (2007) likens HCI's evolution to "a moving target," a cliché that captures the nature of mobile computing evaluations as well as the challenges therein. Kjeldskov and Graham (2003) surveyed mobile HCI research papers from 2000 to 2002 and called out a lack of focus on mobile evaluation methods. They recognized that a large proportion of mobile computing evaluations were still occurring in the lab, despite acknowledgement from the HCI community that techniques commonly applied to the evaluation of computer use in static settings would need to evolve and more emphasis should be placed on natural setting research (Johnson, 1998).

2.2.1 The Lab vs. The Field

For researchers and practitioners who are intent on evaluating mobile computing tasks, there was (and still is) a tension between the choice of the lab and the field study due to their tradeoffs. The lab has a high degree of experimental control, it places lesser demands on practitioners' time and resources, data collection is enhanced in the lab (dedicated facilities can even be built) and lab studies are generally easier to replicate (Delikostidis, Fritze, Fechner, & Kray, 2015; Kjeldskov & Graham, 2003; Zhang & Adipat, 2005). However, the lab suffers from its deficiency in recreating the context of use for a given task, which reduces the ecological validity of studies and calls into question the real-world legitimacy of findings obtained from such evaluations. On the other hand, the field study exposes context of use factors that are important to task outcomes (e.g., quirks of the real task environment, mobility, divided attention)—all of which facilitate authentic behaviors in participants and hence increase ecological validity (Kjeldskov & Graham, 2003; Nielsen, 1998; Nielsen, Overgaard, Pedersen, Stage, & Stenild, 2006; Zhang & Adipat, 2005). The drawbacks of field studies include: Greater time and resource commitments from practitioners, complicated data collection, reduced experimental control, and study replication (i.e., the reproducibility of results) poses a challenge. Delikostidis et al. (2015) summarize the tradeoffs between these methods by stating, "Key disadvantages of either method are the inverse of the other method's key benefits…" (p. 258).

Given the legacy of HCI evaluations, which were once broadly focused on personal computing tasks that occurred in static settings, and the familiarity of researchers and practitioners with the methods of that era, it is reasonable that surveys of mobile HCI practices have shown that the lab method is predominant (Kjeldskov & Graham, 2003; Kjeldskov & Paay, 2012). The preponderance of lab studies to field studies in mobile HCI contradicts the assumption that, due to the nature of mobile device use, evaluations "*should* be done in the field" (Kjeldskov, Skov, Als, & Høegh, 2004, p. 62). Kjeldskov et al. (2004) tested this assumption by comparing outcomes between lab and field evaluations of the same task. They found that outcomes of their lab evaluation of a mobile, medical data collection task (in a simulated hospital setting) corresponded to field study outcomes. They concluded that a field study was not "worth the hassle." In response, Nielsen et al. (2006) compared the outcomes of lab and field evaluations centered on a different task—one that involved skilled fieldworkers who used a mobile computer to register their use of equipment, materials, mileage and time. Contradicting the findings of Kjeldskov et al., Nielsen et al. concluded that field studies were indeed "worth the hassle." Nielsen et al. found significantly more usability problems during their field evaluation and also recognized problems related to cognitive load and interaction style that were not evident in the lab.

2.2.2 Building on Lab and Field Evaluation Techniques

Though a number of mobile HCI papers have argued for either lab or field evaluations, this dissertation reflects the views of those researchers who have taken the position that both methods have their place (Kjeldskov & Paay, 2012; Kjeldskov & Skov, 2014; Nielsen, 1998). In a longitudinal review of mobile HCI research methods, despite finding that only 5% of multi-method evaluations explicitly combined lab and field studies (as is done in this dissertation), Kjeldskov and Paay (2012) maintain that “lab and field evaluations both have justification, albeit for studying different things, and therefore should be combined and integrated” (p. 75). Furthermore, Nielsen (1998) highlights that learnings from one method can lead to substantial improvements in the other. Both points are embraced in the present work, where field experiments conducted by the author’s research group (Batinov et al., 2016; Whitney et al., 2011; Whitney et al., 2010) were built on earlier lab evaluations (Rusch et al., 2012; Rusch, 2008) and vice versa (Batinov et al., 2013). Kjeldskov and Skov (2014) highlight improvements in both lab and field evaluation techniques that enable both to better converge on a shared goal: To balance ecological validity with experimental control. “From the lab study side our community has arrived at new ways of simulating context, and from the field study side it has arrived at new ways of experimentation in situ” (Kjeldskov & Skov, 2014, p. 7).

Because field studies are generally more difficult to conduct than lab studies, a proposed alternative is to conduct lab studies on mobile computing that incorporate aspects of a task’s context of use, including the mobility of the user, in order to achieve more ecologically valid outcomes while maintaining the benefits of a controlled lab setting (Brade et al., 2017; Delikostidis et al., 2015; Kjeldskov & Stage, 2004; Nielsen, 1998). This dissertation builds on an earlier lab experiment that incorporated photos of a real-world task environment (i.e., a neighborhood) to better contextualize a location-based task for the sake of comparing two digital map interfaces (Rusch et al., 2012; Rusch, 2008). The Rusch et al. (2012) experiment simulated task scenarios in a stationary setting while participants interacted with a digital map in a seated position. However, fieldworkers in the real world would complete the task in a standing position while holding the mobile device, they would be able to walk and freely explore the neighborhood, and they would be responsible for route planning and navigation. Though Rusch et al. simulated some elements of the task environment, it became clear that much of the task's context of use could not be reflected using this method. More "immersive" alternatives to the lab-based approach of Rusch et al. were desired to further evaluate "address verification," which is the location-based task that is central to the present work.

2.3 An Opportunity for the Use of a Virtual Environment

Virtual environments (VEs) are artificial, computer-driven environments that can simulate one's presence in both real and imaginary worlds. VEs play an important role in HCI research because they drive studies that pertain to evaluation (Bhimani, 2017; Brade et al., 2017; Brade, Lorenz, Klimant, Pürzel, & Putz, 2016; Busch, Lorenz, Tscheligi, Hochleitner, & Schulz, 2014; Delikostidis, Fechner, Fritze, AbdelMouty, & Kray, 2013; Delikostidis et al., 2015) and to cognitive inquiry (Durlach et al., 2000; Nathanael, Vosniakos, & Mosialos, 2010; van der Ham, Faber, Venselaar, van Kreveld, & Löffler, 2015; Yuan, Song, & Zhang, 2014).

2.3.1 Why use virtual environments to study a location-based task?

Virtual environments offer a certain level of immersion that can induce cognitive states in users that would otherwise only be achieved in real-world situations. Lombard and Ditton (1997) describe this phenomenon as presence. A great deal of VE research and development is concerned with presence as it applies to visual experiences; thus, VEs are ideal for research involving spatial behavior and cognition. Furthermore, researchers have historically investigated human spatial abilities either by using psychometric tests or by exploring subject behavior in large natural environments, yet there is little empirical evidence that connects the psychometric literature to that of environmental cognition (Waller, 2005). Virtual environments may very well provide a solution to this dilemma. Durlach et al. (2000) outline four kinds of research that utilize VEs to study spatial behavior:

- "VEs are being used as a research tool to help advance fundamental understanding of spatial behavior."
- "VEs are being used to help assess spatial abilities and skills."
- "Because users often find VEs confusing and difficult to navigate (often getting lost in them), efforts are being directed towards the development and evaluation of methods for improving spatial behavior in VEs."
- "Research is being conducted on the use of VEs to improve spatial behavior in the real world" (p. 594).

Waller (2005) suggests that the nature of VEs make them a more logical choice than paper-and-pencil tests for exploring spatial behavior in large-scale environments; this implies that they are also well suited for use in the study of location-based tasks.

Li and Longley (2006) used a CAVE-based immersive virtual environment to evaluate an application that provided location-based services; they determined that “richer and more diverse information” could be gleaned from such test environments to assess users’ wayfinding behaviors, however, no comparison was made to the field in this example. On the other hand, Delikostidis et al. (2015) used an “Immersive Video Environment” as a hybrid solution to evaluate a location-based task and compared their results to those of a field evaluation; this approach enabled them to identify nearly the same number of major usability problems as was seen in the field.

The work presented in *Chapter 4* of this dissertation is most similar to the work of Brade et al. (2017), who compare a CAVE-based immersive virtual environment to that of a field environment—an experimental design similar to that of Delikostidis et al. (2015). Brade et al. compared evaluation outcomes of a mobile computing task in their “virtual field environment” to those of a real field environment with respect to measures of presence, usability, and user experience. Using this approach, Brade et al. were able to achieve a high degree of immersion and interactivity in their virtual environment to better capture the ecological validity afforded to the field evaluation. The present work is also similar to that of Brade et al. in that their task involved pedestrian navigation in a large-scale environment relative to the virtual environments that are frequently cited in such work. Brade et al. concluded that studies in a high-fidelity “virtual field environment” can be concrete alternatives to real field environments for the evaluation of mobile computing products.

2.3.2 Spatial Behavior in Virtual Environments

Studies in computer and software use have indicated that spatial ability explains a large amount of variance in user performance. Such effects have also been observed in studies concerning virtual environments (Durlach et al., 2000). For example, Koh and von Wiegand (1999) conducted a study that required subjects to learn the spatial configuration of a building floor either by studying the actual location, a physical model, a non-immersive virtual environment (VE) or an immersive virtual environment (IVE). They discovered that differences in spatial orientation—as determined by a Guilford and Zimmerman (1947) orientation test—explained more variance in participant performance in the virtual environments than did their treatment conditions (i.e., learning via: route, survey, or composite views). Arnold and Farrell (2003) suggest that such orientation differences stem from two interacting sources: (1) the dissimilarity of virtual and real-world environments and (2) the complexity of wayfinding.

Arnold and Farrell (2003) make an important distinction between VE and real-world navigation by quoting James J. Gibson (1979), a psychologist renowned for his work in visual perception: "one sees the environment not with the eyes but with the eyes-in-the-head-on-the-body-resting-on-the-ground" (p. 205). This quote describes proprioception: one's sense of position and movement with respect to one's own body. Arnold and Farrell (2003) stress that the absence of natural motion and other proprioceptive cues can be problematic in VE studies; they state, "in VEs, participants are effectively disembodied" (p. 659). As a result of this, VE studies have demonstrated that people's perception of distance (Witmer & Kline, 1998) and location (Richardson, Montello, & Hegarty, 1999; Riecke & Wiener, 2006) are subject to error. In fact, missing vestibular and proprioceptive cues have been shown to impair spatial updating during imagined, real, and virtual locomotion (Klatzky & Loomis, 1998).

Advances in the fidelity of displays, graphics rendering, and capabilities that enable navigation in VEs have increased immersion, narrowing the gap between VE and real-world experiences. Riecke, Heyde, and Bülthoff (2004); (2005) for example, observed that visual cues alone were sufficient to trigger spatial updating in a VE driven by a head-mounted display (HMD). Fully-immersive VEs—such as the CAVE used in *Chapter 4* of this dissertation—are among the most effective ways to mimic real world experiences (Arns & Cruz-Neira, 2004; Bowman, Davis, Hodges, & Badre, 1999; Brade et al., 2017; Cruz-Neira, Sandin, & DeFanti, 1993; Li & Longley, 2006; Waller, Hunt, & Knapp, 1998).

Wayfinding, the cognitive aspect of navigation, is closely tied to motion. Given that many cues of motion are missing in VEs, it is useful to understand the degree to which the visual stimulus can sufficiently support tasks that require spatial cognition. Darken and Sibert (1996) used a non-immersive VE to explore connections between map usage and navigation. They demonstrate that principles of real-world wayfinding and spatial knowledge acquisition can be implemented in VEs to support skilled wayfinding behavior. Darken and Sibert suggest that VEs should take advantage of environmental design methodology. For example, the inclusion of urban design elements such as paths, edges, districts, nodes, and landmarks (Lynch, 1960) allow VE users to make spatial connections within their environment. The VR lab that is discussed in this dissertation takes advantage of such cues because the underlying VE is a 3D model (replica) of a real-world neighborhood and faithfully reflects many of its important urban design features.

Map usage techniques for wayfinding should also be considered. Darken and Cevik (1999) define four types of navigation tasks:

- “Targeted search: A searching task in which the target in question is shown on the virtual map.”
- “Primed search: A searching task in which the location of the target is known, but the target does not appear on the virtual map.”
- “Naïve search: A search task in which there is no prior knowledge of the whereabouts of the target in question and the target is not shown on the map.”
- “Exploration: A wayfinding task in which there is no specific target.” (p. 135)

They conclude that for egocentric tasks such as targeted searches, a forward-up map orientation is preferred over north-up. However, a north-up orientation is superior for geocentric tasks such as primed or naïve searches (map planning tasks are similar in nature). Furthermore, Darken and Cevik (1999) report, “Under almost every possible condition, individuals with high spatial abilities will be able to use either type of map better than individuals with low spatial abilities” (p. 139). These findings are congruent with those of Aretz & Wickens (1992).

While it is important that we base VE research design and methodology on specific use cases, it is clear that spatial cognition theory should always be considered. As virtual reality systems advance and techniques are developed to provide experiences that better reflect those of the real world, many of the obstacles that inhibit spatial cognition in VEs will be mitigated or resolved.

2.4 Considerations for the Fieldworker

“The least flexible component of any system is the user.” -Lowell Jay Arthur

Consider the U. S. Decennial Census: It is said to involve the largest mobilization of a workforce short of war (“The census found it could count on NCS,” 2001). Because of this, seasonal Census fieldworkers are unique from the specialists who tend to support the data collection needs of other organizations. Some of the characteristics that differentiate Census fieldworkers include:

- Their demographics vary widely (e.g., culture, gender, age, experience, education).
- They often are recruited, hired, and complete their job within a narrow time frame.
- They likely have limited (if any) experience with Census operations.
- They receive minimal training.

These characteristics make it difficult to generalize an effective design for interfaces that support seasonal Census operations. In fact, there is no “typical” Census fieldworker. Yet, in order to increase the usability of the devices that they might use as tools, UCD emphasizes a focus on relevant user characteristics that can lead to better designs. With this end goal in mind, a major thread of this research involves the classification and evaluation of users based on cognitive ability.

2.4.1 Individual Differences

Many of the works seminal to human-computer interaction (HCI) emphasize a user-centered approach for developing effective systems and software. Designers and engineers are urged to conduct user analysis early so that they may better understand how differences in human ability and predisposition impact the usability of a product (Dillon & Watson, 1996). Murray and Kluckhohn (1948) eloquently capture the consequence of such differences: “Every man is in certain respects (a) like all other men, (b) like some other men, (c) like no other man" (p. 35). By understanding how physiological, psychological, social, and cultural characteristics define and differentiate users, we are better able to create usable artifacts.

People differ considerably in how well they perform a variety of tasks. Such differences span a wide range of abilities. For example, gender, experience, age, navigational strategy, and working memory ability have been found to account for significant variance in navigational performance (Baldwin & Reagan, 2009). With such a vast amount of user characteristics, each of which could potentially impact a given task's outcome, scholars have responded by developing frameworks for accommodating individual differences in users (Egan & Gomez, 1985; Messick, 1976). Egan and Gomez (1985) propose a three-stage approach: (1) relevant user characteristics are isolated; (2) an understanding of the quality and extent of these differences is established; and (3) the task or system is modified accordingly. In a similar fashion, Benyon, Crerar, and Wilkinson (2001) offer two questions that should be answered with respect to individual differences and HCI applications: (1) How do people differ?; and (2) Which differences are pertinent to the HCI application?

Psychological Differences

The physiological and sociocultural categories outlined by Benyon et al. (2001) certainly play an important role in HCI; however, psychological differences bear the greatest influence on the present work. Benyon et al. give prominence to psychological differences in this excerpt:

> Computers are general-purpose machines and using computer systems is, to a large extent, a cognitive activity. HCI involves information processing and is concerned with the acquisition, manipulation, and expression of abstract symbols that signify something else. Most other systems allow for some form of physical interaction, allowing the user to look inside and see how it works. The user of such physical artifacts can employ a broad range of strategies, which are unavailable to the computer user, who must judge the system purely by its external displays. This is the reason why individual differences in cognitive abilities, preferences, and learning styles are so important in HCI. (pp. 21-22)

Benyon et al. assert that despite there being a large number of psychological differences, most fall within three sub-categories: personality, cognitive style, and intelligence.

An understanding of cognitive style and intelligence offer promise with respect to field data collection. Cognitive style can be expressed as the various modes of cognition that manifest themselves through stable attitudes, preferences, or habitual strategies—these determine an individual's approach to perceiving, remembering, thinking, and problem solving (Messick, 1976). Dillon and Watson (1996) consider cognitive style to be the information-processing equivalent of personality. "[Cognitive styles] are used when acquiring new knowledge through perception, its subsequent storage in long-term memory, and finally its application in thought" (Benyon et al., 2001, p. 26). Dimensions such as "field dependence-independence", "holism-serialism", and "reflective-impulsive" exemplify the bipolar nature of cognitive style. Depending on the task, one's affinity to either pole of a particular cognitive style dimension could potentially benefit or detriment performance. Dillon and Watson (1996) suggest that, in addition to personality constructs, cognitive style is hard to distinguish and therefore offers little for the sake of predicting user performance. However, participants of the studies featured in the present work are expected to demonstrate unique approaches to wayfinding, navigation, and the address verification task as a whole; these represent task-specific cognitive styles, which may be tied to abilities that do in fact offer predictive power (e.g., spatial ability). Some examples of such cognitive styles include map usage style (track-up vs. north-up) and heading preference (egocentric vs. allocentric).

Cognitive style is not to be confused with intelligence; both describe cognitive processing, but they differ in measure. Cognitive style is a measure of the form that cognition takes, whereas intelligence measures the capacity and speed of various types of cognitive processing (Benyon et al., 2001). Where cognitive style dimensions are often bipolar, intelligence dimensions are unipolar and value directional; for example, high spatial ability is superior to low spatial ability (Messick, 1976). This is why measures of intelligence are particularly useful for user and task analysis. Differences stemming from a subcomponent of intelligence—spatial ability—are focal to each of the studies discussed in this dissertation.

Spatial Ability

Spatial ability can be defined as “the ability to generate, retain, retrieve, and transform well-structured visual images” (Lohman, 1996, p. 99). Lohman (1996) describes spatial ability components as “pivotal constructs of all models of human abilities.” The isolation of such ability has roots in 20th Century studies of human reasoning and mechanical aptitude. These studies aimed to show that spatial ability is useful in predicting job performance and success in vocational/technical training programs (McGee, 1979). Because spatial ability is a ubiquitous reasoning skill, it has proven to be robust with regard to user classification. Spatial ability has been closely tied to general intelligence (Spearman & Wynn Jones, 1950), working memory capacity (Baddeley & Hitch, 1974), field-independent creativity Lohman (1996), higher-order thinking in STEM domains (Wai, Lubinski, & Benbow, 2009), and computer/software use (Benyon et al., 2001; Egan & Gomez, 1985; Norman, 1994; Pak, Rogers, & Fisk, 2006; Stanney & Salvendy, 1992).

The literature concerning address verification (Murphy & Nusser, 2003; Rusch et al., 2012; Rusch, 2008) suggests that computer and interface designs that account for the spatial ability of users may be promising. This is reasonable considering that in addition to the link with computer and software usage, spatial ability has been tied to individual differences in map usage (Aretz & Wickens, 1992; Darken & Cevik, 1999; Willis, Hölscher, Wilbertz, & Li, 2009) and navigation (Garden, Cornoldi, & Logie, 2002; Roger, Bonnardel, & Le Bigot, 2009; Thorndyke & Hayes-Roth, 1982)—all of which are key to the address verification task. It is also believed that spatial cognition and decision making are important factors for the address verification task, whereby fieldworkers must visually compare GIS-based materials and maps (which contain housing unit data points that may or may not be geospatially accurate) to the topography of their assignment area. For example, in two ethnographic studies, Murphy and Nusser (2003) showed that the use of digital navigation aids by Census Bureau employees led to "improved average performance as well as more consistent performance across addresses in an assignment" (p. 5). Furthermore, the "Virginia study" results indicate that spatial ability (as measured by psychometric tests) was positively correlated to the number of completed assignments. The work of Murphy and Nusser (2003) suggests that a user-centered design approach, which accounts for individual differences in spatial ability, may help to improve the designs of field data collection systems. The Rusch et al. (2012) findings further support this conclusion.

Ekstrom et al. (1976) describe spatial visualization as the ability to mentally restructure or rotate an image or configuration of spatial patterns. While numerous other definitions of spatial visualization exist, McGee (1979) asserts:

> All involve the ability to mentally manipulate, rotate, twist, or invert a pictorially presented stimulus object. The underlying ability seems to involve a process of recognition, retention, and recall of a configuration in which there is movement among the internal parts of the configuration [Thurstone, 1938] or the recognition, retention, and recall of an object manipulated in three-dimensional space [French, 1951] or which involves the folding or unfolding of flat patterns [Ekstrom et al., 1976]. (p. 893)

Norman (1994) cites spatial visualization as "the primary cognitive factor driving differences in performance using computers" (p. 195). Spatial visualization has also been identified as the ability most critical to people's interpretation and integration of map information (Aretz

& Wickens, 1992). This subcomponent of spatial ability appears to be critical to address verification. In work closely related to that of this dissertation, Batinov (2017) demonstrates that participants' spatial visualization ability can be identified based on their digital map usage patterns as they carry out the address verification task. Batinov (2017) examined map usage patterns of participants who completed the task in three separate environments: A non-immersive lab, an immersive lab, and the field (i.e., the real world). His work suggests that once such differences are identified, software designs can then adapt to users in order to better accommodate them. This premise of accommodating differences in participants' spatial visualization ability through adaptive interface designs is demonstrated in the follow-up work of Patanasakpinyo (2017) and Patanasakpinyo et al. (2018).

Classifying Fieldworkers Using Psychometric Tests

Research in psychology has proven that we can quantify mental abilities with the aid of cognitive tests. Spearman (1904) developed statistically correlated tests that ushered in a general form of intelligence where a single factor (*g*)—general intelligence—is common to all intellectual activities, while a number of sub-factors (*s*) are further associated with distinct tasks. While Spearman focused on similarities in cognitive scores, Thurstone (1938) focused on statistical differences; this approach allowed for the identification of seven primary mental abilities: Perceptual speed, memory, verbal meaning, spatial ability, numerical ability, inductive reasoning, and verbal fluency. In 1963, the Educational Testing Service (ETS) introduced a kit of factor-referenced tests (French, Ekstrom, & Price, 1963). The experiments of this dissertation utilize several tests from this kit to probe and differentiate individuals based on cognitive aptitude; these tests include: Visualization (VZ), visual memory (MV) and perceptual speed (P-2). A separate measure of perspective-taking ability (Kozhevnikov, Motes, Rasch, & Blajenkova, 2006) is also used. A re-analysis of ten years of multivariate intelligence studies (Carroll, 1993) further validates the efficacy of such an approach to user classification.

The aforementioned strides in intelligence classification are important because they allow us to identify, validate, and establish tests for mental abilities that have been shown to predictably influence user performance on various tasks. These mental abilities evidently influence the specific field data collection task (address verification) of concern in this dissertation (Batinov et al., 2015; Batinov, et al., 2011; Batinov et al., 2013; Batinov, 2017; Batinov et al., 2016; Meng et al., 2015; Murphy & Nusser, 2003; Thitivatr et al., 2018; Patanasakpinyo, 2017; Rusch, 2008; Rusch et al., 2012; Whitney et al., 2011; Whitney et al., 2010).

Spatial Ability and Software

Benyon et al. (2001) state, “Of all the components of intelligence, spatial ability is the one that has been most frequently studied in connection with software use” (p. 25). Spatial ability has been positively correlated to users’ navigational efficiency with file systems (Vicente et al., 1987) and hypertext (Chen & Rada, 1996). Database and World Wide Web search task performance (Benyon, 1993; Downing, Moore, & Brown, 2005; Zhang & Salvendy, 2001) also show correlations to spatial aptitude. More recent work has shown that spatial ability is useful in characterizing users during software evaluations (Batinov et al., 2013; Batinov, 2017; Batinov et al., 2016; Campbell, 2011; Whitney et al., 2011) as well as in accommodating individual differences in users through the software interface (Chellappan, 2012; Chellappan & Miller, 2014; Patanasakpinyo et al., 2018; Patanasakpinyo, 2017). Because tasks such as these involve the interpretation of computer-based information, their dependency on spatial visualization can primarily be attributed to users’ interaction with interface elements. Applications that require individuals to interpret and act on environmental cues will place additional demands on spatial visualization ability and will likely add some level of spatial orientation to the mix of relevant cognitive processes. Furthermore, such applications often use maps to represent spatial information.

Digital maps introduce new challenges to the way that we acquire spatial knowledge. They are abstract visualizations that lack the tactility afforded to paper maps. Furthermore, maps depicted by mobile devices are constrained by screen real estate, thereby forcing users to rely on operations such as panning and zooming to achieve desired views (Büring, Gerken, & Reiterer, 2006). The side effects are apparent: Studies that contrast digital and paper map usage indicate that digital maps contribute to poorer route and survey-based knowledge (Ishikawa, Fujiwara, Imai, & Okabe, 2008; Münzer, Zimmer, Schwalm, Baus, & Aslan, 2006; Willis et al., 2009). This is likely a consequence of the fragmented availability and focus of relevant spatial cues in situations where map information is used to support navigation. Willis et al. (2009) note, “Mobile spatial applications tend to work on a model where the individual is not encouraged to plan a spatial task, i.e., to conceptualize a schematic form of information and then act on it, but instead to proceed in a task where information is delivered incrementally…” (p. 109).

In an NSF report on GIS services (Hecht & Kucera, 2000), it is suggested that researchers (1) “Extend the promise of cognitive research to make geographic information technologies more accessible to inexperienced and disadvantaged users” and (2) “Support research on multiple representations/interfaces focused on task-specific (procedural) workflow classes” (p. 2). The work presented in this dissertation is in line with these recommendations; it is grounded in spatial cognition theory and continues work that has focused on the use of computer-assisted survey instruments to complete location-based tasks

(Batinov et al., 2011; Nusser, Goodchild, Clarke, & Miller, 2004; Nusser, Miller, Clarke, & Goodchild, 2002; Nusser & Fox, 2002; Rusch et al., 2012; Rusch, 2008; Whitney et al., 2011). Egan and Gomez (1985) propose a three-stage approach to handling individual differences: They should be (1) isolated, (2) assayed, and (3) accommodated. The application of this methodology has prompted the enhancement and redesign of numerous kinds of interfaces, resulting in improved software user performance on a variety of tasks (Benyon, 1993; Downing et al., 2005; Patanasakpinyo et al., 2018; Patanasakpinyo, 2017; Stanney & Salvendy, 1992; Vicente et al., 1987). There is far greater potential in accommodating individual differences in spatial ability for software that supports field surveying because such tasks are spatial in nature and typically require the use of digital maps.

2.5 Cognitive Considerations for The Task

Failing to adequately address user-centered design (UCD) at the onset of a project can be catastrophic, even when proper UCD practices are subsequently adopted. A deep understanding of users and their tasks is key to requirements-gathering. In order to create more usable artifacts, "design teams are urged to perform user and task analyses at the earliest stages of product development" (Dillon & Watson, 1996, p. 1). The central task of the three studies outlined in this research is known as "address verification". During address verification, the fieldworker carries at least two materials: (1) a *list* that denotes the residences to be verified and (2) a *map* that displays the geospatial location of all "known" residences in the area. The fieldworker uses these materials for:

- *Route planning and navigation* (How do I find my assigned addresses?).
- *Decision-making* (Does what I see on the ground differ from my map? How?).

- *Updating geospatial information* (What changes need to be made to update or correct errors that are reflected by my materials?).

This section will cover several aspects of cognition that are relevant to the address verification task as well as approaches that have been used in similar work to take these elements into consideration.

2.5.1 The Acquisition of Spatial Knowledge

During address verification, decisions are made based on the environmental configuration of housing units and any landmarks that are used to anchor them. It becomes necessary for the fieldworker to acquire an appropriate level of spatial knowledge about their assigned area. In a comprehensive review of allocentric and egocentric spatial representations, (Klatzky (1998) designates reference frames as a means to represent entity locations in space. Individuals use egocentric reference frames (ERF) to encode things based on their own position (e.g., to my right), whereas allocentric reference frames, also termed geocentric or world reference frames (WRF), allow for encoding that is irrespective to self-position (e.g., to the east). Spatial knowledge is built around these reference frames in various scenarios—the two most common are (1) through direct exposure to the environment and (2) through the use of maps. The following section outlines the underlying processes behind these two methods.

2.5.2 Acquiring Spatial Knowledge via Environment

The research of Hart and Moore (1973) suggests that we establish spatial frames of reference based on three developmental stages: (1) egocentric orientation, (2) fixed frame of reference orientation, and (3) coordinated frame of reference orientation. During egocentric orientation, we orient objects in our environment based on our own position. During the fixed frame of reference stage, we orient ourselves based on the fixed locations of environmental objects. Finally, during the coordinated frame of reference stage, we acquire an allocentric orientation, i.e., we no longer rely on referencing based on our own position. Coordinated frames of reference allow us to perceive many possible routes to destinations based on a comprehensive knowledge of environmental objects that can be described abstractly (e.g., cardinal directions). In the experiments of this dissertation, it is likely that participants will not be able to operate based on coordinated frames of reference because they will have limited to no prior knowledge of the study area (based on screening procedures) and the study exercise is limited in its duration.

The work of Siegel and White (1975) proposes the Landmark-Route-Survey (LRS) model, which consists of three phases that are defined by: (1) landmark recognition, (2) route knowledge, and (3) survey knowledge. During landmark recognition, we declare salient objects in the environment but we do not yet integrate them into a cognitive map. As we develop route knowledge, we make connections between landmarks and path intersections, we cluster them, and then we link them via topological relationships. In order to finally establish survey knowledge, we develop coordinated frames of reference within and across these clustered landmarks and path intersections. This cognitive process is complementary to that proposed by Hart and Moore (1973).

Egocentric and allocentric perspectives, landmark recognition, and the use of reference frames enable individuals to create cognitive maps so that they may conceptualize their surroundings. Location-based tasks require the use of such cognitive maps, which is atypical to most software usage scenarios. Modern field surveyors must rely on spatial orientation to learn their environments, while simultaneously applying spatial visualization skills to align digital maps with newly acquired environmental knowledge. How then do we bridge route-based knowledge with that which is acquired from maps?

2.5.3 Acquiring Spatial Knowledge via Maps

If you were traveling with a group, would you want to hold the map or would you pass it to someone else? Your answer may very well depend on your spatial visualization and perspective taking abilities. Aretz and Wickens (1992) conducted two experiments to explore the cognitive processes that affect how people mentally rotate map displays. They determined that in order to achieve navigational awareness during egocentric tasks, individuals must perform mental rotations so that their map stays congruent with their forward view of the environment, i.e., the top of the map always reflects one's heading rather than north. Aretz and Wickens (1992) found that the time required to perform this mental rotation is roughly a linear function of the rotation angle of the map with respect to its forward-up (track-up) alignment. As mental rotation becomes increasingly difficult during a map-driven task, other strategies may be employed to resolve misalignments between the map and the environment. For example, researchers have observed frequent map rotations by individuals completing address verification tasks in spatially complex environments (Batinov et al., 2011; Batinov et al., 2016; Whitney et al., 2011; Whitney et al., 2010).

Darken and Peterson (2001) assert that if a map is used before navigation, for example during a geocentric task such as route planning, the mental transformations outlined by Aretz and Wickens (1992) are not necessary (note: we are likely to orient maps north-up in these situations). There is preliminary evidence that links spatial ability to one's tendency to perform this kind of planning using maps for such surveying tasks (Batinov et al., 2011; Batinov et al., 2016; Whitney et al., 2011; Whitney et al., 2010). One explanation for this phenomenon might be that spatial knowledge that is acquired during map-based planning must still be translated and transformed during subsequent navigation. One's proficiency and confidence in his/her ability to carry out geocentric planning, to store the plan in memory, and to later perform mental transformations could very well affect whether or not they decide to plan at the onset. This theory is supported by Aretz and Wickens (1992) conclusion that, "When an image must be held in memory, mental rotation becomes more difficult and may not be the desired strategy" (p. 325).

2.6 Conclusion

User and task analysis approaches from the literature have been applied to the experiments of this dissertation. From the user analysis angle, this research takes people's spatial abilities into account to examine differences in behavior, task performance and workflows. In the literature, the consideration of users' spatial ability with respect to computer and software use has frequently revealed connections to task outcomes. Furthermore, task analysis has shown that the skills that are necessary to complete the address verification task require some degree of spatial reasoning. In order to verify a map's accuracy, users are expected to interpret the map, plan routes, navigate on foot, and compare homes seen in a neighborhood to their depictions on the map. In identifying the components of the address verification task, we see an interconnection between users' spatial ability and task demands. The literature on wayfinding, navigation, the acquisition of spatial knowledge, and their connections to map use are germane to the present work.

The literature has shown that evaluation methods for mobile computing tasks such as this, whereby the context of use is important to task outcomes, should be carefully considered. In order to better balance ecological validity with experimental control, mobile

HCI practitioners continue to advance both field and lab evaluation methods. In this dissertation, both methods are used in accordance with the literature to examine and conduct experiments on the address verification task—in comparison, Kjeldskov and Paay (2012) reported that only 5% of the mobile HCI studies that were reviewed in their longitudinal survey explicitly combined field and lab-based methods.

The experiments covered in this dissertation are a synthesis of both lab and field methodologies, with each building on the learnings of the prior studies on the address verification task. The literature demonstrates that this approach enables researchers to advance mobile HCI methods while offering deeper insights into phenomena of interest (Nielsen, 1998). The field experiments discussed in *Chapter 3* combine the methodological strengths of both a prior lab experiment and field study. For example, the quantitative techniques used to examine task performance in the earliest lab experiment were combined with qualitative techniques used in the earliest field study to better understand participant behavior and workflows. Furthermore, in *Chapter 4*, a virtual reality lab is used to better simulate the environmental context of the field, hence increasing the resulting study's ecological validity—this is an improvement to the approach used in the earliest lab experiment to simulate the task environment. Outcomes of the VR lab evaluation are compared to a field evaluation that shares its study design—both of which build on the mixed-methods approach used in the field experiments that are discussed in *Chapter 3*.

CHAPTER 3 – TWO FIELD EXPERIMENTS ON A LOCATION-BASED TASK

3.1 Introduction

In this chapter, two experiments are discussed: *Field Experiment 1 (FE1)* and *Field Experiment 2 (FE2)*. These field experiments combined the methodological strengths of two prior studies—*Rusch's Lab Experiment* and the *Ethnographic Field Study*—which are discussed in the *Earlier Studies Section*. As was done in the *Ethnographic Field Study*, we conducted the field experiments in an authentic task environment—a real neighborhood. In addition to collecting qualitative data on participants as they completed the task, we recorded quantitative data as had been done in *Rusch's Lab Experiment*. Quantitative measures included task performance data, measures of participants' spatial visualization ability and several other cognitive abilities. *Table 3.1* provides an overview of the field experiments discussed in this chapter as well as the earlier, related studies.

Table 3.1 *Overview of Field Experiments Discussed in Chapter 3.*

	Research Method					**Data Method**		**Environmental Context**			**Spatial Viz.**		**Map Type**	
	Setting		**Type**											
Study or Experiment	Field	Lab	Ethnography	Experiment	Observ. Study	Qualitative	Quantitative	Natural setting	Photo-based	Immersive VE	Tested on it.	Screened on it.	Digital	Paper
Rusch's Lab Experiment[1]		X		X			X		X		X		X	
Ethnographic Field Study[2]	X		X			X		X					X	
Field Experiment 1 (oblique streets)	X			X	X	X	X	X			X	X		X
Field Experiment 2 (oblique + grid)	X			X	X	X	X	X			X	X		X
VR/Field Experiment (oblique streets)	X	X		X	X	X	X	X		X	X	X	X	

[1] A prior lab experiment used for comparison (Rusch et al., 2012; Rusch, 2008); related, but not an experiment covered in this dissertation.
[2] A prior field study used for comparison; related, but not a study covered in this dissertation.

Field Experiment 1 is a smaller study ($N = 10$) that was situated in a neighborhood consisting of a complex network of streets that ran and intersected at odd, oblique angles.

Field Experiment 2 is a larger study ($N = 26$). For *FE2*, we retained half of the study area used in *FE1* and added a new portion where the streets were arranged in a grid-like (orthogonal) pattern that is more typical of neighborhood street layouts. The six address verification scenarios for *FE2* were divided between the two neighborhood halves. The goal of modifying the study area for *FE2* was to examine if differences in street layout had a noticeable impact on participants' behaviors and task outcomes.

Relationships between the spatial ability of participants and their task outcomes were further explored in *FE1* and *FE2*. Participants were screened into two groups based on their cognitive test scores: A high spatial visualization ability group (HighVZ) and a low spatial visualization ability group (LowVZ). Additionally, rather than use the mobile devices and digital maps of the prior studies, *FE1* and *FE2* participants were given paper maps and written materials to complete the address verification scenarios. This decision was made so that the task could be better understood before bringing in technologies that might affect study outcomes via pre-existing software designs and workflows.

3.2 Organization

In this chapter, the overarching research questions for both experiments are presented. The methods and results of *FE1* and *FE2* are described and discussed in sequence, followed by a conclusion for the overall work.

3.3 Research Questions

1. What can be learned about the address verification task by using a mixed-methods approach?
2. How do people use printed maps and written materials to complete the address verification task?

3. How will differences in participants' spatial visualization ability affect task outcomes?
4. How will differences in the neighborhood layout affect task outcomes?

3.4 Field Experiment 1: Address Verification, a Paper Map, and Oblique Streets

3.4.1 Method

This section covers the method used in Field Experiment 1 (*IRB ID: 09-386*). Subsections address the topics of participant recruitment and screening, the address verification task, study materials, the study neighborhood, and the study procedure.

Participant Recruitment and Screening

University students responded to posters placed around campus and to ads posted on a student web portal. A total of 39 students applied to the study. Study applicants completed psychometric tests that assessed spatial visualization ability (*VZ-2*) and visual memory (*MV-2*) (Ekstrom et al., 1976), as well as perspective taking ability (Kozhevnikov et al., 2006). Applicants also completed a survey that collected demographic information and assessed their knowledge of surrounding neighborhoods—those who were familiar with the study neighborhood were not considered. Out of the thirty-nine applicants, ten participants were selected. Stratified sampling was used, where the five highest and the five lowest scoring participants on the spatial visualization tests were divided into a high spatial visualization group (HighVZ) and a low spatial visualization group (LowVZ). All applicants signed informed consent forms at the onset of the study and received separate gift cards for their participation in the screening and the field exercise appointments. *Field Experiment 1* documentation can be referenced in *Appendix A*.

The Address Verification Task

During the address verification task, participants walked through a neighborhood to find and verify six residential addresses with the aid of a paper map. Upon locating a residence, participants were to confirm whether its existence and location were accurately reflected on the map based on the label and placement of the residence's "map spot" as shown in *Figure 3.1*.

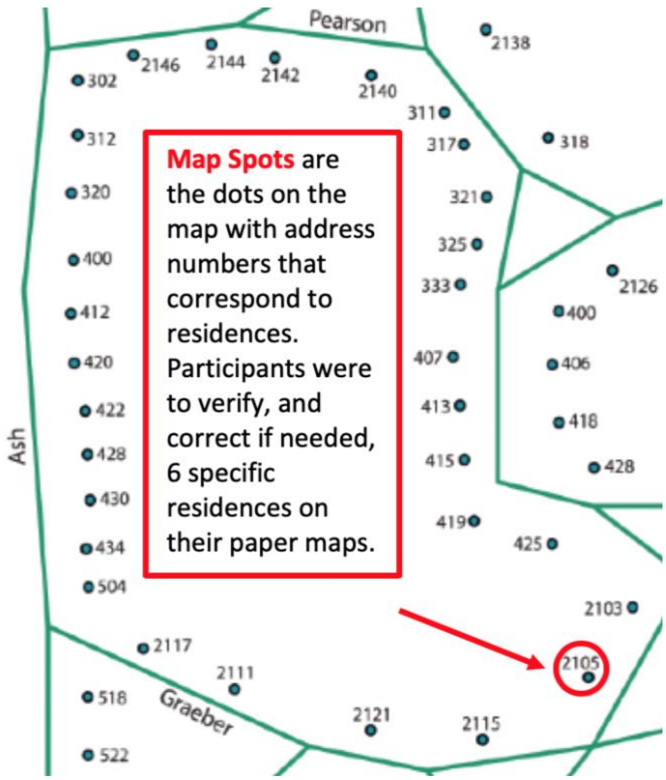


Figure 3.1 *A "map spot" is shown and a definition is provided.*

In scenarios where participants identified a discrepancy between an assigned residence and its corresponding map spot, they were instructed to correct the map. The possible map editing scenarios are shown in *Table 3.2*. Key task performance metrics for the study include participants' task time, distance traveled, and the frequency of their address verification errors.

Figure 3.2 *Diagram exemplifying a participant with materials*.

Table 3.2 *Possible Address Verification Scenarios and Their Descriptions*.

Scenario	**Description**
Address confirmed	Residence found in neighborhood is correctly shown on map; no change is necessary to the corresponding map spot.
Address added	Residence found in neighborhood is not on map; corresponding map spot should be added.
Address removed	Residence shown on map was not found in neighborhood; corresponding map spot should be removed.
Address moved	Residence shown on map was found in a different neighborhood location; existing map spot should be moved to the correct location.

Materials

The map given to participants (see *Figure 3.3*) was generated by combining Topologically Integrated Geographic Encoding and Referencing System (TIGER\Line) Shapefiles using ESRI's ArcGIS Desktop© software. The map consisted of two layers of information: (1) a street layer that showed the streets and their respective labels and (2) an address layer that depicted each residential address as a map spot accompanied by an address number. Each participant was provided a clipboard with a paper map of the neighborhood attached to the front side and attached to the back was a printed list of six residences that were to be verified. Participants were given a multi-colored pen so that they could update their map, add features (e.g., landmarks) and notes with color-coding if preferred. Participants were also outfitted with an audio recording device.

To introduce errors to the study map given to participants, residences (i.e., map spots) were deliberately added, removed, or incorrectly placed on the map to create six specific address verification scenarios (see *Table 3.3*; *Figure 3.3*). The modifications made to the study map are as follows:

- Two residences were removed; participants were to add them to the map to correct it.
- Two residences were added; participants were to remove them from the map.
- Two addresses were moved; participants were to move them to their correct locations on the map.

Table 3.3 *Address Verification Scenarios for Field Experiment 1*.

Scenario	Map Edit Required	Count
Residence is correctly reflected on map	No change to the map (spot) needed	0
Residence is new or missing (not on map)	Map spot added to map	2
Residence no longer exists, but is on map	Map spot removed from map	2
Residence is incorrectly placed on the map	Map spot moved to correct location	2

Observers used both a GPS-enabled smart phone and written materials to collect quantitative and qualitative data. The GPS provided estimates of participants' speed and distance. Latitude and longitude coordinates were logged. A JAVA program was written to convert each participant's coordinates into query strings per the Google® Static Maps API—this allowed us to generate rasterized maps of each participant's path using the Google® Maps service. A questionnaire was given to participants (see *Appendix A*) at the end of the exercise to gain insight into map use, planning behavior, and task difficulty.

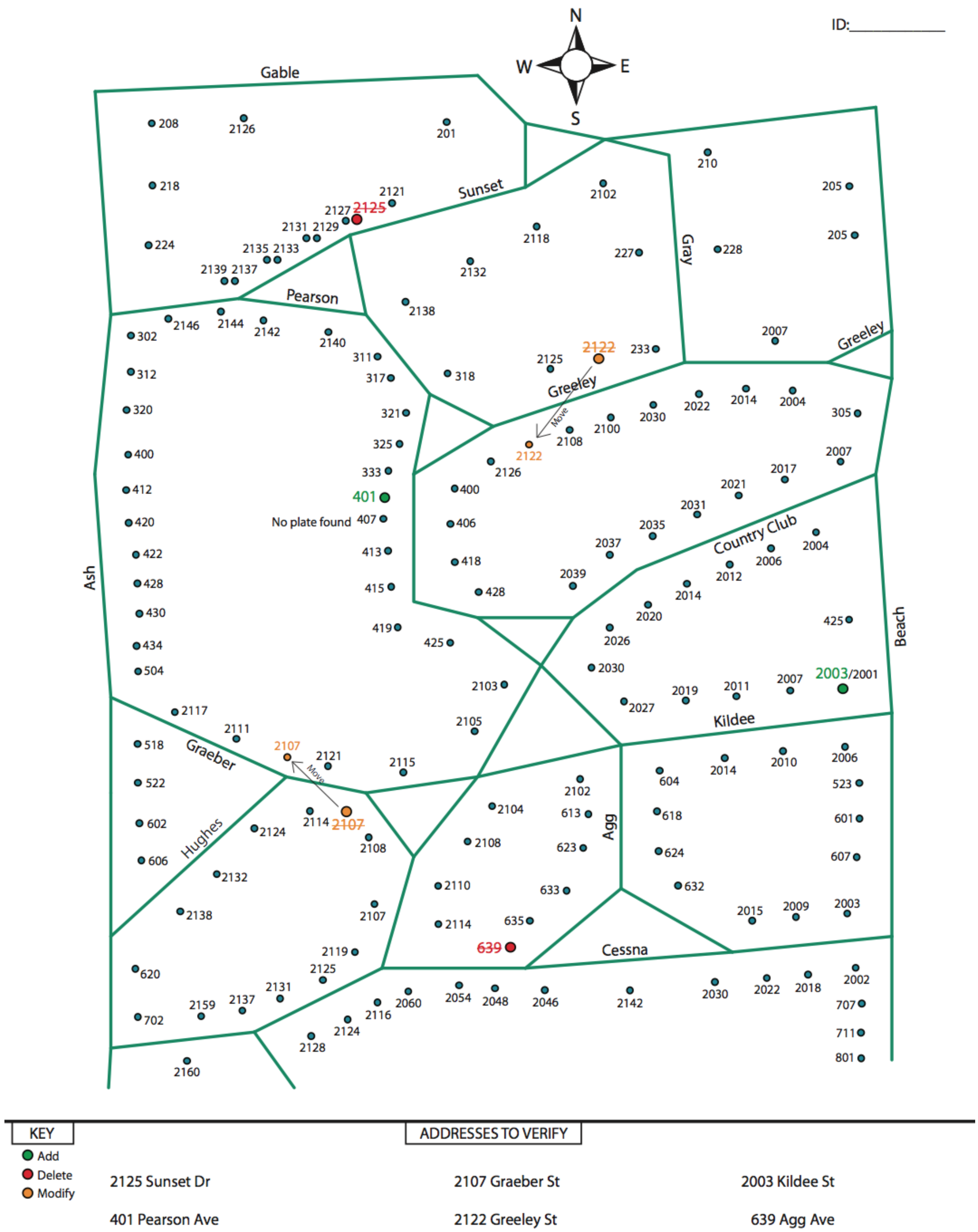


Figure 3.3 *Field Experiment 1 map with a key and indicators to show the corrections that participants should have made.*

The Study Neighborhood

The study neighborhood was selected after a review of residential areas using an online map service. Close attention was paid to the street layouts, the distribution and variety of residences, as well as proximity to the university. A residential area that would challenge participants in terms of orientation, navigation, and the address verification task was preferred. A neighborhood was selected near the university to ease logistics and the study area was confined to a particularly challenging cross section—approximately 1.3 square miles.

The residential area was predominantly made up of oblique, non-uniform streets that seldom ran parallel to the cardinal directions. Many of the residential streets formed three-way intersections known as Y-junctions, which can be more challenging to interpret than the more typical T-junction intersection. The streets often ran and intersected at odd angles that were not orthogonal like other residential areas, adding difficulty in terms of participants' understanding of their location, bearing and direction of travel. There was variety in the types of residences (e.g., homes, duplexes, multi-unit buildings) and landmarks (e.g., a treehouse, playground, and large stadium). The mix and spread of occluding landmarks such as trees and fences also resulted in spotty line of sight from various vantage points. Overall, these features presented participants with notable difficulty in acquiring spatial knowledge and making task-related decisions.

Procedure

A researcher accompanied each participant to the study neighborhood. The task was explained along with the think aloud protocol. Participants were equipped with a digital recorder and microphone. Each participant was given three training addresses to verify, which allowed them to ask questions and to get comfortable with the think aloud protocol. At the end of the training, the participant and observer reviewed the outcomes for the three training addresses to ensure that the procedure and objectives were clear. The participant and observer then returned to the starting location to begin the main exercise. The observer turned on the participant's voice recorder and started a GPS application to record the participant's route. The participant was given a clipboard that had a map of the study neighborhood attached to the front and on the back was a list of six addresses to verify. Each participant was shadowed by a researcher who carried coding sheets to record observations. The main exercise was completed after participants verified all six of the assigned residences. At the end of the exercise, participants answered an exit questionnaire on the task (see *Appendix A*).

3.4.2 Experimental Design

A between-subjects design was used. The task outcomes of two groups of participants were examined: Those who were screened into a high spatial visualization ability group (HighVZ) and those of a low spatial visualization ability group (LowVZ). This was achieved through stratified sampling, whereby participants whose VZ scores fell within predefined thresholds for high- and low-scorers were assigned to either group.

3.4.3 Hypothesis

We hypothesized that participants of the high spatial visualization group (HighVZ), given their spatial cognition advantages, would demonstrate superior overall task performance compared to their low spatial visualization counterparts (LowVZ). This hypothesis was tested per the three task performance variables, whereby:

a) Lower *task time* is better.

b) Shorter *distance traveled* is better.

c) Fewer *address verification errors* are better.

We also anticipated evidence of differences between the two participant groups with respect to task-related behaviors, strategies, and workflows.

3.4.4 Variables Used in Analysis

Four categories of data were captured during the screening and field exercise: (1) cognitive test scores, (2) task performance metrics, (3) coded data from participants' written materials, and (4) coded data from the observations that were recorded during the field study. This section covers the variables stemming from these categories.

Cognitive Test Scores

Participants were given cognitive tests prior to the field exercise and their scores were calculated. The tests are as follows:

1) *Spatial Visualization (VZ-2)* – (Ekstrom et al., 1976).
2) *Perspective Taking* – (Kozhevnikov et al., 2006).
3) *Visual Memory (MV-2)* – (Ekstrom et al., 1976).

Task Performance

The following task performance metrics were captured during the field exercise:

1) *Task time* – Task completion time for the address verification exercise.
2) *Distance traveled* – The distances that participants traveled to complete the address verification task as determined by a GPS device that was carried by an observing researcher.
3) *Address verification errors* – The total number of errors that participants made (or corrections that they failed to make) with respect to the address verification task and the materials that they were given and told to annotate.

Participants' Written Materials

Participants received two printed sheets: (1) A paper map of the neighborhood that depicted the residences within and (2) a list of residences that they were asked to verify. During the field study, participants took notes on these materials using a multi-colored pen. These materials were collected, analyzed and coded based on the types of notes that participants jotted down.

Recorded Observations

One observer was present while participants completed the field study. Coding sheets were used to record participants' behaviors and any relevant think aloud responses. These written materials were compiled and the observations were analyzed, categorized, and quantified.

3.4.5 Results (FE1)

Descriptive statistics for the cognitive test scores are shown in *Table 3.4*. Descriptive statistics for the task performance variables are shown in *Table 3.5*. In terms of task performance, the HighVZ group consistently outperformed the LowVZ group with shorter mean task time, shorter distance traveled, and fewer address verification errors.

Table 3.4 *Descriptive Statistics for Cognitive Test Scores by VZ Group (FE1).*

Performance Variable	HighVZ Group			LowVZ Group		
	n	Mean	SD	*n*	Mean	SD
Spatial Visualization	5	16.4	1.3	5	5.4	1.7
Visual Memory	5	17.7	2.0	5	15.4	4.6
Perspective Taking	5	24.9	1.1	5	10.4	4.8

Table 3.5 *Descriptive Statistics for Task Performance Variables by VZ Group (FE1).*

Performance Variable	**HighVZ Group**			**LowVZ Group**		
	n	**Mean**	**SD**	***n***	**Mean**	**SD**
Task time (min.)	5	32.2	7.0	5	47.8	16.4
Distance traveled (mi.)	5	1.1	0.1	5	1.6	0.4
Address verification errors	5	0.8	0.4	5	1.6	0.9

Correlations of Cognitive Test Scores and Task Performance

A Pearson product-moment correlation coefficient was computed to assess the relationship between *task time* and *spatial visualization* scores. There was a negative correlation between the two variables, $n = 10$, $r = -.65$, $p = .041$. Participants with higher spatial visualization scores tended to finish the task sooner than those who scored lower. Visual memory and perspective-taking ability were not significant as predictors of task time. A scatterplot summarizes the results (see *Figure 3.4*).

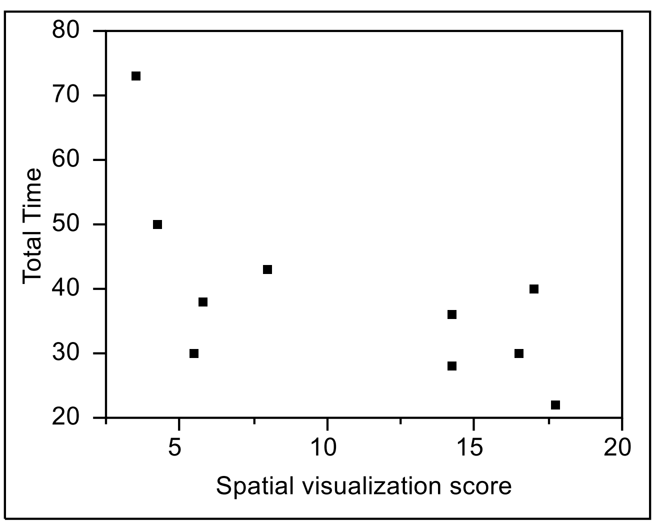


Figure 3.4 *Scatterplot of task time and spatial visualization score (VZ).*

Participants' paths were generated via a GPS device that was carried by an observer. *Distance traveled* was negatively correlated to *spatial visualization* scores (n = 10, r = -.70, p = .025). The scatterplot is shown in *Figure 3.5*. Distance traveled was also negatively correlated to *perspective taking* scores and yielded nearly identical statistics before rounding (n = 10, r = -.70, p = .025); a scatterplot summarizes the results (see *Figure 3.6*). Spatial visualization and perspective taking scores were highly correlated (n = 10, r = .83, p = .003). These results imply that participants with higher spatial visualization and perspective taking scores tended to travel shorter routes. This is illustrated in *Figure 3.7*, which exemplifies the routes taken by a high-scoring participant and a low-scoring participant. Visual memory was not significant as a predictor of travel distance.

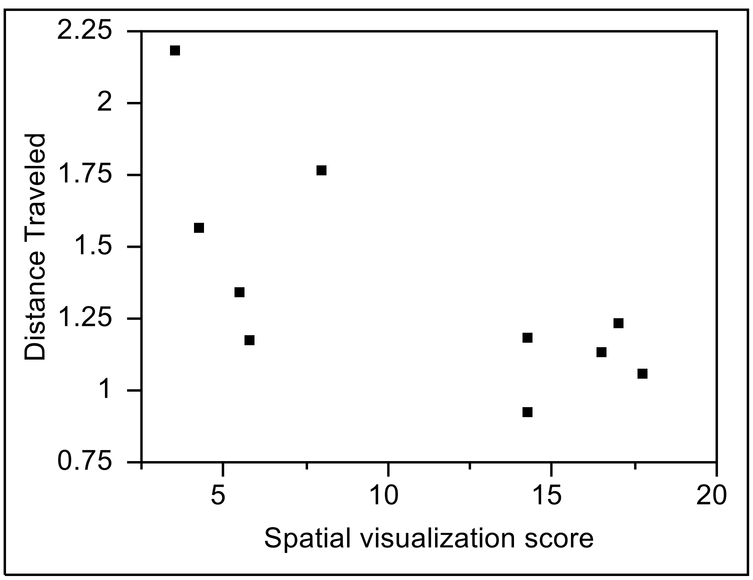


Figure 3.5 *Scatterplot of distance traveled and spatial visualization score (VZ).*

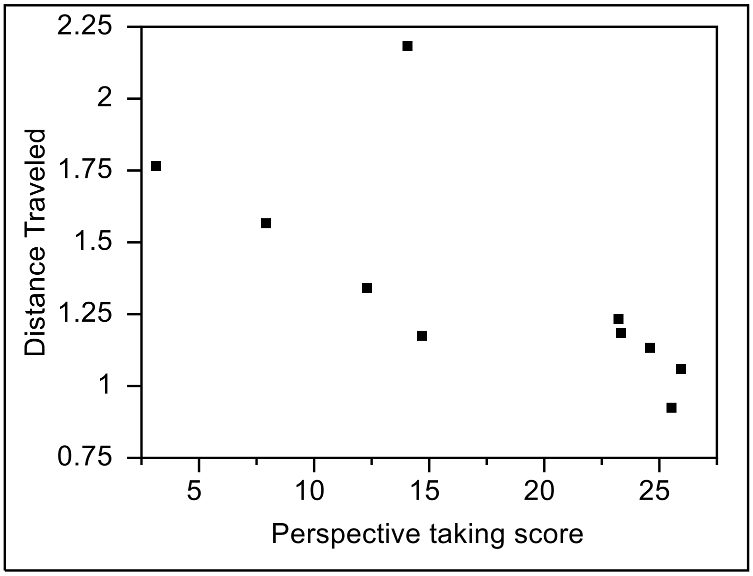


Figure 3.6 *Scatterplot of distance traveled and perspective taking score (PT).*

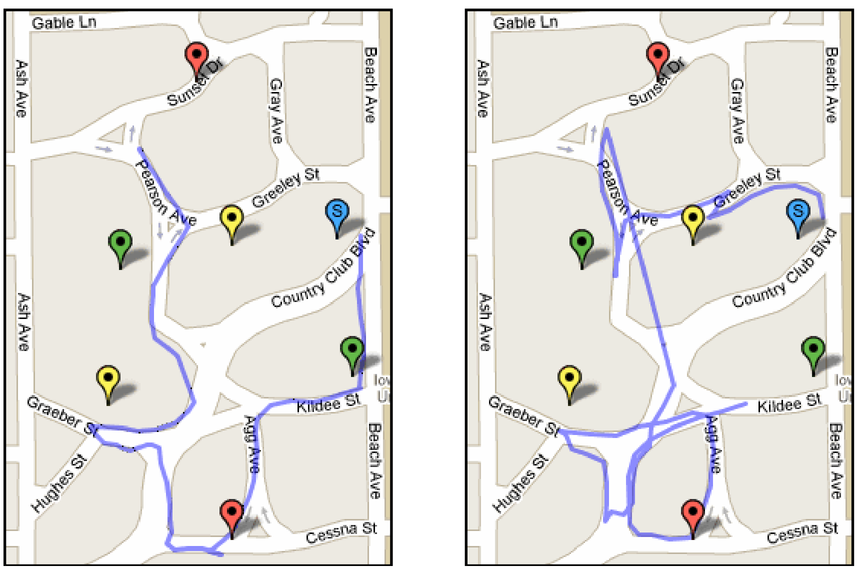


Figure 3.7 *Routes of high spatial visualization group (example on left) were more efficient than those of low group (right).*

Address verification errors were negatively correlated to spatial visualization scores (n = 10, r = -.63, p = .046) as shown by the scatterplot in *Figure 3.8*. This indicates that higher-scoring participants tended to make fewer errors when verifying residences.

Perspective-taking and visual memory scores were not significant as predictors of verification errors. One participant solved all six scenarios correctly. The nine erring participants failed to correctly verify a residence in at least one move scenario, where an address that existed on their map was found to be in a different location in the neighborhood—this is an indicator that move scenarios are particularly troublesome for participants. The two participants with the lengthiest task completion times also made the most mistakes. However, the fastest participants did not make the fewest mistakes.

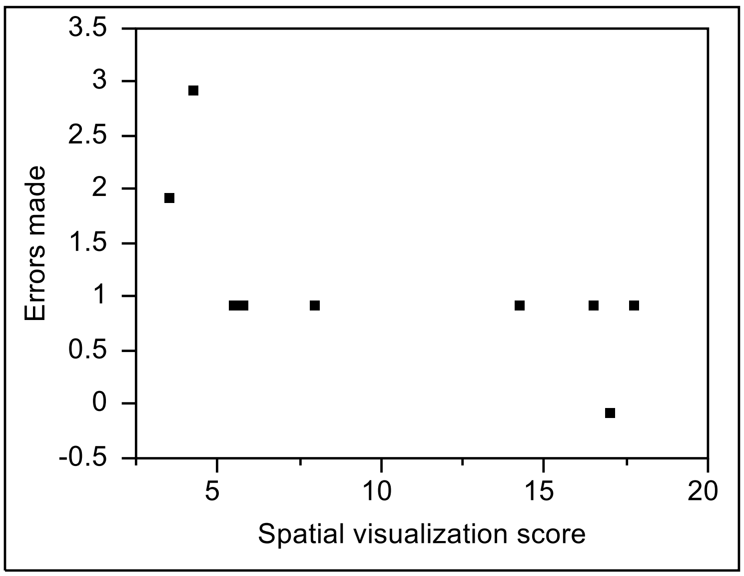


Figure 3.8 *Scatterplot of address verification errors and spatial visualization score (VZ).*

Notes from Participants' Written Materials

Participants used a multicolored pen to take notes on the paper map and the printed address list in order to verify residences and to correct the map when necessary. These materials were collected and analyzed to identify the various types of notes that participants jotted down. The note types found on participants' maps are as follows:

1) *Starting position* – True if participant marked their origin on the map.
2) *Residences to verify* – True if participant highlighted the assigned residences on the map.

3) *Verification order* – True if participant numbered the assigned residences on the map based on the order that he or she intended to verify them (i.e., a route).
4) *Completed verifications* – True if participant used the map to mark the addresses of assigned residences that were verified.
5) *Verification details* – True if participant added notes to the map that were related to needed map corrections (e.g., "residence was in the wrong location").
6) *Move indicators* – True if participant recognized an assigned residence was incorrectly positioned on their map and subsequently drew indicators (e.g., an arrow) denoting the correct location.
7) *Confirmed addresses* – True if participant verified addresses that were not a part of the assigned list.
8) *Streets of assigned residences* – True if participant highlighted the streets and/or street labels of assigned residences on the map.
9) *Additional street labels* – True if participant added extra street labels to the map.
10) *Confusing areas* – True if participant highlighted areas of the neighborhood that were confusing.
11) *Landmarks* – True if participant added indicators for landmarks in the neighborhood.
12) *Traced route* – True if participant traced the route that they followed.
13) *Color coding* – True if participant used color-coding for different types of notes written on the map.

An example of the notes that one participant jotted down on the study map is shown in *Figure 3.9*.

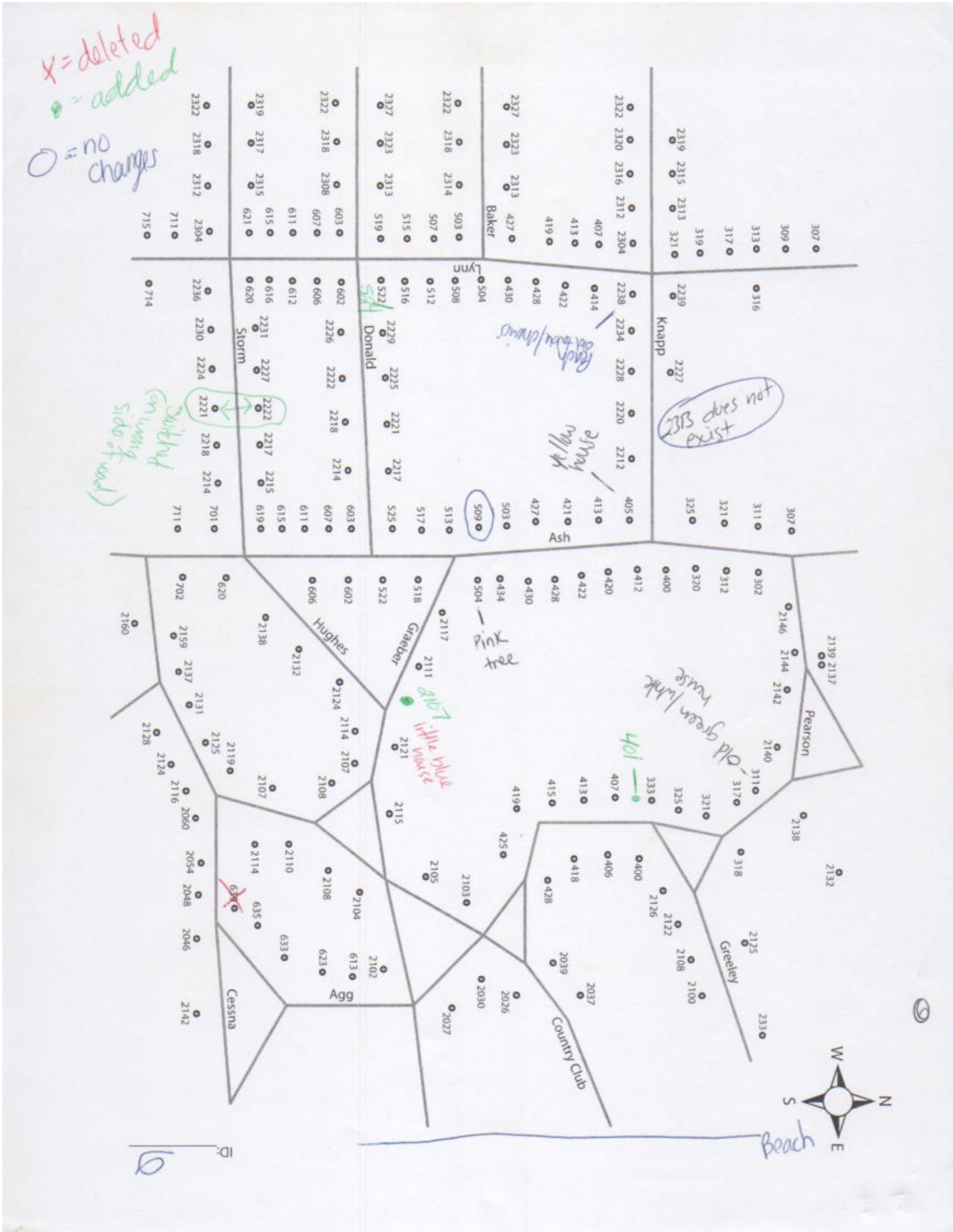


Figure 3.9 *Example of a participant's map by which color-coded notes served as task aids and as instructions on how to correct the map.*

Figure 3.10 shows the types of notes found on the study maps that were given to participants in addition to the percentage of participants that used each type—participant percentages are broken out by spatial ability group. The types of notes with the highest participant utilization are associated with particulars that supported the verification of the assigned residences. The highest percentages of participants tended to:

1) Keep track of the residences that were already verified.
2) Specify the correct locations of residences that were incorrectly located on the map (i.e., provide instructions for move/relocation scenarios)
3) Highlight the addresses of residences that they intended to verify in order to keep track and stay organized.

Some types of notes found on the maps were exclusive to either the high- or low-spatial visualization group. Exclusive to the HighVZ group were the participants who marked their starting location on the map (*starting position)*, who traced their route on the map (*traced route*), and who used *color-coding* to differentiate between the different types of notes that they wrote on the map. The LowVZ group had the only participant to draw prominent landmarks onto the map for use with navigation. Participants from neither group used the maps to jot down the order that they planned to verify residences (*verification order*) or comments, instructions and details centered on each verification scenario (*verification details*)—these two types of notes were found exclusively on the lists that participants received.

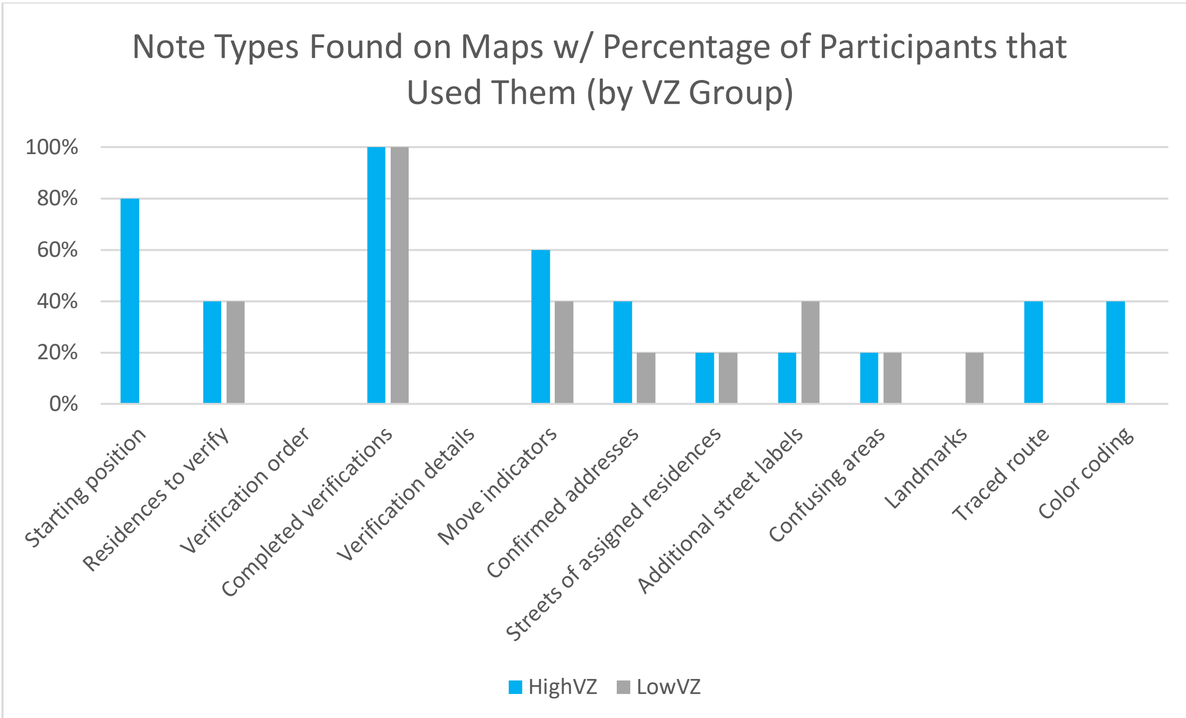
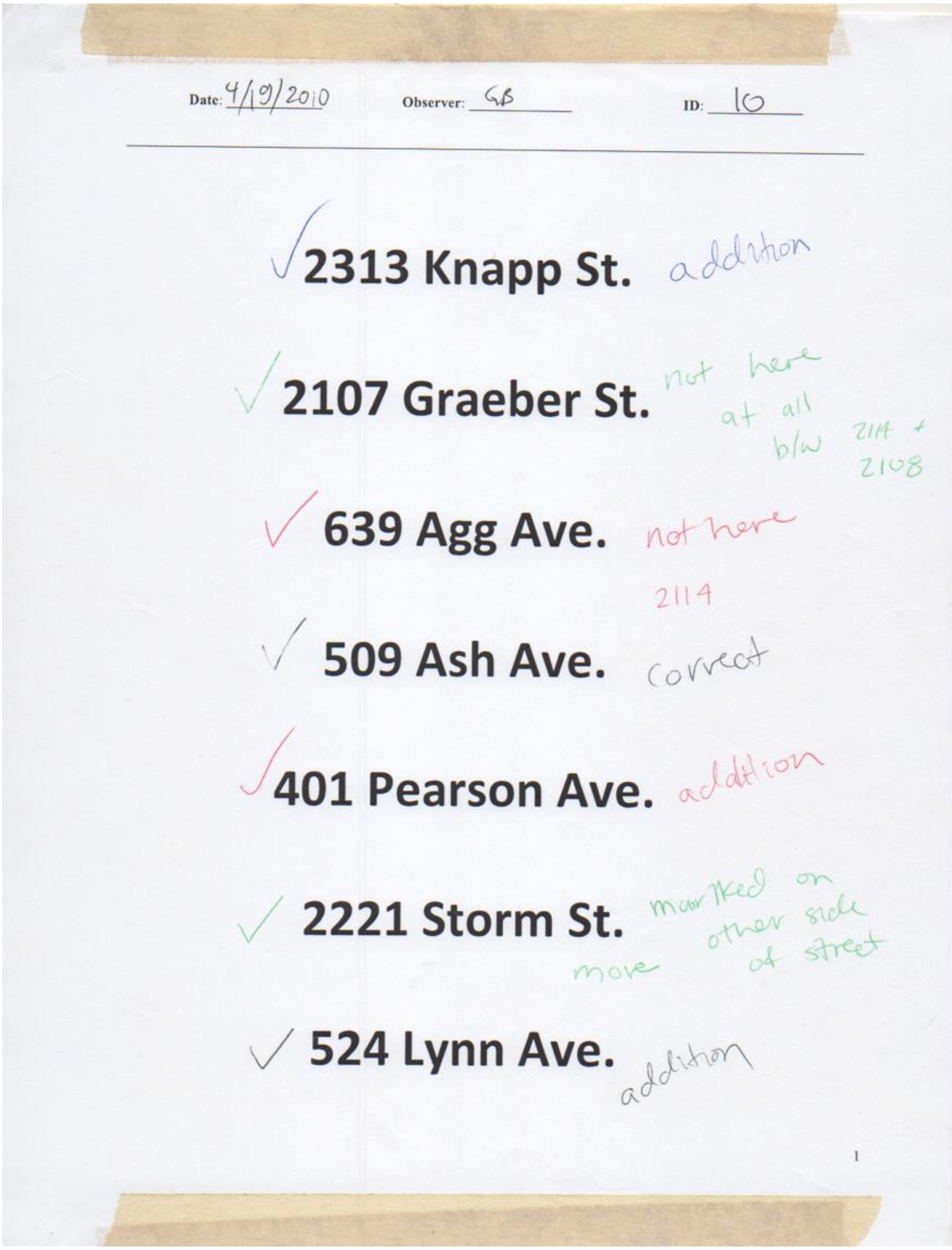


Figure 3.10 *Bar graph showing the types of notes found on FE1 maps and the percentage of participants that used each type (by VZ group).*

Shown on the lists that participants received were six addresses that they were asked to verify. In addition to using the lists for reference, participants also used them as worksheets during the task. An analysis of the lists resulted in the following note types:

1) *Verification order* – True if participant numbered the assigned residences on the list based on the order in which he or she intended to verify them.
2) *Completed verifications* – True if participant used the list to mark the addresses of assigned residences that were verified.
3) *Verification details* – True if participant added notes to the list that were related to the verification of assigned residences (e.g., "residence was in the wrong location").
4) *Color coding* – True if participant used color-coding for different types of notes on list.

It is worth noting that the categories of notes found on the address lists were sometimes also found on the maps, hence, the note types can overlap. An example of the notes that one participant jotted down on his or her assignment list is shown in *Figure 3.11*.



Date: 4/19/2010 Observer: GB ID: 10

2313 Knapp St. addition

2107 Graeber St. not here at all b/w 2114 + 2108

639 Agg Ave. not here 2114

509 Ash Ave. correct

401 Pearson Ave. addition

2221 Storm St. marked on other side of street more

524 Lynn Ave. addition

1

Figure 3.11 *A participant's color-coded notes written on the list of assigned residences to be verified; such notes served as task aids and as instructions on how to correct the map.*

Figure 3.12 shows the types of notes found on the lists accompanied by the percentage of participants that used each type—participant percentages are broken out by spatial ability group. The address lists shared two types of notes with the study map: (1) All participants noted their *completed verifications* on their lists and (2) two of the HighVZ participants used *color coding*. The HighVZ participants were the only to jot down *verification details* or to use *color coding* on their lists.

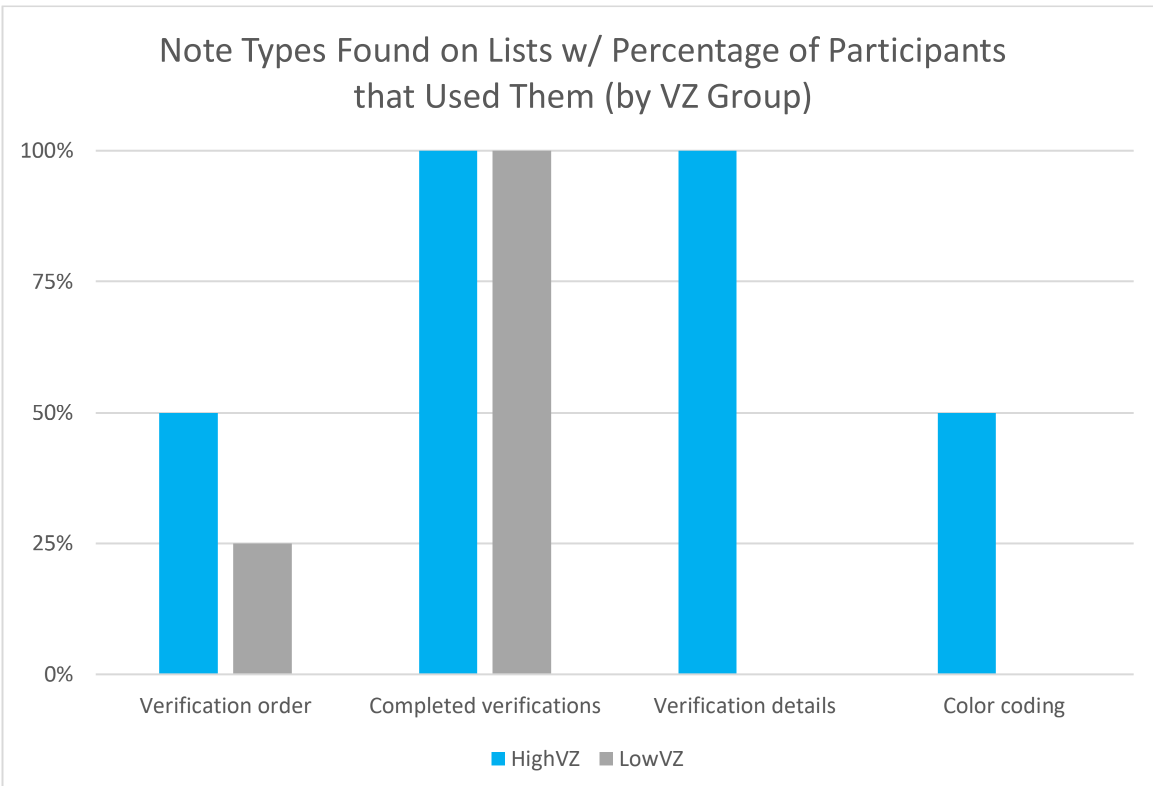


Figure 3.12 *Bar graph showing the types of notes found on FE1 lists along with the percentage of participants that used each note type (by VZ group).*

Generalizing the note-taking tendencies of participants, it appears that participants tended to use the address list as a worksheet that was directed toward the organization, verification, and documentation of the six verification scenarios that were assigned. Participants also used their maps for these purposes, however, the notes found on the map suggest a broader utility, where participants also focused on:

- Tracking landmarks and addresses (that were not assigned for verification) to aid with navigation.
- Tracing travel routes, adding street labels and other visual indicators that assisted them with navigation and ensured that proper corrections were made to the map.

Researcher Observations

During the field exercise, participants verbalized their thought processes related to the task per the think aloud protocol, while an observer recorded their notable behaviors and actions. An exit questionnaire was administered at the conclusion of the field exercise to further probe participants on the task (see *Appendix A*). This section discusses key observations and feedback.

Planning Behaviors

The spatial visualization grouping of participants appeared to be linked to whether or not they devised a thorough plan to verify residences at the onset of the address verification task—we refer to this activity as pre-planning. Pie graphs of the pre-planning tendencies of participants from the HighVZ and LowVZ groups are shown in *Figure 3.13*. Four out of five participants in the HighVZ group pre-planned their route at the beginning of the exercise. Pre-planners aimed to minimize walking distance, oftentimes noting that they wanted to plan an efficient circular path that would lead back to their starting position. Three of the five high-scoring participants wished they had planned better, either by further optimizing the route or by adding more supplemental detail to the map. In the low-scoring group, only one out of five participants pre-planned. The remaining participants began moving through the neighborhood immediately and planned on-the-fly, typically giving priority to the assigned residences that were close in proximity or by using some other method to prioritize.

Retrospectively, the five participants that failed to plan indicated that they would do so if given a second opportunity.

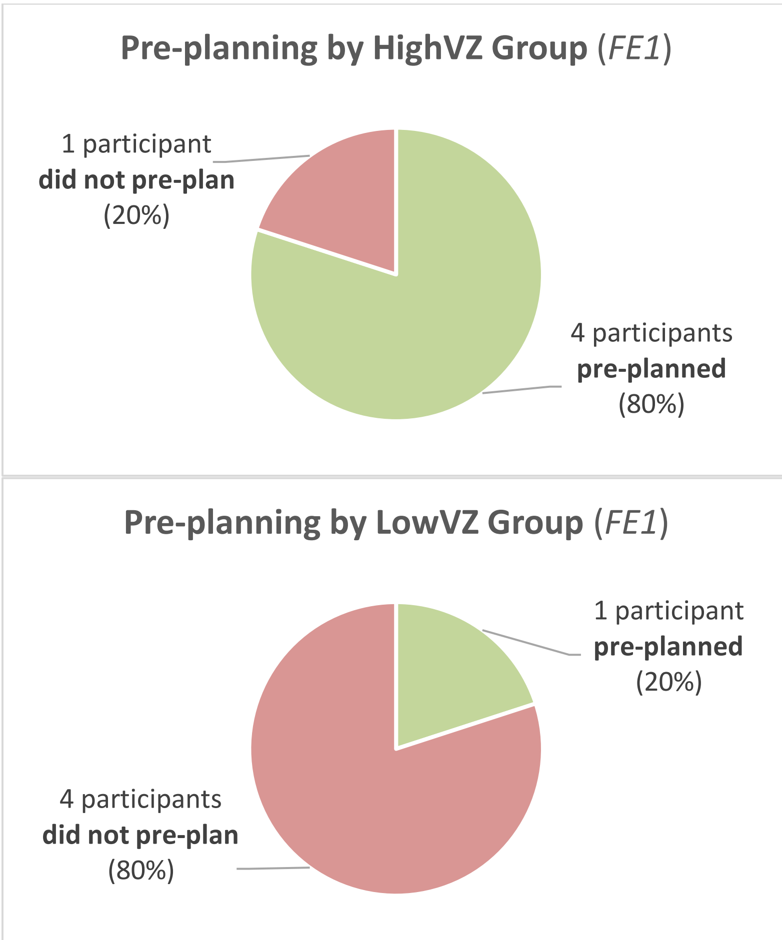


Figure 3.13 *Pie graphs comparing FE1 participants' pre-planning tendencies between the HighVZ group (top) and the LowVZ group (bottom).*

Expression of Heading and Direction

Participants verbalized direction using two different modes: Egocentric (e.g., forward-backward, left-right) and cardinal (e.g., north-south, east-west). The cardinal mode was predominantly used to describe heading, but when describing their immediate surroundings, participants utilized either mode depending on preference and context. One participant was observed using the two modes interchangeably when describing the environment.

Map Orientation

The map rotations of four participants corresponded to shifts in their heading, indicating that the rotations were for alignment purposes. These participants were categorized as track-up map users (i.e. the heading is always oriented forward/up regardless of the cardinal direction). Of our five HighVZ group participants, only one preferred the track-up orientation. Among the remaining four high-scorers who preferred a north-up orientation, two were observed to briefly switch to track-up in confusing areas such as convergent and divergent intersections, indicating that a map orientation switch to track-up may be helpful in situations where cognitive load is high.

Importance of Landmarks and Neighborhood Features

Participants mentioned landmarks such as prominent apartment complexes, Greek-organization dormitories, and prominent buildings associated with the university (e.g., a student center or auditorium). Participants also noted ascending/descending and odd/even address numbering patterns on residential streets. When participants reviewed their maps and noticed that an assigned residence deviated from a recognized pattern, this sometimes served as an early indicator that certain map corrections would be necessary (e.g. the map shows an even-numbered address on an odd-numbered street). Similarly, some participants—especially those that pre-planned—recognized that certain assigned residences were missing from their maps.

Task Difficulty

We asked participants questions regarding the difficulty of the six address verification scenarios. Participants expressed that the add scenarios were the easiest of those encountered; these scenarios involved residences that were confirmed to be in the neighborhood, but were recognized to be missing from the map and needed to be added. The move scenarios (residences that were incorrectly located on the map and needed to be moved) and the remove scenarios (residences that no longer existed and needed to be removed from the map) were thought to be the most difficult situations and required thorough comparisons of the maps against what was seen on the neighborhood grounds. The perceived difficulty of the field exercise was not correlated to cognitive test scores.

The Task Workflow (FE1)

A diagram of the address verification workflow based on observations from *Field Experiment 1* is shown in *Figure 3.14*. The rounded shapes represent either the beginning or ending of the workflow. The rectangles represent the major phases of the task and the arrows indicate their flow. The triangle indicates where the verification decision is made. One iteration through the workflow equates to the verification of a single residence by a participant. We tabulated a total of 59 iterations through the address verification workflow—the HighVZ group verified 30 residences and the LowVZ group verified 29 (one LowVZ participant failed to verify a residence). The first three shaded boxes to the right of the workflow diagram indicate the frequency of actions taken by participants (broken out by participant group) that could be associated with the workflow phases; also shown are the number of residences that were subject to these actions and decisions. The fourth shaded box to the right of the diagram is not associated with a particular phase, rather it shows the frequency of address verification errors made. The remainder of this section will describe the phases of the workflow as well as the actions and behaviors that were observed in each phase.

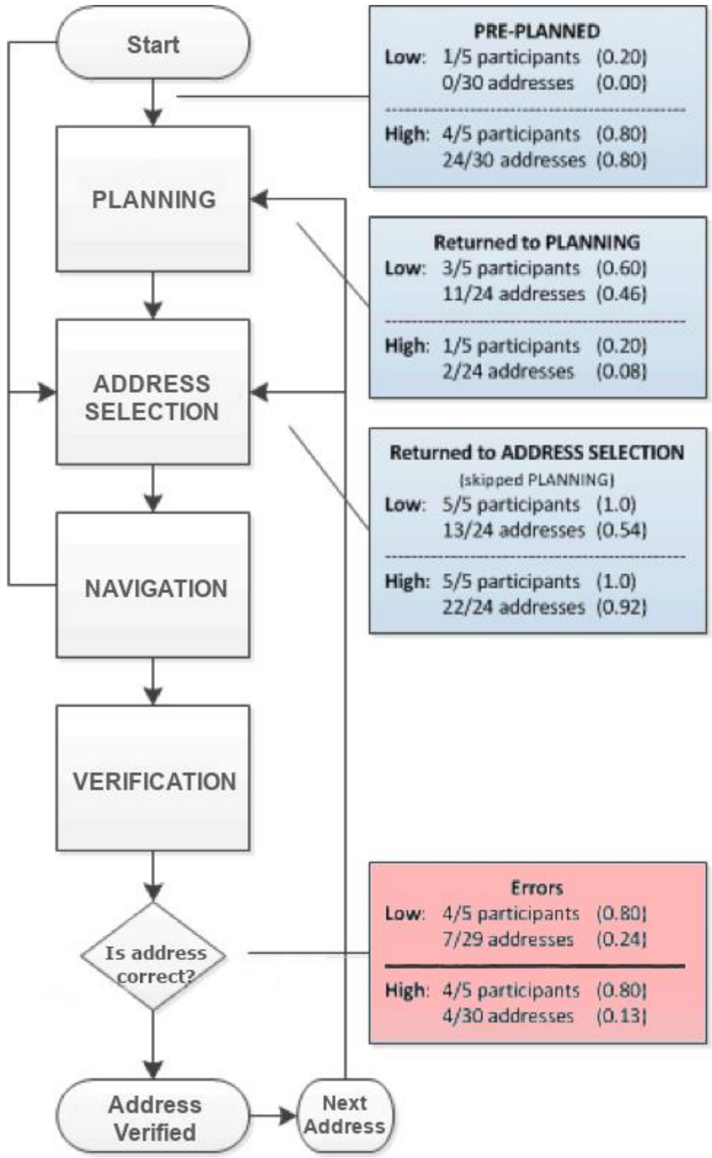


Figure 3.14 *The task workflow diagram for Field Experiment 1.*

Planning Phase

The earliest actions and behaviors observed in participants were centered on planning, thus, the first phase of the workflow is the *Planning Phase*. At the earliest juncture of the Planning Phase, participants had the ability to “pre-plan.” We define pre-planning as a process that occurred at the onset of the task where participants formulated a comprehensive plan for their route and the order in which they intended to verify residences—this minimized or eliminated the need for further planning. Pre-planning was not attributed to participants who revisited planning after having started the task without creating thorough plan at the onset. *Figure 3.14* shows that the HighVZ group more often pre-planned (4/5 participants) and they did so more frequently (24/30 addresses) than the LowVZ group (1/5 participants; 0/30 addresses). One LowVZ participant attempted to pre-plan a route, but abandoned the process after encountering difficulty. Some participants from the LowVZ group (3/5) subsequently developed comprehensive plans after verifying at least one address—this suggests that they recognized the utility of planning after the exercise was already underway.

Scrutiny of the map and the residences that were to be verified during the planning phase was advantageous for some participants as this behavior revealed potential errors early on. For example, some individuals noticed that a residence deviated from a recognized address numbering convention and this served as an early indicator that certain map corrections would be necessary (e.g., the map showed an even-numbered address on an odd-numbered street). Similarly, some participants, especially those that pre-planned, recognized that certain assigned residences were actually missing from their maps.

Address Selection Phase

The *Address Selection Phase* is the second phase of the workflow whereby participants chose the residence that they would verify next. It was, however, the first phase for those participants who skipped or breezed through the Planning Phase. Participants of the LowVZ group used one or more of the following approaches to select residential addresses from their list (ordered by frequency): (1) proximity, (2) a planned route, and (3) haphazardly chosen (e.g., randomly selected the first address from their list). Participants of the HighVZ group most frequently selected addresses based on either (1) a planned route or (2) proximity. No participants from the HighVZ group were observed selecting an address haphazardly—they always followed some sort of method. Those participants who pre-planned and who devised a verification order during the Planning Phase were able to efficiently move to the Address Selection Phase during subsequent iterations through the workflow without further planning.

Navigation Phase

During the *Navigation Phase*, participants used cues from the neighborhood and their maps to reach the vicinity of the residence that was being verified. Each of these actions is graphed in *Figure 3.15*. Navigation behaviors differed noticeably between HighVZ and LowVZ participants. For example, participants would check addresses along the way to reaching their intended destination to ensure that they were on the right track. The LowVZ group was observed checking such addresses approximately three times more (M = 31) than the HighVZ group (M = 10.6). Participants of the LowVZ group also checked street signs and verbally identified neighborhood streets about two and a half times more (M = 20) than participants of the HighVZ group (M = 8.4). The map checking tendencies between the two participant groups did not differ as substantially as the other actions shown in *Figure 3.15*, though the LowVZ group again relied on the map for navigation more (M = 6.6) than did the HighVZ group (M = 5.8). The preponderance of these behaviors during travel among the LowVZ group suggests that LowVZ participants relied more on such details to find their way around the neighborhood.

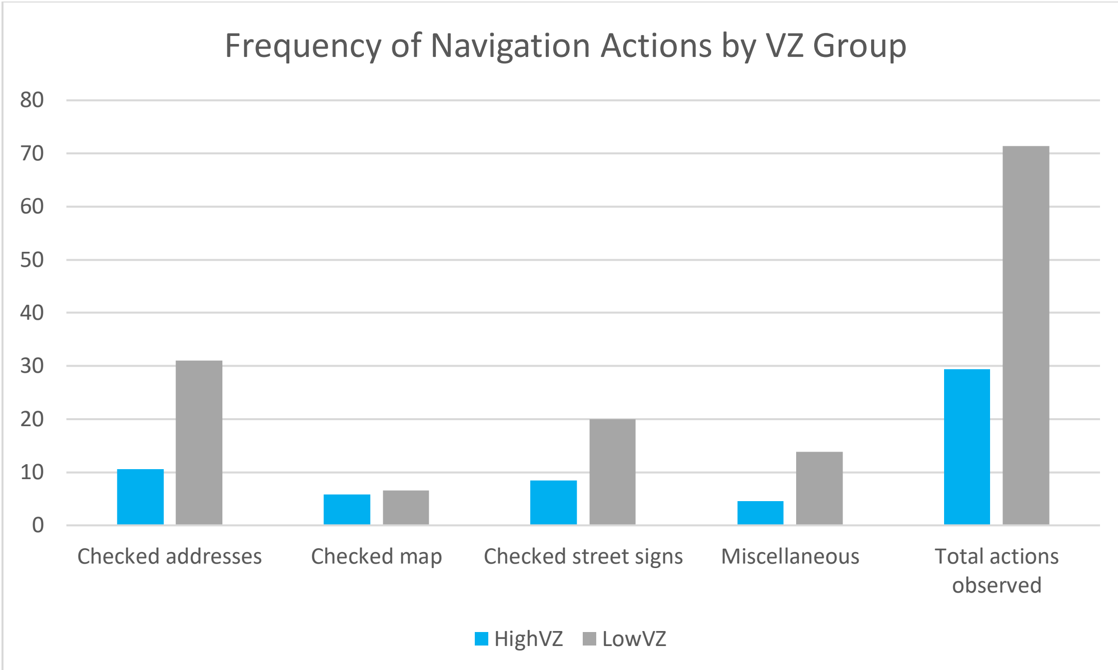


Figure 3.15 *The mean actions of participants during Navigation Phase (by VZ group).*

Participants either preferred a north-up (n = 6) or track-up (n = 4) map orientation, with some north-up participants temporarily switching to track-up in confusing areas of the neighborhood such as at the three-way intersections. Participants' responses to the exit questionnaire indicated that the existing map features were useful during wayfinding and navigation. Participants also expressed that the graphical representation of the "triangles" (i.e., the Y-junctions formed by three-way intersections) served as cues to help them understand their position and bearing in the neighborhood. Participants mentioned landmarks such as prominent apartment complexes, Greek-organization dormitories, and other prominent buildings associated with the university (e.g., a student center or auditorium). When participants were asked how to improve the map, the two most requested map features were additional street labels and landmarks.

Verification Phase

In general, participant behavior changed once they were in the vicinity of an address being verified. During this *Verification Phase*, the participant made a final decision about the correctness of each assigned residence with respect to the map. A key difference in participant behavior during the Verification Phase was the thoroughness with which they scrutinized a variety of cues near the residences of interest—this was done to ensure that each home had the correct address number, was on the proper street and was located in the proper place on the map as seen from the ground. An accompanying observer used a coding sheet to tally each of these verification-centered actions—they are shown in *Figure 3.16*.

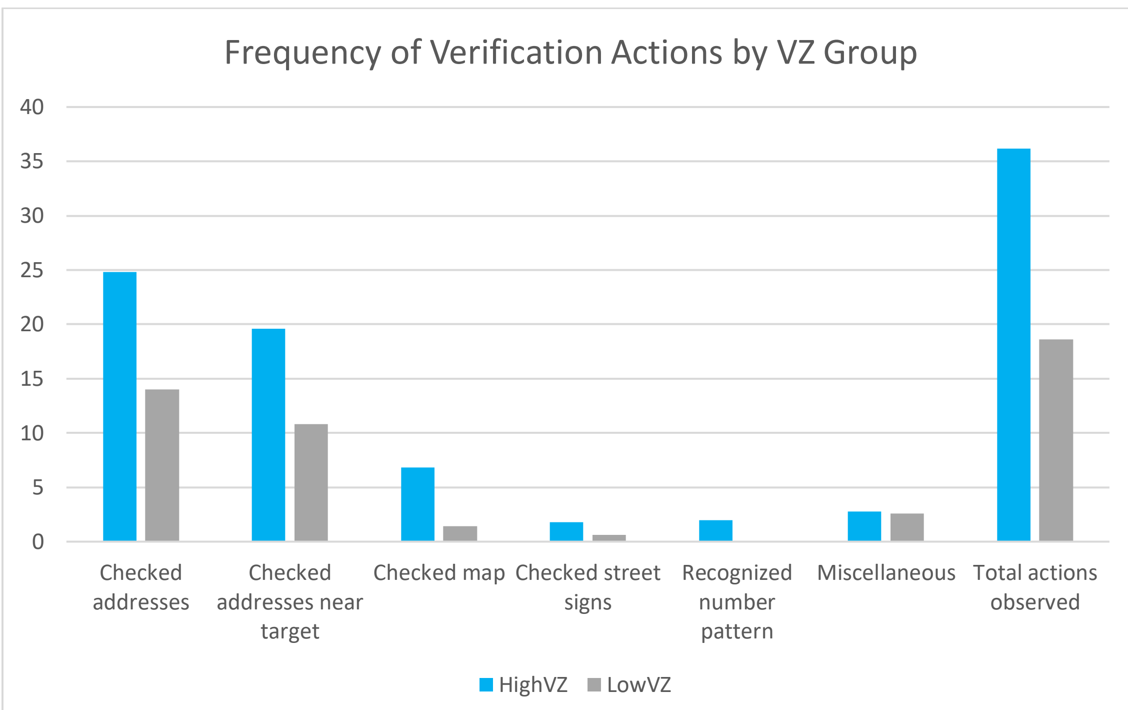


Figure 3.16 *The mean actions of participants during Verification Phase (by VZ group).*

Participants initially checked addresses to pinpoint the precise location of the one that they were verifying. We observed the HighVZ group to check addresses in this manner nearly twice as much (M = 24.8) as the LowVZ group (M = 14). Once reaching the vicinity of an assigned residence, participants oftentimes scrutinized nearby and surrounding addresses very closely, using such addresses as anchors to ensure that the residence that they were verifying was properly numbered, labeled and located. Participants of the HighVZ group checked nearly twice as many of the nearby addresses (M = 19.6) than those of the LowVZ group (M = 10.6).

Further examination of the address-checking behaviors, both before and after an assigned residence was first identified, revealed differences in the thoroughness with which the HighVZ and LowVZ groups utilized nearby addresses to ensure the correctness of the map. The Venn diagrams of the address checking tendencies of the HighVZ *(Figure 3.17*) and LowVZ (*Figure 3.18*) participant groups illustrate these differences.

Some participants checked nearby addresses until they reached the residence that was assigned, at which point they made a verification decision without subsequently confirming other surrounding addresses for precision; the leftmost circle of the Venn diagrams (blue) indicate the percentage of residences that were verified in this manner. The LowVZ group checked more addresses (55.2%) than the HighVZ (43.3%) group prior to reaching their assignments and stopping there.

A couple of participants reached an assigned residence unwittingly, at which point they decided to verify the residence by subsequently checking nearby addresses for precision. The rightmost circle of the Venn diagram associated with the HighVZ group *(Figure 3.17*) indicates that 6.7% of residences were verified in this fashion—there were no observations of the LowVZ group checking nearby addresses after (and only after) identifying an assigned residence.

Participants that were the most thorough in verifying residences tended to check nearby addresses both before identifying an assigned residence and afterward. The inner circle of the Venn diagrams (brown) indicate the percentage of residences that were verified this way. The LowVZ group checked fewer addresses in this fashion (27.6%) than the HighVZ group (46.7%).

Participants that were the least thorough with their address checking during the Verification Phase tended to navigate directly to the residence that they were seeking and once they identified it, they made a decision then and there without checking nearby addresses to ensure correctness. The red-colored circles located to the bottom right of the Venn diagrams show the percentage of participants that were observed verifying residences without any rigorous address checking. A larger percentage of participants from the LowVZ group ignored nearby addresses during verification (17.2%) than did those of the HighVZ group (3.3%).

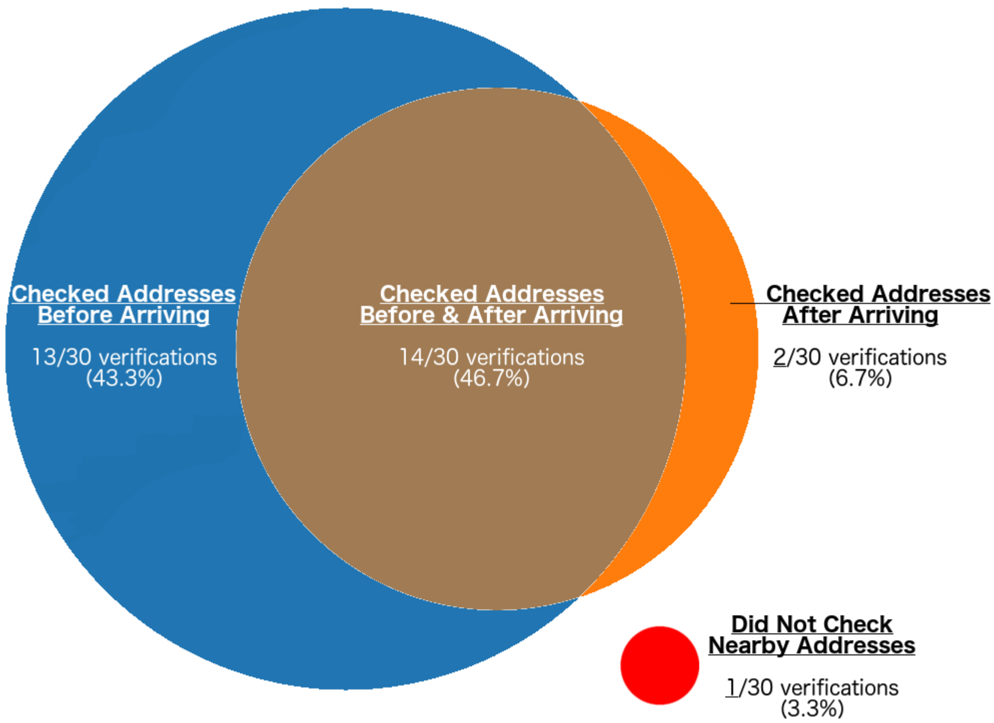


Figure 3.17 *Venn diagram showing the thoroughness of address checking for the HighVZ participant group.*

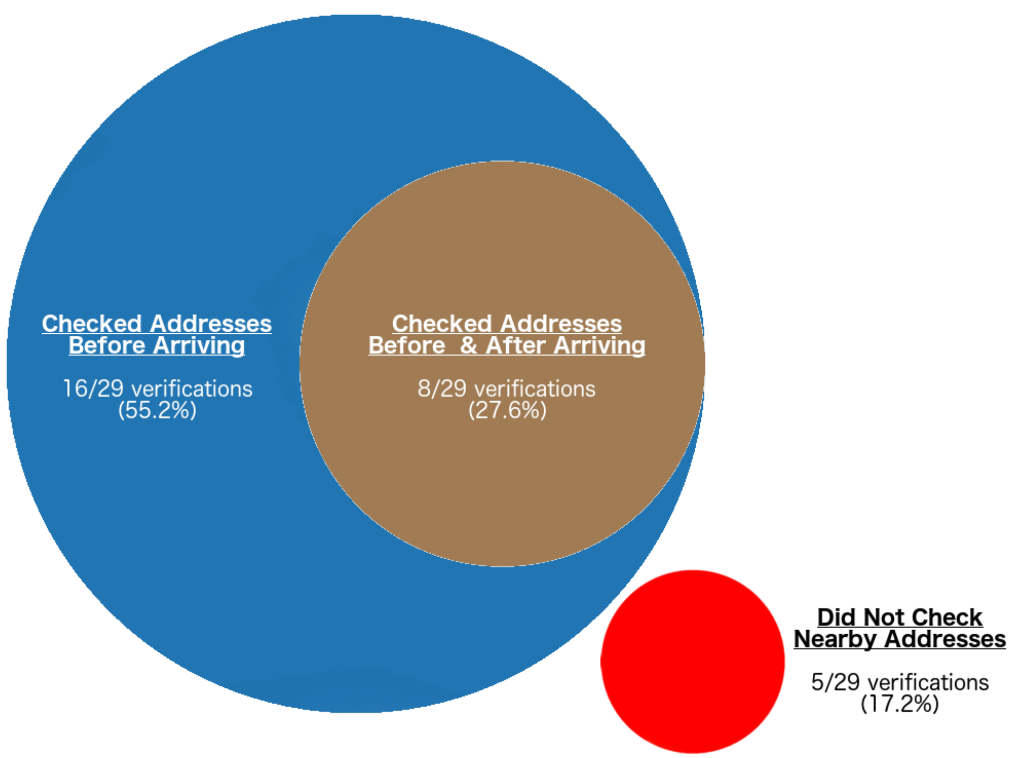


Figure 3.18 *Venn diagram showing the thoroughness of address checking for the LowVZ participant group.*

3.4.6 Discussion

This study served as the first step in our group's exploratory research agenda. We set out to gain a better understanding of how users with either high or low spatial ability tackled the mobile task of address verification in the field—a neighborhood—using a paper map. Our objectives were to vet and refine our methods, the study protocol, and to collect preliminary data that could validate our research direction.

We hypothesized that participants of the HighVZ group, given their spatial cognition advantages, would demonstrate superior overall task performance compared to their LowVZ counterparts as measured by *task time* (shorter is better), *distance traveled* (shorter is better), and *address verification errors* (fewer are better). The quantitative analysis supports this hypothesis. On average, the HighVZ group demonstrated shorter task times, shorter travel distances, and fewer address verification errors than the LowVZ group. The significant Pearson correlations spanning the cognitive test scores and task performance metrics are listed in *Table 3.6*.

Table 3.6 *Significant Pearson Correlations of the Task Performance Variables and Cognitive Test Scores (FE1).*

Variable	**by Variable**	***N***	***r***	***p***
Perspective Taking	Spatial Visualization	10	0.84	<.01
Perspective Taking	distance traveled	10	-0.69	<.05
Spatial Visualization	task time	10	-0.64	<.05
Spatial Visualization	distance traveled	10	-0.69	<.05
Spatial Visualization	verification errors	10	-0.63	.05
task time	distance traveled	10	0.98	<.001

Overall, spatial visualization score shows the most significant correlations out of the cognitive tests (a significant negative correlation to all three of the task performance metrics) followed by perspective taking score, which has a significant negative correlation to travel distances. These results indicate that as participants' spatial visualization and perspective taking abilities increase, so does their tendency to be more efficient at the task per the measures. Perspective taking scores are positively correlated to spatial visualization scores, which suggests that there is a link between the cognitive abilities and underlying processes that these tests probe. Distance traveled is positively correlated to task time, meaning that as participants traveled greater distances, they also tended to spend more time working on the task. Visual memory shows no significant correlations with the task performance metrics or the other cognitive test scores.

The qualitative analysis revealed a task workflow that can be broken down into phases based on participant behaviors and actions. The observational data shows some trends within these phases. During the Planning Phase, HighVZ participants were more likely to pre-plan (4 out of 5 did) than their LowVZ counterparts (1 out of 5 did). The "pre-planners" created very comprehensive plans early-on to locate and verify the six assigned residences; a behavior that reduced task times, travel distances, and helped some participants to identify potential map errors before physically arriving at the locations in question. Based on the frequency of observed actions and behaviors within the phases, the LowVZ participants were more active during the Navigation Phase than the HighVZ group. This was apparently in response to the difficulties presented by navigation, whereby the LowVZ group relied more heavily on cues from the map and environment in order to reach their destinations. However, during the Verification Phase, the HighVZ participants were more active. Whereas the LowVZ group increased their activity in response to the demands of navigation, the increased activity observed in the HighVZ group during verification appeared to be a function of thoroughness rather than difficulty. The HighVZ group scrutinized more cues once they reached the vicinity of the residence being verified. Furthermore, this thoroughness extended to the HighVZ group's address checking behaviors. For example, participants who were observed to be very thorough, checking addresses before and after identifying an assigned residence, did so with more frequency in the HighVZ group. Whereas the least thorough participants, who did not check nearby addresses for precision during the Verification Phase, were more frequently identified within the LowVZ participant group.

We used a mixed-methods approach and have employed techniques that are well-suited for the study of mobile human-computer interaction—these include the field study method, think aloud, user analysis and task analysis. Despite *Field Experiment 1 (FE1)* being relatively small in sample size (N = 10, it has provided a rich set of quantitative and qualitative data, which speaks to the strengths of the methods chosen. The findings evidence differences in task performance and behavior when comparing participants with high and low spatial visualization ability. Based on the outcomes of *FE1*, our team was compelled to conduct a larger, follow-up study that is discussed in the next section.

3.5 Field Experiment 2: Address Verification, A Paper Map, and Grid-Like Streets

3.5.1 Method

This section covers the method used in Field Experiment 2 (*IRB ID: 09-386*). Subsections address the topics of participant recruitment and screening, the address verification task, study materials, the study neighborhood, and the study procedure.

Participant Recruitment and Screening

Twenty-six participants (7 male, 19 female) were selected out of a pool of over 100 college students. Participants responded to posters placed around campus and to ads posted on a student web portal. During screening, participants completed psychometric tests of spatial visualization ability (VZ-2) and visual memory (MV-2) (Ekstrom et al., 1976), as well as perspective taking ability (Kozhevnikov et al., 2006). Participants also completed a survey to ensure that they had limited familiarity with the study neighborhood (see *Appendix B*). Participants signed informed consent forms and received separate gift cards for the study screening and exercise.

The Address Verification Task

Participants were asked to verify the addresses of seven residences in a neighborhood (and to correct any map discrepancies) while thinking aloud. The address verification scenarios that participants encountered are shown in *Table 3.7*. For a more detailed description of the task scenarios, refer to *The Address Verification Task Section*.

Table 3.7 *Address Verification Scenarios for Field Experiment 2*.

Scenario	Map Edit Required	Count
Residence is correctly reflected on map	No change to the map (spot) needed	1
Residence is new or missing (not on map)	Map spot added to map	2
Residence no longer exists, but is on map	Map spot removed from map	2
Residence is incorrectly placed on the map	Map spot moved to correct location	2

Materials

Each participant was given a clipboard with a paper map of the neighborhood attached to the front side (8½ x 11) and a printed list of seven residences to verify was attached to the back. Participants were provided with a multi-colored pen so that they could update their map, add features (e.g., landmarks) and notes. Participants were also outfitted with an audio recording device.

The map given to participants was generated by combining Topologically Integrated Geographic Encoding and Referencing System (TIGER\Line) Shapefiles using ESRI's ArcGIS Desktop© software. The map consisted of two layers of information: (1) a street layer that showed the streets and their respective labels and (2) an address layer that depicted each residential address as a map spot (i.e., a small dot and address number corresponding to a residence). A compass rose was also shown on the map. Participants were encouraged to modify their maps and to take notes during the exercise.

Map spots were deliberately added, deleted, or incorrectly placed on the study map to create the seven address verification scenarios (see *Table 3.7*; *Figure 3.19*). Of the addresses to be verified:

- One was left unmodified; participants were to confirm its location.
- Two addresses were removed; participants were to add them to the map.
- Two addresses were added; participants were to delete them from the map.
- Two addresses were moved; participants were to move them to their correct locations on the map.

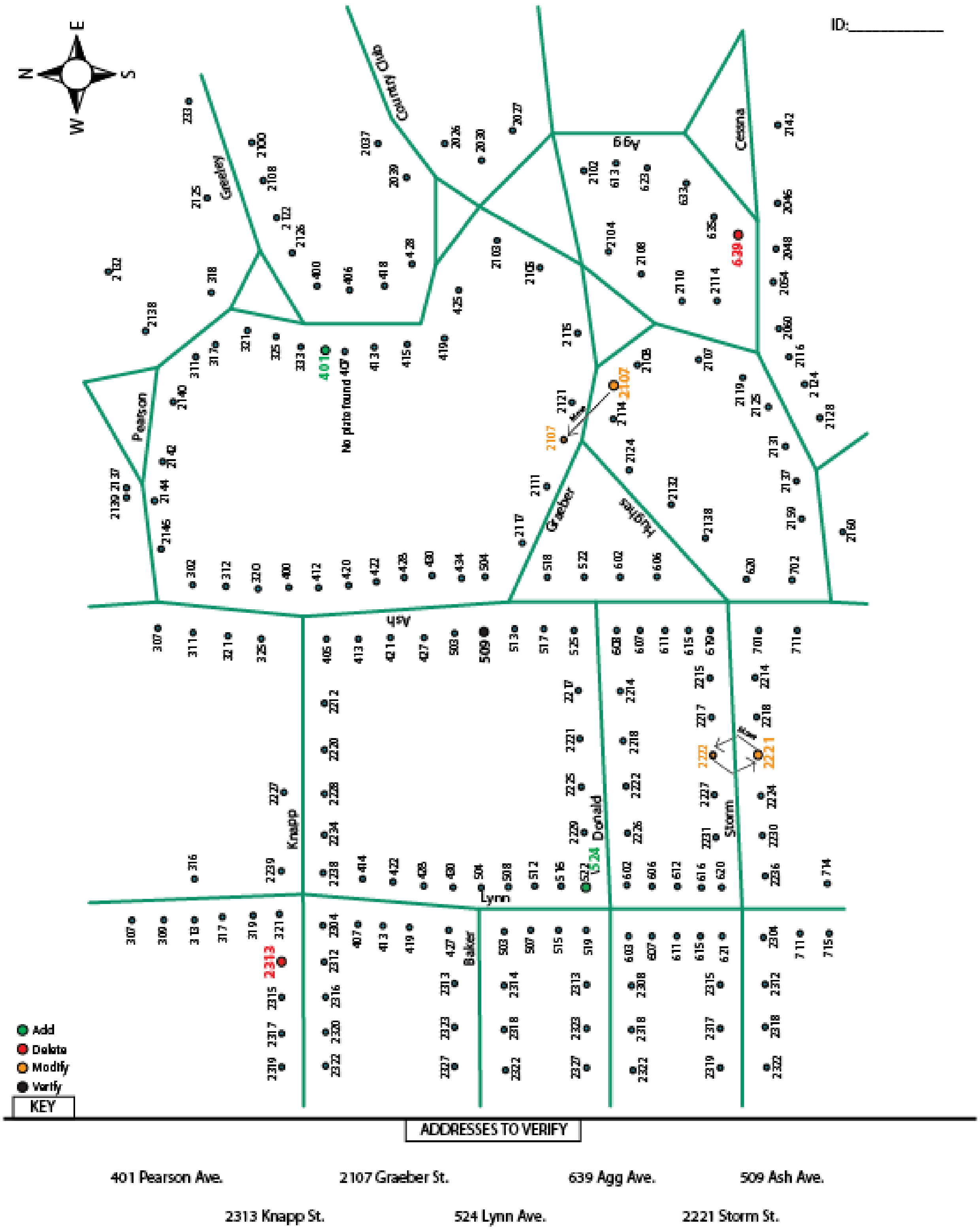


Figure 3.19 *Field Experiment 2 map with a key and indicators to show the corrections that participants should have made.*

Observers used a GPS-enabled smart phone and written materials to collect quantitative and qualitative data. The GPS provided estimates on participants' speed and distance. Latitude and longitude coordinates were logged. A JAVA program was written to convert each participant's coordinates into query strings per the Google® Static Maps API—this allowed us to generate rasterized maps of each participant's path using the Google® Maps service.

A questionnaire was given to participants at the end of the exercise (see *Appendix B*) to gain insight into their map use, planning behavior, and thoughts on task difficulty.

The Study Neighborhood

The field exercise took place in the residential area depicted in *Figure 3.19*. The eastern half of the study neighborhood is what was used primarily for *Field Experiment 1*. It contains oblique, non-uniform streets that seldom run parallel with the cardinal directions. The three-way intersections formed by these streets are Y-junctions, rather than T-junctions. The four-way intersections occur at varying angles. The large triangular medians are another notable feature—they are formed when three Y-junctions interconnect. The streets of this half of the neighborhood are oblique rather than orthogonal like the western half.

The western half of the study neighborhood—a new addition—is laid out in a grid. It is made up of streets that are homogeneous and closely aligned with the cardinal directions. The three-way intersections of this area are exclusively comprised of T-junctions; the 4-way intersections are similarly perpendicular. These features give this area an orthogonal, uniform structure. The resulting blocks are approximately rectangular in appearance.

Field Experiment 1 was conducted in the non-uniform, eastern half of the neighborhood because it was believed to be the most challenging. The research team subsequently included the western half of the neighborhood for *Field Experiment 2* in order to better understand how participants tackled a grid-like, more structured neighborhood layout.

Procedure

A researcher accompanied each participant to the study neighborhood. The task and think aloud protocol were explained. Participants were equipped with a digital recorder and microphone. Participants were given a training map, a list of training addresses and a multicolored pen. During training, participants verified three addresses to get familiar with the task and think aloud protocol. The researcher provided feedback and answered questions during this time. At the end of the training exercise, a map key was given to participants to show the changes that should have been made to their map to accurately reflect what was seen on the grounds of the neighborhood. Any discrepancies between participants' edited training maps and the map key were discussed until participants fully understood the procedure.

After training, participants were walked to a uniform starting location and asked if they had any final questions. Participants were told that the researcher would no longer answer questions and would only speak to get clarification on participants' actions and to remind them to think aloud. Participants were then given the study map and a list of seven addresses to verify. The researcher inititalized a GPS device and started each participant's audio recorder. The researcher shadowed participants with a coding sheet to record observations and notes on participant actions and behaviors. At the end of the exercise, participants answered a questionnaire about the task (see *Appendix B*).

3.5.2 Experimental Design

A between-subjects design was used to test the hypotheses. Participants completed psychometric tests of spatial visualization ability. Stratified sampling was used so that participants whose VZ scores fell within predefined thresholds for high- and low-scorers were assigned to either a HighVZ or LowVZ group. We examined the effects of participants' spatial visualization ability grouping (VZ) on task outcomes.

3.5.3 Hypothesis

We hypothesized that participants of the HighVZ group, given their spatial cognition advantages, will demonstrate superior overall task performance compared to their LowVZ counterparts. This hypothesis was tested per the three task performance variables, whereby:

- Lower *task time* is better.
- Shorter *distance traveled* is better.
- Fewer address verification errors are better.

We also anticipated evidence of differences between the two participant groups with respect to task-related behaviors, strategies, and workflows.

3.5.4 Variables Used in Analysis

Four categories of data were captured during the screening and field exercise: (1) cognitive test scores, (2) task performance metrics, (3) coded data from participants' written materials, and (4) coded data from the observations that were recorded during the field study. This section covers the variables stemming from these categories that were used in the analyses.

Cognitive Test Scores

Participants were given cognitive tests prior to the field exercise and their scores were calculated. The tests are as follows:

1) *Spatial Visualization (VZ-2)* – (Ekstrom et al., 1976).
2) *Perspective Taking* – (Maria Kozhevnikov et al., 2006).

3) *Visual Memory (MV-2)* – (Ekstrom et al., 1976).

Task Performance

The following task performance metrics were captured during the field exercise:

1) *Task time* – Task completion time for the field study.
2) *Distance traveled* – The distances that participants traveled to complete the address verification task as determined by a GPS device that was carried by an observing researcher.
3) *Address verification errors* – The total number of errors that participants made (or corrections that they failed to make) with respect to the address verification task and the materials that they told to annotate.

Participants' Written Materials

Participants received two printed sheets: (1) A paper map of the neighborhood that depicted the residences within and (2) a list of residences that they were asked to verify. During the field study, participants took notes on these materials using a multi-colored pen. These materials were collected, analyzed and coded based on the types of notes that participants jotted down. The note types that were originally identified in *Field Experiment 1* were re-used and did not require modification. The note types found on participants' maps are as follows:

1) *Starting position* – True if participant marked the starting location on the map.
2) *Residences to verify* – True if participant highlighted the assigned residences on the map.
3) *Verification order* – True if participant numbered the assigned residences on the map based on the order in which he or she intended to verify them.

4) *Completed verifications* – True if participant used the map to mark the addresses of assigned residences that were verified.
5) *Verification details* – True if participant added notes to the map that were related to the verification of assigned residences (e.g., “residence was in the wrong location”).
6) *Move indicators* – True if participant recognized an assigned residence was incorrectly positioned on the map and subsequently drew indicators of the correct location.
7) *Confirmed addresses* – True if participant verified addresses that were not a part of the assigned list.
8) *Streets of assigned residences* – True if participant highlighted the streets and/or street labels of assigned residences on the map.
9) *Additional street labels* – True if participant added extra street labels to the map.
10) *Confusing areas* – True if participant highlighted areas of the neighborhood that were confusing.
11) *Landmarks* – True if participant added indicators for landmarks seen in the neighborhood.
12) *Traced route* – True if participant traced the route that they followed.
13) *Color coding* – True if participant used color-coding for different types of notes added to the map.

Analysis of the notes taken by participants on their address lists resulted in the following note types:

1) *Verification order* – True if participant numbered the assigned residences on the list based on the order in which he or she intended to verify them.
2) *Completed verifications* – True if participant used the list to mark the addresses of assigned residences that were verified.
3) *Verification details* – True if participant added notes to the list that were related to the verification of residences (e.g., "residence was in the wrong location").
4) *Color coding* – True if participant used color-coding for different types of notes on list.

It is worth noting that the categories of notes found on the address lists were also found on the maps, hence, the note types overlap between the two materials.

Recorded Observations

The recorded audio coming from participants' think aloud was transcribed and coded for subsequent quantification and analysis. Below are definitions of the significant coded observations stemming from these analyses:

1) *Identified errors early* – Some participants, early on, recognized problems with the residences that they were asked to verify before reaching their respective locations within the neighborhood (e.g., during planning a participant recognized an odd-numbered address on an even numbered street).
2) *Chose residences based on proximity* – When verifying residences, some participants would queue the address nearest to them for verification.
3) *Preferred to use cardinal directions* – During wayfinding and navigation, some participants demonstrated a preference to describe their heading and spatial relationships in terms of the cardinal directions (i.e., north-south, east-west) as opposed to an egocentric perspective (e.g., "to my left").

3.5.5 Method of Analysis

Pearson product-moment correlation coefficients were computed to assess the relationships between the task performance variables and participants' cognitive test scores. Behaviors were categorized via the analysis of qualitative data found in the observer coding sheets, the coded think aloud transcripts, participant annotations on the provided maps and address lists, and participants' responses to the field exercise questionnaire. Behavioral variables that could be quantified across participants were tested against the cognitive test scores and the performance variables. Pair-wise correlations and a two-tailed Welch's *t* test were computed when appropriate. The available records did not allow for coding of some measures, so the number of observations per variable can be fewer than 26.

3.5.6 Results (FE2)

Descriptive statistics for the cognitive test scores are shown in *Table 3.8*. Descriptive statistics for the task performance variables are shown in *Table 3.9*. In terms of task performance, the HighVZ group consistently outperformed the LowVZ group with shorter mean task times, shorter distance traveled, and fewer address verification errors.

Table 3.8 *Descriptive Statistics for FE2 Cognitive Test Scores (by VZ Group).*

Performance Variable	HighVZ Group			LowVZ Group		
	n	Mean	SD	*n*	Mean	SD
Spatial Visualization	13	16.8	2.1	13	9.0	1.8
Visual Memory	13	22.1	1.7	13	17.6	4.3
Perspective Taking	13	21.8	4.3	13	17.4	3.3

Table 3.9 *Descriptive Statistics for FE2 Task Performance Variables (by VZ Group).*

Performance Variable	HighVZ Group			LowVZ Group		
	n	Mean	SD	*n*	Mean	SD
Task time (min.)	13	35.5	5.1	13	44.1	9.9
Distance traveled (mi.)	11	1.36	.17	10	1.63	.15
Address verification errors	13	1.3	1.4	13	1.6	0.9

Correlations of Task Performance and Cognitive Test Scores

Parallel boxplots showing task time for the two participant groups are shown in *Figure 3.20*. Task time was negatively correlated to cognitive test scores on spatial visualization ($N = 26$, $r = -.44$, $p = .02$) and perspective-taking ($N = 26$, $r = -.51$, $p = .01$). These results indicate that participants with higher spatial visualization or perspective-taking ability tended to finish the exercise significantly faster.

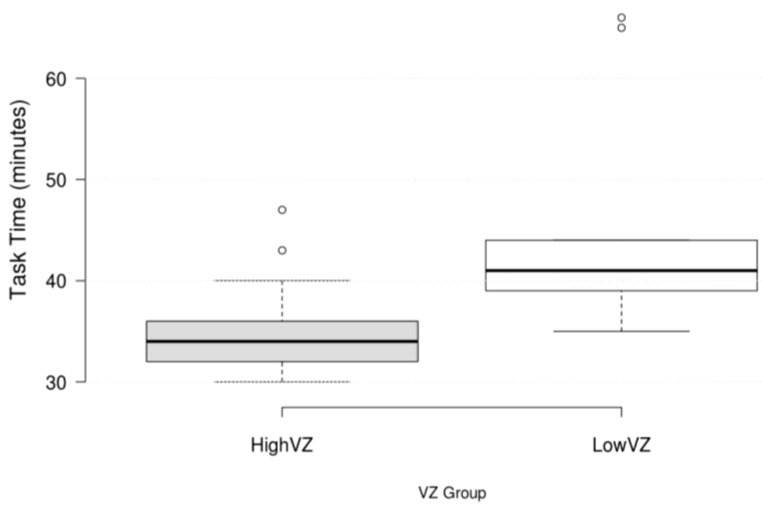


Figure 3.20 *Parallel boxplots of task time by participant group*.

Parallel boxplots showing distance traveled for the two participant groups are shown in *Figure 3.21*. Distance traveled was negatively correlated with spatial visualization test scores ($N = 21$, $r = -.65$, $p < .00$), thus, participants with higher spatial visualization ability tended to travel shorter distances to complete the task. Distance traveled showed a positive correlation with task time ($N = 21$, $r = .47$, $p < .05$), indicating that participants who traveled greater distances to complete the task also spent more time doing so.

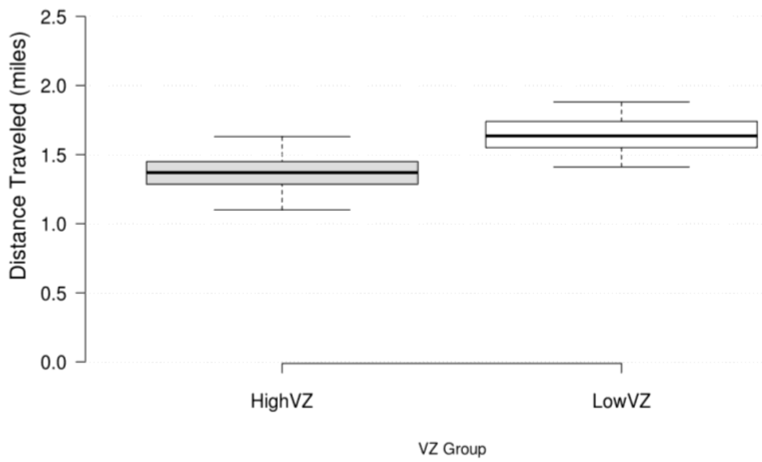


Figure 3.21 *Parallel boxplots of distance traveled by participant group*.

Participants from the HighVZ group made fewer *address verification errors* on average (*n* = 13, M = 1.3, SD = 1.4) than did those of the LowVZ group (*n* = 13, M = 1.6, SD = 0.9), however, address verification errors showed no significant correlation to participants' cognitive ability test scores. Parallel boxplots showing the address verification errors for the two participant groups are shown in *Figure 3.22.*

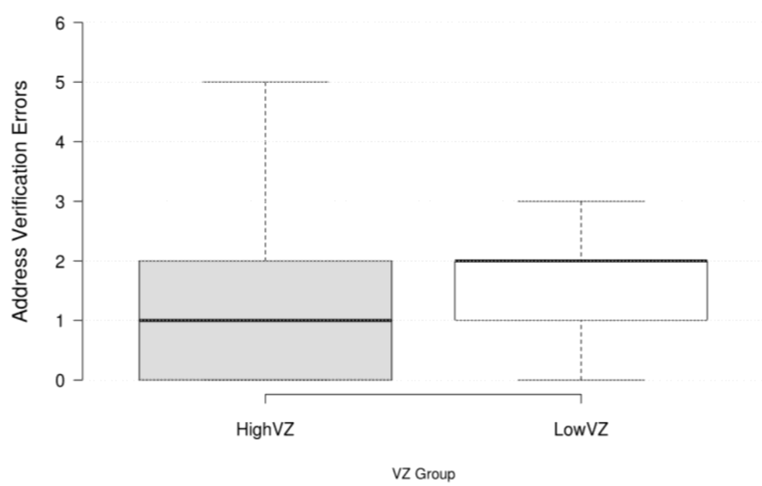


Figure 3.22 *Parallel boxplots of address verification errors by participant group*.

Visual memory test scores showed no significant correlation to any of the performance variables.

Notes from Participants' Written Materials

Figure 3.23 shows the types of notes found on the study maps that were given to participants in addition to the percentage of participants that used each type—participant percentages are broken out by spatial ability group. Consistent with the findings from *Field Experiment 1*, the types of notes with the highest participant utilization pertain to keeping track of the verification of the assigned residences.

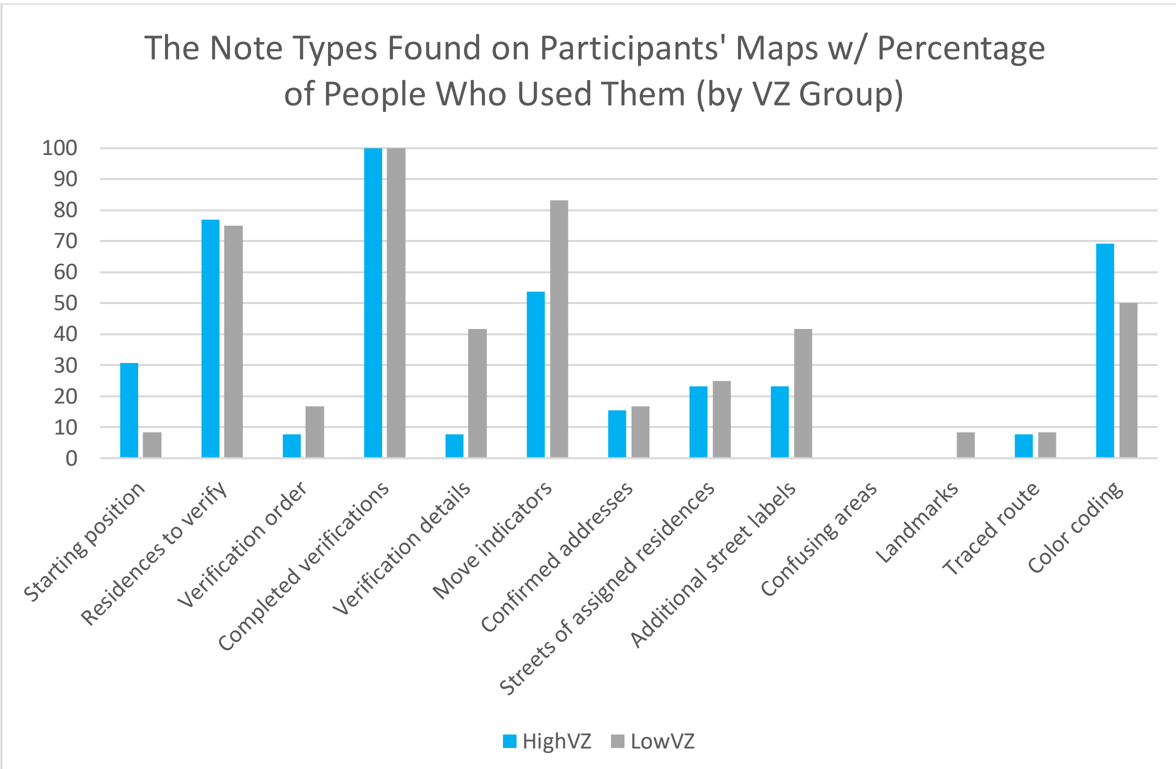


Figure 3.23 *Bar graph showing the types of notes found on the FE2 study maps and the percentage of participants that used each note type (by VZ group).*

The highest percentages of participants tended to:

- Keep track of the residences that were already verified.
- Specify the correct locations of residences that were incorrectly located on the map (i.e., provide instructions for move/relocation scenarios).

- Highlight the addresses of residences that they intended to verify in order to keep track and stay organized.

Members of the LowVZ group were the only participants to draw prominent *landmarks* onto the map to aid with navigation—this finding is consistent with *Field Experiment 1*. Contrary to *Field Experiment 1*, participants from neither group used the maps to note the *confusing areas* that they encountered.

Figure 3.24 shows the types of notes found on the lists in addition to the percentage of participants that used each type—participant percentages are broken out by spatial ability group. The four note types identified on the address lists were also found on the study maps. In order of participant usage, the address lists were used to capture: (1) *Completed verifications*, (2) *color coding*, (3) *verification details*, and (4) the *verification order* that participants planned to use for the assigned residences.

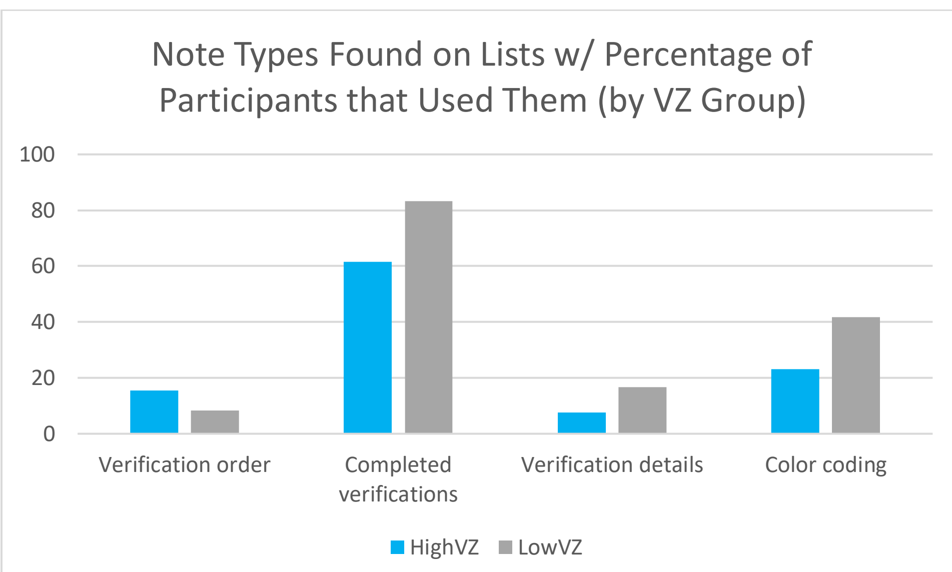


Figure 3.24 *Bar graph showing the types of notes found on the FE2 printed list of addresses assigned to participants along with the percentage of participants that used each note type (by VZ group).*

In terms of the overall frequency of annotations, participants from the LowVZ group took slightly more notes on the paper maps (n = 12, M = 4.8, SD = 1.5) than did those of the HighVZ group (n = 13, M = 4.2, SD = 1.8). Similarly, the printed lists of residences to be verified were also annotated slightly more by the LowVZ group (n = 12, M = 1.5, SD = 1.0) than by the HighVZ group (n = 13, M = 1.1, SD = 0.9). Consistent with the *Field Experiment 1* findings on participants' note taking behaviors, people tended to use the address list as a worksheet that was directed solely toward the organization, verification, and documentation of the six assigned residences. Participants used their maps for these purposes as well, however, the notes found on the maps suggest a broader utility, where participants also focused on:

- Tracking landmarks and addresses, including those that were not assigned for verification, to aid with navigation.
- Tracing travel routes, adding street labels and other visual indicators that assisted them both with navigation and to ensure that the proper corrections were made to the map.

Researcher Observations

Correlations of Coded Observations and Cognitive Test Scores

A think aloud protocol was used and the study audio was recorded analyzed, coded, and quantified. This section will cover the coded variables that are statistically significant.

Instances where participants *identified errors early* were positively correlated to their spatial visualization scores (n = 21, r = .44, p = .05) and their perspective-taking scores (n = 25, r = .49, p = .01). This suggests that participants with higher spatial visualization and perspective taking ability were more likely to identify errors associated with the assigned residences on the maps and lists that they were given—this happened early-on and prior to them physically identifying the residences (typically during planning). Instances where participants *chose residences based on proximity* were positively correlated to their spatial visualization scores ($n = 21$, $r = .45$, $p = .04$). This indicates that participants with higher spatial visualization ability consistently chose residences to verify from their list based on whichever was closest at a given time—this behavior contributed to reduced overall task times and travel distances. Instances where participants *preferred to use cardinal directions* were positively correlated to their perspective-taking scores ($n = 23$, $r = .51$, $p = .01$). This suggests that participants with higher perspective-taking ability were more likely to describe their heading and other spatial relationships from an allocentric frame of reference, where the cardinal directions (i.e., north, south, east, west) were typically used in participants' descriptions rather than terms of an egocentric nature (e.g., "behind me", "to my left").

Planning Behaviors

In *Field Experiment 1*, there appeared to be a relationship between participants' spatial visualization grouping and their tendency to devise a thorough plan at the onset of the field study—we refer to this behavior as pre-planning. Those who pre-planned were likely to be more efficient in completing the task. In *Field Experiment 1*, participants from the HighVZ group demonstrated a greater tendency to plan ahead when compared to the LowVZ group. This finding is confirmed with a larger sample in the present study. Pie graphs of the pre-planning tendencies of participants from the HighVZ and LowVZ participant groups are shown in *Figure 3.25*. Fewer participants from the LowVZ group pre-planned (30.8%; 4 of 13) compared to the HighVZ group (76.9%; 10 of 13).

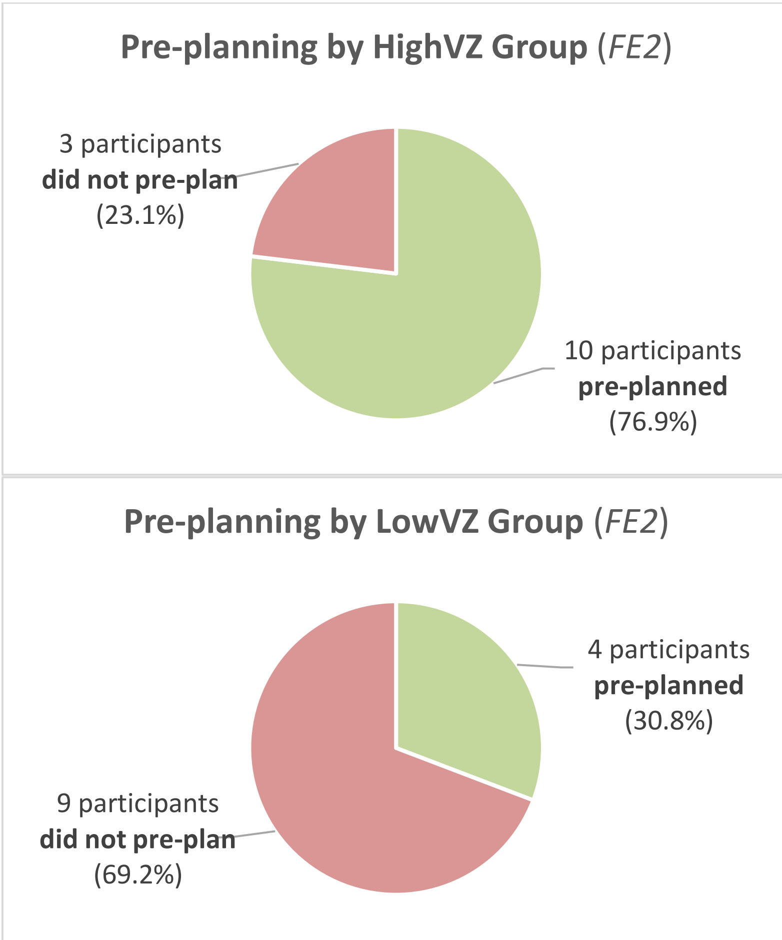


Figure 3.25 *Pie graphs comparing FE2 participants' pre-planning tendencies between HighVZ group (top) and LowVZ group (bottom).*

Expression of Heading and Direction

Participants verbalized direction using two different modes: Egocentric (e.g., forward-backward, left-right) and cardinal (north-south, east-west). The cardinal mode was predominantly used to describe heading, but when describing their immediate surroundings, participants utilized either mode depending on preference and context.

Map Orientation

The majority of participants preferred a track-up map orientation, where their map rotations during the field study corresponded to shifts in their heading; in other words, their heading or direction of travel was always in the forward direction when referencing their maps. Track-up is commonly a default setting in many navigation applications because it eliminates the need for mental rotations and more thorough understanding of the environment by users. The track-up preference showed up for both the HighVZ and LowVZ participant groups. Consistent with *Field Experiment 1* findings, some participants who exhibited a preference for a north-up map were observed to briefly switch to track-up in confusing areas such as convergent and divergent intersections—this suggests that a switch to track-up may be helpful in situations where cognitive load is high.

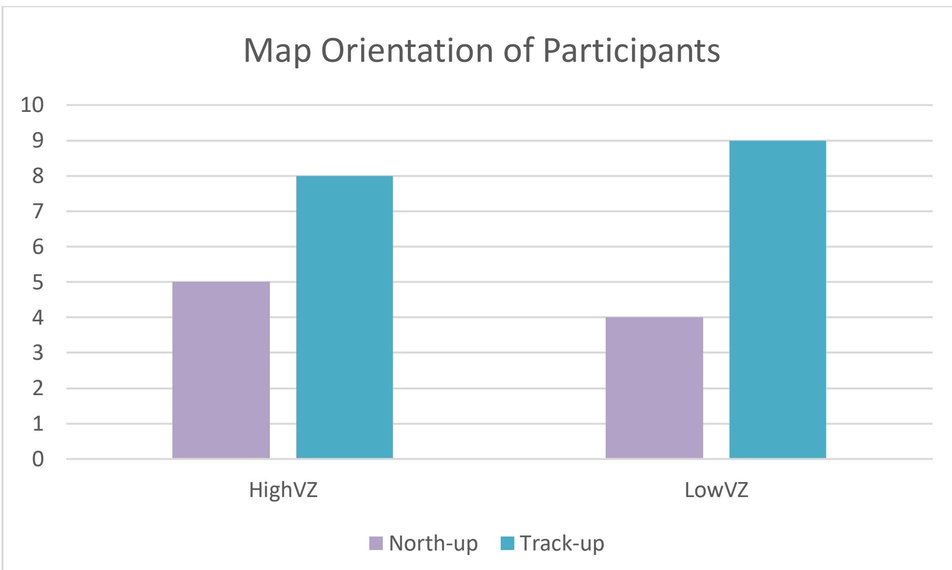


Figure 3.26 *Map orientation of participants (by VZ group).*

Importance of Landmarks and Neighborhood Features

One of the most obvious groups of features on participants' maps were the triangles formed by 3-way street intersections that were located in the eastern half of the neighborhood—these triangles served as landmarks on the map that participants referenced to aid with navigation. Participants also spoke to the use of landmarks that were not shown on the map such as prominent apartment complexes, Greek-organization dormitories, and prominent buildings associated with the university (e.g., a student center or auditorium). At least one participant was observed adding landmarks to the map for navigation purposes. Participants also referenced the map to observe numbering patterns in residences; for example, they recognized ascending/descending addresses, or arrangements where odd and even addresses were on opposite streets. In addition to navigational aids, these various cues served as anchors to help participants ensure that their maps were correct with respect to the assigned residences.

Task Difficulty

We asked participants questions regarding the difficulty of verifying the six residences that they were assigned. In general, responses were similar to those received during *Field Experiment 1*. The difficulty of the move scenarios was most obvious, where participants had to correct an assigned residence by moving a pre-existing map spot from one location on the map to another. For example, some participants arrived at the physical location of an assigned residence but recognized that the map was in error since it indicated no such map spot at that location (it was actually incorrectly placed on the map). In this situation, some participants indicated that a new address needed to be added to the map because they were unaware that the address already existed—i.e., participants corrected the map using an “add” workflow rather than a “move” workflow. Similarly, sometimes participants reached the physical location of a residence that was incorrectly placed on the map and after seeing that it was not there, they indicated that the map needed to be corrected by removing the address altogether; ideally, they should have spent more time to identify the residence’s correct location and then update the erroneous map spot—i.e., participants corrected the map using a “remove” workflow rather than a “move” workflow.

Participants expressed that the easiest verification scenarios involved only the confirmation of an assigned residence (i.e., no map correction was required); followed by the add scenarios (i.e., a residence identified in the field was missing from the map) and remove scenarios (i.e., a residence shown on the map was absent in the field). Lastly, researchers observed that the western half of the neighborhood (with a grid orientation) appeared to be easier for participants to navigate and work compared to the eastern half where streets oftentimes ran and connected at angles that were not orthogonal and aligned with the cardinal directions.

The Task Workflow (FE2)

A diagram of the address verification workflow is shown in *Figure 3.27*. In addition to the shapes that were used previously for the *Field Experiment 1* workflow, the components of the workflow are differentiated using color-coding. The rounded, oval shapes colored in blue represent either the beginning or ending of the workflow. The orange-colored rectangles represent the major phases of the task and the arrows indicate their flow. It is worth noting that the present study's workflow diagram calls out the "Address Selection" process (beige-colored rectangle in diagram) as a part (or product) of the "Planning Phase" rather than a phase in and of itself as it was described for *Field Experiment 1*. The red triangle represents where the verification decision is made. One iteration through the workflow equates to the verification of a single residence by a participant. Reference *Field Experiment 1* (see *The Task Workflow Section (FE1)*) for a more detailed description of participants' behaviors as they move through the workflow.

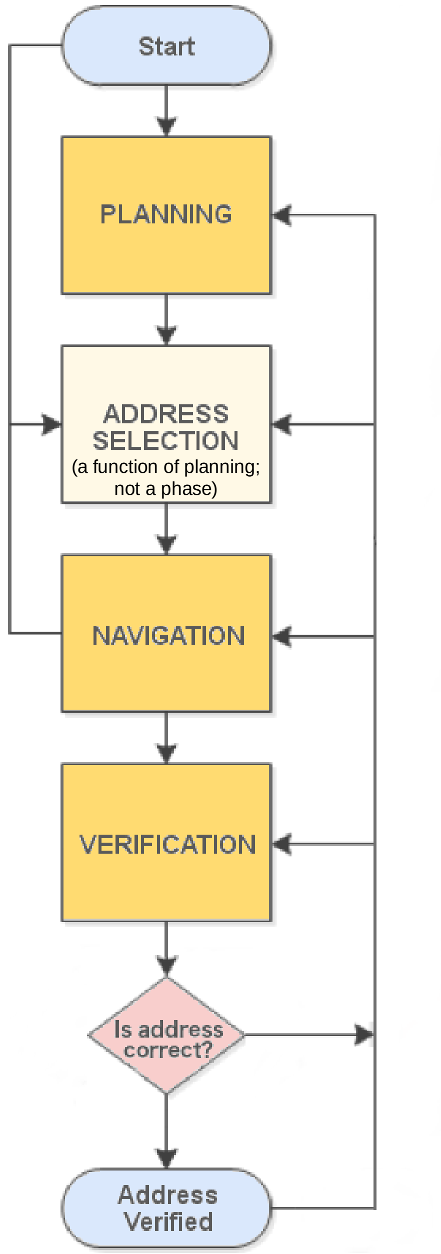


Figure 3.27 *The task workflow diagram for Field Experiment 2.*

3.5.7 Discussion

In the previous experiment, we used the field study method, a think aloud protocol, user analysis and task analysis to explore how people with high or low spatial visualization ability executed a location-based task. The *Field Experiment 1 (FE1)* findings suggest that there are differences in task performance and behavior between the two groups. To corroborate these results and to potentially gain new insights, *FE2* was conducted with a similar design—the key differences being its larger sample size (*N* = 26) and its incorporation of a new section of the study neighborhood that had a grid-like layout.

We again hypothesized that participants of the HighVZ group, given their spatial cognition advantages, would demonstrate superior overall task performance compared to their LowVZ counterparts per the task performance measures. The quantitative analysis continues to support this hypothesis. On average, the HighVZ group demonstrated shorter task times, shorter travel distances, and fewer address verification errors than the LowVZ group, which is consistent with *FE1*.

The significant Pearson correlations spanning the cognitive test scores and task performance metrics are listed in *Table 3.10*. Showing similarities to the *FE1* findings, spatial visualization score has the most significant correlations out of the cognitive tests (a significant negative correlation to two of the three task performance metrics) followed by perspective taking score, which has a significant negative correlation to task time. These results indicate that as participants' spatial visualization and perspective taking abilities increase, so does their tendency to be more efficient at the task per the correlated measures.

In *FE1*, we reported a correlation between address verification errors and spatial visualization score, however, *FE2* yielded no significant correlations between verification errors and other measures. This may be explained by the potential for Type I errors in the *FE1* correlations, which were based on a smaller sample (*N* = 10). Another explanation can be seen in researcher observations and participants' responses to the exit questionnaire, where the western, grid-like half of the *FE2* neighborhood was noted to have a lesser degree of difficulty when compared to the more confusing eastern half—the eastern half made up the entirety of the *FE1* study area. Spatial visualization scores were positively correlated to perspective taking and visual memory scores, which suggests that there is some interconnection in the cognitive abilities and underlying processes that these tests probe. Task time is positively correlated to distance traveled, meaning that as participants traveled greater distances, they also tended to spend more time working on the task.

Table 3.10 *Significant Pearson Correlations Between FE2 Task Performance Variables and Cognitive Test Scores*.

Variable	**by Variable**	***N***	***r***	***p***
Visual Memory	Spatial Visualization	26	.54	<.01
Perspective Taking	Spatial Visualization	26	.44	<.05
Spatial Visualization	task time	26	-.44	<.05
Perspective Taking	task time	26	-.51	<.01
Spatial Visualization	distance traveled	21	-.65	<.05
task time	distance traveled	21	.46	<.05

Based on an analysis of participants' think aloud transcripts, there were several behaviors that showed statistically significant differences between the two participant groups. HighVZ participants were more likely to identify map errors during review of the map and prior to arriving at the physical location of an address. LowVZ participants were observed to detect fewer potential errors upfront. HighVZ participants, for the sake of efficiency, were observed to more frequently use proximity as the basis for selecting their next residence to verify. LowVZ participants more frequently used other strategies; for example, some participants were observed selecting residences based on their printed order (which was randomized) rather than based on a preconceived plan or optimized route. Lastly, HighVZ participants were more likely than their LowVZ counterparts to describe spatial relationships using the cardinal directions (e.g., north, south) rather than using egocentric reference frames (e.g., “to my left” or “behind me”).

The *FE1* qualitative analysis led to a proposed task workflow that was broken down into phases based on participants’ behaviors and actions. The *FE2* observations are compatible with the task workflow and associated behaviors that were presented in *FE1*, thus, only slight modifications were made to the task workflow diagram. For example, in terms of the workflow phases, the “pre-planners” that were identified in the Planning Phase of *FE2* showed similar behavior to the *FE1* pre-planners. These pre-planners were observed creating comprehensive plans early-on, enabling them to more efficiently complete the task. In *FE2*, we also observed that HighVZ participants were more likely to pre-plan (76.9%; 10 out of 13 did) than LowVZ participants (30.8%; 4 out of 13 did).

3.6 Conclusion

A goal of this research was to utilize techniques in mobile human-computer interaction to more effectively evaluate a location-based task in the field—one which requires pedestrian navigation in addition to the use of a map. The task is complex, mobile, and heavily depends on the user’s environmental context and one’s ability to interpret the layout and structural properties of the environment. In response to these challenges, *Field Experiment 1 (FE1)* and *Field Experiment 2 (FE2)* both leverage a mixed-methods approach, and as such we have demonstrated our ability to collect and analyze rich sets of quantitative and qualitative data coming from the field.

By screening participants based on their spatial visualization ability, which has a well-established connection to task performance in computing applications (Campbell, 2011; Egan & Gomez, 1985; Kozhevnikov et al., 2013; Vicente et al., 1987; Zhang & Salvendy, 2001), we have effectively narrowed the design space. In observing the HighVZ and LowVZ participants "in the wild," we were able to identify a task workflow that can be broken down into phases by which distinct sets of behaviors are associated. Furthermore, we demonstrate that differences in participants' workflow and task performance can be linked to their spatial ability. This approach has resulted in an abundance of data from which user needs can be identified and software requirements can be drawn.

CHAPTER 4 – THE VR/FIELD EXPERIMENT

4.1 Introduction

In the numerous field studies that we have conducted, there were times when adverse weather, extreme outdoor temperatures and precipitation (especially during the frigid winters) forced our group to postpone research. The studies were situated in a neighborhood and they involved navigation, so an untimely construction zone, closed roads, or other unforeseen obstacles could have confounded our data or forced us to toss it out. In the field, environmental factors such as these are out of the researcher's control.

The appeal of lab studies is that they can be conducted in well-controlled environments. In the lab, environmental factors that can confound study results are minimized; this enables researchers to establish clearer cause-and-effect relationships in service of some theory or hypothesis. Laboratory settings also facilitate more consistent study protocols, robust data collection and precision in repeated measurements, thus, they are easier to replicate. Field studies, on the other hand, promote more natural behaviors and better reflect what would occur in a real-world setting (i.e., greater ecological validity than lab studies). The realism afforded to field studies also allows for richer sets of data to be collected. However, the disadvantages of field studies include: Greater burden on researchers to plan, resource and execute the studies; reduced experimental control; challenging data collection procedures; and less replicability than lab studies. In other words, "Key disadvantages of either method are the inverse of the other method's key benefits…" (Delikostidis et al., 2015, p. 258).

In the earlier work, Rusch et al. (2012) implemented a desktop simulation to study the relationships between people's spatial ability and their proficiency at completing the address verification task. The methodological strengths of this earlier work include the experimental control of the lab, the replicability of the study design as well as the ease and reliability of data collection. From a practical standpoint, *Rusch's Lab Experiment* (RLE) was also safer for all parties involved and easier to manage logistically. The drawbacks, however, were evident in the reproduction of the study environment and its context. Photos of the neighborhood were shown to participants so that they had vantage points that could be used for orientation and decision-making, however, much of the physical and cognitive aspects of spatial orientation and pedestrian navigation were absent. These limitations create uncertainty with regard to the generalizability of such a study's findings.

The field experiments from the previous chapter produced rich datasets that revealed distinct phases of the task in which participants' behaviors and actions could be grouped—we refer to these as the planning, navigation, and verification phases (see *The Task Workflow Section (FE1)*; *The Task Workflow Section (FE2)*. However, in the prior lab experiment, because our ability to contextualize the task environment was constrained, the scope of the evaluation was limited to a single phase of the task (the verification phase); the other two phases were our blind spots. And while there was merit in the lab experiment's inclusion of photo-based vantage points to immerse participants in the task environment, it also became clear from our field studies that—by further incorporating the environmental context of the neighborhood—we could paint a more complete picture of the dynamics that occur between the user, the mobile device, and the environment.

These previous studies raise the question: Can we capitalize on the ecological validity associated with the field study designs, while retaining much of the experimental control and reproducibility of a lab-based study? This question is one inspiration for the *VR/Field Experiment (VRFE)*. An overview of the *VRFE* is presented in *Table 4.1*.

Table 4.1 *Overview of the VR/Field Experiment Discussed in Chapter 4.*

	Research Method					**Data Method**		**Environmental Context**			**Spatial Viz.**		**Map Type**	
	Setting		**Type**											
Study or Experiment	Field	Lab	Ethnography	Experiment	Observ. Study	Qualitative	Quantitative	Natural setting	Photo-based	Immersive VE	Tested on it.	Screened on it.	Digital	Paper
Rusch's Lab Experiment[1]		X		X			X		X		X		X	
Ethnographic Field Study[2]	X		X			X		X					X	
Field Experiment 1 (oblique streets)	X			X	X	X	X	X			X	X		X
Field Experiment 2 (oblique + grid)	X			X	X	X	X	X			X	X		X
VR/Field Experiment (oblique streets)	X	X		X	X	X	X	X		X	X	X	X	

[1] A prior lab experiment used for comparison (Rusch et al., 2012; Rusch, 2008); related, but not an experiment covered in this dissertation.
[2] A prior field study used for comparison; related, but not a study covered in this dissertation.

4.2 Research Questions

In the *VR/Field Experiment*, we compare the outcomes of a rigorous field study to its VR lab equivalent. In the VR lab, the task is simulated using an immersive virtual environment that is a replica of the field study neighborhood; the field study design and protocol are also replicated in the lab. The research questions for the present work are as follows:

1. Can a neighborhood that is large in scale be accurately reflected in an immersive virtual environment?
2. Can a field study that involves pedestrian navigation, use of a mobile device and digital map be replicated in a VR lab?
3. Can data be generated and collected in a VR lab with comparable quality and reliability to that of a mobile field study?

4. How does task performance compare between the field and the VR lab?
5. How do individual differences in spatial ability affect task performance and behavior? In the field? In the VR lab? How do the study results compare?
6. How do people use a digital map and software to complete the address verification task? In the field? In the VR lab?

4.3 Method

4.3.1 Participant Recruitment and Screening

During recruitment, study applicants from both the university and nearby towns responded to posters placed around campus and to bulletins posted in local communities and town centers. Ads were also placed in the local newspaper and to a student web portal. For the screening phase, 124 applicants completed cognitive tests that assessed spatial visualization ability (*VZ-2*), visual memory (*MV-2*), perceptual speed (*P-2*) (Ekstrom et al., 1976), and perspective taking ability (Kozhevnikov et al., 2006).

The spatial visualization ability test was used to screen participants into high- and low-ability groups. Stratified random sampling was used, where participants with high spatial visualization scores (VZ-2 score ≥ 15 out of a possible 20) and low spatial visualization scores (VZ-2 score ≤ 9) were randomly assigned to one of the two treatment environments—this yielded fifteen HighVZ group participants and seventeen LowVZ group participants.

Participants also completed a survey to ensure that they had limited knowledge of the study area (see *Appendix C*). Out of the 124 applicants who were screened, thirty-two participants (14 males, 18 females) were selected to complete the field task. Participants signed informed consent forms at the onset of the study and received separate gift cards for their participation in the screening and the field task (if selected). No participants had prior experience with virtual reality (*IRB ID: 10-075*).

4.3.2 The Address Verification Task

During the address verification task, participants walked through a neighborhood to find and verify six residential addresses with the aid of a mobile device and digital map. Upon locating a residence, participants were to confirm whether its existence and location was accurately reflected on the map based on the label and placement of the residence's "map spot". In scenarios where participants identified a discrepancy between an assigned residence and its corresponding map spot, they were instructed to correct the map. The possible map editing scenarios are shown in *Table 4.2*. Key performance metrics for the address verification task include the frequency of address verification errors, task time, and the distance traveled by participants.

Figure 4.1 *Photo showing a person using a mobile device to verify a residence.*

4.3.3 The Point-to-Origin Task

All participants were escorted to the same starting location at the beginning of the study. While completing the address verification task, participants were interrupted and asked to point to this starting location on two separate occasions: Once at the midpoint of the address verification task (i.e., half of the addresses were verified) and once again at the end. For the first pointing test, each participant was escorted to a uniform location in the neighborhood and faced due north. They were then asked to point to their starting location. A compass was used to record the angle of each participant's pointing direction—measured as the angle of the arc formed between magnetic north and the direction that the participant was pointing toward. This process was repeated for the second pointing test. During the first pointing test, participants were still relatively close to the starting point of the study. In the case of the second pointing test, participants had generally spent more time walking the neighborhood and they were farther away from their origin.

The key metric for the pointing tests is participants' pointing error. The pointing error was calculated by subtracting the angle measure of a participant's pointing direction from the ground truth angle of the starting location. The ground truth angles associated with the two pointing tests were calculated using GPS coordinates and were compared against the researchers' compass measurements for accuracy.

Table 4.2 *Address Verification Scenarios for the VR/Field Experiment.*

Scenario	**Map Edit Required**	**Count**
Residence is correctly reflected on map	No change to the map (spot) needed	0
Residence is new or missing (not on map)	Map spot added to map	0
Residence no longer exists, but is on map	Map spot removed from map	1
Residence is incorrectly placed on the map	Map spot moved to correct location	5

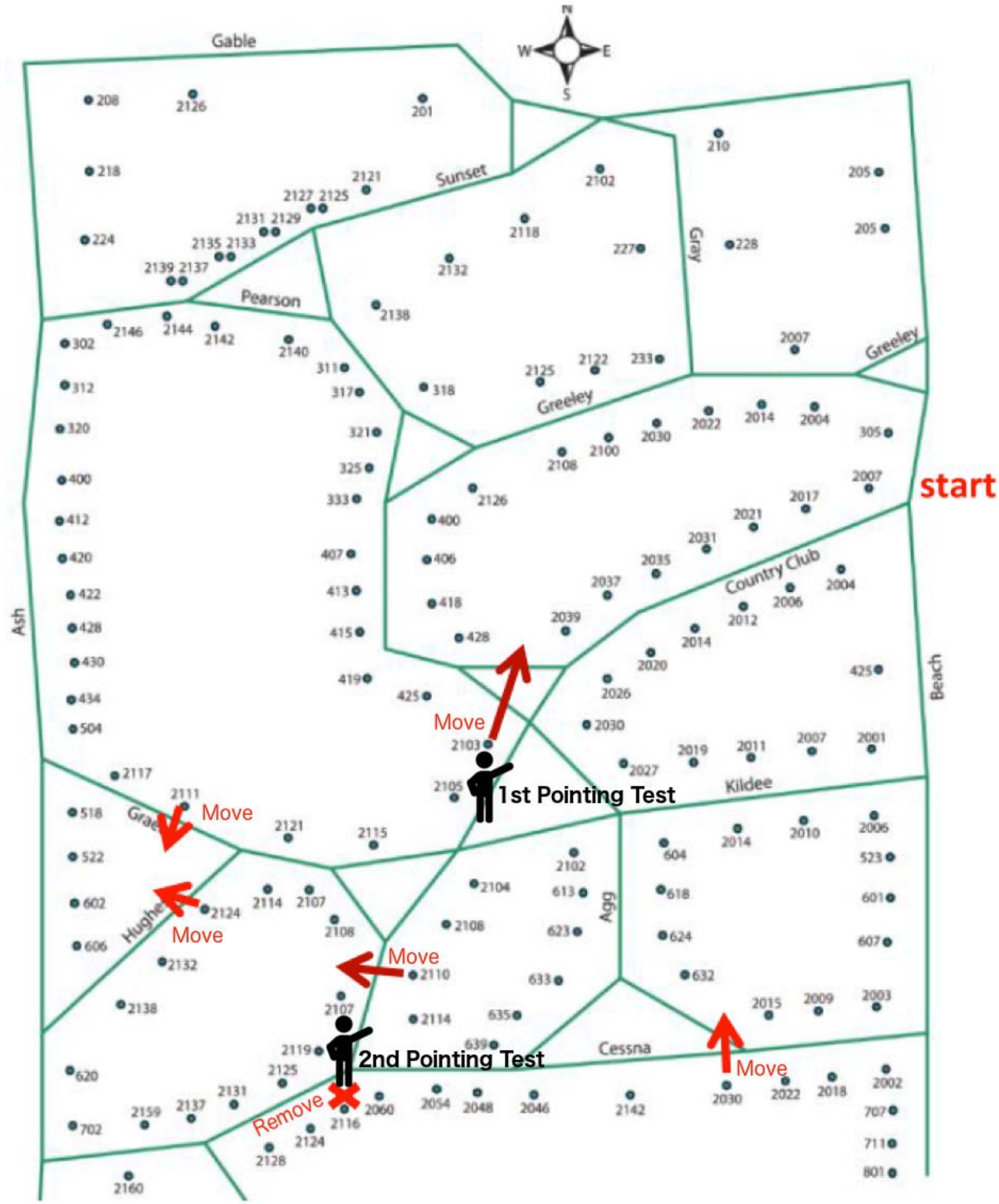


Figure 4.2 *Diagram showing expected map corrections and location of pointing tests in the study neighborhood*.

4.3.4 The Mobile Device and Study Software

The mobile device was set up to resemble prototype systems that were being tested by the U.S. Census Bureau at the time. Participants used a Pharos Traveler 535x paired with a stylus to operate the study software (see *Figure 4.3*). The software and digital map provided the core functions necessary to verify the residences. Participants could manipulate the digital map using pan and zoom controls. If the location of an assigned address needed to be updated on the map, participants could edit its corresponding “map spot”. Map spots are the dots on the map with address numbers that correspond to residences.

Using the software, a participant could select one of the six residences to be verified from a dropdown list. Once an address was selected, it became active and the participant could add or remove its associated map spot at any time. Once an address was selected and made active, the software was able to log all operations associated with the address for later analysis; these include: Taps on the screen, the addition and removal of map spots, and button presses. The ‘Reset map’ button undid all zooms and pans, returning the map to its original state at the beginning of the study. The ‘Reset question’ button undid any changes made to the active address’s map spot. The ‘Submit’ button saved the participant’s changes to the map (if any were made) for the selected address and the software flagged the address as verified.

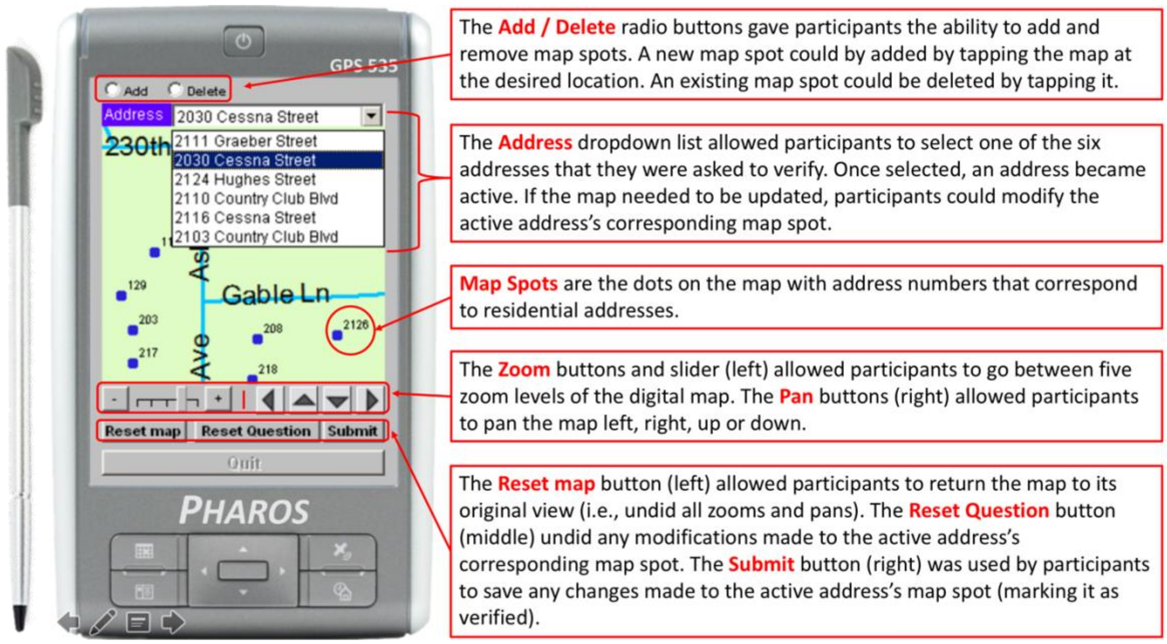


Figure 4.3 *Diagram of mobile device and software functions used for address verification*.

The digital map of the study neighborhood was generated by combining TIGER\Line Shapefiles using ESRI's ArcGIS Desktop© software (see *Figure 4.3*). The digital map included two layers of information: 1) a street layer that shows the streets and their respective labels and 2) an address layer that depicted each residential address as a map spot (i.e., a small dot) accompanied by an address number. The accuracy of residential addresses shown on the map was verified by researchers who walked the study neighborhood during review of the map. Graphics editing software was used to make final touches to the study map before it was segmented and integrated into the study application. The map did not incorporate GPS or other map aids that utilize the device's radios or onboard sensors.

4.3.5 Real-World Treatment: A Nearby Neighborhood

Core to the task of address verification is one's ability to interpret a map and locate addresses in a residential area. The study location was an important consideration because environmental cues and landmarks are essential to the underlying decision-making processes.

Google Maps™ was used to identify several potential neighborhoods for the study. Close attention was paid to the street layouts, the distribution and variety of residences, as well as proximity to the university. We preferred a neighborhood that would challenge participants in terms of wayfinding, navigation, and the address verification task itself.

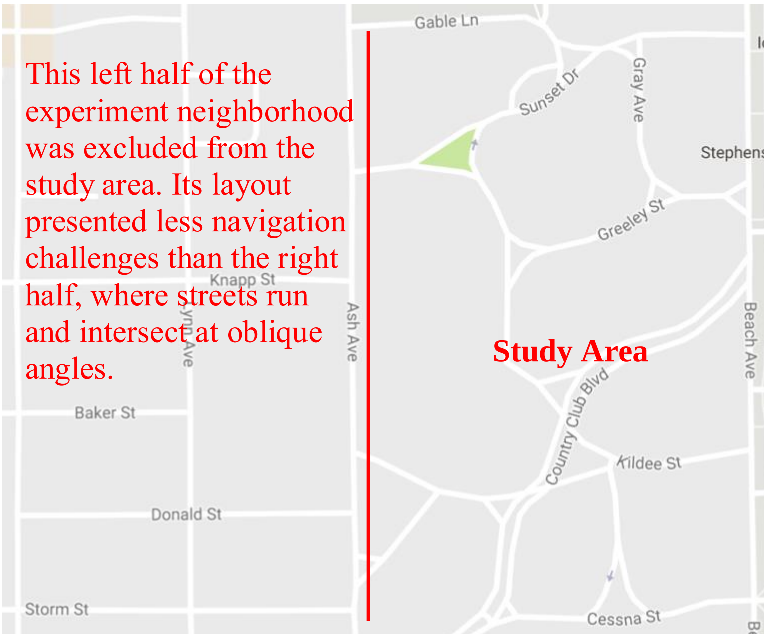


Figure 4.4 *Map showing neighborhood study area (right); it is not orthogonal like left half.*

A neighborhood was selected nearby campus and the study area was confined to a particularly demanding cross section that spanned approximately 1.3 square miles. It was comprised of many non-uniform streets that seldom ran parallel to the cardinal directions. Many of the residential streets formed three-way intersections known as Y-junctions, which can be more challenging to interpret than the more typical T-junction intersection. The streets often ran and intersected at odd angles that were not orthogonal like other residential areas—this added difficulty in participants' understanding of their location, bearing and direction of travel. There was variety in the types of residences (e.g., homes, duplexes, multi-unit buildings) and landmarks (e.g., a treehouse, playground, and large stadium). The mix and spread of occluding landmarks such as trees and fences also resulted in spotty line of sight from various vantage points. Overall, these features presented participants with notable difficulty in making task-related decisions and acquiring spatial knowledge.

4.3.6 Virtual Reality Treatment: A Neighborhood 3D Replica

A significant challenge in executing the experiment was creating the virtual reality treatment (VRT). It is an interactive, 3D replica of the real-world study neighborhood that is discussed in the previous section. The VRT was comprised of three key technical components: (1) the CAVE hardware interface, (2) the 3D model of the neighborhood, and (3) the software application that tied it all together.

CAVE Hardware Interface

The C6, located at Iowa State University's Virtual Reality Applications Center (VRAC), was the lab space in which the virtual reality treatment was administered. The C6 is a Cave Automatic Virtual Environment (CAVE), an integration of hardware and software that enabled users to interact with the immersive virtual environment (IVE) in a room that spanned 1000 cubic feet (10x10x10 ft.). A cluster of computers generated stereoscopic images—left eye and right eye—that were projected onto the four walls, ceiling, and floor of the space. Participants wore a pair of shutter glasses that wirelessly synced with the C6 projectors so that each eye received the correct stereoscopic image; this allowed participants to perceive a 3-dimensional virtual environment. The C6 was outfitted with a wireless, optical tracking system so that a participant's location within the space could be determined based on sensors situated on the shutter glasses. The tracking system enabled a primary user—the participant—to trigger the walking interface (i.e., stepping farther away from the center of the space accelerated one's virtual walking speed). The tracking system also outputted the participant's location in the physical space so that the projected, stereoscopic images would show the proper perspective of the virtual neighborhood. Altogether, the CAVE implementation afforded several key interactions that were not feasible with alternative VR systems such as those based on head-mounted displays or other display configurations.

- Participants could simultaneously see and interact with the mobile device, the digital map, the virtual neighborhood, and attending researchers in a shared, immersive virtual environment.
- Participants could use both hands to hold the mobile device and manipulate the digital map and address verification software because the interface for locomotion enabled hands-free travel through the virtual neighborhood.
- Attending researchers could directly observe participants' interactions with the mobile device and virtual environment. They could also see and talk to each other and interact with the participant when necessary.

Figure 4.5 *Photo of researcher in the VR lab standing at a street intersection within the virtual neighborhood.*

3D Model of the Study Neighborhood

A significant challenge in readying the virtual reality treatment (VRT) was designing a sufficiently high-fidelity 3D replica of the study neighborhood, which spanned approximately 1.3 square miles. Several months were spent designing the 3D model from scratch. SketchUp™ was selected as the primary 3D modeling application for its strengths in rapid architectural modeling, the creation of large-scale 3D environments, its support for the geo-location of models, its education-friendly licensing, and its large repository of community-shared 3D assets.

Prior to designing the 3D model, the team spent significant time walking the real-world neighborhood, taking photographs, and creating maps of the streets and residences. Effective planning and iterative design were critical in recreating a virtual environment of this size. There were multiple iterations of the VE design. We initially focused on the most important urban design elements such as the paths, edges, nodes, and landmarks (Lynch, 1960) of the neighborhood and surrounding area. We started the design by creating the terrain, the network of streets, and by placing pre-existing 3D models of houses, buildings, and other structures in their corresponding locations. To do this efficiently and accurately, we overlaid satellite imagery of the neighborhood in SketchUp™ to draw and extrude the terrain and streets. We placed homes based on their locations as indicated by the satellite imagery (see *Figure 4.6*)—this also ensured that the key features of the neighborhood were accurately geo-located.

Figure 4.6 *Satellite imagery aided w/ modeling and geo-locating neighborhood features.*

During a walk-through of an earlier iteration of the VR neighborhood, members of the research team who frequently walked the real neighborhood noticed incongruences. For example, they could see across much greater distances because objects like trees and fences, which limit one's line of sight, were not present at the time. They recognized that memorable landmarks were also missing such as several large university buildings that were viewable from a distance. Other structures that were prominent while walking in the real-world neighborhood were also missing (e.g., a playground that was centrally located). Overall, the early VE felt barren in comparison to the real thing. To get the best results out of the VRT, such features were added to reflect the real-world as closely as possible. We frequently referenced Google Street View™ during the design process to add fine touches to the model for realism (see *Figure 4.7*).

Figure 4.7 *Panoramas of the real-world (left) were referenced to fine tune the model (right).*

Over dozens of iterations, prominent buildings and landmarks were added to the 3D model as well as hundreds of trees, traffic signs, fire hydrants, and we even went so far as to create a realistic sky with the sun positioned to the west to approximate the time of day—at least one participant used this as a directional queue during the study. Altogether, a balanced combination of geo-located residences, core urban design features and embellishments made the VE feel more authentic.

Design time and computational power were constraints in reproducing certain neighborhood landmarks and features in sufficient detail. The VE also lacked auditory feedback such as an outdoor soundscape. Additionally, the activity that one would expect while walking a neighborhood (e.g., traffic, passersby, etc.) was absent.

VR Application Software

The software for the VRT was written in C++. It served multiple functions, including interfacing with the CAVE hardware, loading the 3D model of the neighborhood, facilitating locomotion within the VE, and logging participants' performance metrics (i.e., time, distance traveled, and routes). The software application was built using VR Juggler, a cross-platform, open source virtual reality application development framework (Cruz-Neira, Bierbaum, Hartling, Just, & Meinert, 2002). The 3D model was handled using OpenSceneGraph, an open source 3D graphics toolkit and API. We implemented a hands-free locomotion interface that enabled users to move through the virtual neighborhood with relative ease while simultaneously being able to interact with the mobile device and digital map. The locomotion interface is described in *Figure 4.8*. The top speed of locomotion was restricted to approximate the maximum walking speeds that were encountered in the real-world.

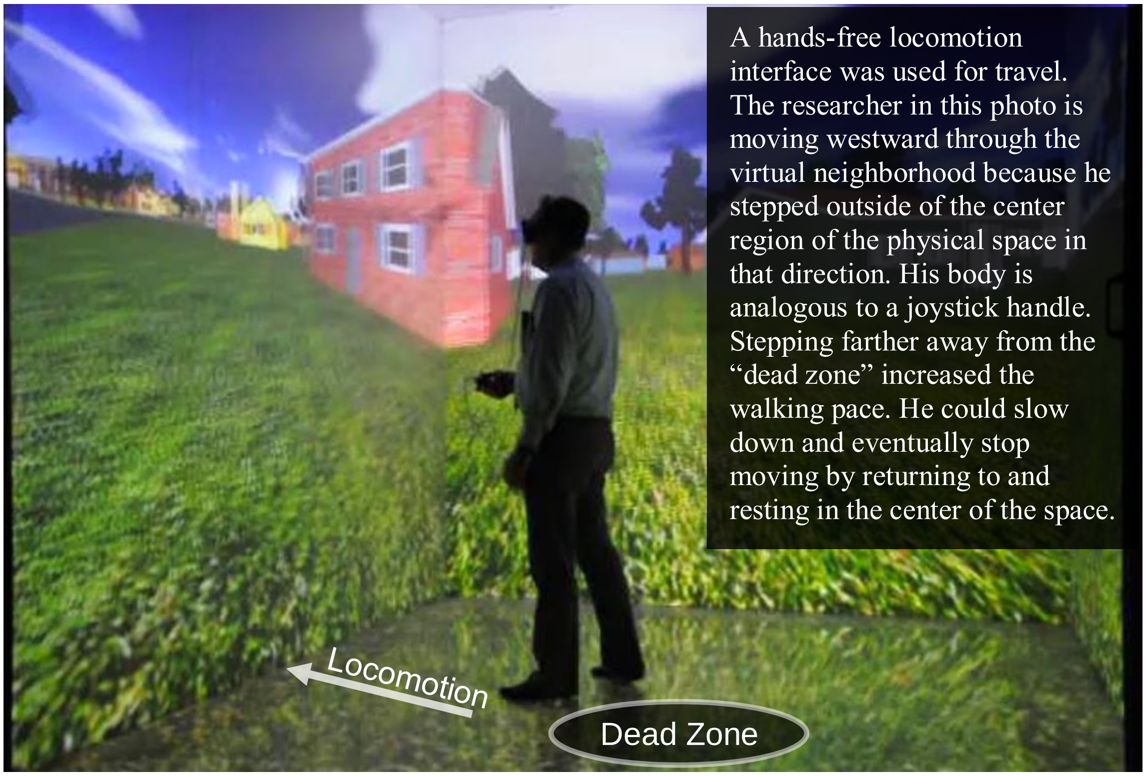


Figure 4.8 *Researcher uses the locomotion interface to traverse the virtual neighborhood.*

4.3.7 Procedure

Prior to the address verification exercise, participants were randomized to either the real-world treatment (RWT) or the virtual reality treatment (VRT).

For the RWT, two researchers accompanied each participant to the study neighborhood. The task was explained along with the think aloud protocol. Participants were told that they should travel to each of the assigned residential addresses to determine whether their digital map accurately reflected what was seen on the grounds of the neighborhood. An assigned address needed to either be added to the map, removed, moved to a new location, or confirmed to be correct in its existing location (i.e., no map change). Participants were told to only correct the addresses that they were assigned, which showed up in a drop-down list, and to ignore other potential errors on the map. Participants were then taught how to edit the map spots associated with residences on the digital map and were given instruction on the think aloud protocol and the verbalization of their thoughts.

The map software was started in training mode and participants were asked to locate and verify three training addresses in the immediate vicinity. Observers answered procedural questions and provided feedback on the quality of the think aloud. At the end of the training, the participant and observers reviewed the outcomes for the three training addresses to ensure that the procedure and objectives were clear. Observers answered any final questions the participants had and explained that they would refrain from talking during the exercise, other than to prompt the participant to keep verbalizing or to ask follow-up questions on behaviors or actions.

Observers then returned the participant to the starting location, switched the map software to experiment mode, and started an audio recorder worn by the participant. A GPS-enabled tracking app was started and carried by one of the observers to record the routes taken by the participant. During the exercise, each participant was shadowed by two researchers who carried coding sheets to record their observations. After participants verified half of the assigned addresses, they were taken to a uniform location and asked to point to the starting location of the study while an observer recorded the pointing angle. The pointing error was later measured as the angle (in degrees) formed by the arc between participants' pointing direction and the actual direction of the origin. The main exercise was completed after participants verified all six of the assigned residences. After verifying the sixth address, each participant was asked a second time to point to their origin, this time from a different location and the pointing angle was again recorded. Participants answered an exit questionnaire on the task (see *Appendix C*) at the end of the exercise, observers' written notes were filed, and participants' route data from a GPS logging app was saved.

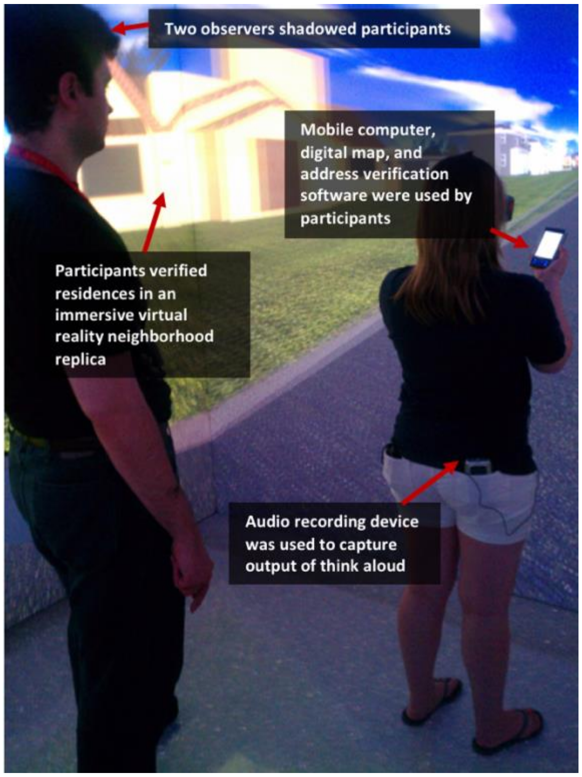


Figure 4.9 *Photo of VR treatment w/ observer (left) shadowing participant (right)*.

The procedure for the VRT was nearly identical to that of the RWT. Two researchers accompanied each participant to the C6 facility at Iowa State University's Virtual Reality Applications Center (VRAC). In addition to training participants on the task and use of the handheld device, they were also briefed on virtual reality and given instructions on use of the CAVE and the "virtual walking" interface. Participants were coached on the same three training addresses that were used for the RWT while observers ensured that the procedure and objectives were clear. Before proceeding to the main exercise, observers also checked to see if participants experienced any sickness symptoms that stemmed from use of the VR system and locomotion method. For the primary exercise, observers used a remote control connected to the CAVE to log participants' task times, their elapsed verification times for the assigned residences, and the routes that they traveled. The shadowing of participants in the VRT by observers required much less walking, with observers diligently maintaining a position behind the participant so as not to distract them from the task. The procedures described for the RWT were followed in terms of completion of the primary exercise, the pointing task, and the disposition of study data.

4.4 Experimental Design

A between-subjects design was used to test the hypotheses. We examined the effects of the environment condition (ENV) combined with participants' spatial visualization ability grouping (VZ) on task outcomes. To compare a field task occurring in a real-world setting to one simulated in an immersive virtual environment, participants were randomly assigned to two treatments of the environment factor: (1) The real-world treatment (RWT) and (2) the virtual reality treatment (VRT). Prior to being assigned to either of the two environments, participants were screened based on their scores on psychometric tests of spatial visualization ability. Stratified random sampling was used, where participants whose VZ scores fell within predefined thresholds for high- and low-scorers were assigned to either a high spatial visualization group (HighVZ) or a low spatial visualization group (LowVZ). Pairs were then taken from the HighVZ and LowVZ groups and were randomized together to give participants from each group an equal chance of being assigned to either of the two treatments.

Comparisons were first made based on the ENV factor alone—irrespective of participants' spatial visualization groupings. For these comparisons, the HighVZ and LowVZ participant groups were merged into a single group of participants for each treatment environment. This yielded two experimental groups based on the treatment environments:

1. **RWT Group** – the combination of high and low spatial visualization participant groups who completed tasks in the real-world neighborhood (i.e., RWT+HighVZ combined with RWT+LowVZ).
2. **VRT Group** – the combination of high and low spatial visualization participant groups who completed tasks in the virtual reality neighborhood replica (i.e., VRT+HighVZ combined with VRT+LowVZ).

Further comparisons consider both the ENV factor and the participants groups in combination, which resulted in the following four experimental groups:

3. **RWT+HighVZ** – participants of the high spatial visualization group who completed the field tasks in the real-world neighborhood.
4. **RWT+LowVZ** – participants of the low spatial visualization group who completed the field tasks in the real-world neighborhood.
5. **VRT+HighVZ** – participants of the high spatial visualization group who completed the field tasks in the virtual reality neighborhood replica
6. **VRT+LowVZ** – participants of the low spatial visualization group who completed the field tasks in the virtual reality neighborhood replica.

4.5 Variables Used for Analyses

This section covers the variables stemming from the categories that were used in the analyses. Three categories of dependent variables were captured in both study treatments: (1) task performance metrics, (2) digital map operations, and (3) recorded observations via the attending researchers' notes.

4.5.1 Task Performance Variables

The following task performance metrics were captured in both treatments for later analysis:

Address Verification Task Variables

1) *Task time* – Participants' task completion time for the address verification exercise.
2) *Distance traveled* – A GPS device was used to determine the distance that participants traveled in the real-world neighborhood. The immersive virtual environment (IVE) was georeferenced so that a subroutine could calculate and output participants' travel distances in the virtual neighborhood.
3) *Address verification errors* – Participants' total number of address verification errors were determined using both the study software logs and the observer notes pertaining to each verification decision.

Pointing Task Variables

4) *1st point-to-origin test error* – Participants were asked to point to their starting location after they verified half of the assigned addresses. The error angles, in degrees, were recorded.
5) *2nd point-to-origin test error* – Participants were asked a second time to point to the starting location after verifying their last address. The error angles, in degrees, were recorded.

4.5.2 Digital Map Operation Variables

Participants used the study software to manipulate the digital map and to edit map spots that corresponded to the addresses being verified. The digital map operations were saved to log files by the software; they are as follows:

1) *Pans* – Participants had the ability to pan the map up, down, left, or right

2) *Zooms* – Participants could switch between five different zoom levels.
3) *Pan limit reached* – During panning, if participants reached the bounds of the map and continued to pan in the bounded direction, the software displayed a notification.
4) *Zoom limit reached* – During use of the map zoom feature, participants were notified when they attempted to go beyond the minimum or maximum zoom levels that were available.
5) *Address added* – If a participant reached the physical location of an assigned address and concluded that the address was not present on the map, the instruction was to correct the map by adding the corresponding map spot in the proper location.
6) *Address removed* – If a participant traveled to the location of an assigned address as indicated by the map and he or she concluded that the address no longer existed, the instruction was to correct the map by removing the corresponding map spot.
7) *Map resets* – Participants could reset the map view to its original state at the onset of the address verification exercise—effectively undoing all pan and zoom operations.

4.5.3 Recorded Observations

Two observers were present while participants completed the field tasks. The observers used coding sheets to record participants' behaviors and any relevant think aloud responses. These written materials were compiled and the observations were analyzed, categorized, and quantified. A total of 113 variables were created for the categories that were identified from the coded observations (see *Appendix D*).

4.6 Hypotheses

We expect that participants of the real-world treatment (RWT) will outperform their virtual reality treatment (VRT) counterparts due to the advantages of having an authentic task environment. Three sets of hypotheses were generated to test for differences in task performance between the RWT group and the VRT group:

H1) We hypothesized that participants of the RWT group would outperform those of the VRT group per the five task performance variables (A – E), whereby:

A. Lower *task time* is better.
B. Shorter *distance traveled* is better.
C. Fewer *address verification* errors are better.
D. Smaller pointing angles are better for the *1st point-to-origin test error*.
E. Smaller pointing angles are better for the *2nd point-to-origin test error*.

H2) For the high spatial visualization group, we hypothesized that participants of the real-world treatment (RWT+HighVZ) would outperform those of the virtual reality treatment (VRT+HighVZ) per the five task performance variables (A – E).

H3) Similarly, for the low spatial visualization group, we hypothesized that participants of the real-world treatment (RWT+LowVZ) would outperform those of the virtual reality treatment (VRT+LowVZ) per the five task performance variables (A – E).

In the first three sets of hypotheses, we expect significant differences in participants' task performance between the real-world treatment and the virtual reality treatment. In a similar fashion, we expect significant differences in task performance between the high spatial visualization (HighVZ) and low spatial visualization (LowVZ) participant groups. In our previous field studies on the task in a real neighborhood, participants who were screened into HighVZ and LowVZ groups used a paper map rather than a digital map to verify residences. The prior results suggest that in terms of overall task performance, individuals with high spatial visualization ability typically outperform those with low ability. In the present experiment, we expected comparable results in the real-world treatment environment and we anticipated that the virtual reality treatment environment, assuming that it has sufficient ecological validity, would yield results similar to the real-world treatment. Therefore, hypotheses were generated for each treatment environment to test for differences in task performance between HighVZ and LowVZ groups. These hypotheses are as follows:

H4) For the real-world treatment environment, we hypothesized that participants of the high spatial visualization group (RWT+HighVZ) would outperform those of the low spatial visualization group (RWT+LowVZ) per the five task performance variables (A – E).

H5) Similarly, for the virtual reality treatment environment, we hypothesized that participants of the high spatial visualization group (VRT+HighVZ) would also outperform those of the low spatial visualization group (VRT+LowVZ) per the five task performance variables (A – E).

To complement the task performance analyses between treatment groups, a set of hypotheses was generated to test for differences in participants' use of the digital map operations between treatments:

H6) We hypothesized that participants' digital map operations in the RWT would differ from those of the VRT group per the seven map operation variables (F – L):

F. Fewer *pans* tend to indicate better task performance.

G. Fewer *zooms* tend to indicate better task performance.

H. Fewer *pan limit reached* occurrences tend to indicate better task performance.

I. Fewer *zoom limit reached* occurrences tend to indicate better task performance.

J. Greater *address added* occurrences tend to indicate better task performance.

K. Greater *address removed* occurrences tend to indicate better task performance.

L. Fewer *map resets* tend to indicate better task performance.

H7) For the high spatial visualization group, we hypothesized that participants' digital map operations in the real-world treatment (RWT+HighVZ) would differ from those of the virtual reality treatment (VRT+HighVZ) per the map operation variables (F, G, J, K). Note: Three map operation variables were excluded (H, I, L) because they lacked the necessary frequencies to test the hypotheses against crosses of the participant groups and treatment environments.

H8) Similarly, for the low spatial visualization group, we hypothesized that participants' digital map operations in the real-world treatment (RWT+LowVZ) would differ from those of the virtual reality treatment (VRT+LowVZ) per the map operation variables (F, G, J, K).

Sets of hypotheses were also generated for each treatment environment to test for differences in participants' use of the digital map operations between the high and low spatial visualization groups. These hypotheses are as follows:

H9) For the real-world treatment environment, we hypothesized that high spatial visualization group participants (RWT+HighVZ) would differ in their digital map operations from the low spatial visualization group (RWT+LowVZ) per the map operation variables (F, G, J, K).

H10) Similarly, for the virtual reality treatment environment, we hypothesized that high spatial visualization group participants (VRT+HighVZ) would differ in their digital map operations from the low spatial visualization group (VRT+LowVZ) per the map operation variables (F, G, J, K).

4.7 Results

The results of the quantitative analyses are presented in this section. The impact of these results on the hypotheses will be elaborated on in the *Discussion Section*.

4.7.1 Task Performance Comparisons Between Treatment Environments

Descriptive statistics for the task performance variables pertaining to the whole of participants in each treatment environment are shown in *Table 4.3*.

Table 4.3 *Descriptive Statistics for the Task Performance Variables by Treatment.*

	Real World				**Virtual Reality**			
Performance Variable	***n***	**Mdn**	**Mean**	**SD**	***n***	**Mdn**	**Mean**	**SD**
Task time (min.)	16	30.5	36.6	15.0	14	41.5	37.1	9.1
Distance traveled (mi.)	15	1.1	1.3	.6	15	1.3	1.3	.3
Address verification errors	16	2.5	3.1	2.0	14	2.5	2.5	1.3
1st point-to-origin error (deg.)	14	11.5	14.7	14.0	14	19.5	37.6	54.5
2nd point-to-origin error (deg.)	16	14.0	17.3	16.0	14	23.5	35.4	43.8

A series of analyses were performed to determine if there were significant differences in task performance between participants of the real-world treatment (RWT) group and those of the virtual reality treatment (VRT) group. Inspection of the histograms and Q-Q Plots for *task time* and *distance traveled* indicated that the distributions were not approximately normal for both the RWT and VRT (see *Table 4.4*), thus, a Mann-Whitney *U* test (Mann &

Whitney, 1947) was selected for the comparisons. The distribution of *address verification errors* appeared to be approximately normal, so the Shapiro-Wilk test (Shaphiro & Wilk, 1965) was used to further confirm the assumption of normality. The distributions of the *1st point-to-origin test error* and *2nd point-to-origin test error* were skewed to the right, so a log transformation was applied to normalize both variables and the Shapiro-Wilk test of normality was used to confirm the normality assumption for the transformed data. The Shapiro-Wilk test results are shown in *Table 4.4*.

Table 4.4 *Results of Shapiro-Wilk Test of Normality for Performance Variables by Treatment.*

	Real World			Virtual Reality		
Performance Variable	***n***	***W***	***p***	***N***	***W***	***p***
Task time	16	.799	.003*	14	.856	.027*
Distance traveled	15	.715	.000*	15	.922	.207
Address verification errors	16	.915	.139	14	.906	.140
Log(1st point-to-origin test error)	14	.943	.464	14	.953	.603
Log(2nd point-to-origin test error)	16	.964	.729	14	.963	.777

* $p < .05$; rejected H_0: Distribution is normal for specified variable and treatment group.

Levene's test for equality of variances (Levene, 1960) was used to confirm the assumption of homogeneity of variance for *address verification errors*, the *1st point-to-origin test error* and the *2nd point-to-origin test error* (see *Table 4.5*). For *address verification errors* and the *1st point-to-origin test error*, we rejected the null hypothesis of equality of variances, thus, we selected a Welch's *t*-test to compare means between treatments. We failed to reject the null hypothesis for the *2nd point-to-origin test error*, in which case an independent-samples *t*-test was chosen to compare means.

Table 4.5 *Results of Levene's Test for Equality of Variances for Task Performance Variables.*

Performance Variable	Test Statistics		
	F	**d.f.**	***p***
Address verification errors	4.42	28	.045*
Log(1st point-to-origin test error)	4.91	26	.036*
Log(2nd point-to-origin test error)	1.12	28	.298

* $p < .05$; rejected H_0: Treatments have equal variances for specified variable.

We expected the RWT group participants to have shorter *task times* than those of the VRT group (shorter time is better). The parallel boxplot medians (see *Figure 4.10*) indicate that participants of the RWT group did take less time on average to verify addresses (Mdn = 30.5) than did those of the VRT group (Mdn = 41.5). There was no significant difference in the distributions of participants' task times between the RWT group and the VRT group (Mann–Whitney $U = 96.0$, $n_{rw} = 16$, $n_{vr} = 14$, $p = .522$). The boxplot interquartile ranges are similar between the RWT (IQR = 17.7) and the VRT (IQR = 17.0). The boxplots differ in that the RWT task times are skewed to the right, have a greater spread and an outlier beyond the 70-minute mark; whereas task times for the VRT are left-skewed with a tighter spread and no outliers present.

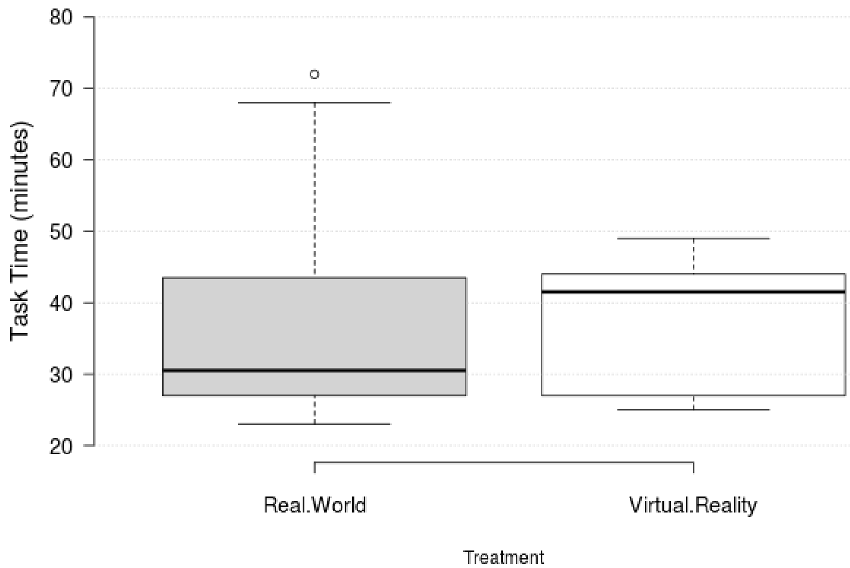


Figure 4.10 *Parallel boxplots of task time by treatment*.

We expected the RWT group participants to travel shorter distances than those of the VRT group in order to complete the task (shorter distance is better). The parallel boxplot medians for *distance traveled* (see *Figure 4.11*) indicate that participants of the RWT group did travel shorter distances on average (Mdn = 1.1) than did those of the VRT group (Mdn = 1.3). However, there was no significant difference in the distributions of participants' travel distances between the RWT group and the VRT group (Mann–Whitney $U = 83.0$, $n_{rw} = 15$, $n_{vr} = 15$, $p = .230$). The interquartile ranges of the boxplots are similar between the RWT (IQR = .3) and the VRT (IQR = .4). The boxplots differ in that the RWT travel distances are skewed to the right and have two outliers beyond the 2.5-mile mark; whereas travel distances for the VRT are closer to the center with a single outlier short of the 2-mile mark. The presence of more extreme outliers for the RWT may indicate that participants who struggled with navigation in the real-world neighborhood were penalized to a greater degree than those who struggled in the VRT.

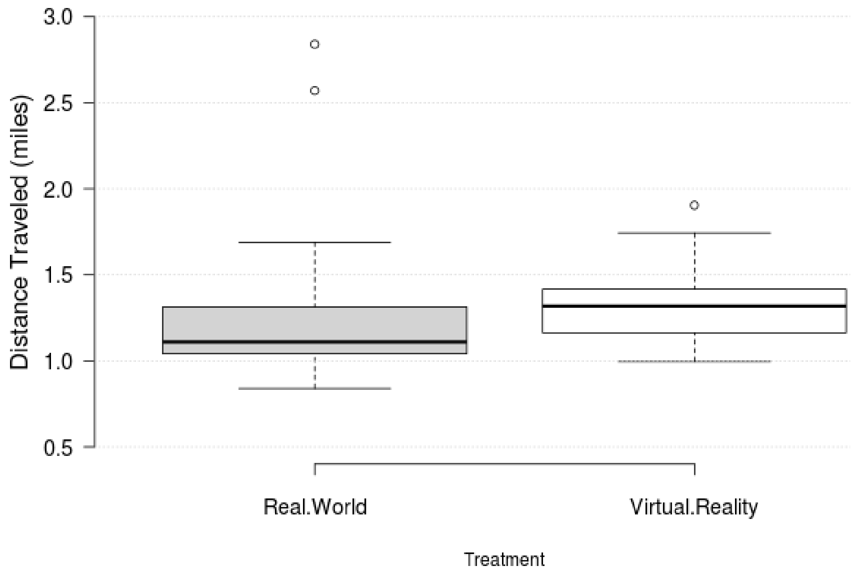


Figure 4.11 *Parallel boxplots of distance traveled by treatment*.

We expected the RWT group participants to make fewer *address verification errors* than those of the VRT group (fewer errors is better). The parallel boxplot medians (see *Figure 4.12*) indicate that participants of the RWT group unexpectedly made more verification errors (M = 3.1, SD = 2.0) than did those of the VRT group (M = 2.5, SD = 1.3). A Welch's *t*-test was calculated and it was determined that there was no significant difference in verification errors between treatments ($t(26) = .933$, $n_{rw} = 16$, $n_{vr} = 14$, $p = .360$). The address verification errors shown in the RWT boxplot have a greater interquartile range (IQR = 3.0) and spread when compared to the VRT (IQR = 2.0)—bear in mind that these boxplots can appear misleading considering the narrow range of potential verification errors, which was from 0 to 6.

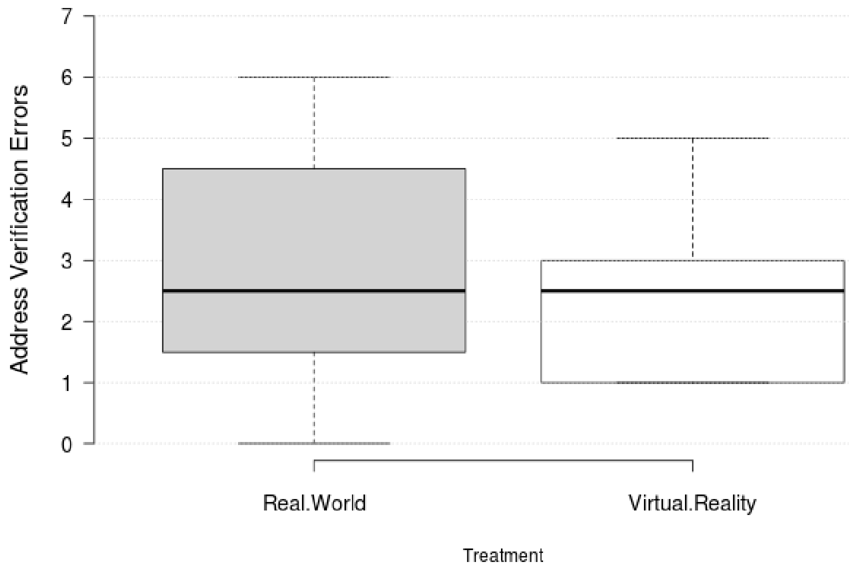


Figure 4.12 *Parallel boxplots of address verification errors by treatment.*

Participants were asked to point to their starting location midway through the field exercise and the error angle was measured in degrees (*1st point-to-origin test error)*. We expected the RWT group participants to make less error on the 1st point-to-origin test than those of the VRT group (smaller error angles are better). The parallel boxplot medians (see *Figure 4.13*) indicate that participants of the RWT group did err less pointing to their starting location (M = 14.7˚, SD = 14.0˚) than did those of the VRT group (M = 37.6˚, SD = 54.5˚). A Welch's *t*-test was calculated to compare the pointing error angles between the treatment groups. Prior to the *t*-test, a logarithmic transformation was applied to the error data for both treatments to normalize their distributions. There was no significant difference in pointing errors between treatments ($t(21.6) = .696$, $n_{rw} = 14$, $n_{vr} = 14$, $p = .503$). The boxplots do reveal noticeable differences in the interquartile ranges of the RWT (IQR = 9.0) and VRT (IQR = 27.7). Another noticeable difference can be seen in the outliers of the two treatments, where the VRT has extreme outliers in comparison to the RWT; these outliers indicate that participants from the VRT who struggled the most pointing to their origin were almost completely turned around, with pointing error angles that exceeded 140˚. The presence of these extreme outliers might imply that difficulty with spatial orientation and path integration among participants of the VRT was exacerbated by the simulated environment and the walking interface of the virtual reality lab.

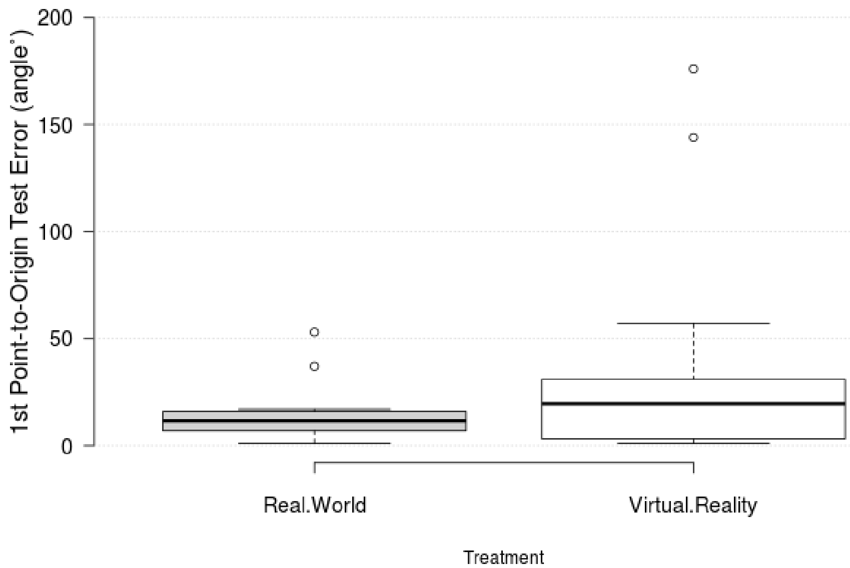


Figure 4.13 *Parallel boxplots of 1st point-to-origin test error by treatment.*

Participants were asked to point to their starting location a second time at the end of the field exercise and the error angle was measured in degrees (*2nd point-to-origin test error)*. We expected the RWT group participants to make fewer errors on the 2nd point-to-origin test than those of the VRT group (smaller error angles are better). The parallel boxplot medians (see *Figure 4.14*) indicate that participants of the RWT group did err less pointing to their starting location (M = 17.3˚, SD = 16.0˚) than did those of the VRT group (M = 35.4˚, SD = 43.8˚). An independent-samples *t*-test was calculated to compare the pointing error angles between the treatment groups. Prior to the *t*-test, a logarithmic transformation was applied to the error data for both treatments to normalize their distributions. There was a significant difference in pointing errors between treatments for the 2nd point-to-origin test ($t(28) = 2.05$,

$n_{rw} = 16$, $n_{vr} = 14$, $p = .050$). Unlike the previous pointing test, the interquartile ranges were similar between the RWT and the VRT. There is also an extreme outlier in the VRT, which is similar to outliers seen in the VRT for the 1st point-to-origin test, where outlying participants errantly thought that the bearing of the origin was in the opposite direction of its true location.

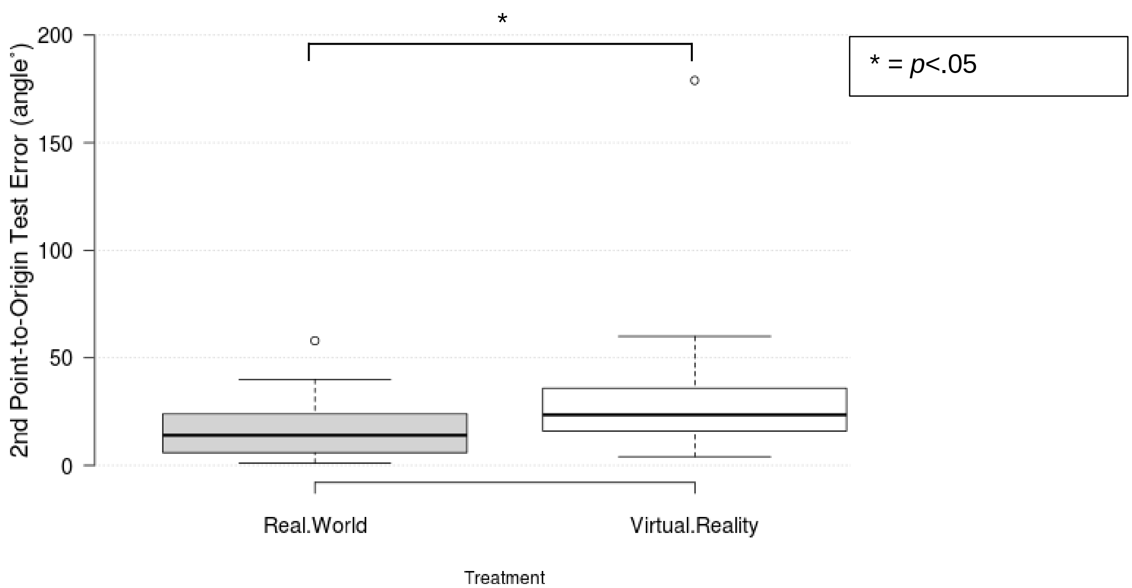


Figure 4.14 *Parallel boxplots of 2nd point-to-origin test error by treatment*.

4.7.2 Task Performance Comparisons Between Crosses of the Treatment Environments and Participant Groups

The following analyses involve comparisons of the task performance variables that are associated with the four experimental groups that are crosses of the treatment environments (RWT/VRT) and the participant groups (HighVZ/LowVZ). We examined parallel box plots, key statistics and statistical tests. A Mann-Whitney *U* test was chosen for statistical comparisons due to its robustness in situations when the normality of a sample cannot be assumed. The associated descriptive statistics are shown in *Table 4.6*.

Table 4.6 *Descriptive Statistics for the Task Performance Variables by Treatment and VZ Group.*

	Performance Variable	HighVZ Group *n*	HighVZ Group Mdn	HighVZ Group Mean	HighVZ Group SD	LowVZ Group *n*	LowVZ Group Mdn	LowVZ Group Mean	LowVZ Group SD
Real World	Task time (min.)	8	29.0	28.1	3.5	8	43.5	45.0	17.5
	Distance traveled (mi.)	7	1.0	1.0	.15	8	1.3	1.6	.7
	Address verification errors	8	2.0	2.1	1.4	8	4.5	4.0	2.1
	1st point-to-origin test error (deg.)	8	11.5	14.6	9.8	7	10.0	14.1	17.7
	2nd point-to-origin test error (deg.)	8	10.5	12.4	8.7	8	15.5	22.3	20.4
Virtual Reality		***n***	**Mdn**	**Mean**	**SD**	***n***	**Mdn**	**Mean**	**SD**
	Task time (min.)	7	27.0	31.7	9.1	9	43.0	44.1	7.4
	Distance traveled (mi.)	7	1.3	1.4	.3	9	1.3	1.2	.3
	Address verification errors	7	2.0	2.1	1.2	9	3.0	2.9	1.3
	1st point-to-origin test error (deg.)	7	15.0	32.0	50.8	9	12.0	36.0	55.3
	2nd point-to-origin test error (deg.)	7	17.0	39.0	62.2	9	29.0	42.1	38.8

When comparing like participant groups, we expected the RWT group participants to have shorter *task times* than those of the VRT group (shorter time is better). The parallel boxplot medians indicate that the RWT+HighVZ group unexpectedly had a slightly higher median task time of 29 minutes compared to the VRT+HighVZ group's median task time of 27 minutes. The distributions of the two treatments did not differ significantly (Mann–Whitney $U = 25.5$, $n_{rw} = 8$, $n_{vr} = 7$, $p = .818$). The RWT+LowVZ group also had a slightly higher median task time of 43.5 minutes compared to the VRT+LowVZ group's median task time of 43.0 minutes. The distributions of the two treatments did not differ significantly (Mann–Whitney $U = 35.0$, $n_{rw} = 8$, $n_{vr} = 9$, $p = .960$). The boxplots also show a striking difference in the spread and interquartile ranges of the RWT+LowVZ group (IQR = 32.5 min.) when compared to the VRT+LowVZ group (IQR = 5.5 min.).

We expected the HighVZ group to have shorter task completion times than the LowVZ group in both treatments. The parallel boxplot medians (see *Figure 4.15*) indicate that this was true in both cases. The RWT+HighVZ group had a median task time of 29.0 minutes compared to the 43.5-minute median task time of the RWT+LowVZ group; the distributions of the two groups differed significantly (Mann–Whitney $U = 11.5$, $n_{high} = 8$, $n_{low} = 8$, $p = .036$). The VRT+HighVZ group had a median task time of 27 minutes compared to the 43-minute median task time of the VRT+LowVZ group; the distributions of the two groups also differed significantly (Mann–Whitney $U = 8.5$, $n_{high} = 7$, $n_{low} = 9$, $p = .017$).

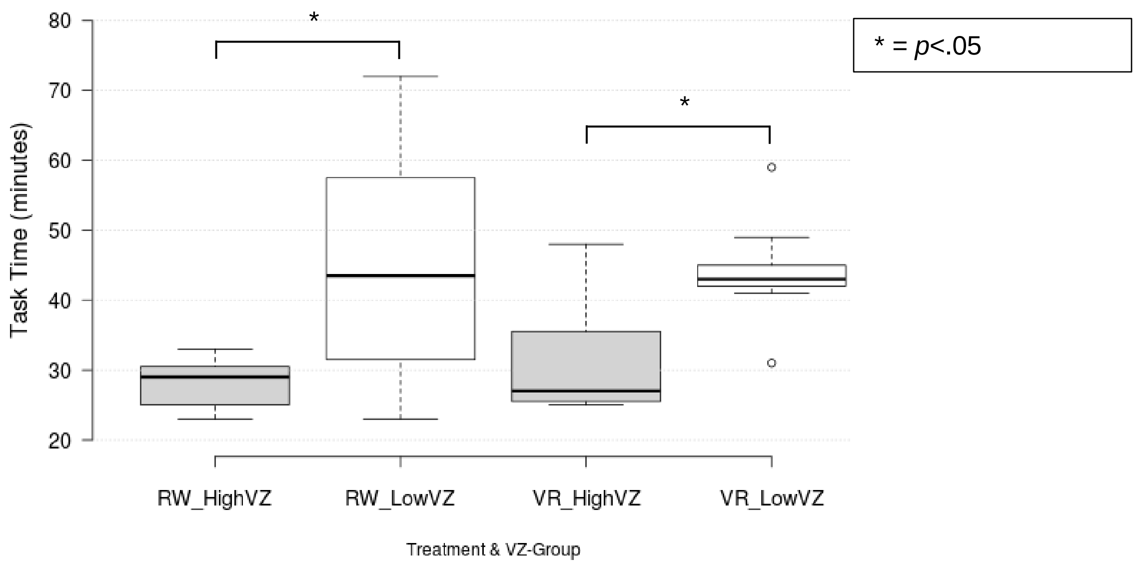


Figure 4.15 *Parallel boxplots of task time for treatments by VZ group.*

There is a pattern in the task time boxplots that will also be seen in the boxplots of other task performance variables where the treatment environments and spatial visualization groups have been crossed. Notice that the location of the HighVZ group boxplots (darker shade) tend to always be lower than those of the LowVZ group. In general, this trend suggests that regardless of the treatment environment, the HighVZ group tends to outperform the LowVZ group; this is because lower values can be associated with better performance for each of the task performance variables.

For *distance traveled*, we expected the RWT group participants to travel shorter distances than those of the VRT group in order to complete the task (shorter distance is better). The boxplot medians indicate that the RWT+HighVZ group had a shorter median travel distance of 1.02 miles compared to the VRT+HighVZ group's median travel distance of 1.33 miles; the distributions of the two treatments differed significantly (Mann–Whitney $U = 5.0$, $n_{\mathrm{rw}} = 7$, $n_{\mathrm{vr}} = 7$, $p = .015$). The RWT+LowVZ group unexpectedly had a slightly higher median travel distance of 1.29 miles compared to the VRT+LowVZ group's median travel distance of 1.28 miles. However, for the LowVZ group there was no significant difference between the two treatments (Mann–Whitney $U = 26.0$, $n_{\mathrm{rw}} = 8$, $n_{\mathrm{vr}} = 9$, $p = .363$). The spread and interquartile range of the RWT+LowVZ group (IQR = 1.24 mi.) is far greater than the VRT+LowVZ group (IQR = .41 mi.) or any other groups for that matter.

We expected the HighVZ group to travel shorter distances than the LowVZ group in both treatments. The parallel boxplot medians (see *Figure 4.16*) indicate that this was true for the RWT, however, for the VRT the medians were similar between the HighVZ and LowVZ groups. The RWT+HighVZ group had a median travel distance of 1.02 miles, which was shorter than the RWT+LowVZ group's median travel distance of 1.29 miles; the distributions of the two groups differed significantly (Mann–Whitney $U = 7.0$, $n_{\mathrm{high}} = 8$, $n_{\mathrm{low}} = 7$, $p = .018$). Notice that the VRT+LowVZ group's boxplot is not in a higher location than the VRT+HighVZ group's boxplot as was expected. The VRT+HighVZ group had a median travel distance of 1.33 miles, which was slightly farther than the

VRT+LowVZ group's median travel distance of 1.28 miles; the distributions of the two groups did not differ significantly (Mann–Whitney U = 22.0, n_{high} = 7, n_{low} = 9, p = .342).

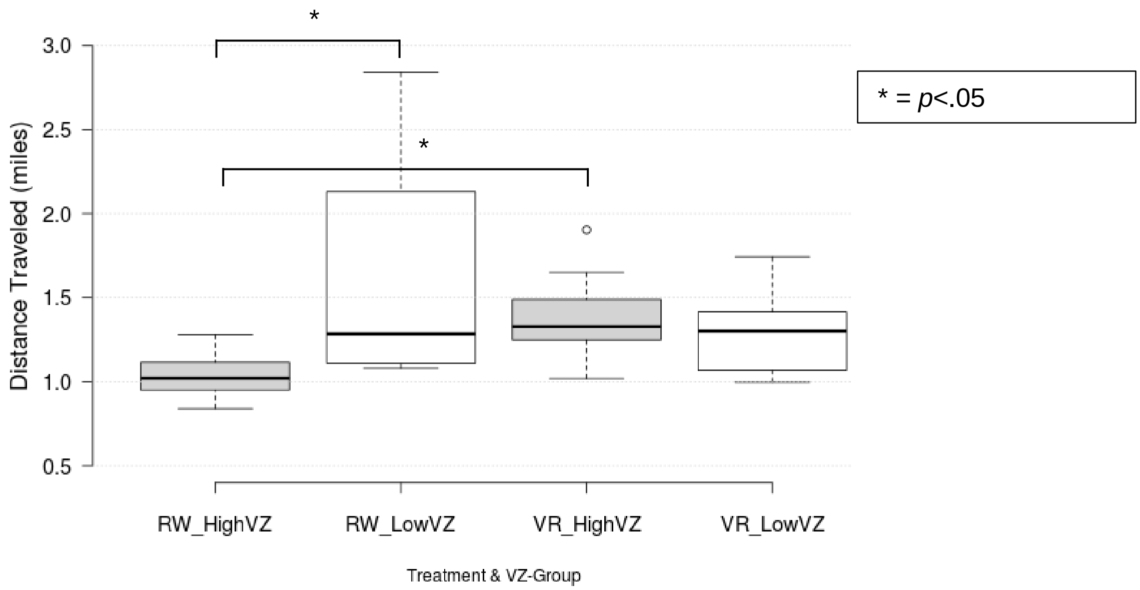


Figure 4.16 *Parallel boxplots of distance traveled for treatments by VZ group.*

We expected the RWT group participants to make fewer *address verification errors* than those of the VRT group (fewer errors is better). The boxplot medians indicate that the RWT+HighVZ group and the VRT+HighVZ group both had a median error frequency of 2.0 errors; the distributions of the two treatments did not differ significantly (Mann–Whitney U = 27.5, n_{rw} = 8, n_{vr} = 7, p = 1.00). The RWT+LowVZ group had a median error frequency of 4.5 errors, which was unexpectedly higher than the VRT+LowVZ group's median error frequency of 3.0 errors; there was no significant difference between treatments for the LowVZ group (Mann–Whitney U = 23.5, n_{rw} = 8, n_{vr} = 9, p = .250). Notice that the RWT+LowVZ group boxplot has a greater spread and interquartile range (IQR = 4.5) than the VRT+LowVZ group boxplot (IQR = 2.0) or those of any other experimental groups for that matter.

We expected the HighVZ group to make fewer address verification errors than the LowVZ group in both treatments. The boxplots exemplify the HighVZ participants' superior performance in both treatment environments. The RWT+HighVZ group had a median error frequency of 2.0 errors, which was lower than the RWT+LowVZ group's median error frequency of 4.5 errors (see *Figure 4.17*), however, the distributions of the two groups did not differ significantly (Mann–Whitney $U = 16.0$, $n_{\text{high}} = 8$, $n_{\text{low}} = 8$, $p = .103$). The VRT+HighVZ group had a median error frequency of 2.0 errors, which was lower than the VRT+LowVZ group's median error frequency of 3.0 errors, however, the distributions of the two groups did not differ significantly (Mann–Whitney $U = 21.0$, $n_{\text{high}} = 7$, $n_{\text{low}} = 9$, $p = .289$).

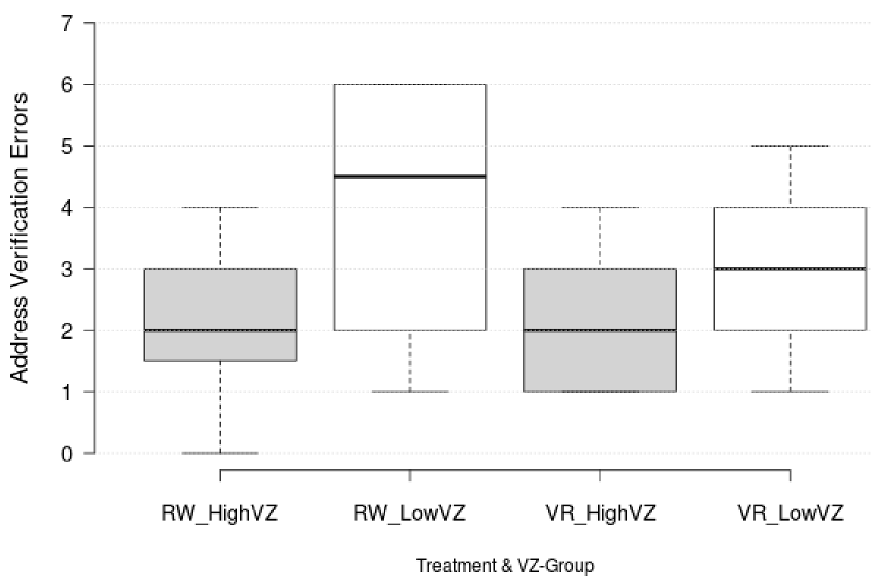


Figure 4.17 *Parallel boxplots of address verification errors for treatments by VZ group.*

We expected the RWT group participants to make fewer errors on the 1st point-to-origin test than those of the VRT group (smaller error angles are better). The parallel boxplot medians indicate that the RWT+HighVZ group had a smaller median pointing error angle of 11.5° when compared to the VRT+HighVZ group's median pointing error angle of 15.0°; the distributions of the two treatments did not differ significantly (Mann–Whitney U = 27.0, n_{rw} = 8, n_{vr} = 7, p = .952). The RWT+LowVZ group's median pointing error angle of 10.0° was also smaller than the VRT+LowVZ group's median pointing error angle of 12.0°; there was no significant difference between treatments for the LowVZ group (Mann–Whitney U = 22.0, n_{rw} = 7, n_{vr} = 9, p = .522).

We expected the pointing error of the HighVZ group to be lower than that of the LowVZ group for both treatments. However, the parallel boxplot medians (see *Figure 4.18*) indicate that the opposite was true for both treatments and LowVZ participants outperformed their HighVZ counterparts. The RWT+HighVZ group had a median pointing error angle of 11.5°, which was slightly larger than the RWT+LowVZ group's median pointing error of 10.0°, however, the distributions of the two groups did not differ significantly (Mann–Whitney U = 20.5, n_{high} = 8, n_{low} = 7, p = .418). Similarly, the VRT+HighVZ group had a median pointing error angle of 15.0°, which was slightly larger than the VRT+LowVZ group's median pointing error of 12.0°, however, the distributions of the two groups did not differ significantly (Mann–Whitney U = 28.5, n_{high} = 7, n_{low} = 9, p = .795).

For the pointing tests, we expected less errors from the HighVZ group than from the LowVZ group. It is interesting that, to the contrary, the LowVZ group erred slightly less for both treatment environments; the outcome for the 2nd point-to-origin test is indeed different. Also note the patterns in the outliers for the treatment environments and the spatial visualization groups. The presence of extreme outliers for the VRT group boxplots (the rightmost two) may indicate signs of pronounced disorientation from those participants; they errantly thought that the bearing of the origin was in the opposite direction of its true location in the simulated environment.

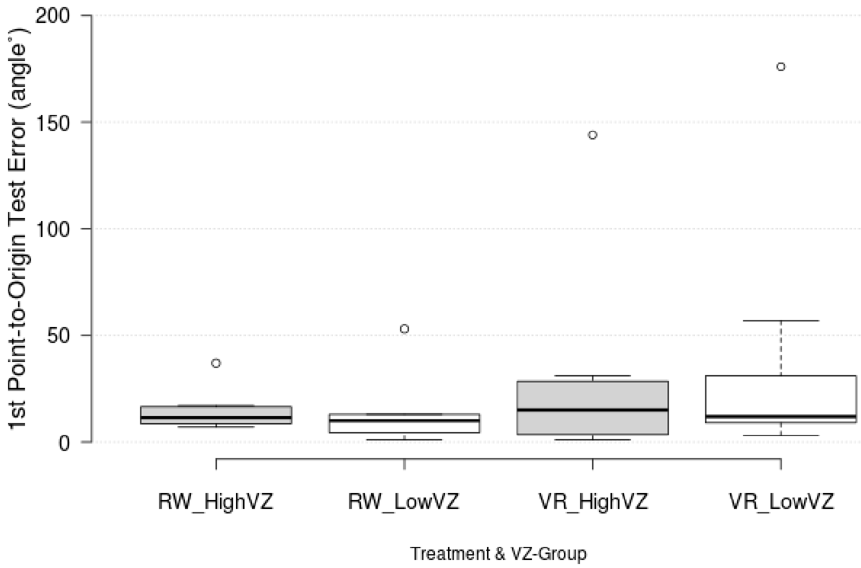


Figure 4.18 *Parallel boxplots of 1st point-to-origin test error for treatments by VZ group.*

Participants were asked to point to their starting location a second time at the end of the field exercise and the error angle was measured in degrees (*2nd point-to-origin test error)*. We expected the RWT group participants to make fewer errors on the 2nd point-to-origin test than the VRT group (smaller error angles are better). As expected, the parallel boxplot medians indicate that the RWT+HighVZ group had a smaller median pointing error angle of 10.5˚ when compared to the VRT+HighVZ group's median pointing error angle of 17.0˚; the distributions of the two treatments did not differ significantly (Mann–Whitney U = 17.5, $n_{rw} = 8$, $n_{vr} = 7$, $p = .246$). The RWT+LowVZ group's median pointing error angle of 15.5˚ was also smaller than the VRT+LowVZ group's median pointing error angle of 29.0˚; there was no significant difference between treatments for the LowVZ group (Mann–Whitney $U = 21.5$, $n_{rw} = 8$, $n_{vr} = 9$, $p = .177$).

We expected the HighVZ group's pointing error to be lower than that of the LowVZ group in both treatments. The parallel boxplot medians (see *Figure 4.19*) indicate that this was true for both the RWT and the VRT. For the real-world treatment, the RWT+HighVZ group had a median pointing error angle of 10.5˚, which was smaller than the RWT+LowVZ group's median pointing error of 15.5˚, however, the distributions of the two groups did not differ significantly (Mann–Whitney $U = 25.0$, $n_{\text{high}} = 8$, $n_{\text{low}} = 8$, $p = .497$). For the virtual reality treatment, the VRT+HighVZ group had a median pointing error angle of 17.0˚, which was smaller than the VRT+LowVZ group's median pointing error angle of 29.0˚, however, the distributions of the two groups did not differ significantly (Mann–Whitney $U =$ 17.0, $n_{\text{high}} = 7$, $n_{\text{low}} = 9$, $p = .139$).

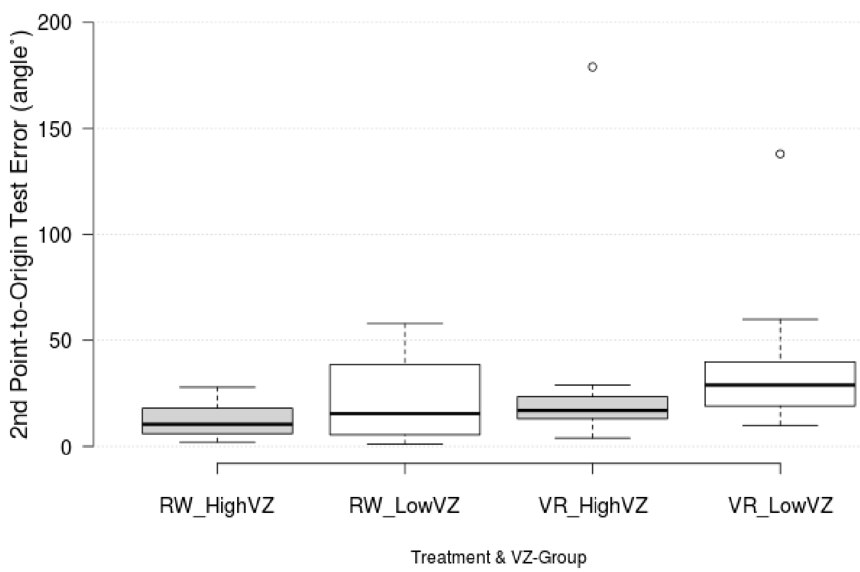


Figure 4.19 *Parallel boxplots of 2nd point-to-origin test error for treatments by VZ group.*

The boxplots exemplify the patterns that we expected for the treatment environments and the spatial visualization groups, where the RWT group boxplots (the leftmost two) have a lower location than those of the VRT group, and the HighVZ group boxplots (shaded) have a lower location than those of the LowVZ group. The presence of extreme outliers for the VRT group indicate that those participants errantly thought that the bearing of the origin was in the opposite direction of its true location in the simulated environment.

4.7.3 Digital Map Operation Comparisons Between Treatment Environments

Participants' digital map operations were logged by the study software and analyzed to determine if there were significant differences between the RWT group and the VRT group. Descriptive statistics for this dataset are shown in *Table 4.7.*

Table 4.7 *Descriptive Statistics for Map Operation Variables by Treatment.*

	Real World				**Virtual Reality**			
Performance Variable	***n***	**Mean**	**SD**	**Mdn**	***n***	**Mean**	**SD**	**Mdn**
Pans	15	45.2	30.8	40.0	15	70.0	35.2	68.0
Zooms	15	30.3	13.9	28.0	15	26.5	16.8	29.0
Pan limit reached	15	1.1	.9	1.0	15	1.9	2.0	2.0
Zoom limit reached	15	.53	.9	0	15	.1	.4	0
Addresses added	15	4.8	1.8	5.0	15	4.8	1.3	5.0
Addresses removed	15	2.4	1.9	2.0	15	3.9	1.7	4.0
Map resets	15	2.3	3.1	0	15	0.7	1.0	0

Initial inspection of the histograms and Q-Q Plots for the map operation variables indicated that, in each case, the distributions of either one or both treatment groups did not appear to be approximately normal. Thus, a Mann-Whitney *U* test was chosen for statistical comparisons because of its robustness when normality of the data cannot be assumed. The Mann-Whitney *U* test was used to determine whether the distributions of scores between treatments were significantly different for the following variables: *Pans*, *zooms*, *pan limit reached*, *zoom limit reached*, *addresses added*, *addresses removed*, and *map resets*.

Participants had the ability to pan the map up, down, left, or right (pans). We expected the RWT group participants to pan the map less than those of the VRT group (less panning implies better performance). The parallel boxplot medians (see *Figure 4.20*) show that participants of the RWT group did pan their maps less on average (Mdn = 40.0) than did those of the VRT group (Mdn = 68.0). There was a significant difference in the distribution of participants' pans between the RWT group and the VRT group (Mann–Whitney U = 54.5, n_{rw} = 15, n_{vr} = 15, p = .017). The boxplots reflect the significant differences between the treatments, as can be seen in the spread and interquartile ranges of the RWT group (IQR = 15.0) and the VRT group (IQR = 39.0).

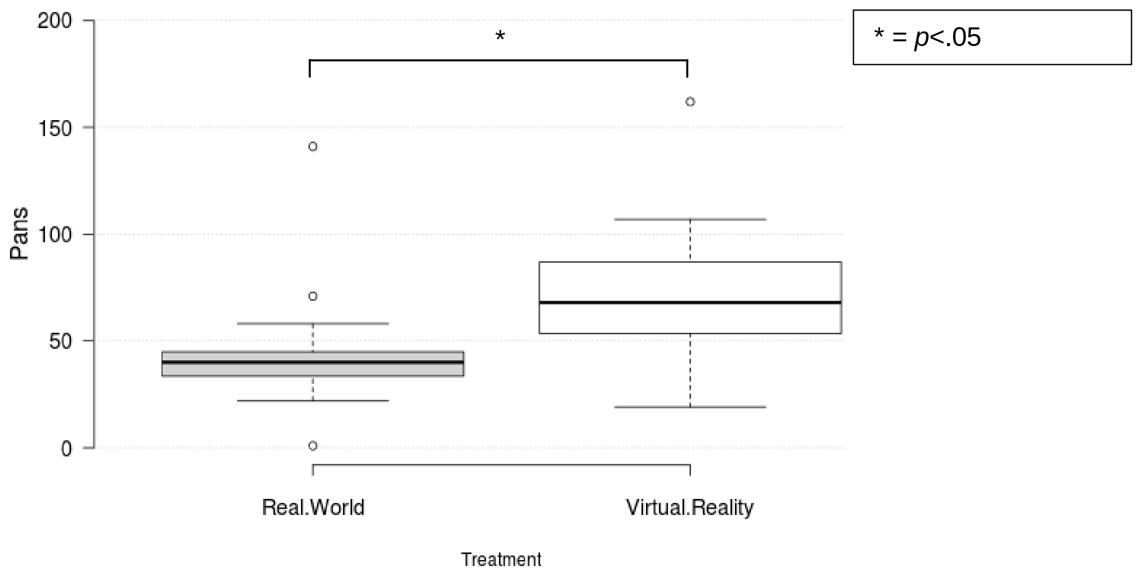


Figure 4.20 *Parallel boxplots of map pan operations (pans) by treatment.*

Participants could switch between five different zoom levels during their interaction with the digital map (*zooms*). We expected the RWT group participants to use the zoom function less than those of the VRT group (less zooming implies better performance). The parallel boxplot medians in *Figure 4.21* show that participants of the RWT group did change the zoom level of their maps slightly less on average (Mdn = 28.0) than did participants of the VRT group (Mdn = 29.0). There was no significant difference in the distribution of participants' *zooms* between the RWT group and the VRT group (Mann–Whitney U = 95.0, n_{rw} = 15, n_{vr} = 15, p = .478). The boxplots show similar interquartile ranges between the

RWT group (IQR = 27.0) and the VRT group (IQR = 25.5); the VRT group also has a greater spread.

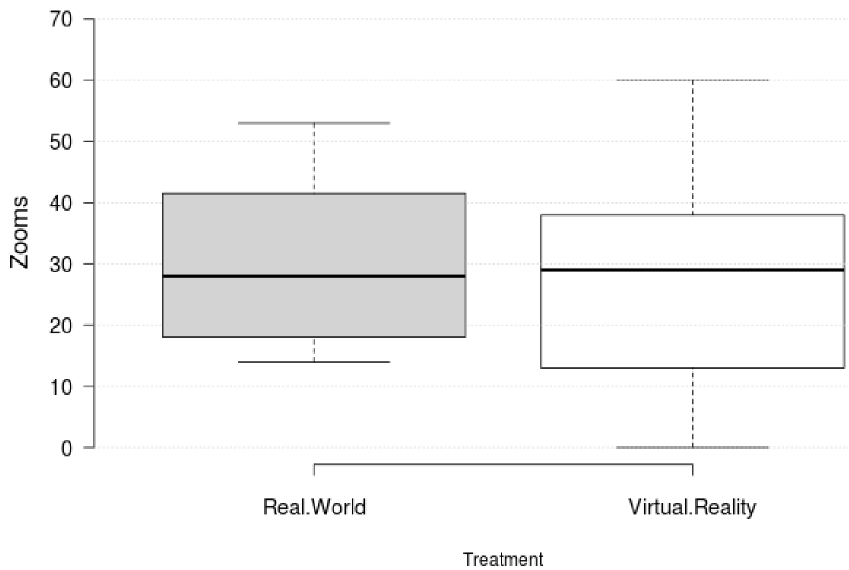


Figure 4.21 *Parallel boxplots of map zoom operations (zooms) by treatment.*

During panning, if participants reached the bounds of the map and continued to pan in the bounded direction, the software displayed a notification to prevent further attempts (*pan limit reached*). We expected the RWT group participants to have less occurrences of reaching the pan limit than those of the VRT group (less pan limiting occurrences implies better performance). The parallel boxplot medians in *Figure 4.22* show that participants of the RWT group did attempt to go beyond the bounds of the map slightly less on average (Mdn = 1.0) than did participants of the VRT group (Mdn = 2.0). There was no significant difference in the distribution of participants' *pan limit reached* occurrences between the RWT group and the VRT group (Mann–Whitney U = 89.0, n_{rw} = 15, n_{vr} = 15, p = .342).

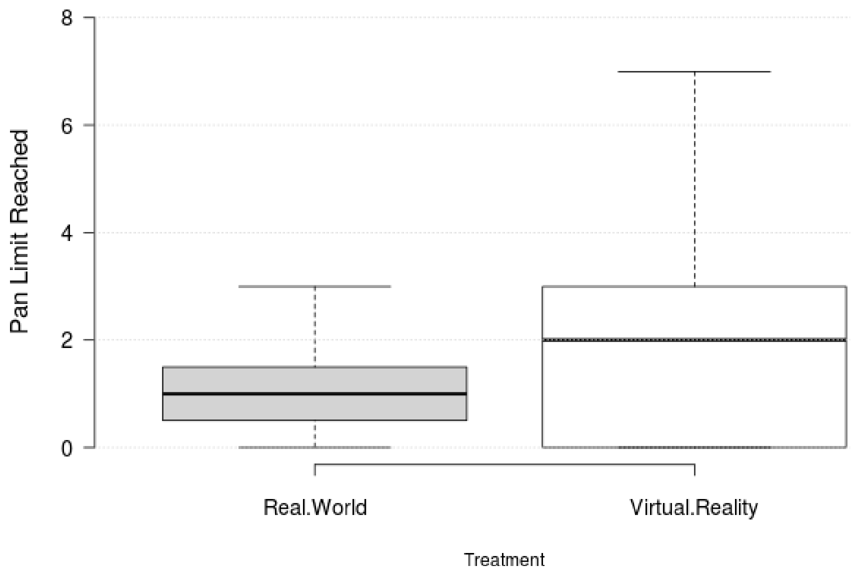


Figure 4.22 *Parallel boxplots of blocked pan operations (pan limit reached) by treatment.*

During use of the map zoom feature, participants were notified when they attempted to go beyond the minimum or maximum zoom levels that were available (*zoom limit reached*). We expected the RWT group participants to have less occurrences of reaching the zoom limits than those of the VRT group (fewer zoom limiting occurrences implies better performance). The parallel boxplot medians in *Figure 4.23* show that participants of the RWT group were notified of reaching a zoom limit slightly more on average (Mdn = 0, M = .53) than were participants of the VRT group (Mdn = 0, M = .1). There was no significant difference in the distribution of participants' *zoom limit reached* occurrences between the RWT group and the VRT group (Mann-Whitney $U = 88.0$, $n_{rw} = 15$, $n_{vr} = 15$, $p = .317$).

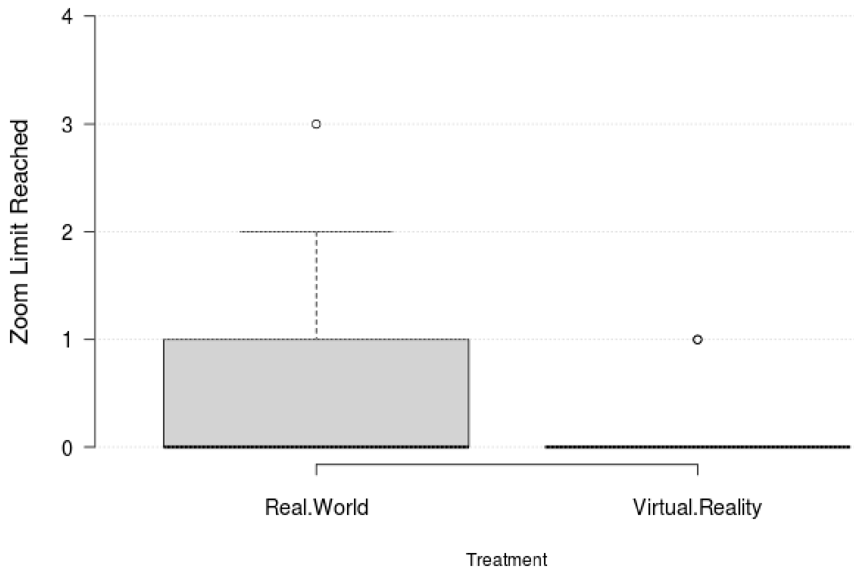


Figure 4.23 *Parallel boxplots of blocked zoom operations (zoom limit reached) by treatment.*

If a participant reached the physical location of an assigned address and concluded that the address existed but was not present on the map, the instruction was to correct the map by adding the corresponding map spot for the address to the proper location (*addresses added*). We expected the RWT group participants to add more addresses than those of the VRT group (more address additions imply better performance). The parallel boxplot medians in *Figure 4.24* show that participants of the RWT group added the same number of addresses on average as did those of the VRT group (Mdn = 5.0). There was no significant difference in the distribution of participants' *addresses added* between the RWT group and the VRT group (Mann-Whitney $U = 99.5$, $n_{rw} = 15$, $n_{vr} = 15$, $p = .603$). The spreads and interquartile ranges of the RWT group (IQR = 2.0) and the VRT group (IQR = 1.75) boxplots also show similarities.

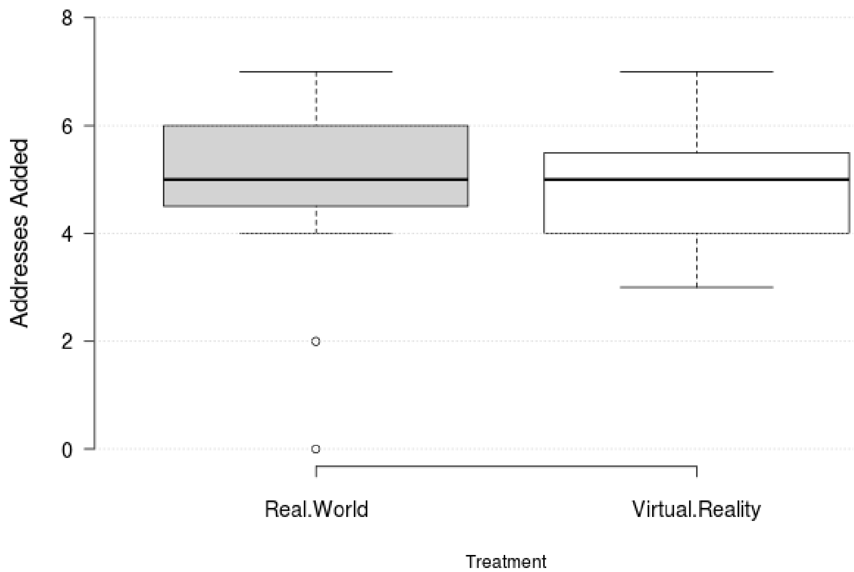


Figure 4.24 *Parallel boxplots of operations to add addresses (address added) by treatment.*

If a participant traveled to the location of an assigned address as indicated by the map and he or she concluded that the address no longer existed, the instruction was to correct the map by removing the corresponding map spot for the address (*addresses removed*). We expected the RWT group participants to remove more addresses than those of the VRT group (more address removals imply better performance). The parallel boxplot medians in *Figure 4.25* show that participants of the RWT group removed less addresses on average (Mdn = 2.0) than those of the VRT group (Mdn = 4.0). There was a significant difference in the distribution of participants' *addresses removed* between the RWT group and the VRT group (Mann-Whitney $U = 63.0$, $n_{rw} = 15$, $n_{vr} = 15$, $p = .042$).

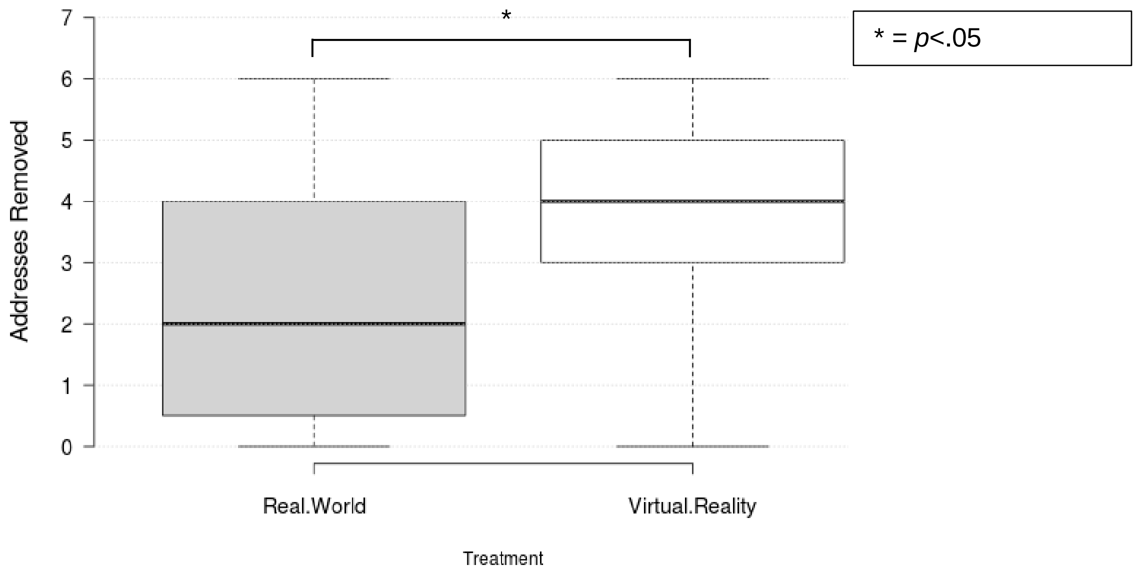


Figure 4.25 *Parallel boxplots of address removal operations (address removed) by treatment.*

Participants could reset the map view to its original state, effectively undoing all pan and zoom operations (*map resets*). We expected the RWT group participants to reset the map less than those of the VRT group (less map resets imply better performance). The parallel boxplot medians in *Figure 4.26* show that participants of the RWT group reset their map more times on average (Mdn = 0, M = 2.3) than did those of the VRT group (Mdn = 0, M = 0.7). There was no significant difference in the distribution of participants' *map resets* between the RWT group and the VRT group (Mann-Whitney U = 87.0, n_{rw} = 15, n_{vr} = 15, p = .298).

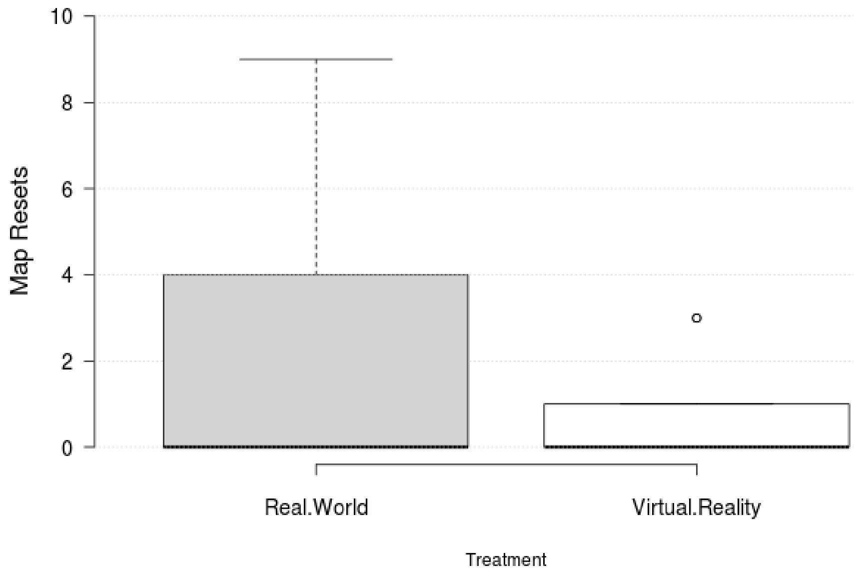


Figure 4.26 *Parallel boxplots of map reset operations (map resets) by treatment.*

4.7.4 Digital Map Operation Comparisons Between Crosses of the Treatments and Participant Groups

The following analyses involve comparisons of the digital map operations associated with the four experimental groups that are crosses of the treatment environments (RW/VR) and the participant groups (HighVZ/LowVZ). Only the four most frequently used digital map operations were examined (*pans*, *zooms*, *addresses added*, *addresses removed*) due to low frequencies for all other map operations pertaining to the crossed experimental groups. We examined parallel box plots, key statistics and statistical tests. A Mann-Whitney *U* test was chosen for statistical comparisons due to its robustness in situations when the normality of a sample cannot be assumed. Descriptive statistics for this dataset are shown in *Table 4.8*.

Table 4.8 *Descriptive Statistics for Digital Map Operation Variables by Treatment and VZ Group.*

	Performance Variable	HighVZ Group				LowVZ Group			
Real World		***n***	**Mdn**	**Mean**	**SD**	***n***	**Mdn**	**Mean**	**SD**
	Pans	7	44.0	42.7	8.6	8	37.0	47.4	42.7
	Zooms	7	21.0	27.0	15.5	8	35.0	33.3	12.6
	Addresses added	7	6.0	5.7	1.0	8	5.0	4.0	0.7
	Addresses removed	7	4.0	3.4	1.6	8	1.0	1.5	1.8
Virtual Reality		***n***	**Mdn**	**Mean**	**SD**	***n***	**Mdn**	**Mean**	**SD**
	Pans	7	68.0	58.1	29.1	9	62.0	76.7	37.8
	Zooms	7	21.0	19.9	13.0	9	36.0	31.6	17.3
	Addresses added	7	5.0	4.7	1.3	9	5.0	4.9	1.4
	Addresses removed	7	5.0	4.3	1.7	9	4.0	3.3	1.7

When comparing like VZ groups, we expected the RWT group participants to pan the map less than those of the VRT group (less panning implies better performance). The parallel boxplot medians (see *Figure 4.27*) indicate that the RWT+HighVZ group panned their maps less on average (Mdn = 44.0) than did those of the VRT+HighVZ group (Mdn = 68.0); the distributions of the two treatments did not differ significantly (Mann–Whitney U = 15.0, n_{rw} = 7, n_{vr} = 7, p = .250). The RWT+LowVZ group also panned their maps less on average (Mdn = 37.0) than did those of the VRT+LowVZ group (Mdn = 62.0); the distributions of the two treatments differed significantly (Mann–Whitney U = 35.0, n_{rw} = 8, n_{vr} = 9, p = .038).

We expected participants of the HighVZ group to pan the map less than those of the LowVZ group in both treatments. Surprisingly, the parallel boxplots associated with the VRT groups have a higher location than their RWT group counterparts in the cases of both the HighVZ and LowVZ participant groups. The boxplot medians indicate that the RWT+HighVZ group unexpectedly panned their maps more on average (Mdn = 44.0) than did those of the RWT+LowVZ group (Mdn = 37.0); the distributions of the two treatments did not differ significantly (Mann–Whitney $U = 20.0$, $n_{high} = 7$, $n_{low} = 8$, $p = .384$). The VRT+HighVZ group also unexpectedly panned their maps more on average (Mdn = 68.0) than did those of the VRT+LowVZ group (Mdn = 62.0); the distributions of the two treatments did not differ significantly (Mann–Whitney $U = 26.5$, $n_{high} = 7$, $n_{low} = 9$, $p = .631$).

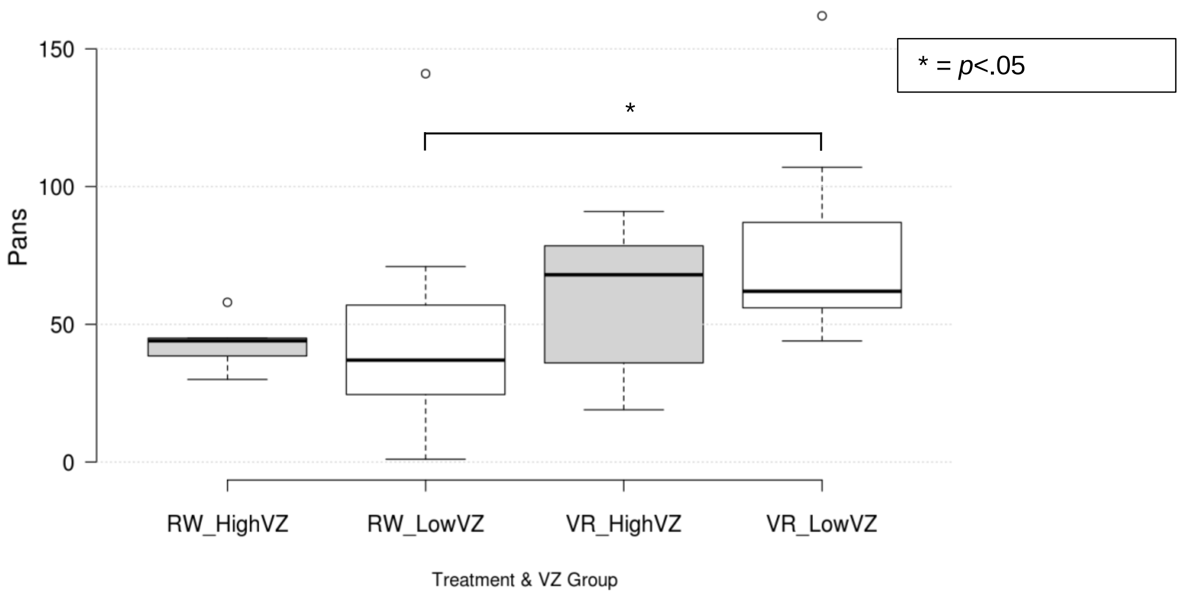


Figure 4.27 *Parallel boxplots of map pan operations (pans) by treatment and VZ group.*

The parallel boxplots for *zooms* show striking similarities across the treatment environments and the spatial visualization participant groups.

We expected the RWT group participants to use the zoom function less than those of the VRT group (less zooming implies better performance). The parallel boxplot medians (see *Figure 4.28*) show that the RWT+HighVZ group and the VRT+HighVZ group participants changed the zoom level of their maps the same number of times on average (Mdn = 21.0); the distributions of the two treatments did not differ significantly (Mann–Whitney $U =$ 17.5, $n_{rw} = 7$, $n_{vr} = 7$, $p = .407$). The RWT+LowVZ group changed the zoom level of their maps slightly less on average (Mdn = 35.0) than did the VRT+LowVZ group (Mdn = 36.0);

the distributions of the two treatments did not differ significantly (Mann–Whitney U = 34.5, n_{rw} = 8, n_{vr} = 9, p = .920).

We expected participants of the HighVZ group to use the zoom function less than those of the LowVZ group in both treatments. The parallel boxplot medians indicate that the RWT+HighVZ group did use the zoom function less on average (Mdn = 21.0) than did the RWT+LowVZ group (Mdn = 35.0); the distributions of the two treatments did not differ significantly (Mann–Whitney U = 19.0, n_{high} = 7, n_{low} = 8, p = .327). The VRT+HighVZ group also used the zoom function less on average (Mdn = 21.0) than did the VRT+LowVZ group (Mdn = 36.0); the distributions of the two treatments did not differ significantly (Mann–Whitney U = 16.5, n_{high} = 7, n_{low} = 9, p = .126).

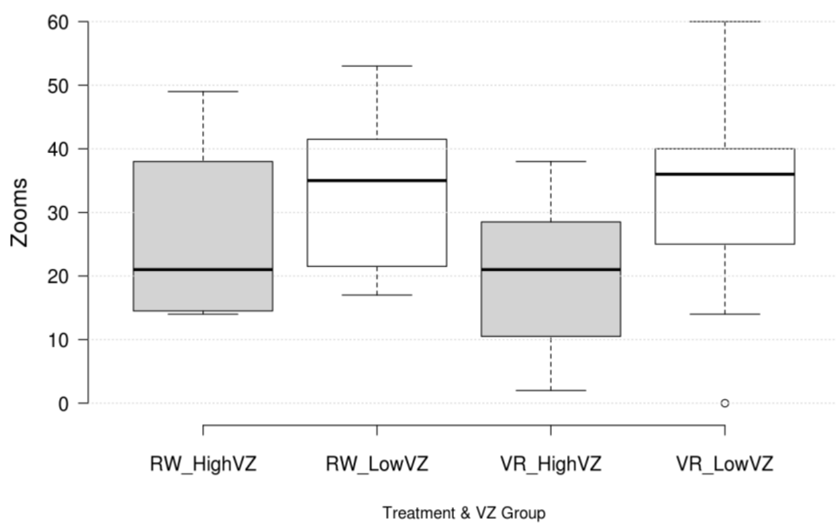


Figure 4.28 *Parallel boxplots of map zoom operations (zooms) by treatment.*

We expected the RWT group participants to add more addresses than those of the VRT group (more address additions imply better performance). The parallel boxplot medians (see *Figure 4.29*) show that participants of the RWT+HighVZ group did add more addresses on average (Mdn = 6.0) than did those of the VRT+HighVZ group (Mdn = 5.0); the distributions of the two treatments did not differ significantly (Mann-Whitney U = 12.0, n_{rw} =

7, n_{vr} = 7, p = .126). The RWT+LowVZ group and the VRT+LowVZ group unexpectedly added the same number of addresses on average (Mdn = 5.0); the distributions of the two treatments did not differ significantly (Mann-Whitney U = 27.5, n_{rw} = 8, n_{vr} = 9, p = .441).

We expected participants of the HighVZ group to add more addresses than those of the LowVZ group in both treatments. The parallel boxplot medians indicate that the RWT+HighVZ group did add more addresses on average (Mdn = 6.0) than did the RWT+LowVZ group (Mdn = 5.0); the distributions of the two treatments differed significantly (Mann–Whitney U = 10.5, n_{high} = 7, n_{low} = 8, p = .049). Unlike the RWT outcome, the VR+HighVZ group boxplot does not have a higher location than that of the VR+LowVZ group as was expected. Furthermore, the VRT+HighVZ and the VRT+LowVZ group unexpectedly added the same number of addresses on average (Mdn = 5.0); the distributions of the two treatments did not differ significantly (Mann–Whitney U = 28.0, n_{high} = 7, n_{low} = 9, p = .749).

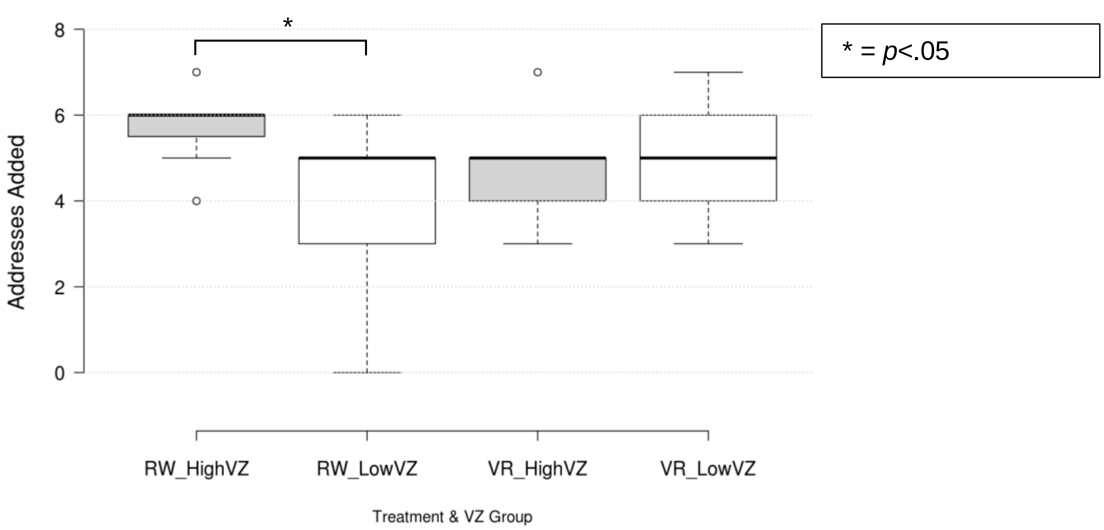


Figure 4.29 *Parallel boxplots of operations to add addresses (address added) by treatment.*

We expected the RWT group participants to remove more addresses than those of the VRT group (more address removals imply better performance). The parallel boxplots (see

Figure 4.30) indicate that the opposite is true for both the HighVZ and LowVZ participant groups. The medians show that participants of the RWT+HighVZ group unexpectedly removed fewer addresses on average (Mdn = 4.0) than did those of the VRT+HighVZ group (Mdn = 5.0); the distributions of the two treatments did not differ significantly (Mann-Whitney $U = 16.5$, $n_{rw} = 7$, $n_{vr} = 7$, $p = .337$). The RWT+LowVZ group also unexpectedly removed fewer addresses on average (Mdn = 1.0) than did the VRT+LowVZ group (Mdn = 4.0); the distributions of the two treatments did not differ significantly (Mann-Whitney $U = 17.0$, $n_{rw} = 8$, $n_{vr} = 9$, $p = .075$).

We expected participants of the HighVZ group to remove more addresses than those of the LowVZ group in both treatments. This expectation is exemplified by the HighVZ group's boxplots, which have a higher location than those of the LowVZ group. The parallel boxplot medians indicate that the RWT+HighVZ group did remove more addresses on average (Mdn = 4.0) than did the RWT+LowVZ group (Mdn = 1.0); the distributions of the two treatments did not differ significantly—though the p-value is suggestive (Mann–Whitney $U = 12.0$, $n_{high} = 7$, $n_{low} = 8$, $p = .073$). The VRT+HighVZ group also removed more addresses on average (Mdn = 5.0) than did the VRT+LowVZ group (Mdn = 4.0); the distributions of the two treatments did not differ significantly (Mann–Whitney $U = 20.5$, $n_{high} = 7$, $n_{low} = 9$, $p = .267$).

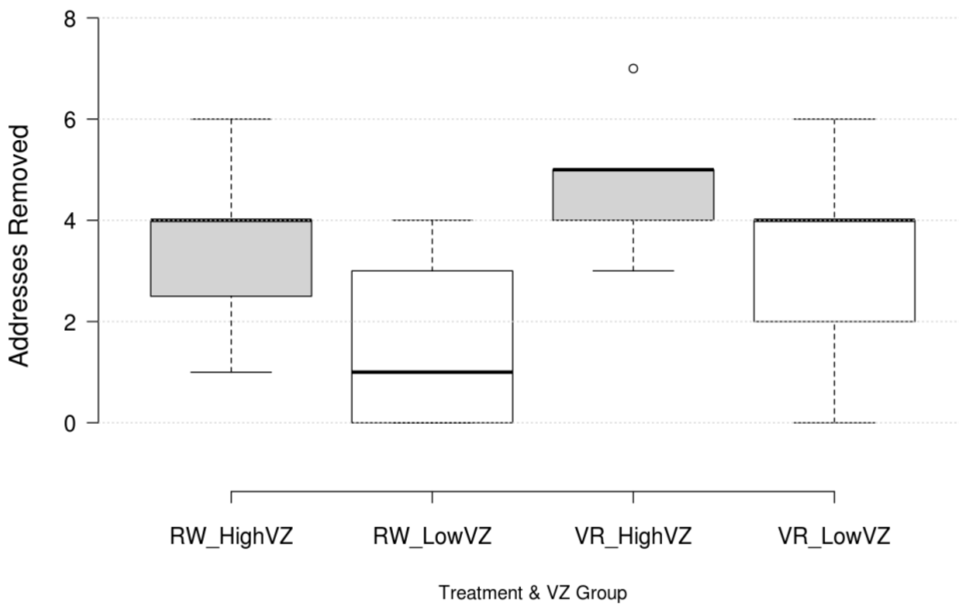


Figure 4.30 *Parallel boxplots of address removal operations (address removed) by treatment.*

4.7.5 The Task Workflow and Researcher Observations

Participants in the real-world treatment and virtual reality treatment followed the same think aloud protocol as they verified residential addresses. They were accompanied by two researchers who observed and recorded their thought processes, actions, and behaviors. The observations were subsequently coded, analyzed, and a task workflow was generated. The resulting workflow is discussed in this section along with a description of the important participant behaviors that were observed.

A diagram of the address verification workflow based on observations from both treatment environments is shown in *Figure 4.31*. This task workflow holds for all address verification scenarios encountered during this experiment. The components of the workflow are differentiated based on shape and color-coding. The rounded, oval shapes colored in blue represent either the beginning or ending of the task workflow. The orange-colored rectangles represent the major phases of the task and the arrows indicate their flow; the lighter-shaded rectangles (beige) are important parts (or outputs) of the *Planning Phase*. The red triangles represent the major decision points of the workflow. One iteration through the workflow equates to the verification of a single residence by a participant.

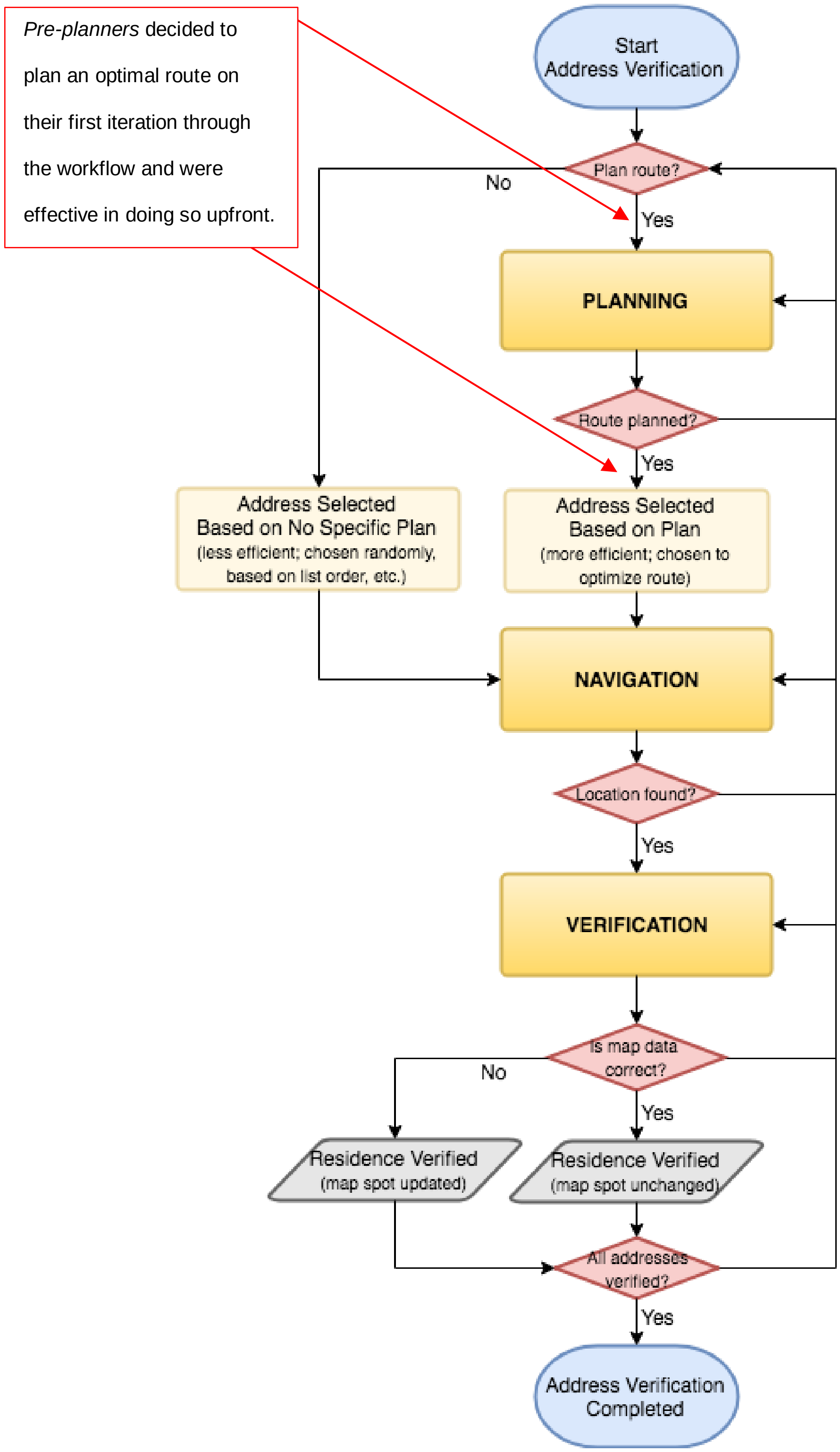


Figure 4.31 *The VRFE task workflow diagram for address verification.*

Planning Phase Behaviors

Some of the earliest actions and behaviors observed in participants were centered on planning. At the earliest juncture of the *Planning Phase*, the most effective planners tended to “pre-plan”. Pre-planning is a behavior that has previously shown up in our field studies on the use of a paper map for address verification; it is defined as the process where, at the onset of the address verification task, participants formulate a comprehensive plan for their route and the order in which they intend to verify residences. In the previous studies, the tendency for participants to pre-plan has been linked to their spatial ability and their task efficiency. Pie graphs comparing participants’ pre-planning tendencies between treatment environments are shown in *Figure 4.32*. For the RWT, 4 out of 16 participants pre-planned (25%) compared to 12 out of 16 participants (75%) who did not. Pre-planning in the VRT was slightly lower, where 3 out of 16 participants were observed pre-planning (19%) compared to 13 out of 16 participants (81%) who did not appear to pre-plan.

In contrast to the pre-planners, some participants very quickly selected a residence to verify at the start of the exercise and moved on to navigation with very little planning if any at all. In general, the participants who failed to pre-plan tended to do limited planning up front (sometimes because of difficulty or frustration); their planning activities were spread across various legs of their routes. Many participants frequently revisited the Planning Phase throughout the exercise to adjust an existing plan or to further optimize. Effective planners generally exited the Planning Phase after selecting a destination that was believed to be optimal in terms of the overall route. The least effective planners haphazardly chose their destinations and some participants were even observed to select residences based on the order that they were listed in the dropdown list of the software, which was randomized and not optimal. This is consistent with observations from the previous studies, where some participants were observed to initially base their routes on the listed order of assigned residences per their printed materials.

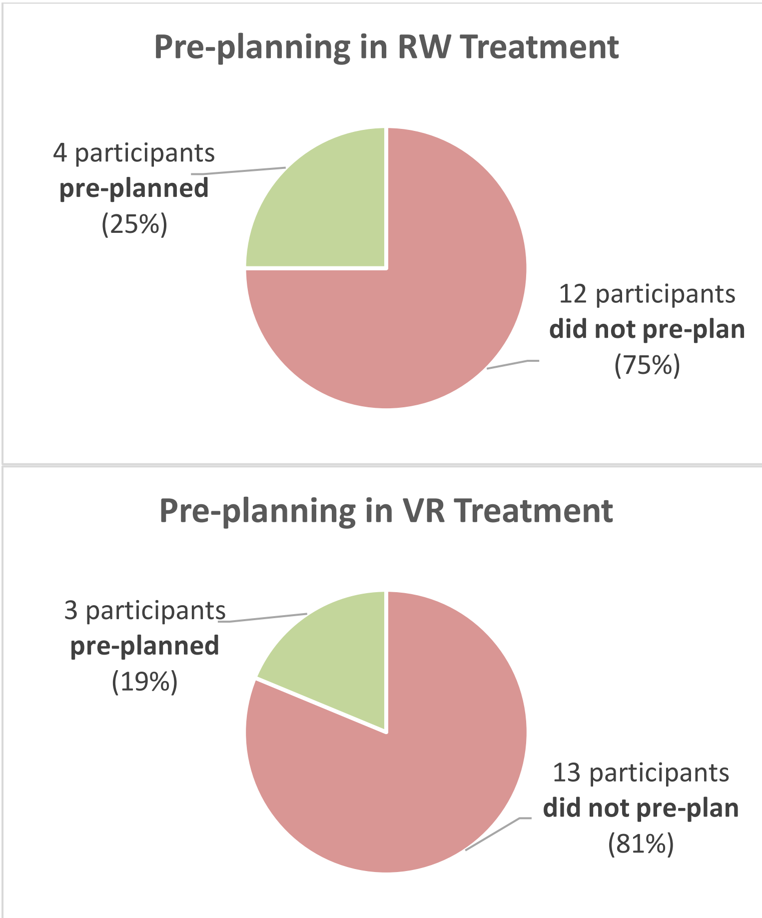


Figure 4.32 *Pie graphs comparing participants' pre-planning tendencies between the RW treatment (top) and the VR treatment (bottom).*

The very use of a handheld computer seemed to influence participants' planning behaviors as well. We observed a markedly drastic decrease in participants' pre-planning compared to the prior address verification studies. Another general observation is that fewer participants, if any, planned while they walked. This behavior, to plan on the move, was occasionally observed in the participants of *Field Experiment 1* and *Field Experiment 2*, however, these participants used paper maps. Another software usage pattern of interest was participants' tendency to pan frequently while viewing the digital map at its outer zoom levels; this facilitated better tracking of streets and potential routes during planning.

Navigation Phase Behaviors

During the *Navigation Phase*, participants checked cues from the environment and the map to reach their intended destination. Effective navigators tended to:

- Pay attention to street signs, landmarks (including homes along the way) and relate them to the map.
- Work through confusion and difficulty by making sense of environmental cues and map features to reorient themselves.
- Pan the map as they traveled, but not excessively.
- Limit their use of the digital map's zoom function, oftentimes sticking to one or two zoom levels that facilitated easier tracking of their position, the network of streets and their routes.
- Less frequently run into the map's panning and zooming limits; less frequently reset the map.

Less effective navigators tended to:

- Misinterpret their direction and heading.
- Wander due to difficulty interpreting the map and/or environment.
- Switch to a different residence for verification when difficulty and/or frustration was too great.
- Struggle with use of the software and digital map.
- Excessively rotate the map (or their bodies while holding the map steady) when disoriented.

- Either sparsely check and pan their maps or excessively do so out of confusion and frustration.
- More frequently run into the panning and zooming limits of the map and sometimes use the map reset feature to start over with the navigation process.
- More frequently review the map and revise their routes.

Verification Phase Behaviors

The behavior of participants changed once they reached the vicinity of the address being verified. During this *Verification Phase*, participants made a final decision about the correctness of each assigned residence with respect to the map. A key difference in participant behavior during the Verification Phase was the increased thoroughness with which they scrutinized a variety of cues near the residences being verified—this was done to ensure that each home had the correct address number, was on the proper street and was located in the proper place on the map with respect to what was seen on the grounds of the neighborhood. Participants who were most effective during the Verification Phase tended to:

- Change their attitude regarding the veracity of the map once reaching their destination; during navigation, participants seemed to trust the map without question, whereas during verification the most effective participants approached the map with skepticism (e.g., they assumed that the placement of assigned residences were incorrect until verified and proven otherwise).
- Demonstrate an increased thoroughness in terms of their scrutiny of the map and environmental cues—especially in checking the position of the neighboring residences of those that were verified.
- Pan and rotate the map more frequently during the verification process.

- Modify the map spots more frequently, with an increased and balanced use of the functions to "add" and "remove" map spots.
- Avoid resetting the map, which returned the view of the map to its default state.
- Have a better grasp of their direction and heading and made fewer errors of judgement in this regard.

4.7.6 Identifying Usability Issues

The address verification software that participants used was developed specifically for this research. The software could be viewed as a minimum viable product (MVP) in the development sense because it included only those features that were necessary to complete the task and to log participants' actions at the interface. Though usability testing was not a primary objective of this experiment, the pre-production nature of the software helped us to better understand how usability issues might be identified during evaluations conducted within the treatment environments. The attending researchers drew two key conclusions when reflecting on the usability issues that surfaced:

1) Based on their use of the field study and think aloud methods, both environments were well-suited for participants and observers to identify a variety of usability issues centered on the mobile computer and its software.
2) The virtual reality treatment was exceptional in several respects:
    i. For participants, it was of sufficient fidelity to draw out behaviors and interactions that reflected real-world usage patterns without the drawbacks that come with traveling substantial distances on foot in a real neighborhood (e.g., physical exertion, adverse weather, safety concerns and distractions).

ii. For observers, it was much easier to shadow participants in the VR lab. Observers were readily able to closely examine participants' interactions with the mobile device and display. The real-world treatment environment, on the other hand, raised concerns by researchers that they were not able to directly observe many of participants' interactions due to the physical and cognitive demands of facilitating the study and shadowing participants while traveling on foot. In other words, researchers felt that they could not always sufficiently keep up with participants in the real-world—this was not a problem in the VRT.

Table 4.9 lists usability issues that were identified in both environments through (a) researchers' direct observation of participant interactions with the mobile device and (b) participants who called out issues either during the think aloud process or in response to the exit questionnaire.

Table 4.9 *Usability Issues Identified During VR/Field Experiment.*

Category	Usability Issue	Description
Planning	Difficulty referring to residence list	The full list of assigned residences could only be viewed when the user tapped a drop-down list to select one to be verified; the user was not able to otherwise view them all at once—this negatively impacted planning.
Map Use	Changing map orientation	There was no way for the user to change the orientation of the map based on a north-up or track-up preference; physical rotation of the mobile device was not ideal.
	Orientation of street labels	The street labels did not rotate with the map when the user physically rotated the device.
	Visibility of street labels	The street labels were fixed and were not always visible (i.e. they were off screen sometimes) depending on the cross section of the map being viewed and the user's zoom level; labels should travel with their respective streets as views change.
Map Spot Functionality	Moving a map spot is not intuitive	To move a map spot requires two actions: 1) the existing map spot is removed, 2) a new map spot is added to the intended location; moving a map spot should be a one-step process; if the user adds a map spot that unknowingly already exists on the map, the previous map spot is retained rather than replaced and still requires removal by the user to correct the map.
	Possibility of duplicate map spots	Multiple map spots could be added for the same address; each address should have only one map spot.
	Map spots are difficult to tap	The map spot icons (and their hotspots) were too small for some participants to select and manipulate with the stylus; this led to false negative taps.
	Recognizability of active map spot	It was difficult to differentiate, from all other map spots, the map spot associated with the active residence being verified; visual indicators such as size and color could help.
	Visibility of closely grouped map spots	When the map spots were closely grouped, it was difficult to see the corresponding address numbers because they overlapped.

4.7.7 Comparisons of Observations Recorded in Treatments

A comparison was made of the observations that were recorded in each of the treatment environments. The total number of observations recorded in each treatment as well as their variation offers some insight into the effectiveness of the treatments. The observation totals, by treatment, are shown in *Figure 4.33*. Combined, the two treatments resulted in 1954 total observations that were captured in researchers' written notes. For the real-world treatment (RWT), 1167 observations were captured, which accounted for 60% of the study total. The virtual reality treatment (VRT) yielded 787 observations, accounting for 40% of the total.

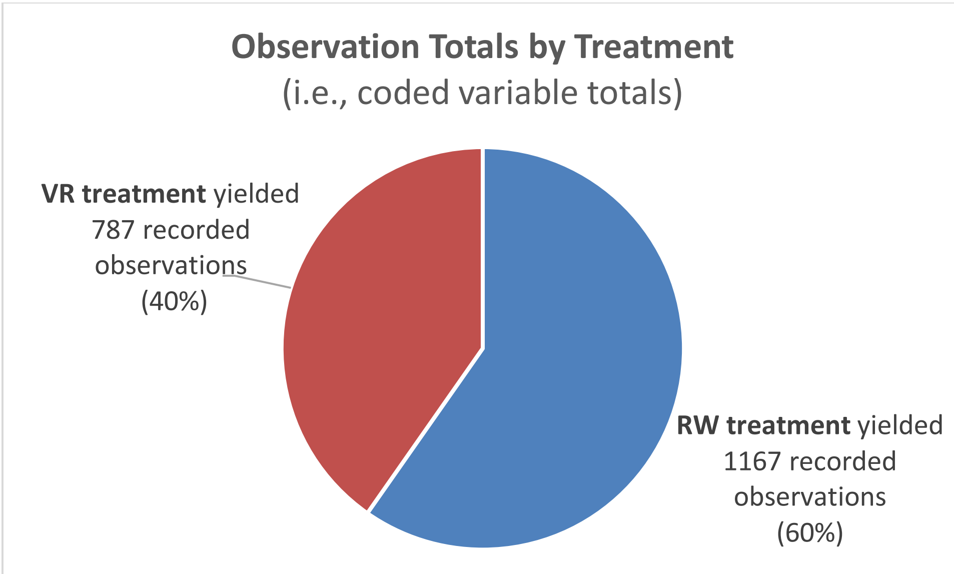


Figure 4.33 *Graph of total number of recorded observations by treatment.*

The 1954 coded observations coming from both treatments were combined, randomly analyzed and classified into distinct types. A total of 113 unique observation types were identified during the analysis. The 1167 coded observations of the RWT could be sorted into 74 of the 113 types, accounting for 65% of the variation in observation types. The 787 observations of the VRT could be sorted into 98 of the 113 types, which accounted for 87% of the variation in observation types. Of the 113 total observation types, 59 types (52%) were common to both treatments, 15 types (13%) were exclusive to the RWT, and 39 types (35%) were exclusive to the VRT.

Figure 4.34 shows the observation categories by treatment group. Overall, more than half of the observation categories were shared between the RWT and VRT. Despite the VRT yielding fewer total observations (787) than the RWT (1167), it was comprised of more distinct observation types (35%) than the RWT (13%).

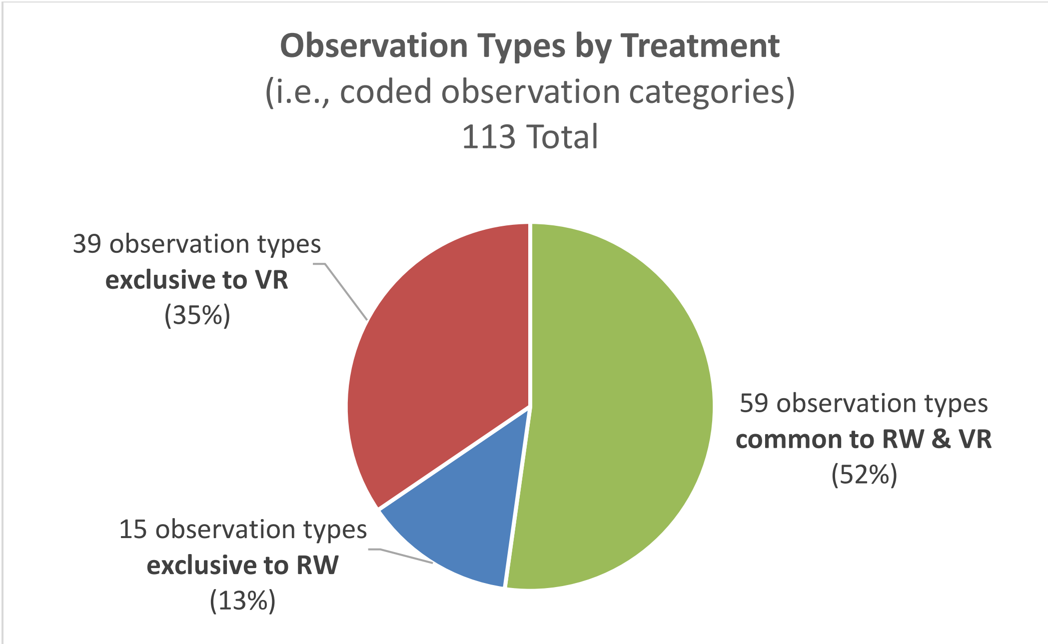


Figure 4.34 *Graph of coded observation types by treatment.*

4.7.8 Observer Feedback

Observers were asked to compare and contrast their data collection experiences in the real-world treatment (RWT) and virtual reality treatment (VRT). For the RWT, observers noted the benefits of conducting the field experiment in a real-world setting. Such benefits included the authenticity of the environment, its contextual richness, and the ecological validity of conducting the field experiment in an actual neighborhood. However, observers noted that the benefits inherent to the real-world field study came with tradeoffs. Among the challenges encountered in the real-world, observers cited:

1. Difficulty keeping up with and shadowing participants on foot.
2. Difficulty observing participants' interactions with the mobile device and environment.
3. Difficulty recording observations while staying mobile.

4. Potentially long travel distances and travel times, which can be physically exerting for both participant and observer.
5. Difficulty ensuring participant safety while multitasking during the exercise.
6. Inability to conduct the exercise in poor weather conditions. For example, participants were unable to effectively use the stylus and operate the handheld computer in rain or extremely cold temperatures.

The VRT mitigated the challenges listed above. By confining the study exercise to the VR lab space (10 cubic feet), the required walking was drastically reduced. The simulated walking effectively enabled participants to travel substantial distances in the virtual neighborhood simply by taking a few steps away from the center of the simulator floor. The accompanying observer needed only to stay out of the way of the participant during the exercise (see *Figure 4.9*).

Observers noted that the minimized walking, the absence of traffic and other safety hazards, and the ease of shadowing participants in the simulator altogether reduced the burden of multitasking that was prevalent in the field environment. The VRT facilitated closer observation of participants' behaviors and actions. Observers also noted that they could more closely scrutinize participants' interactions with the handheld device (e.g., errors and usability issues were easier to recognize).

In summary, the VRT enabled observers to devote more attention to participants as they completed the task. This resulted in a richer set of observations that could be further analyzed.

4.8 Discussion

In this experiment, we examine and compare the outcomes of two evaluations of a location-based task. One evaluation was conducted in a real-world setting where participants walked a neighborhood to verify six residential addresses with a mobile device and digital map—the real-world treatment (RWT). The second evaluation was conducted in a VR lab, in which the real-world task environment was simulated and the field study protocol was replicated—the virtual reality treatment (VRT). In the two treatment environments, participants who were screened into high (HighVZ) and low spatial visualization ability (LowVZ) groups completed a couple of tasks: (1) They verified six neighborhood residences using address verification software and (2) they pointed to their starting location twice during the exercise (once halfway through and a second time at the end). Quantitative and qualitative data were collected and analyzed. This section covers results that are relevant to the research questions, hypothesis tests, and key findings of the experiment.

4.8.1 Feasibility of Replicating a Field Experiment Using VR

Initial research questions were concerned with the feasibility of using a virtual reality lab to replicate a field study. Could we create a VR lab to duplicate the rigorous, mixed-method field study procedure that was carried out in the field? In both of the treatment environments of this experiment, participants followed the same study protocol and completed the same tasks. Task performance metrics were captured in both environments in addition to observations that were recorded by attending researchers. In our prior field studies on the task, participants had the ability to:

- Walk a neighborhood that was large in extent (spans approximately 1.3 square miles).

- Seamlessly use a mobile device during interactions with the environment.
- Think aloud as they completed the tasks, answering occasional questions from attending researchers.

Observers had the ability to:

- Facilitate the study and ensure that a specific protocol was followed.
- Train participants on the task.
- Shadow participants and take notes on their think aloud, behaviors, mobile device use and software use.
- Occasionally ask participants clarifying questions regarding behaviors and actions that are relevant to a task outcome or approach.

To enable these important capabilities in the VR lab, careful attention was paid to the selection of the underlying VR technology, the fidelity and accuracy of the virtual environment, the interaction design, and the data collection procedures. *Table 4.10* summarizes key features of the VR lab that enabled our team to create an immersive and functional substitute to a real-world field study experience.

Table 4.10 *Features of the VR Lab that Made a Field Study Feasible.*

Feature	Description
6-sided CAVE	- Facilitated immersion in virtual environment as participants completed the task. - Enabled use of separate mobile device while viewing IVE (difficult w/ HMD). - Enabled a shared VR experience between participants and researchers.
Hi-fi., 3D Model Neighborhood Replica	- Streets, signage, buildings, residences, and other important landmarks were incorporated into the 3D model to closely reflect the real neighborhood. - Geolocation of 3D model facilitated accurate locations and distances.
Hands-free Locomotion	- In a confined lab space, participants traversed a virtual neighborhood using a locomotion interface that allowed them to travel at walking speeds. - The walking interface allowed for hands-on interaction with the mobile device. - Participants quickly acclimated to the “your body is the joystick” metaphor.
Reproduced Data Collection Procedures	- Two researchers shadowed each participant as was done in the RW field study. - Distance, time, and participants’ pointing angles were accurately recorded. - Notes were taken on participants’ think aloud, behaviors, and task outcomes.

4.8.2 Quantitative Analysis

In the *Results Section*, hypothesis test statistics were intermingled with other results and boxplots pertaining to participants' task performance and digital map operations. For clarity, this section further discusses the hypothesis test outcomes and provides additional graphs and tables to summarize the key findings.

The ten hypotheses of this experiment are grouped as follows:

Group 1) Between-treatment comparisons (RWT vs. VRT) of participants' task performance (*H1*, *H2*, *H3*) and digital map operations (*H6*, *H7*, *H8*).

Group 2) Between-participant group comparisons (HighVZ vs. LowVZ) of participants' task performance (*H4*, *H5*) and digital map operations (*H9*, *H10*).

Based on the above groupings, *Table 4.11* organizes and lists the hypotheses and provides descriptions of each for reference.

Table 4.11 *Hypotheses Tested for the VR/Field Experiment.*

Group	Hypo-thesis	Description of Relationship	Based on
Group 1 RW vs. VR	H1	The RWT group will outperform the VRT group on the tasks.	Task performance variables
	H2	The RWT+HighVZ group will outperform the VRT+HighVZ group.	
	H3	The RWT+LowVZ group will outperform the VRT+LowVZ group.	
	H6	Map usage of the RWT group will differ from the VRT group.	Digital map operation variables
	H7	Map usage of RWT+HighVZ group will differ from VRT+HighVZ.	
	H8	Map usage of RWT+LowVZ group will differ from VRT+LowVZ.	
Group 2 HighVZ vs. LowVZ	H4	The RWT+HighVZ group will outperform the RWT+LowVZ group.	Task performance variables
	H5	The VRT+HighVZ group will outperform the VRT+LowVZ group.	
	H9	Map usage of RWT+HighVZ group will differ from RWT+LowVZ.	Digital map operation variables
	H10	Map usage of VRT+HighVZ group will differ from VRT+LowVZ.	

Task performance variables: Task time, distance traveled, address verification errors, 1st point-to-origin test, 2nd point-to-origin test.
Digital map operation variables: Pans, zooms, pan limit reached, zoom limit reached, address added, address removed, map resets.

Task Performance Outcomes: RWT vs. VRT

We expected that the real-world treatment participants (RWT) would outperform those of the virtual reality treatment (VRT) due to the benefits of an authentic task environment. Hypothesis tests were run to determine if there were significant differences in participants' task performance between treatments. *Table 4.12* and *Table 4.13* summarize the results of hypothesis tests associated with *H1*. The expectation was that the real-world treatment group (RWT) would outperform the virtual reality treatment group (VRT) per the five task performance variables. Only the *2nd point-to-origin test error* showed statistically significant differences between the two treatments. As expected, the error angles made by participants as they pointed to their origin for the second and final pointing test were greater for the VRT group (M = 35.4°, SD = 43.8°) than for the RWT group (M = 17.3°, SD = 16.0°). Although great pains were taken to simulate walking, to incorporate landmarks, and to accurately recreate the layout and scale of the real-world neighborhood, participants took advantage of these features only through visual input and through the optic flow of the virtual environment; limited proprioceptive cues were available through the CAVE and locomotion interface, thus, the spatial updating that one takes advantage of during normal walking was not available in the CAVE. The literature frequently cites the importance of inputs from natural walking (e.g., proprioception, kinesthetic and vestibular information) to facilitate spatial orientation and path integration (Chance, Gaunet, Beall, & Loomis, 1998; Klatzky & Loomis, 1998; Ruddle, Volkova, & Bülthoff, 2011; Waller & Hodgson, 2013). The pointing errors that participants made in the VRT could be attributed to a lack of these inputs. There were no significant differences between the VRT and the RWT for the *1st point-to-origin test error*, which suggests that participants made pointing errors to a lesser degree earlier in the exercise, however, as more time passed and as participants traveled deeper into the neighborhood, the discrepancy grew in terms of their ability to spatially update their position within the neighborhood with respect to the starting location; this is a likely explanation as to why there were significant differences between the two treatments for the second pointing test and not the first.

Table 4.12 *Summary of Mann-Whitney U Tests Comparing Task Performance Variables Between Treatments.*

	Real World		Virtual Reality		Test Statistics		
Performance Variable	***n***	**Mean Rank**	***n***	**Mean Rank**	***U***	***Z***	***p***
Task time (min.)	16	14.5	14	16.6	96	-.64	.522
Distance traveled (mi.)	15	13.5	15	17.5	83	-1.20	.230

* $p < .05$; rejected H_0: There is no significant difference in the distribution of scores between the RWT & VRT.

Table 4.13 *Summary of Welch's and Independent t-tests Comparing Task Performance Variables Between Treatments.*

	Real World		VR		Test Statistics		
Performance Variable	**Mean**	**SD**	**Mean**	**SD**	***t***	***d.f.***	***p***
Address verification errors	3.1	2.0	2.5	1.3	.933	26.0	.360
1st point-to-origin test error	14.7	14.0	37.6	54.5	-	-	-
2nd point-to-origin test error	17.3	16.0	35.4	43.8	-	-	-
Log(1st point-to-origin test error)	1.01	.42	1.15	.68	.696	21.6	.503
Log(2nd point-to-origin test error)	1.03	.48	1.37	.39	2.05	28.0	.050*

* $p < .05$; rejected H_0: $\mu_{rw} = \mu_{vr}$.
Welch's *t*-test used for *address verification errors* and *1st point-to-origin test error*.
Independent-samples *t*-test used for *2nd point-to-origin test error*.

Table 4.14 summarizes the results of hypothesis tests associated with *H2* and *H3*. The expectation of *H2* is that, for the high spatial visualization group, participants who completed the tasks in the real-world treatment (RWT+HighVZ) would outperform those who did so in the virtual reality treatment (VRT+HighVZ) per the five task performance variables. Only *distance traveled* showed statistically significant differences between the two treatments for the HighVZ group. The distances participants traveled were significantly greater for the VRT+HighVZ group (Mdn = 1.33 mi.) than for the RWT+HighVZ group (Mdn = 1.02 mi.). Observers in the VRT noted a key issue with the locomotion interface where many participants "drifted" unintentionally; in other words, they traveled at low speeds in a given direction even though they intended to come to a stop—this was due to those participants being off-center of the locomotion dead zone within the CAVE. These occurrences were most common when participants interacted with the mobile device and software. During such interactions, VRT participants typically came to a physical stop, but they did not pay attention to the unintended locomotion that was happening within the virtual neighborhood. It will be important to minimize or eliminate the drifting occurrences in future uses of the VR lab so that the calculated travel distances are more accurate.

The expectation of *H3* is that, for the low spatial visualization group, participants who completed the tasks in the real-world treatment (RWT+LowVZ) will outperform those who did so in the virtual reality treatment (VRT+LowVZ) per the task performance variables. There were no task performance measures that showed statistically significant differences between treatment environments for the low spatial visualization group.

Table 4.14 *Summary of Mann-Whitney U Tests Comparing Task Performance Variables Between Treatments (by VZ Group).*

	Performance Variable	**Real World**		**Virtual Reality**		**Test Statistics**		
HighVZ Group		***n***	**Mean Rank**	***n***	**Mean Rank**	***U***	***Z***	***p***
	Task time	8	7.7	7	8.4	25.5	-.23	.818
	Distance traveled	7	4.7	7	10.3	5.0	-2.43	.015*
	Address verification errors	8	8.1	7	7.9	27.5	0	1.00
	1st point-to-origin test error	8	8.1	7	7.9	27.0	.06	.952
	2nd point-to-origin test error	8	6.7	7	9.5	17.5	-1.16	.246
LowVZ Group		***n***	**Mean Rank**	***n***	**Mean Rank**	***U***	***Z***	***p***
	Task time	8	8.9	9	9.1	35.0	-.05	.960
	Distance traveled	8	10.3	9	7.9	26.0	-.91	.363
	Address verification errors	8	10.6	9	7.6	23.5	-1.15	.250
	1st point-to-origin test error	7	7.6	9	9.2	22.0	.64	.522
	2nd point-to-origin test error	8	7.2	9	10.6	21.5	1.35	.177

* $p < .05$; rejected H_0: There is no significant difference in the distribution of scores between the RWT & VRT.

Out of the fifteen tests that were run against the address verification and pointing task variables, two (13%) showed statistically significant differences ($p < .05$). Fewer overall significant differences were identified in the hypothesis tests than was expected. The boxplots from *Section 4.7.1* show a trend where the RWT group participants tend to have lower averages across the task performance metrics; this includes shorter task times, shorter travel distances, and smaller error angles for the pointing tests (i.e., better performance). Only *address verification errors* showed similar averages between the two treatments. A reasonable explanation is that although the boxplots show evidence of treatment effects, the hypothesis tests did not have the statistical power needed to detect differences between the treatment environments. Furthermore, the features included in the VR lab (see *Table 4.10*) altogether minimized the treatment effects that otherwise would have had a stronger impact in a lab setting of inferior ecological validity.

Figure 4.35 is a graph that summarizes which treatment groups had the best average task performance across the variables that comprise the address verification task and both pointing tasks. Notice that for the address verification task, the RWT group and the VRT group were nearly equal in terms of how often either group demonstrated better average task performance over the other. However, the pointing tasks—where participants were prompted to point to their origin on two occasions during the study—resulted in the RWT group outperforming the VRT group over 80% of the time. This may indicate that participants of the VRT had greater difficulty with spatial orientation and path integration (an important subcomponent of the address verification task), which could be explained by the lack of proprioceptive, kinesthetic and vestibular inputs in the VR lab (Chance et al., 1998; Klatzky & Loomis, 1998; Ruddle et al., 2011; Waller & Hodgson, 2013).

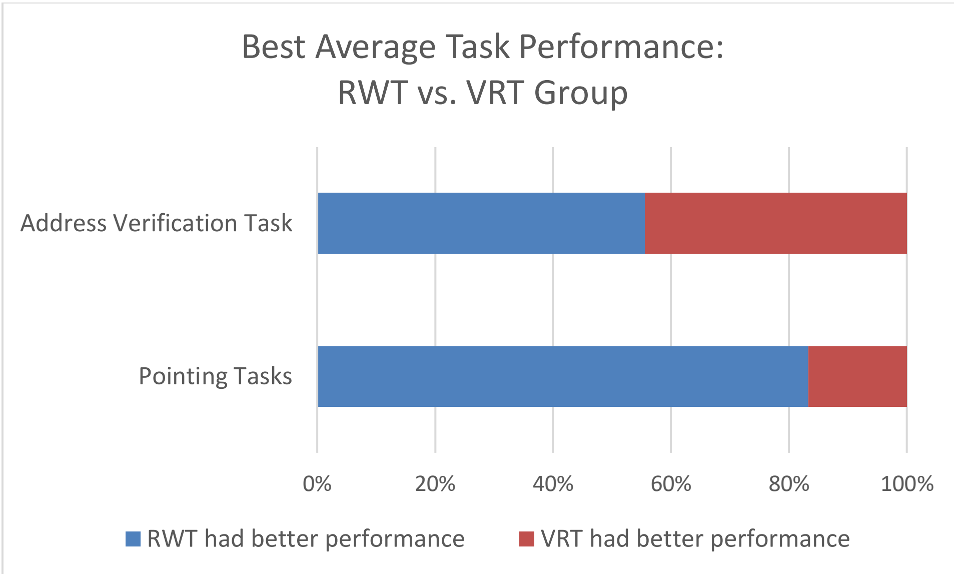


Figure 4.35 *Graph showing participant groups (based on treatment) with the best average task performance across the variable categories.*

Task Performance Outcomes: HighVZ vs. LowVZ

We expected that participants of the high spatial visualization group (HighVZ), due to their spatial cognition advantages, would outperform those of the low spatial visualization group (LowVZ) in both treatment environments. *Table 4.15* summarizes the results of hypothesis tests associated with *H4* and *H5*. The expectation of *H4* is that, for the real-world treatment, participants of the high spatial visualization group (RWT+HighVZ) will outperform those of the low spatial visualization group (RWT+LowVZ) per the five task performance variables. Two out of the five task performance variables showed statistically significant differences between VZ groups: *Task time* and *distance traveled*.

Task times were significantly greater for the RWT+LowVZ group (Mdn = 43.5) than for the RWT+HighVZ group (Mdn = 29.0). This indicates that in the RWT, participants from the LowVZ group took longer to complete the address verification task than their HighVZ counterparts. The distances traveled by participants were significantly greater for the RWT+LowVZ group (Mdn = 1.29 mi.) than for the RWT+HighVZ group (Mdn = 1.02 mi.).

This indicates that participants from the LowVZ group took longer to complete the address verification task than their HighVZ counterparts. The HighVZ group has demonstrated superior task performance over the LowVZ group for both task time and distance traveled in the real world. These results are consistent with the task performance outcomes discussed in the *FE1 Results Section* and the *FE2 Results Section* of Chapter 3.

The expectation of *H5* is that, for the virtual reality treatment, participants of the high spatial visualization group (VRT+HighVZ) will outperform those of the low spatial visualization group (VRT+LowVZ) per the five task performance variables. Task times were significantly greater for the VRT+LowVZ group (Mdn = 43.0) than for the VRT+HighVZ group (Mdn = 27.0). Task time is the only performance metric that showed statistically significant differences between VZ groups in both treatment environments. It is unusual that the VRT showed significance for *task time*, but not *distance traveled*; these two variables have shown a strong positive correlation in the RWT of this experiment and in our prior address verification studies. A review of the boxplots for distance traveled (see *Figure 4.16*) shows the expected trend between the RWT+HighVZ (shorter distances) and RWT+LowVZ groups, however, for the VRT we see similar median travel distances between the HighVZ and LowVZ groups. This may be the result of the drifting effect that was observed in the locomotion interface of the VR lab.

Table 4.15 *Summary of Mann-Whitney U Tests Comparing Task Performance Variables Between VZ Groups by Treatment.*

	Performance Variable	HighVZ Group		LowVZ Group		Test Statistics		
Real World		***n***	**Mean Rnk**	***n***	**Mean Rnk**	***U***	***Z***	***p***
	Task time	8	5.9	8	11.1	11.5	-2.10	.036*
	Distance traveled	7	5.0	8	10.6	7.0	2.37	.018*
	Address verification errors	8	6.5	8	10.5	16.0	-1.63	.103
	1st point-to-origin test error	8	8.9	7	6.9	20.5	.81	.418
	2nd point-to-origin test error	8	7.6	8	9.4	25.0	-.68	.497
Virtual Reality		***n***	**Mean Rnk**	***n***	**Mean Rnk**	***U***	***Z***	***p***
	Task time	7	5.2	9	11.1	8.5	2.38	.017*
	Distance traveled	7	9.9	9	7.4	22.0	-.95	.342
	Address verification errors	7	7.0	9	9.7	21.0	1.06	.289
	1st point-to-origin test error	7	8.1	9	8.8	28.5	.27	.795
	2nd point-to-origin test error	7	6.4	9	10.1	17.0	1.48	.139

* $p < .05$; rejected H_0: There is no significant difference in the distribution of scores between the HighVZ & LowVZ groups.

Out of the ten hypothesis tests that were run to determine if there were differences in task performance between the HighVZ and LowVZ groups, three (30%) showed statistically significant differences ($p < .05$) between spatial visualization groups. The results indicated that in the real-world treatment: (1) HighVZ participants traveled significantly shorter distances than the LowVZ group and (2) HighVZ participants had significantly shorter task completion times than the LowVZ group. In the virtual reality treatment: (3) HighVZ participants also traveled significantly shorter distances than the LowVZ group. These three outcomes are in accordance with the hypotheses. Again, fewer overall significant differences were identified in the hypothesis tests than was expected. This was unforeseen given how often spatial visualization ability has been linked to users' proficiency on a variety of computing tasks. To delve deeper, we graphed the participant groups with the best average task performance across the variables of the experiment (see *Figure 4.36*). Notice that the

HighVZ group demonstrated better average task performance—more than 75% of the time—across the two categories of task performance variables. This implies that effects associated with the participant groups' spatial visualization ability may exist, but the statistical power of the hypothesis tests was insufficient to discern significant differences between the groups.

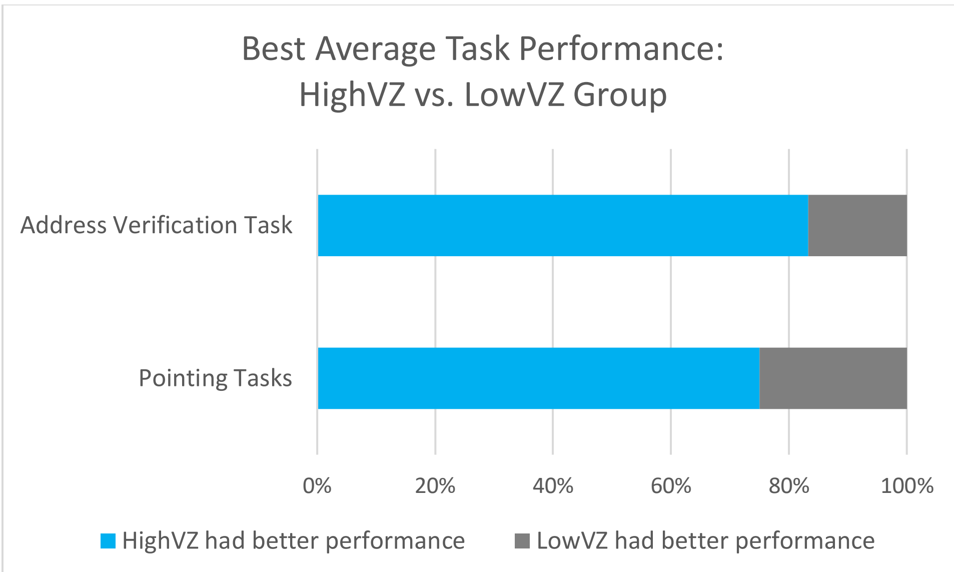


Figure 4.36 *Graph showing participant groups w/ best average task performance across the variable categories.*

Further evidence of task performance differences between the participant groups can be seen in many of the boxplots where the participant groups and treatment environments have been crossed. In such cases, the locations of the HighVZ participants' boxplots for a given variable tend to be higher or lower than those of the LowVZ group (depending on which location indicates better task performance); this is generally true in the cases of both treatments. For example, *Figure 4.15* shows boxplots of participants' *task times* for the crossed experimental groups. In both treatment environments, the HighVZ group demonstrates shorter task completion times (i.e., lower boxplot locations and medians) than the LowVZ group. The boxplots for *address verification errors* and the *2nd point-to-origin test error* show similar relationships. *Distance traveled*, which did not exhibit this relationship, may have been negatively impacted by the drifting anomalies that were identified in the locomotion interface.

Digital Map Operation Outcomes: RWT vs. VRT

Table 4.12 summarizes the results of hypothesis tests associated with *H6*. The expectation was that the RWT group's digital map operations would be significantly different from the VRT group. Two out of the seven digital map operations showed statistically significant differences: *Pans* and *addresses removed*.

The panning operations were significantly greater for the VRT group (Mdn = 68.0) than for the RWT group (Mdn = 40.0) as was expected. Thus far, we have associated increased panning with poorer performance. Boxplots of the panning operations—broken down by crosses of the treatment environments and participant groups (see *Figure 4.27*)—reveal several trends. Notice that in both treatment environments, the HighVZ group has logged greater median panning operations than the LowVZ group. This would typically be an indicator that increased panning may be a function of thoroughness (or some other favorable strategy) rather than confusion given the fact that the HighVZ group has consistently outperformed the LowVZ group across the task performance metrics in our studies. However, while the HighVZ groups do have greater median panning operations, the LowVZ groups have greater means, which can be seen in the right-skewed boxplots in addition to the presence of outliers that have extreme panning frequencies (> 100); these trends are in line with observations that indicated that increases in participants' panning can be attributed to confusion and difficulty. Such behavior was first discussed in *Field Experiment 1* of the previous chapter, where we observed an increase in LowVZ participants' panning operations due to them having difficulty with navigation and during other situations that caused confusion (see *Navigation Phase Section*); similar behaviors are discussed in this chapter's task workflow section covering the *Navigation Phase behaviors*.

Figure 4.37 shows the mean frequency of digital map operations across the various experimental groups. Notice that the RWT participants have a significantly lower number of panning operations on average than the VRT participants. Furthermore, the LowVZ participants pan more on average than the HighVZ participants. Taken together, these graphs imply that participants who belong to the LowVZ group and those who complete the tasks in the VRT are likely to pan their maps more, which is a behavior that we have shown to be a function of increased difficulty. Thus, it makes sense that the VRT+LowVZ group has logged the greatest number of average pans than any other experimental group; it is plausible that the greater degree of difficulty encountered by this group can be sourced to (1) the lower spatial visualization ability of the participants and (2) performance-hindering treatment effects of the virtual reality environment.

The six verification scenarios are comprised of one address that needed to be removed and five that needed to be moved—remember, the move scenarios require participants to both add a map spot and to remove the pre-existing map spot for each scenario. Thus, in order for participants to correct their maps without errors, they would have to remove all six map spots (i.e., the software should log six *address removed* operations) and add five addresses back to the map in the correct locations per the move scenarios. This explains why in most cases, the more addresses that participants added and removed for this experiment, the fewer errors they tended to make. The frequency with which participants removed addresses from their maps was significantly greater for the VRT group (Mdn = 4.0) than for the RWT group (Mdn = 2.0). Though the VRT environment was designed to replicate the real neighborhood with high fidelity, it simply did not have all of the features and complexities of the real-world neighborhood, many of which could negatively impact participants' ability to make map corrections (e.g., occluding objects, traffic, and other activities that might compete with the objective). This may explain the superior performance of the VRT group—they had less distractions and sources of confusion while locating residences and correcting the map in the virtual neighborhood, thus, it was easier for them to identify and remove addresses that were incorrectly placed on the map.

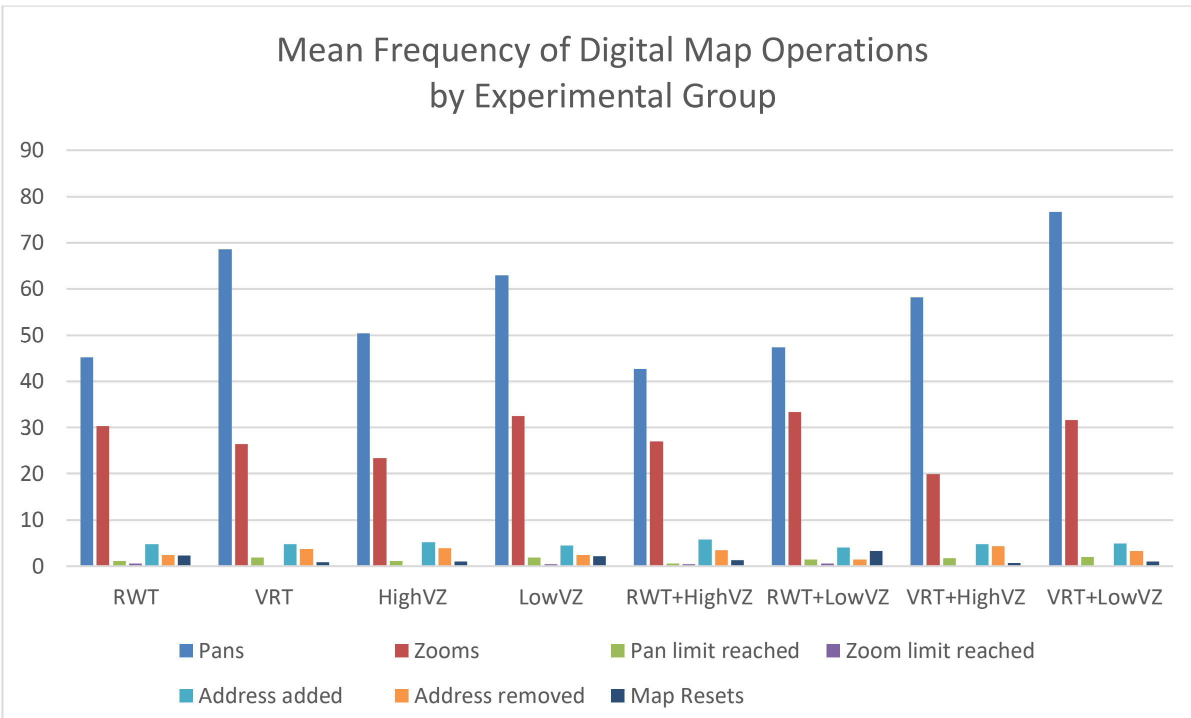


Figure 4.37 *Graph showing average frequency of participants' digital map operations by experimental group.*

Table 4.16 *Summary of Mann-Whitney U Tests Comparing Map Operation Variables Between Treatments.*

Performance Variable	Real World		Virtual Reality		Test Statistics		
	n	**Mean Rank**	***n***	**Mean Rank**	***U***	***Z***	***p***
Pans	15	11.6	15	19.4	54.5	-2.38	.017*
Zooms	15	16.7	15	14.3	95.0	.71	.478
Pan limit reached	15	13.9	15	17.1	89.0	-.95	.342
Zoom limit reached	15	17.1	15	13.9	88.0	1.00	.317
Addresses added	15	16.4	15	14.6	99.5	.52	.603
Addresses removed	15	12.2	15	18.8	63.0	-2.03	.042*
Map resets	15	17.2	15	13.8	87.0	1.04	.298

* $p < .05$; rejected H_0: There is no significant difference in the distribution of scores between the RWT & VRT.

Table 4.17 summarizes the results of hypothesis tests associated with *H7* and *H8*. The expectation of *H7* is that, for the high spatial visualization group, participants digital map operations in the real-world treatment (RWT+HighVZ) would be significantly different from those who complete the tasks in the virtual reality treatment (VRT+HighVZ) per the subset of digital map operation variables. There were no digital map operations from the subset that showed statistically significant differences between the two treatments for the HighVZ group.

The expectation of *H8* is that, for the low spatial visualization group, participants who completed the tasks in the real-world treatment (RWT+LowVZ) would have significantly different map operations than those who did so in the virtual reality treatment (VRT+LowVZ). Only one out of the subset of map operations, *pans*, showed statistically significant differences between the RWT and the VRT for the LowVZ group. As expected, the panning operations were significantly greater for the VRT+LowVZ group (Mdn = 62) than for the RWT+LowVZ group (Mdn = 37). As stated before, the LowVZ participants in both treatments used the panning operation more frequently on average, which implies that higher panning frequencies signify poorer task performance for participants in general. In accordance with the conclusion drawn for *H6*, it seems that participants from the VRT have greater difficulty with the task as it relates to map use as can be seen by the greater frequency of pans from the VRT+LowVZ group when compared their RWT+LowVZ counterparts.

Table 4.17 *Summary of Mann-Whitney U Tests Comparing Map Operation Variables Between Treatments (by VZ Group).*

	Performance Variable	Real World		Virtual Reality		Test Statistics		
HighVZ Group		***n***	**Mean Rank**	***n***	**Mean Rank**	***U***	***Z***	***p***
	Pans	7	6.1	7	8.9	15.0	-1.15	.250
	Zooms	7	8.5	7	6.5	17.5	.83	.407
	Addresses added	7	9.3	7	5.7	12.0	1.53	.126
	Addresses removed	7	6.4	7	8.6	16.5	-0.96	.337
LowVZ Group		***n***	**Mean Rank**	***n***	**Mean Rank**	***U***	***Z***	***p***
	Pans	8	6.3	9	11.4	14.0	2.07	.038*
	Zooms	8	9.2	9	8.8	34.5	-0.10	.920
	Addresses added	8	7.9	9	9.9	27.5	0.77	.441
	Addresses removed	8	6.6	9	11.1	17.0	1.78	.075

* $p < .05$; rejected H_0: There is no significant difference in the distribution of scores between the RWT & VRT.

Digital Map Operation Outcomes: HighVZ vs. LowVZ

Table 4.18 summarizes the results of hypothesis tests associated with *H9* and *H10*. The expectation of *H9* is that, in the real-world treatment, participants of the high spatial visualization group (RWT+HighVZ) will have a significantly different frequency of map operations than those of the low spatial visualization group (RWT+LowVZ) per the subset of digital map operation variables. Only one out of the subset of map operation variables showed statistically significant differences between VZ groups: *Address added*. The frequency with which participants added addresses to their maps was significantly greater for the RWT+HighVZ group (Mdn = 6.0) than for the RWT+LowVZ group (Mdn = 5.0). Participants should have added a total of five addresses to their maps in order to correct them. Generally speaking, the more addresses that participants added, the fewer verification errors they tended to make. The address verification error boxplots show that the HighVZ participant groups also have fewer address verification errors in both treatment environments (see *Figure 4.17*). It would seem that a greater frequency of *address added* operations tends to indicate better task performance, thus, it could be argued that the RWT+HighVZ group has demonstrated superior task performance to the RWT+LowVZ group as was expected.

The expectation of *H10* is that, in the virtual reality treatment, participants of the high spatial visualization group (VRT+HighVZ) will have a significantly different frequency of map operations than those of the low spatial visualization group (VRT+LowVZ) per the subset of digital map operation variables. There were no digital map operations from the subset that showed statistically significant differences between the two participant groups in the virtual reality treatment.

Table 4.18 *Summary of Mann-Whitney U Tests Comparing Frequently Used Map Operation Variables Between VZ Groups by Treatment.*

	Performance Variable	**HighVZ Group**		**LowVZ Group**		**Test Statistics**		
Real World		***n***	**Mean Rank**	***n***	**Mean Rank**	***U***	***Z***	***p***
	Pans	7	9.1	8	7.0	20.0	-0.87	.384
	Zooms	7	6.7	8	9.1	19.0	0.98	.327
	Addresses added	7	10.5	8	5.8	10.5	-1.97	.049*
	Addresses removed	7	10.3	8	6.0	12.0	-1.79	.073
Virtual Reality		***n***	**Mean Rank**	***n***	**Mean Rank**	***U***	***Z***	***p***
	Pans	7	7.8	9	9.1	26.5	0.48	.631
	Zooms	7	6.4	9	10.2	16.5	1.53	.126
	Addresses added	7	8.0	9	8.9	28.0	0.32	.749
	Addresses removed	7	10.0	9	7.3	20.5	-1.11	.267

* $p < .05$; rejected H_0: There is no significant difference in the distribution of scores between the HighVZ & LowVZ groups.

4.8.3 Qualitative Analysis

The mixed-method design of the studies enabled us to examine qualitative data to complement the quantitative analysis. Observations from the RWT and the VRT together were used to devise a task workflow for address verification as was done in our prior studies. This is the third and most comprehensive workflow generated for the task. Observations from both the real-world study and the VR lab study revealed behaviors that were germane to participants' ability to effectively plan, navigate, and verify residences within the neighborhood using the mobile device and map-based address verification software. Both treatment environments enabled researchers to identify software usability issues. In this regard, the VR lab was particularly effective because participants were able to use the actual mobile device and software to perform tasks in a virtual environment that provided enough context to draw out real-world behaviors and actions.

Observers lauded the VR lab because it was easier for them to shadow participants, observe behavior and discern participants' interactions with the software without the distractions that were encountered in the real-world environment. There was also a key difference in the richness of observations that were recorded in the two treatment environments. Despite a greater number of total observations being recorded in the RWT (1167) when compared to the VRT (787), the VRT yielded more distinct observation types (35% of the total) than the RWT (13%). This can be attributed to the increased attention that observers were able to direct toward participants in the VR lab. When observers were in the RWT they recorded a greater number of overlapping and redundant notes, whereas in the VRT observers were more selective about the notes that they took. Aside from these differences, more than half of the total number of observation types recorded during this experiment (52%) were identified in both the RWT and the VRT, which shows alignment between the qualitative observations that were captured by observers in both treatments.

4.8.4 Observer Feedback

When observers were asked to compare and contrast their data collection experiences between the two treatment environments, they praised the authenticity of the RWT while noting drawbacks such as the lack of experimental control, the inability to work in adverse weather conditions, the difficulty of shadowing participants on foot (across travel distances that sometimes approached three miles), and safety concerns. The VRT mitigated these challenges and facilitated a greater degree of experimental control, yet it retained important characteristics of the real-world environment and enabled key participant-observer interactions, thus, improving the ecological validity of a lab study.

4.9 Conclusion

In this experiment, we have shown that a comprehensive, mixed-methods field study on mobile computing that is conducted in a real-world setting can be replicated in a VR lab with worthwhile results. The VR lab was driven by a CAVE and a high-quality 3D model replica of the study neighborhood, which enabled key interactions between participants, the task environment, a mobile computer, and attending researchers who were able to shadow participants and record observations as was done in the real-world setting. Rich sets of quantitative and qualitative data were collected in both treatment environments, allowing for the same manner of task analysis and user analysis to be done.

The use of stratified random sampling in both treatment environments enabled us to observe how two distinct user groups (i.e., participants with high and low spatial visualization ability) approached the location-based task of address verification. Task performance trends and behavioral observations associated with the participant groups were oftentimes consistent between the two treatment environments; our results were also consistent with findings from the prior field experiments (*FE1* and *FE2*). In general, this experiment has shown that participants with high spatial visualization ability demonstrate superior task performance and more effective strategies while completing their tasks.

When comparing the treatment environments, participants of the real-world treatment typically outperformed their virtual reality treatment counterparts (based on task performance averages); however, many measures did not show statistically significant differences between the treatments—this is thought to be due to small treatment effects, which speaks to the fidelity that we were able to achieve in the VR lab. Furthermore, by looking at crosses of the treatment environments (RWT/VRT) and the participant groups (HighVZ/LowVZ), we were better able to compare and contrast field study outcomes. Comparing study outcomes between the crossed experimental groups not only allowed us to identify and validate certain results (e.g., significantly faster task completion times for high spatial visualization participants in both treatments), but it also helped us to identify issues with the VR lab (e.g., erroneous travel distance calculations due to “drifting”) that should be addressed in future lab evaluations.

In a similar experiment, researchers compared outcomes of a mobile computing task in a “virtual field environment” to those of a real field environment with respect to measures of presence, usability, and user experience (Brade et al., 2017). In our approach, in-depth comparisons were made between outcomes of the two treatment environments based on (a) statistical analysis and hypothesis tests of users’ quantitative task performance metrics, (b) user and task analysis of participants who were screened into either high or low spatial visualization ability groups, and (c) findings from qualitative observations of participants by researchers who shadowed them in both environments. Whereas Brade et al. (2017) saw enough parallels across measures of presence, usability, and user experience between their treatment environments to prompt them to validate the virtual environment, we saw enough parallels in participants' task performance, in their patterns of mobile computer and software use, in their approaches to the task (and their strategies), as well as in researchers' data collection experiences in both environments—these outcomes allow us to reach a level of confidence in using our VR lab as a means to thoroughly study a mobile computing task of this nature. Rather than conduct a field study and a lab study to realize the distinct benefits of both (a resource intensive endeavor), we have shown that these methods can be combined into a single lab study using immersive virtual reality to get the best of both worlds.

CHAPTER 5 – DISCUSSION

5.1 Introduction

This dissertation has covered various lab and field evaluation methods that were used to examine location-based tasks that involve the use of a map. *Table 5.1* lists the relevant research activities and their characteristics. This chapter summarizes key findings across the studies, it reflects on the various methods that have been employed, applications and future work are discussed, followed by a conclusion.

Table 5.1 *Overview of All Studies on Location-based Tasks (listed chronologically).*

	Research Method					Data Method		Environmental Context			Spatial Viz.		Map Type	
	Setting		Type											
Study or Experiment	Field	Lab	Ethnography	Experiment	Observ. Study	Qualitative	Quantitative	Natural setting	Photo-based	Immersive VE	Tested on it.	Screened on it.	Digital	Paper
Rusch's Lab Experiment[1]		X		X			X		X		X		X	
Ethnographic Field Study[2]	X		X			X		X					X	
Field Experiment 1 (oblique streets)	X			X	X	X	X	X			X	X		X
Field Experiment 2 (oblique + grid)	X			X	X	X	X	X			X	X		X
VR/Field Experiment	X	X		X	X	X	X	X		X	X	X	X	

[1] A prior lab experiment used for comparison (Rusch et al., 2012; Rusch, 2008); related, but not an experiment covered in this dissertation.
[2] A prior field study used for comparison; related, but not a study covered in this dissertation.

5.2 Task Outcomes

5.2.1 Spatial Ability and Task Outcomes

Rusch's Lab Experiment was the first to take participants' cognitive abilities into account (Batinov et al., 2015; Rusch et al., 2012; Rusch, 2008). Prior to the experiment, participants were tested on their spatial visualization, logical reasoning, and perspective-taking abilities. The results indicated that participants with higher spatial visualization scores tended to have faster task completion times and less frequent digital map operations (i.e., pans, zooms, and map resets). In *Rusch's Lab Experiment*, a stationary room was used to simulate a very small portion of the real-world task. In the actual scenarios, users would need to interpret a map, navigate a neighborhood on foot, and verify the correctness of residential addresses on a map. In the subsequent studies (with the exception of the *Ethnographic Field Study*), measures of spatial ability became more central to study designs and outcomes. Our group was interested in evaluating two specific types of users: Those with high spatial visualization ability (HighVZ) and those with low spatial visualization ability (LowVZ). Stratified random sampling was used to isolate HighVZ and LowVZ participants in *Field Experiment 1 (FE1)*, *Field Experiment 2 (FE2)*, and the *VR/Field Experiment (VRFE)*.

In *FE1*, participants navigated a neighborhood to find six residences and to ensure that they were correctly shown on a paper map. We hypothesized that participants who were screened into the high spatial visualization group (HighVZ) would demonstrate superior task performance and more effective strategies than those of the low spatial visualization group (LowVZ). This generally proved to be the case—not only for *FE1*, but also for *FE2* as well as the two treatment environments used in the *VRFE*. This meant that—regardless of the type of map that participants used (paper/digital), or the street layout of the study neighborhood (grid-like vs. oblique), or whether or not participants were in the field or the VR lab treatment environment—the participants with high spatial visualization ability typically outperformed and out-strategized their LowVZ counterparts on the address verification task across the studies. This is an outcome that we have generally come to expect; one that could be used in similar location-based tasks to better drive the evaluation and development of systems and interfaces.

In HCI and related disciplines, there are benefits to studying novice users (e.g., to gauge a system's intuitiveness or learnability) as well as benefits to studying experts (e.g., to gauge a system's depth and adaptability). However, this is difficult with a task such as ours (address verification) because it is, by nature, a seasonal task where fieldworkers have historically been very diverse in their backgrounds, education, experience, age, and other demographics. The reality is that advanced and expert users are difficult to come by for such a task. However, through our field studies and lab experiments, we have shown that researchers can gain valuable insights by examining how users who differ in spatial ability approach the underlying tasks. Though the HighVZ and LowVZ user groups do not equate to expert and novice users in our studies, there were discernable differences between their uptake of the task, their task performance and their behavioral patterns. These characteristics enabled us to identify aspects of the task workflow and interface design that were both productive and counterproductive for participants. This approach, of considering participants' spatial visualization ability, could be used to better evaluate and engineer solutions for mobile computing systems and interfaces designed to support location-based tasks that involve map use.

5.2.2 The Environmental Context and Task Outcomes

The layout of the study neighborhood was manipulated throughout the field experiments. In *Field Experiment 1 (FE1)*, the task environment consisted of a subsection of the study neighborhood that was difficult to navigate. This area was chosen deliberately to challenge participants. The study neighborhood was modified in *Field Experiment 2 (FE2)* to include a section of streets and residences that were more grid-like in their layout—the *FE2* neighborhood had a grid-like half located to the west and to the east was the more confusing area originally used in *FE1* (see *Figure 5.1*). Based on researchers' observations and participants' responses to exit questionnaires, we learned that the grid-like half of the neighborhood was easier and less confusing for participants to work than the non-uniform half.

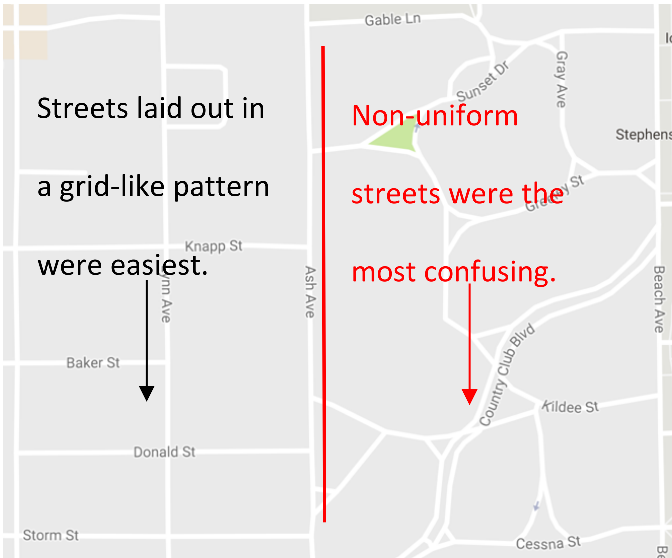


Figure 5.1 *This map exemplifies the grid-like layout of streets in the western half of the FE2 neighborhood compared to the non-uniform, eastern half, which proved to be more difficult for participants during the tasks.*

Participants struggled with spatial orientation and their interpretation of the map at intersections in the eastern half, where they would spend more time reviewing their maps. During such times, participants were observed to repeatedly rotate their maps and/or bodies at intersections to reorient themselves—this behavior is consistent with literature that has shown that participants can switch between map alignment strategies when disrupted under cognitively demanding conditions (Aretz & Wickens, 1992). Similar behaviors were observed in both treatment environments of the *VR/Field Experiment (VRFE)*, which was situated in the confusing area of the neighborhood set to the east. Thus, changes in the difficulty of the task environment were expressed in participants' behaviors and task performance outcomes.

The environmental context was also an important consideration in terms of the research methods employed. In *FE1* and *FE2*, we shadowed participants who walked a real neighborhood to complete a series of location-based tasks. By evaluating participants in the actual task environment, we got a better sense of their approach. For example, a task workflow was identified in each of the field experiments, which described behaviors related to participants' planning, navigation and the manner in which they verified residential addresses while traversing the study neighborhood. In contrast, *Rusch's Lab Experiment (RLE)* revealed very little about participants' task workflow due to the setting: Participants used a mobile computer and map-based software while seated at an office desk rather than while walking a neighborhood—during *RLE*, interaction with the task environment was limited to photos of neighborhood vantage points that participants referenced. There was a great deal of experimental control during *RLE*, but that control came at the expense of researchers' ability to observe real-world behavior and task performance outcomes in a natural setting. This tradeoff between the experimental control of the lab and the ecological validity of the field was addressed in the *VRFE*.

For the *VRFE*, we used an immersive virtual environment to better simulate the task environment in a lab. An intuitive locomotion interface was used by participants to traverse a 3D replica of the study neighborhood, during which time participants could freely explore their virtual surroundings and interact with map-based software that was displayed on a mobile device that they carried. This approach allowed us to maintain a high degree of experimental control in a lab setting, while drawing out real-world behaviors. Similar to the field experiments, the rich environmental context of the VR lab exposed participants' task workflows, which was something that the less-immersive office lab was unable to do. Many observations from the field studies were also evident in the VR lab, such as the tendency for participants to come to a stop when interacting with their maps or the negative effects that the smaller digital maps had on participants' map usage and planning behavior—these behaviors are consistent with other studies on map use (Ishikawa et al., 2008).

5.2.3 Map Use and Task Outcomes

Certain map usage preferences and behaviors were consistent across the research. The majority of participants tended to prefer the map in a track-up orientation throughout the studies. However, both north-up and track-up map users were observed rotating their maps to reorient themselves at confusing intersections and in difficult areas of the neighborhood. Participants indicated that maps should include more street labels because it was sometimes difficult for them to keep track of street names at intersections and across long routes. Participants also indicated that the inclusion of prominent landmarks on the maps would have been helpful—for example, the paper map users of *Field Experiment 1 (FE1)* and *Field Experiment 2 (FE2)* took it upon themselves to draw landmark indicators onto their maps.

A noticeable increase in participant map usage was observed during navigation, especially when participants lost their bearings. We also observed an increase in map usage when participants were in the final process of verifying a residential address on their map; this was oftentimes motivated by participants' desire to be thorough and accurate, rather than in response to an increase in task difficulty. In general, participants in the high spatial visualization group (HighVZ) had fewer interactions with their maps than did those participants from the LowVZ group; this is likely because HighVZ participants tended to be more efficient when using their maps for planning, navigation, and other task-related activities.

In *Rusch's Lab Experiment* and the *Ethnographic Field Study*, participants used digital maps to complete the location-based tasks. In *FE1* and *FE2*, paper maps were used so that participants would not be restricted by existing software designs and task workflows. A practical consideration when we switched to paper maps was that we lost the software logging features that were available during use of a digital map. Thus, observers recorded notes and took tallies of participants' paper map usage in *FE1* and *FE2*; participants were also surveyed on map usage during the exit questionnaires. In the *VR/Field Experiment (VRFE)*, we returned to the use of digital maps, in the field and in an immersive virtual environment, to complete a similar series of location-based tasks. This gave us an opportunity to compare and contrast the behaviors and task outcomes of the digital map users from the *VRFE* to those of the paper map users from *FE1* and *FE2*.

Some *FE1* and *FE2* participants were observed using the paper maps to create comprehensive plans and routes at the start of the task; we referred to this behavior as pre-planning. Participants with high spatial visualization ability (HighVZ), for example, were observed to more frequently pre-plan and as a result they were generally more efficient. We observed a significant decrease in participants' tendency to pre-plan when we began using the digital maps for the *VRFE*.

Three bar graphs are shown in *Figure 5.2*. The top graph shows the percentage of participants that used paper maps to pre-plan in *FE1* and *FE2*—more than 50% of participants did so. However, when we switched to the digital maps of the *VR/Field Experiment*, the pre-planning frequency of participants dropped from over 50% to approximately 20% in the field (see *Figure 5.2*; middle graph) and in the virtual reality treatment (bottom graph). This reduction in participants' pre-planning frequency indicates that the digital maps were less conducive to pre-planning behaviors. This may be explained by the smaller size of the digital map views, which limited participants ability to see the entirety of the study area as well as their ability to thoroughly plan routes as was afforded to paper map users, who were better able to survey the neighborhood; the work of Ishikawa et al. (2008) demonstrated similar negative impacts that mobile device screens can have on map use.

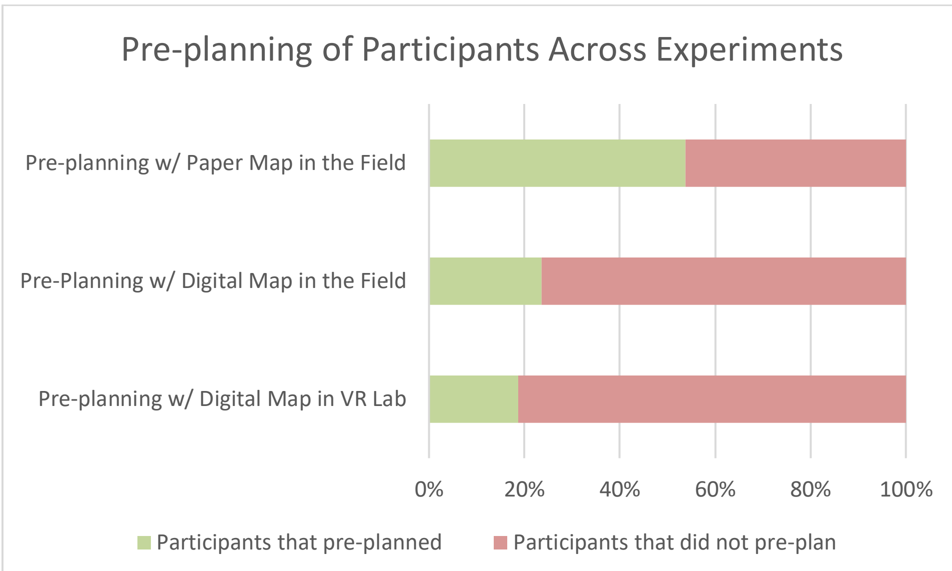


Figure 5.2 *Graphs of participants (across studies) who used either a paper map or digital.*

In addition to demonstrating the effect that digital maps had on pre-planning behaviors, *Figure 5.2* also exemplifies the utility of the VR lab. A goal of building the VR lab was to assess its ecological validity when compared to the field. In a hypothetical scenario, we might have been curious about the impact that moving from a paper map to a digital map would have on participant behavior. *Figure 5.2* shows that the study administered in the VR lab (bottom graph) was able to reflect the effect that digital maps had on participants' pre-planning tendency in the field (middle graph). In both treatment environments of the *VRFE*, there was alignment in the behaviors of the digital map users, who showed a significant reduction in their pre-planning tendencies when compared to the paper map users of the prior field experiments (top graph); this outcome is a testament to the improved ecological validity of the VR lab.

The type of map that participants used had further implications on their actions and behaviors. With paper maps, participants had the ability to better survey the neighborhood (i.e., the entire map could be viewed at once), they had the ability to jot down notes directly on the map and they could trace their routes and add visual indicators to the map (e.g., landmarks) to aid them in completing the task. The rigidity of the digital maps prevented such behaviors in participants. The most frequently used operations for digital map users were the pan and zoom functions. Some digital map users were observed to switch between certain zoom and pan levels depending on their activity. For example, planners and navigators tended to prefer less magnification of the map in order to survey the neighborhood, whereas participants were also observed to zoom-in during the final verification of residences to ensure the correctness of their maps. Paper map users were not burdened by the tedium of the pan and zoom functions, whereas digital map users relied on them in order to achieve views that were conducive to planning, navigation, and verification activities.

5.3 Methodological Considerations

This research highlights the tension that exists when choosing between a lab evaluation or a field evaluation to conduct user research on a location-based task. In the labs that were built for this research, there was the challenge of incorporating the task's environmental context into a lab setting. The studies and experiments conducted in the field presented a different set of challenges (e.g., logistics, mobile data collection and safety). In this section, we compare and contrast: (1) the field setting (i.e., neighborhood) that was used for the field studies; (2) the photo-driven, office lab used in *Rusch's Lab Experiment*; and (3) the VR lab used in the *VR/Field Experiment*.

Table 5.2 summarizes our experiences with respect to the research settings and their characteristics.

Table 5.2 *Characteristics of the Research Methods and Settings Used for This Research.*

	Ecological Validity	Experimental Control	Difficulty of			Resourcing/ Cost ($)
			Preparation	Execution	Data Collection	
Field Setting[1]	Highest	Lowest	High	Highest	Highest	High
Office Lab[2]	Lowest	High	Lowest	Lowest	Lowest	Lowest
VR Lab[3]	High	High	Highest	Moderate	Low	High

[1] Field Setting: Represents the real-world neighborhoods that were used in the field studies.
[2] Office Lab: Represents the photo-based lab that was used in (Rusch et al., 2012; Rusch, 2008) to simulate the task in an office.
[3] VR Lab: Represents the CAVE-based immersive virtual environment that was built for the *VR/Field Experiment*.

5.3.1 Ecological Validity

In terms of ecological validity, the field setting was best suited for evaluating the address verification task because of the task's high dependency on the environmental context—participants' actual behaviors (e.g., planning, navigation) could be observed in an authentic task environment during the field studies. The challenge of the lab studies, on the other hand, was to provide sufficient environmental context (in a static laboratory setting) to draw out real-world behaviors and task outcomes. Between the two lab settings, the VR lab showed much more promise in its ability to immerse participants in a replica of the task environment (a virtual neighborhood) so that real-world behaviors and even some of the expected task performance outcomes rang true.

5.3.2 Experimental Control

Experimental control was high in the office lab because of its static setting. The address verification scenarios used in the office lab followed a strict order, and even now the experiment can be faithfully reproduced in an office of our choosing. Experimental control was also quite high with the VR lab because it too was conducted in a static setting where conditions, for the most part, can be replicated today using a CAVE. In some respects, the experimental control of the VR lab is much greater; for example, if we were interested in comparing study outcomes based on the presence (or absence) of certain landmarks in the study neighborhood, it would be relatively trivial to simply remove them from the virtual neighborhood and re-run the study—this is something that could not be done using the other methods. Of the three approaches, the field study gave us the least amount of control; we were at the mercy of factors such as adverse weather, construction zones in the study neighborhood, inquiries from curious neighbors, and many other potentially confounding factors stemming from the environment.

5.3.3 Difficulty of Implementation

The overall difficulty of executing a user study is a critical consideration for researchers. Our experiences are broken down into three subcategories: The difficulty of preparation, execution, and data collection.

Preparation

Rusch's Lab Experiment required the least amount of preparation out of the three methods. However, it did involve software engineering, computer hardware and networking expertise to display photos of the study neighborhood (synced with the scenarios shown on the mobile device) within the office lab. The field studies were the second most difficult to prepare for as they required a substantial amount of time, effort, and travel to create a robust study protocol that researchers could follow in a mobile context. Preparation of the VR lab was, by far, the most difficult. It involved multiple disciplines, multiple design iterations, testing, and validation over the course of months to ensure its accuracy, proper function and reliability. If such a lab was already built, then of course the preparation required to use it for an evaluation would be much less.

Execution

Once study preparations were complete and participants could be run, the office lab proved to be the least difficult in terms of execution. Only one researcher was necessary to administer the experiment and to ensure that software and hardware were functioning properly—little to no technical intervention was required as everything was set up in a closed-loop network. Logistics were relatively simple—the researcher could conveniently schedule multiple participants on the same day, all of whom met the researcher at the same location. The task was introduced and explained, then participants sat down at a desk and went through several training scenarios. After any clarifying questions were answered, the remaining effort was left to the participant, who could complete the task in a relatively comfortable, seated position at an office desk.

The VR lab study was the second most difficult to execute. There was considerable technical complexity in bringing up the CAVE projectors, launching the simulation application on the cluster, ensuring that the stereo glasses were charged and functioning properly with the optical trackers, and so on and so forth. If this all happened without a hitch, which was usually the case, then both researchers and participants tended to appreciate the convenience of the CAVE-based VR lab. Two or more researchers would be ideal to manage the technical complexity of the lab, to administer the study protocol and to shadow participants as they completed the location-based tasks in a simulated environment.

The VR lab shined during participants' training and during completion of the study tasks. The hands-free locomotion interface was intuitive enough not to distract participants from the task itself and it actually reduced participants' and observers' physical burden of having to walk significant distances through the study neighborhood. Safety was also a non-factor, aside from researchers ensuring that participants did not inadvertently run into the physical walls of the CAVE due to the effects of immersion—which made them feel as if they were in a large neighborhood rather than a 10x10x10-foot lab space. For researchers, there was no need to watch for traffic or other safety hazards and the burden of taking notes while walking behind participants—which proved to be difficult in the field—was all but eliminated. By reducing or eliminating such encumbrances, researchers could direct more of their attention toward participants' think aloud stream, their behaviors and their interactions with the mobile device and environment.

The field study was the most difficult of the three to execute. Bear in mind that during the winter, we could not run participants at all and once the bitter midwestern cold subsided, we were still reliant on favorable weather conditions so as not to confound study results. If the weather permitted, there were then logistical challenges in transporting all parties and materials to the study neighborhood. Two researchers were necessary to administer a strict study protocol—while on foot—in a neighborhood nearby campus. A substantial time and effort commitment were necessary from all parties. The required walking, which oftentimes approached two miles, could physically exhaust both participants and researchers—this was especially true for researchers on days that multiple appointments were scheduled. Altogether, the complexity of facilitating the study while constantly moving was daunting in the field.

Data Collection

Data collection was easiest in the office lab, which was automated; the researcher only needed to initialize the experiment. The VR lab was also automated so that task times, travel distances, and software usage were logged with little effort. Data collection was the most difficult during the field studies, where the burden was on researchers to record task times, travel distances and routes. The field studies required researchers to transport and travel with data collection equipment such as smartphones, GPS, compasses, stop watches and written materials. Furthermore, the field studies required a great deal of multitasking from attending researchers, all while they were expected to shadow participants on foot and record observations. The burden of walking was significantly reduced in the VR lab, which enabled researchers to focus more on observing participants and less on keeping up with them.

5.3.4 Resourcing / Cost

A final consideration is the overall cost of using each of the research methods. The field study cost was high. Major drawbacks of the field studies include the need for more personnel and greater time commitments, increased equipment costs and more complex logistics. Additionally, location-based studies such as ours can raise safety concerns in the field and they are subject to cancellation or delays on account of bad weather or inaccessible study areas—these factors can affect study schedules and further increase the cost of research.

The VR Lab also has a high cost due to the implementation that we chose. Like many CAVEs, in order to stay up and running, the facility that we used required a dedicated lab space that was shared with other researchers; it required clusters of specialized computers, expensive projectors, complex software, and the availability of technical experts for troubleshooting if anything went awry. These factors tend to make CAVEs rare and less accessible to researchers, which can factor into costs.

The office lab, on the other hand, required significantly less preparation, it was comparatively low in its complexity, and it could be staffed and replicated with relative ease—all of which contribute to a reduction in the overall cost.

5.4 Limitations

The experiments had one independent variable in common: The participant groups. In each experiment, participants were screened into a high spatial visualization group (HighVZ) and a low spatial visualization group (LowVZ). Researchers were able to focus on the outcomes of participants with polar opposite spatial visualization abilities, however, this excluded participants who fell somewhere in-between, which limits the generalizability of the results. All of the experiments make use of the think aloud method, which could have created unnatural situations that may have biased or altered user behavior and thought processes—possibly influencing some of the study findings. The observations recorded by attending researchers for each of the experiments were also susceptible to these effects in addition to the bias of researchers' interpretation; though researchers did review coding schemas for their observations on multiple occasions, they lacked the time and resources to ensure intercoder reliability. The field settings of the studies were subject to variations in weather, differences in participants' walking behaviors, and other confounding factors stemming from the neighborhood.

A CAVE-based virtual reality lab was used in the *VR/Field* Experiment. No participants had prior experience with VR technology, thus, they required basic training on use of the technology and the locomotion interface. Participants were noted to be excited and engaged due to the novelty of VR—such factors could have influenced task performance and behavior as has been cited in related work (Brade et al., 2017). Another limitation of the VR lab is that it limits participants' exploration of the neighborhood to visual input and limited body movement. Specifically, the lack of ambient sounds, the noise of the CAVE and its projectors, and the inability of participants to walk naturally—these may have limited the fidelity and the degree to which participants felt immersed within the virtual environment. Lastly, access to a CAVE, its cost and the resourcing required can be prohibitive to researchers who want to go beyond the traditional lab environment to conduct user evaluations on a location-based task.

5.5 Applications

The methodology used in the present work can be applied to evaluations of similar location-based tasks. We used a rigorous, mixed-methods approach, whereby a quantitative dataset on participants' task performance and map usage was collected as well as a qualitative dataset on their thinking and behaviors (i.e., the think aloud method, direct observation, and questionnaires); together, these were used to better understand user behavior and task outcomes. We have demonstrated that this approach is effective in evaluating tasks that require the use of a map—paper or digital—in the field or in a lab setting.

In the preceding work, *Rusch's Lab Experiment (RLE)* presented an alternative to field studies, whereby a lab method was used to study a location-based task in a controlled setting—it incorporated views of the task environment in a non-immersive fashion. In the present work, we discuss the construction and use of an immersive, virtual reality lab for the evaluation of location-based tasks. Such a method can be used in mobile computing evaluations to simulate task environments in a controlled lab setting, while retaining much of the ecological validity that is appreciated in the field. Furthermore, by building a VR lab and replicating a mixed-methods field study concerning a location-based task—and by comparing outcomes of the two evaluation techniques—we have demonstrated an approach that can be used to assess and validate labs of a similar nature and purpose.

5.6 Future Work

We plan to continue our development of the VR lab so that we are better able to simulate real-world environments and replicate field evaluations of location-based tasks. This means that we will continue to refine our field and lab evaluation methods so that apt comparisons can be made between outcomes of the two. In addition to increasing the fidelity of the virtual environment and further automating data collection, we plan to add features that will allow us to simulate GPS on the mobile devices and digital maps that are brought into the lab.

From an implementation standpoint, VR systems based on head-mounted displays (HMD) such as the Oculus Rift™ or the HTC Vive™ present interesting alternatives to the CAVE-based approach that we have adopted due to their accessibility and reduced cost. However, interaction design challenges related to the collaboration of participants and observers during shared use of such technologies will need to be worked out. Furthermore, the capabilities required to incorporate and track external mobile devices and other artifacts for HMD-based VR experiences will have to be resolved before field studies such as ours can be replicated using these technologies.

5.7 Conclusion

In this research, we use lab and field study methods to evaluate a location-based task that involves map use. It has been reported that only 5% of mobile HCI research explicitly combines field and lab-based methods (Kjeldskov & Paay, 2012). In doing so, we have not only gained insights into the task itself, but we have also gained insights into the strengths and weaknesses of the methods used.

The lab evaluation method was used in the earlier work concerning the address verification task, by which the task environment was simulated (photos were used) so that participants could complete a series of address verification scenarios. In *Rusch's Lab Experiment (RLE)*, participants' task performance and digital map operations showed correlation to their spatial visualization and perspective-taking abilities (Batinov et al., 2015; Rusch et al., 2012; Rusch, 2008). In a complementary field study, we used ethnographic methods whereby we observed government fieldworkers execute a task similar to address verification. *The Ethnographic Field Study (EFS)* made evident the richness and complexity of a location-based task similar to our own. In this research, we built on the learnings from these prior studies. We conducted three more experiments on address verification. We examined the differences between participants with high spatial visualization (HighVZ) and low spatial visualization ability (LowVZ) across the experiments and we learned that participants in the HighVZ group typically outperformed and out-strategized their LowVZ counterparts. We identified a task workflow and associated behaviors that were consistent across experiments, to which participants' spatial visualization ability could be linked.

The experiments all shared a mixed-methods design that built on learnings from the previous studies. Rich quantitative and qualitative data sets helped us to recognize a variety of phenomena across experiments. For example, when we switched participants (from using paper maps) to using digital maps in our experiments, we noticed that this had an adverse effect on their planning behavior. We recognized that participants struggled in areas of the study neighborhood whereby the street layouts deviated from the more familiar grid-like patterns seen in typical neighborhoods; participants also struggled at confusing intersections. We recognized that during times of confusion and difficulty, participants tended to check and rotate their maps more frequently; in the case of the digital map users, they demonstrated noticeably more zoom and significantly more panning operations. These such behaviors showed links to participants' task performance and their spatial visualization ability grouping.

In our group's early work, we recognized the strengths of evaluating a location-based task in an office-based lab in which the task environment was simulated. The lab approach had a high degree of experimental control, strengths in quantitative analysis, and its relatively simple implementation was easily replicated. However, the office-based lab method fell short in terms of its ability to reflect the rich context of location-based tasks—its ecological validity is also questionable. The VR lab proposed in this dissertation represents the best of both worlds: The experimental control of the lab and the ecological validity of the field. Using the VR lab, we were better able to immerse participants in the task environment so that real-world behaviors and even some of the expected task performance outcomes rang true. This was enabled by our CAVE-based implementation, which allowed for: Hands-free locomotion in an immersive simulation of the study neighborhood; use of external devices and tools; automated data collection; and the ability for observers and participants to interact in a shared, virtual environment.

The VR lab showcased in the present work represents a middle-ground for the evaluation of location-based tasks, where the trade-offs between the lab method and the field method are not so severe. However, this brand of lab comes with its own trade-offs: It is not cheap or accessible for the majority of mobile HCI practitioners. Commercially available

VR, however, is relatively cheap and advancements continue to be made in key interaction techniques such as room-scale experiences, the incorporation of external artifacts, and collaborative/shared virtual environments. As these new and more immersive tools are incorporated into HCI evaluations, there will be a need to assess, compare and validate them for use in mobile HCI research. This dissertation presents such an approach, which can be applied to mobile HCI research concerning the evaluation of location-based tasks that involve the use of a map.

REFERENCES

Aretz, A. J., & Wickens, C. D. (1992). The mental rotation of map displays. *Human Performance*, *5*(4), 303.

Arnold, P., & Farrell, M. J. (2003). Embodiment and spatial behavior in virtual environments: Comments on Durlach et al. (2000). *Presence: Teleoperators & Virtual Environments*, *12*(6), 658–662. https://doi.org/10.1162/105474603322955941

Arns, L., & Cruz-Neira, C. (2004). Effects of physical and virtual rotations and display device on users of an architectural walkthrough. *2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry*. Singapore: ACM. https://doi.org/10.1145/1044588.1044608

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baldwin, C. L., & Reagan, I. (2009). Individual differences in route-learning strategy and associated working memory resources. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(3), 368–377. https://doi.org/10.1177/0018720809338187

Batinov, G. I. (2017). *Computer detection of spatial visualization in a location-based task.*

Batinov, G. I., Whitney, K. C., Nusser, S. M., Miller, L. L., & Ashenfelter, K. T. (2011). Decision models for an address verification field task. In *26th International Conference on Computers and Their Applications (CATA)* (pp. 9–14). New Orleans, LA.

Batinov, G., Rusch, M., Meng, T., Whitney, K., Patanasakpinyo, T., Miller, L., & Nusser, S. (2015). Understanding map operations in location-based surveys. In *8th International Conference on Advances in Computer-Human Interaction* (pp. 144–149). Lisbon, Portugal.

Batinov, G., Whitney, K., & Miller, L. (2016). Identifying workflow differences based on spatial ability. In *31st International Conference on Computers and Their Applications (CATA)*.

Batinov, G., Whitney, K., Miller, L., Nusser, S., Stanfill, B., & Ashenfelter, K. T. (2013). Evaluating the impact of spatial ability in virtual and real world environments. *6th International Conference on Advances in Computer-Human Interactions (ACHI)*, (c), 274–279.

Benyon, D. (1993). Accomodating individual differences through an adaptive user interface. In M. Schneider-Hufschmidt, T. Kuhme, & U. Malinowski (Eds.), *Adaptive User Interfaces: Principles and Practice* (pp. 149–165). Amsterdam: Elsevier Science.

Benyon, D., Crerar, A., & Wilkinson, S. (2001). Individual differences and inclusive design. In C. Stephanidis (Ed.), *User interfaces for all: concepts, methods, and tools*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bhimani, A. (2017). *Feasibility of Using Virtual Reality in Requirements Elicitation Process*.

Bowman, D. a., Davis, E. T., Hodges, L. F., & Badre, A. N. (1999). Maintaining spatial orientation during travel in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, *8*(6), 618–631. https://doi.org/10.1162/105474699566521

Brade, J., Lorenz, M., Busch, M., Hammer, N., Tscheligi, M., & Klimant, P. (2017). Being there again – Presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. *International Journal of Human Computer Studies*, *101*(September 2016), 76–87. https://doi.org/10.1016/j.ijhcs.2017.01.004

Brade, J., Lorenz, M., Klimant, P., Pürzel, F., & Putz, M. (2016). Virtual Reality for Industrial Application-The Influence of Presence and its Relation to Usability and User Experience. In *Advanced Materials Research* (Vol. 1140, pp. 537–544). Trans Tech Publ.

Büring, T., Gerken, J., & Reiterer, H. (2006). Usability of overview-supported zooming on small screens with regard to individual differences in spatial ability. *Working Conference on Advanced Visual Interfaces*. Venezia, Italy: ACM. https://doi.org/10.1145/1133265.1133310

Busch, M., Lorenz, M., Tscheligi, M., Hochleitner, C., & Schulz, T. (2014). Being there for real: presence in real and virtual environments and its relation to usability. In *Proceedings of the NordiCHI 2014: The 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (pp. 117–126). https://doi.org/10.1145/2639189.2639224

Campbell, S. G. (2011). *Users' spatial abilities affect interface usability outcomes*. University of Maryland, College Park.

Carroll, J. (1993). *Human cognitive abilities: A survey of factor analytical studies*. Cambridge: Cambridge University Press.

Carroll, J. M. (2009). Human computer interaction. (M. Soegaard & R. F. Dam, Eds.), *Encyclopedia of Human-Computer Interaction*. Aarhus, Denmark: The Interaction-Design.org Foundation.

Chance, S. S., Gaunet, F., Beall, A. C., & Loomis, J. M. (1998). Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence*, *7*(2), 168–178.

Chellappan, S. K. (2012). *An object-oriented approach to maps*. Digital Repository@ Iowa State University.

Chellappan, S., & Miller, L. (2014). Representing primary and secondary features in survey maps with objects. In *29th ISCA Conference on Computers and Their Applications (CATA)*. Las Vegas, NV.

Chen, C., & Rada, R. (1996). Interacting with hypertext: a meta-analysis of experimental studies. *Hum.-Comput. Interact.*, *11*(2), 125–156. https://doi.org/10.1207/s15327051hci1102_2

Coursaris, C. K., & Kim, D. J. (2011). A meta-analytical review of empirical mobile usability studies. *Journal of Usability Studies*, *6*(3), 117–171.

Cruz-Neira, C., Bierbaum, A., Hartling, P., Just, C., & Meinert, K. (2002). VR Juggler-An Open Source platform for virtual reality applications. In *40th AIAA Aerospace Sciences Meeting & Exhibit* (p. 754).

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. *20th Annual Conference on Computer Graphics and Interactive Techniques*. Anaheim, CA: ACM. https://doi.org/10.1145/166117.166134

Darken, R. P., & Cevik, H. (1999). Map usage in virtual environments: Orientation issues. In *1999 IEEE Virtual Reality Conference* (pp. 133–140).

Darken, R. P., & Peterson, B. (2001). Spatial orientation, wayfinding, and representation. In K. Stanney (Ed.), *Handbook of Virtual Environment Technology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Darken, R. P., & Sibert, J. L. (1996). Navigating large virtual spaces. *International Journal of Human-Computer Interaction*, *8*(1), 49–71. https://doi.org/10.1080/10447319609526140

Delikostidis, I., Fechner, T., Fritze, H., AbdelMouty, A. M., & Kray, C. (2013). Evaluating mobile applications in virtual environments: A survey. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, *5*(4), 1–19.

Delikostidis, I., Fritze, H., Fechner, T., & Kray, C. (2015). Bridging the gap between field- and lab-based user studies for location-based services. In *Progress in Location-Based Services 2014* (pp. 257–271). Springer.

Dillon, A., & Watson, C. (1996). User analysis in HCI—the historical lessons from individual differences research. *International Journal of Human-Computer Studies*, *45*, 619–637.

Downing, R. E., Moore, J. L., & Brown, S. W. (2005). The effects and interaction of spatial visualization and domain expertise on information seeking. *Computers in Human Behavior*, *21*(2), 195–209. https://doi.org/10.1016/j.chb.2004.03.040

Durlach, N., Allen, G., Darken, R., Garnett, R. L., Loomis, J., Templeman, J., & Wiegand, T. E. von. (2000). Virtual environments and the enhancement of spatial behavior: Towards a comprehensive research agenda. *Presence: Teleoperators and Virtual Environments*, *9*(6), 593–615. https://doi.org/doi:10.1162/105474600300040402

Egan, D., & Gomez, M. (1985). Assaying, isolating, and accommodating individual differences in learning a complex skill . In R. Dillon (Ed.), *Individual Differences in Cognition* (pp. 173–217). New York: Academic Press.

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

French, J., Ekstrom, R., & Price, L. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.

Garden, S., Cornoldi, C., & Logie, R. H. (2002). Visuo-spatial working memory in navigation. *Applied Cognitive Psychology*, *16*(1), 35–50. https://doi.org/10.1002/acp.746

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Grudin, J. (2007). A moving target: The evolution of HCI. In G. Salvendy (Ed.), *The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications* (2nd ed.). Lawrence Erlbaum Associates.

Guilford, J. P., & Zimmerman, W. S. (1947). *The Guilford-Zimmerman aptitude survey*. Palo Alto: Consulting Psychologists Press.

Hart, R. A., & Moore, G. T. (1973). The development of spatial cognition: A review. In R. M. Downs & D. Stea (Eds.), *Image and environment* (pp. 246–288). Chicago: Aldine Publishing Company.

Hecht, L., & Kucera, B. (2000). *Toward improved geographic information services within a digital government: Report of the NSF digital government initiative geographic information systems workshop*. *Proceedings of The 2000 Annual National Conference on Digital Government Research*. Digital Government Society of North America.

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology*, *28*(1), 74–82. https://doi.org/10.1016/j.jenvp.2007.09.002

Johnson, P. (1998). Usability and mobility: Interactions on the move. In *Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices*.

Kjeldskov, J., & Graham, C. (2003). A Review of Mobile HCI Research Methods. *Human-Computer Interaction with Mobile Devices and Services.*, 317–335. https://doi.org/10.1007/978-3-540-45233-1_23

Kjeldskov, J., & Paay, J. (2012). A longitudinal review of mobile HCI research methods. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services* (pp. 69–78). New York, NY, USA: ACM. https://doi.org/10.1145/2371574.2371586

Kjeldskov, J., & Skov, M. B. (2014). Was it Worth the Hassle? Ten Years of Mobile HCI Research Discussions on Lab and Field Evaluations. *Acm*. https://doi.org/10.1145/2628363.2628398

Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Mobile Human-Computer Interaction-MobileHCI 2004* (pp. 61–73). Springer.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies*, *60*(5–6), 599–620.

Klatzky, R. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K. Wender (Eds.), *Spatial Cognition* (Vol. 1404, pp. 1–17). London: Springer-Verlag. https://doi.org/10.1007/3-540-69342-4_1

Klatzky, R., & Loomis, J. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological …*, *9*(4), 293–298.

Koh, G., & von Wiegand, T. E. (1999). Use of virtual environments for acquiring configurational knowledge about specific real-world spaces: I. preliminary experiment. *Presence: Teleoperators & Virtual Environments*, *8*(6), 632.

Kozhevnikov, M., Hayes, D., & Kozhevnikov, M. (2013). The role of spatial immersion for tasks based on the use of egocentric frames of reference. In *HCI International 2013-Posters' Extended Abstracts* (pp. 708–712). Springer.

Kozhevnikov, M., Motes, M., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, *20*, 397–417.

Levene, H. (1960). Robust tests for equality of variances. *Contributions to Probability and Statistics*, *1*, 278–292.

Li, C., & Longley, P. (2006). A test environment for location-based services applications. *Transactions in GIS*, *10*(1), 43–61.

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Mahwah, NJ: Lawrence Erlbaum Associates.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2), 0. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lynch, K. (1960). *The image of the city*. Cambridge: MIT Press.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, *86*(5), 889–918. https://doi.org/10.1037/0033-2909.86.5.889

Meng, T., Hua, M., Gao, Y., Batinov, G., Rusch, M., Whitney, K., & Miller, L. (2015). Understanding situations in map-based applications. In *30th International Conference on Computers and Their Applications* (pp. 425–430).

Messick, S. (1976). *Individuality in learning*. San Francisco: Jossey-Bass.

Münzer, S., Zimmer, H. D., Schwalm, M., Baus, J., & Aslan, I. (2006). Computer-assisted navigation and the acquisition of route and survey knowledge. *Journal of Environmental Psychology*, *26*(4), 300–308. https://doi.org/10.1016/j.jenvp.2006.08.001

Murphy, E. D., & Nusser, S. M. (2003). Evaluating user interfaces for accommodation of individual differences in spatial abilities and way-finding strategies. In *2nd International Conference on Universal Access in Human-Computer Interaction* (Vol. 4, pp. 1005–1009).

Murray, H. A., & Kluckhohn, C. (1948). *Personality in nature, society, and culture*. Oxford, England: Knopf.

Myers, B. A. (1998). A brief history of human-computer interaction technology. *Interactions*, *5*(2), 44–54. https://doi.org/10.1145/274430.274436

Nathanael, D., Vosniakos, G.-C., & Mosialos, S. (2010). Cognitive task analysis for virtual reality training: The case of CNC tool offsetting. *28th Annual European Conference on Cognitive Ergonomics*. Delft, Netherlands: ACM. https://doi.org/10.1145/1962300.1962350

Nielsen, C. (1998). Testing in the Field. In *Computer Human Interaction, 1998. Proceedings. 3rd Asia Pacific* (pp. 285–290). IEEE.

Nielsen, C. M., Overgaard, M., Pedersen, M. B., Stage, J., & Stenild, S. (2006). It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles* (pp. 272–280). ACM.

Norman, K. L. (1994). Spatial visualization--A gateway to computer-based technology. *Journal of Special Education Technology*, *12*(3), 195.

Nusser, S., Goodchild, M., Clarke, K., & Miller, L. (2004). Geospatial information in complex mobile field settings. *2004 Annual National Conference on Digital Government Research*. Seattle, WA: Digital Government Society of North America.

Nusser, S. M., & Fox, J. E. (2002). *Using digital maps and GPS for planning and navigation in field surveys* (February 8). *Iowa State University Department of Statistics Preprint 2002-2009*. Ames, IA.

Nusser, S., Miller, L., Clarke, K., & Goodchild, M. (2002). *A user-centered framework for next-generation field data collection*. *Proceedings of the 2002 Annual National Conference on Digital Government Research*. Los Angeles, California: Digital Government Society of North America.

Pak, R., Rogers, W. A., & Fisk, A. D. (2006). Spatial ability subfactors and their influences on a computer-based information search task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(1), 154–165. https://doi.org/10.1518/001872006776412180

Patanasakpinyo, T. (2017). *Flattening methods for adaptive location-based software to user abilities*.

PatanasakPinyo, T., Batinov, G., Whitney, K., Sulaiman, A., Miller, L., & Gilbert, S. (2018). Extracting useful features for users with different levels of spatial visualization. In *33rd International Conference on Computers and Their Applications (CATA)* (pp. 86–91). Las Vegas, NV.

Poupyrev, I., Maruyama, S., & Rekimoto, J. (2002). Ambient touch: designing tactile interfaces for handheld devices. *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology*. Paris, France: ACM. https://doi.org/10.1145/571985.571993

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition*, *27*(4), 741–750.

Riecke, B. E., Heyde, M. von der, & Bülthoff, H. H. (2004). Spatial updating in real and virtual environments: contribution and interaction of visual and vestibular cues. *1st Symposium on Applied Perception in Graphics and Visualization*. Los Angeles, California: ACM. https://doi.org/10.1145/1012551.1012553

Riecke, B. E., Heyde, M. Von Der, & Bülthoff, H. H. (2005). Visual cues can be sufficient for triggering automatic, reflexlike spatial updating. *ACM Trans. Appl. Percept.*, *2*(3), 183–215. https://doi.org/10.1145/1077399.1077401

Riecke, B. E., & Wiener, J. M. (2006). Point-to-origin experiments in VR revealed novel qualitative errors in visual path integration. In *3rd Symposium on Applied Perception in Graphics and Visualization* (p. 156). Boston, Massachusetts: ACM. https://doi.org/10.1145/1140491.1140533

Roger, M., Bonnardel, N., & Le Bigot, L. (2009). Improving navigation messages for mobile urban guides: Effects of the guide's interlocutor model, spatial abilities and use of landmarks on route description. *International Journal of Industrial Ergonomics*, *39*(3), 509–515. https://doi.org/10.1016/j.ergon.2008.10.009

Ruddle, R. A., Volkova, E., & Bülthoff, H. H. (2011). Walking improves your cognitive map in environments that are large-scale and large in extent. *ACM Trans. Comput.-Hum. Interact.*, *18*(2), 10:1--10:20. https://doi.org/10.1145/1970378.1970384

Rusch, M. L. (2008). *Relationships between user performance and spatial ability in using map-based software on pen-based devices*. *ProQuest Dissertations and Theses*. Iowa State University.

Rusch, M. L., Nusser, S. M., Miller, L., Batinov, G. I., & Whitney, K. C. (2012). Spatial ability and map-based software applications. In *5th International Conference on Advances in Computer-human Interactions (ACHI)*. Valencia, Spain.

Shaphiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika*, *52*(3), 591–611.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In H. W. Reese (Ed.), *Advances in Child Development and Behavior* (Vol. 10, pp. 9–55). Academic Press.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Spearman, C., & Wynn Jones, L. L. (1950). *Human ability*. London: Macmillan.

Stanney, K. M., & Salvendy, G. (1992). Diversity in field-articulation and its implications for human-computer interface design. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *36*(12), 902–906. https://doi.org/10.1518/107118192786750331

The census found it could count on NCS. (2001). *Workforce*.

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, *14*(4), 560–589. https://doi.org/10.1016/0010-0285(82)90019-6

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, *1*, ix + 121.

van der Ham, I. J. M., Faber, A. M. E., Venselaar, M., van Kreveld, M. J., & Löffler, M. (2015). Ecological validity of virtual environments to assess human navigation ability. *Frontiers in Psychology*, *6*, 637.

Vicente, K. J., Hayes, B. C., & Williges, R. C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *29*(3), 349–359.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817–835. https://doi.org/10.1037/a0016127

Waller, D. (2005). The WALKABOUT: Using virtual environments to assess large-scale spatial abilities. *Computers in Human Behavior*, *21*(2), 243–253. https://doi.org/10.1016/j.chb.2004.02.022

Waller, D., & Hodgson, E. (2013). Sensory contributions to spatial knowledge of real and virtual environments. In F. Steinicke, Y. Visell, J. Campos, & A. Lécuyer (Eds.), *Human Walking in Virtual Environments SE - 1* (pp. 3–26). Springer New York. https://doi.org/10.1007/978-1-4419-8432-6_1

Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperators and Virtual Environments*, *7*(2), 129–143. https://doi.org/doi:10.1162/105474698565631

Whitney, K., Batinov, G., Miller, L., Nusser, S., & Ashenfelter, K. T. (2011). Exploring a Map Survey Task's Sensitivity to Cognitive Ability. *Proceedings of the 2011 International Conference on Advances in Computer-Human Interactions*, 63–68.

Whitney, K. C., Batinov, G. I., Nusser, S. M., Miller, L. L., & Ashenfelter, K. T. (2010). Understanding the user's approach toward location-based field tasks. In *23rd International Conference on Computer Applications in Industry and Engineering 2010, CAINE 2010 - Including SNA 2010 Workshop* (pp. 156–161). Las Vegas, NV.

Willis, K., Hölscher, C., Wilbertz, G., & Li, C. (2009). A comparison of spatial knowledge acquisition with maps and mobile maps. *Computers, Environment and …*, 1–24.

Witmer, B. G., & Kline, P. B. (1998). Judging perceived and traversed distance in virtual environments. *Presence: Teleoperators & Virtual Environments*, *7*(2), 144–167. https://doi.org/10.1162/105474698565640

Yuan, S., Song, S., & Zhang, Y. (2014). Experimental research in urban spatial cognition by using virtual reality technology. *Athens Journal of Technology Engineering*.

Zhang, D., & Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction*, *18*(3), 293–308.

Zhang, H., & Salvendy, G. (2001). The implications of visualization ability and structure preview design for web information search tasks. *International Journal of Human-Computer Interaction*, *13*(1), 75–95.

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

DATE: 28 September 2009

TO: Kofi Whitney
226 Atanasoff Hall

CC: Dr. Les Miller
112 Atanasoff Hall

FROM: Roxanne Bappe, IRB Coordinator
Office for Responsible Research

TITLE: **How do we use a paper map? An exploratory study of spatial ability and decision making.**

IRB ID: **09-386**

Approval Date: 28 September 2009
Date for Continuing Review: 12 September 2010

The Chair of the Institutional Review Board of Iowa State University has reviewed and approved this project. Please refer to the IRB ID number shown above in all correspondence regarding this study.

Your study has been approved according to the dates shown above. To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use the documents with the IRB approval stamp** in your research.
- **Obtain IRB approval prior to implementing any changes** to the study by completing the "Continuing Review and/or Modification" form.
- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Research investigators are expected to comply with the principles of the Belmont Report, and state and federal regulations regarding the involvement of humans in research. These documents are located on the Office for Responsible Research website [www.compliance.iastate.edu] or available by calling (515) 294-4566.

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

LIST OF ATTACHMENTS FOR REVIEW

1) **Study Advertisements** (for recruitment)
2) **Scheduling Scripts** (for scheduling participants)
3) **Informed Consent** (presented prior to Phase 1 screening)
4) **Background Information Questionnaire** (for Phase 1 screening)
5) **Cognitive Tests** (for Phase 1 screening)
6) **Training Script** (for Phase 2 field exercise)
7) **Coding Sheet** (for Phase 2 field exercise)
8) **Field Exercise Tests & Questionnaire** (for Phase 2 field exercise)

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

PARTICIPANTS NEEDED

to verify street addresses in a neighborhood for research study.

Participants must be ISU students, aged 18 or older, and fluent in English. Participants should have minimal exposure to Ames neighborhoods.

Participants will be offered compensation.

For more information or to schedule an appointment, please contact:

KOFI WHITNEY
Email: kwhitney@iastate.edu
Phone: (803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

Newspaper advertisement

Participants Needed for Research Study

We are looking for participants to verify street addresses in a neighborhood. Participants must be ISU students, aged 18 or older, and fluent in English. Participants should have minimal exposure to Ames neighborhoods.

Participants completing the study will be offered compensation.

For more information or to schedule an appointment, please contact:
Kofi Whitney @ kwhitney@iastate.edu or (803) 546-0007.

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

SCHEDULING SCRIPTS

The Doodle Online scheduling tool will be used – http://www.doodle.com
After a student has been scheduled, a follow up email will be sent with appointment information

Screening – Email

Subject: Address Verification Study - Screening

Dear [*Student's Name*],

You have received this email because you have expressed an interest to participate in our Address Verification Study. The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1 hour to complete. You will receive a $10 gift card for participating in the screening.

You may be selected after this screening to later participate in a field exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. Compensation for the field exercise is a $20 gift card.

Click here [*Doodle Scheduling Link*] to schedule a screening appointment.

Thank you on behalf of my research group for participating in this study.

Sincerely,

Kofi Whitney
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Kofi Whitney @ 803-546-0007 or Dr. Les Miller @ 515-294-7934.

Screening – Phone

- "Hello, my name is [Scheduler's Name]. I am calling you because you have expressed an interest to participate in our Address Verification Study. I'd like to remind you that your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential."
- "May I continue?"
- "The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1hour to complete. You will receive a $10 gift card for participating in the screening. You may be selected after this screening to later participate in a field exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. Compensation for the field exercise is a $20 gift card."
- "May I email you the information that I have discussed along with a link that will allow you to schedule your screening appointment?"
- "Thank you for your time."

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010

SCHEDULING SCRIPTS (continued)

Field Exercise – Email

Subject: Address Verification Study - Field Exercise

Dear [*Student's Name*],

You have received this email because you participated in the screening portion of our Address Verification Study. We have reviewed your screening information and would like to invite you to participate in the field exercise portion of the study. This exercise will take approximately 2 hours to complete. You will receive a $20 gift card for your participation.

Please refer to your *Informed Consent* document for additional information regarding the study. You are welcome to contact us if you would like to receive another copy.

Click here [*Doodle Scheduling Link*] to schedule the field exercise.
Note: We may contact you to reschedule if weather conditions are not favorable.

Thank you on behalf of my research group for participating in this study.

Sincerely,

Kofi Whitney
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Kofi Whitney @ 803-546-0007 or Dr. Les Miller @ 515-294-7934.

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

INFORMED CONSENT

The purpose of this research study is to gather information on how individuals compare addresses on a street with information on a paper map.

You will complete a brief background questionnaire followed by 3 cognitive assessments. This will take approximately 1 hour. Some subjects will be contacted at a later date to schedule an appointment for a field exercise. If selected for the field exercise, we will explain the task and train you on the procedures. You will then be transported, via CyRide, to an Ames neighborhood where you will practice the procedures. Next, we will give you a list of addresses to find and verify against a map that we provide. During the field exercise, you will be asked to think aloud as you reason through the task and this will be audio-recorded. This exercise will take approximately 2 hours to complete. There are no known risks to participation other than concerns that are normally associated with walking through a neighborhood.

There are no direct benefits to you as a participant other than the educational experience of being involved in an study. Your participation is helping us learn how we can improve address listing methods. By participating in this study, you will be offered a $10 gift card for the background questionnaire and cognitive assessments. If you are selected for the field exercise, you will be offered an additional $20 gift card for participation. You will need to sign a receipt for both gift cards.

Your participation in this study is completely voluntary and you may withdraw at any time. You may skip any part of this study that makes you feel uncomfortable or withdraw from the study at any time without penalty or loss of benefits to which you may otherwise be entitled.

Records identifying participants will be kept confidential to the extent permitted by applicable laws and regulations and will not be made publicly available. However, federal government agencies, the National Science Foundation, and the Institutional Review Board (a committee that reviews and approves human subject research studies) may inspect and/or copy your records for quality assurance and data analysis. These records may contain private information. To ensure confidentiality to the extent permitted by law, the following measures will be taken:

Data that identifies participants will be kept confidential. The information taken from this session will be assigned a unique code. Your name will not be associated with this information. Only researchers from Iowa State University working on this project will have access to data collected during this study. Study records will be kept confidential under password protected computer files. Data will be retained for two years and then will be destroyed.

The data results from this research may be used for educational or scientific purposes and may be presented at scientific and/or educational meetings or published in professional journals. Results will be released in summary form only with no personal identifying information.

For further information about the study contact Kofi Whitney at (803) 546-0007 or Dr. Les Miller at (515) 294-7588. If you have any questions about the rights of research subjects or research-related injury, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, Office for Responsible Research, (515) 294-3115, 1138 Pearson Hall, Ames, IA 50011.

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent.

Participant's Name (printed)____

______________________________ ____________________
(Participant's Signature) (Date)

INVESTIGATOR STATEMENT

I certify that the participant has been given adequate time to read and learn about the study and all of their questions have been answered. It is my opinion that the participant understands the purpose, risks, benefits and the procedures that will be followed in this study and has voluntarily agreed to participate.

______________________________ ____________________
(Signature of Person Obtaining Informed Consent) (Date)

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010

THIS BOX IS FOR RESEARCHER USE ONLY ID:________ Date:________

BACKGROUND INFORMATION

Please answer each of the questions by circling the appropriate response.

1. How often do you use maps?

1 = Never.
2 = Rarely.
3 = Sometimes.
4 = Frequently.
5 = All the time.

2. Student Status.

1 = Freshman.
2 = Sophomore.
3 = Junior.
4 = Senior.
5 = Graduate Student.

3. Gender.

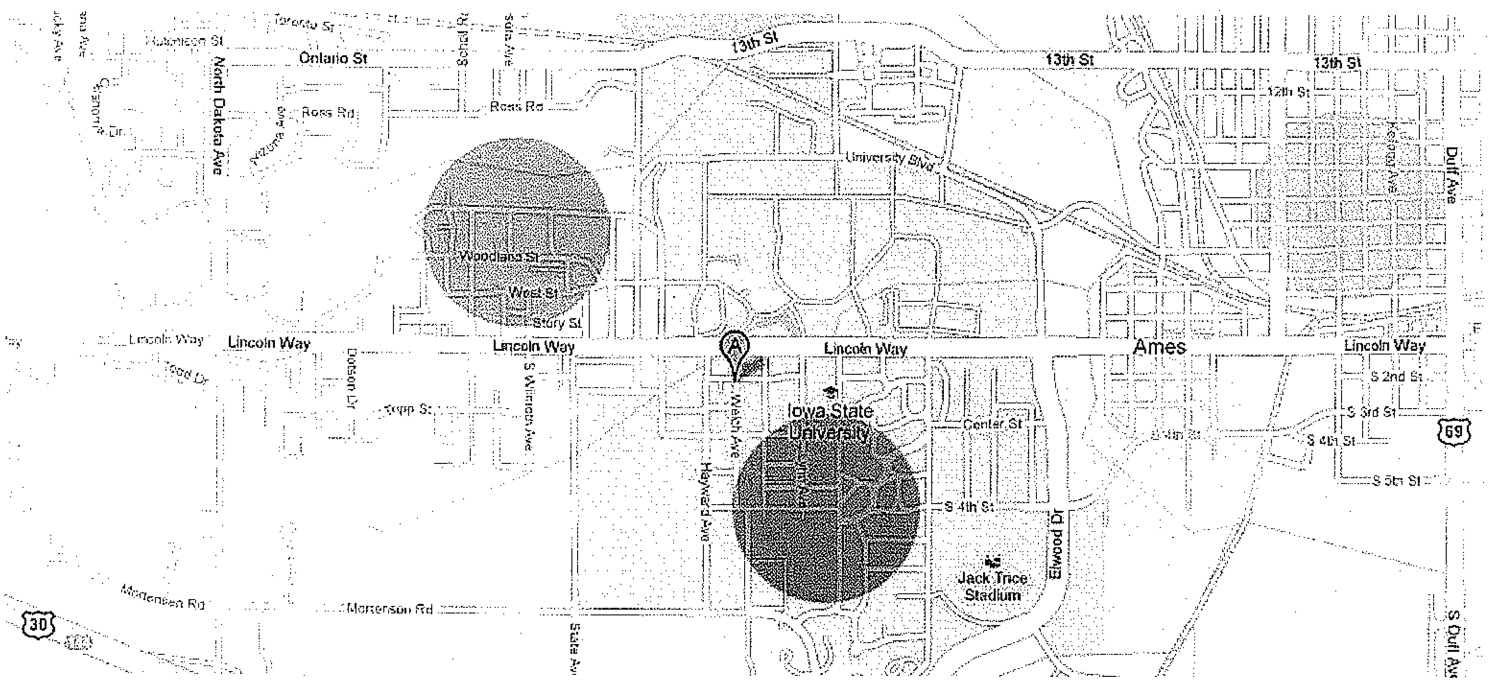
1 = Male.
2 = Female.

4. How long have you lived in Ames?

1 = This is my first semester.
2 = 1-2 years.
3 = 2-3 years.
4 = 3-4 years.
5 = More than 4 years.

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010

BACKGROUND INFORMATION (continued)



Please use the map of Ames above to answer the following questions.

5. How familiar are you with the neighborhood in yellow?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

6. How familiar are you with the neighborhood in red?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

7. How familiar are you with the neighborhood in blue?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

8. How familiar are you with the neighborhood in orange?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

9. How familiar are you with the neighborhood in green?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

VZ VISUALIZATION

Factor

The ability to manipulate or transform the image of spatial patterns into other arrangements

The visualization and spatial orientation factors are similar but visualization requires that the figure be mentally resturctured into components for manipulation while the whole figure is manipulated in spatial orientation. Some researchers think that visualization is a more difficult or more complex and less speeded form of spatial orientation.

Cattell (1971) does not accept visualization as a primary factor. He suggests that it is a second-order factor which includes spatial ability, figural adaptive flexibility, speed of closure, and flexibility of closure. Royce (1973) suggests both primary and higher order visualization factors.

As Carroll (1974) has pointed out, both visualization and spatial orientation require the mental rotation of a spatial configuration in short-term visual memory; visualization requires the additional component of performing serial operations.

Some subjects may employ an analytic strategy in visualization tests and search for symmetry and planes of reflection as clues to the solution. Shepard and Feng (1972) have described the mental processes involved in paper-folding tests.

Identification: Guilford, CFT

References: 3, 7, 13, 18, 21, 39, 45, 52, 58, 60, 67, 69, 73, 79, 85, 86, 90, 91, 93, 100, 103, 108, 110, 121, 130, 143, 153, 156, 161, 164, 165, 174, 197, 200, 207, and 209.

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

Paper Folding Test -- VZ-2

Suggested by Thurstone's Punched Holes. For each item successive drawings illustrate two or three folds made in a square sheet of paper. The final drawing of the folded paper shows where a hole is punched in it. The subject selects one of 5 drawings to show how the punched sheet would appear when fully reopened.

Length of each part: 10 items, 3 minutes

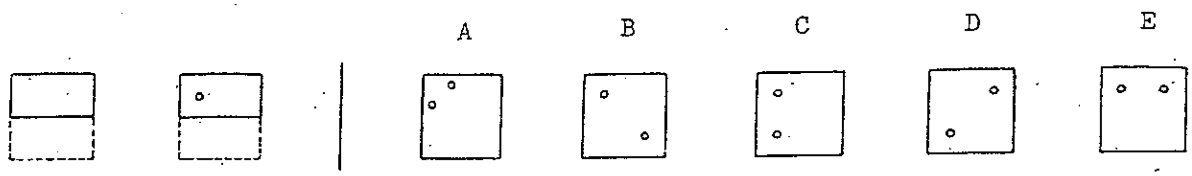
Suitable for grades 9-16

Name ____

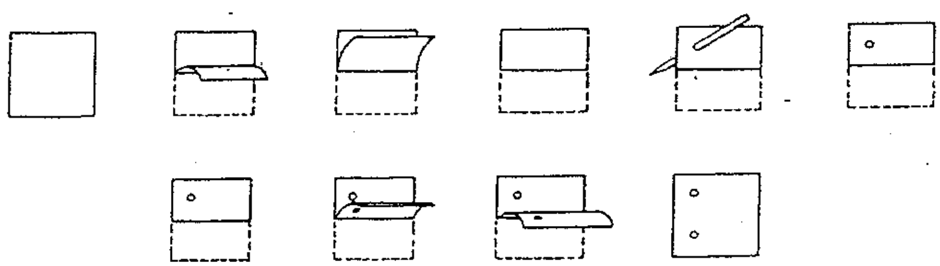
PAPER FOLDING TEST — VZ-2

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures at the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper.)



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.



In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

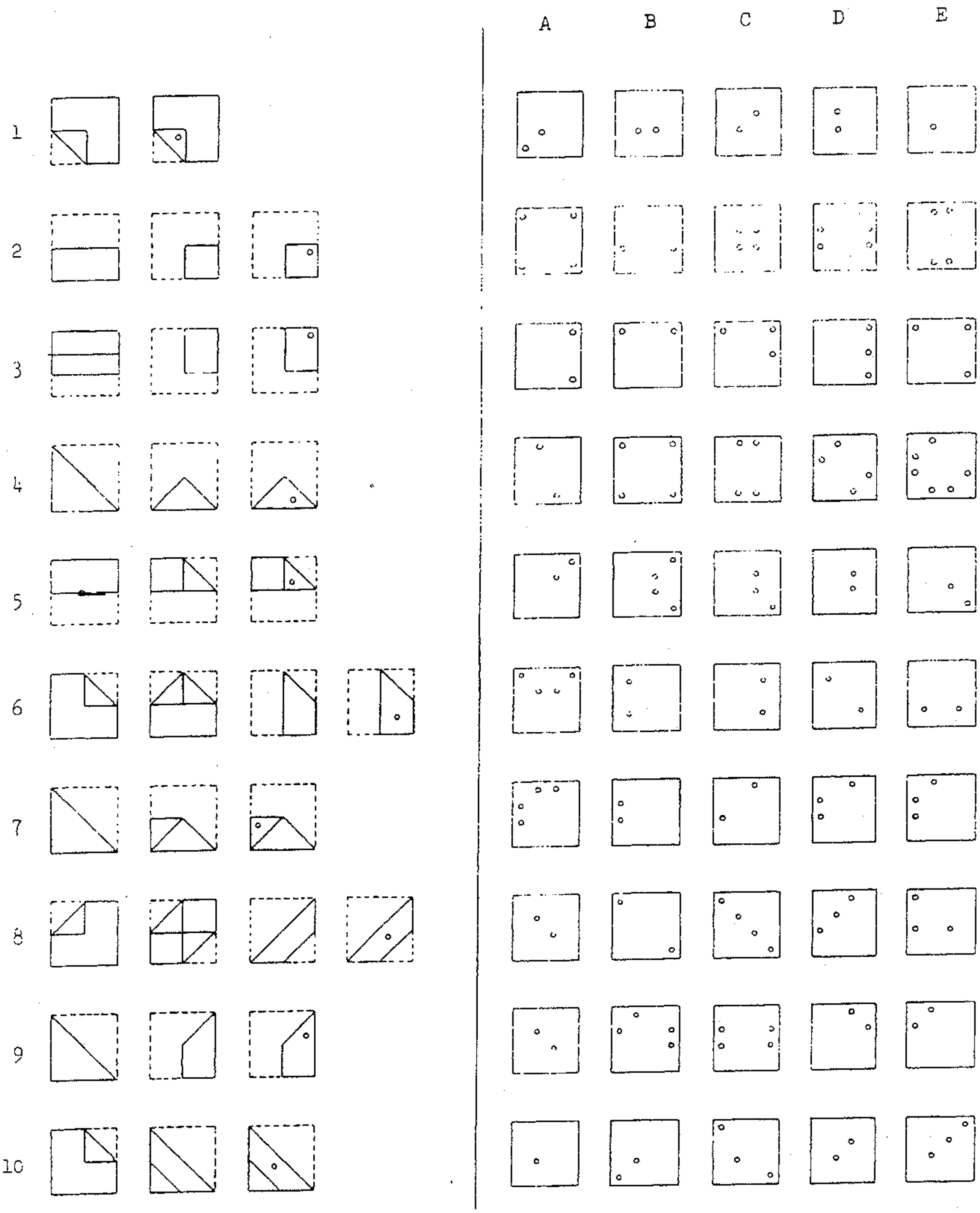
Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will <u>not</u> be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have <u>3 minutes</u> for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, <u>STOP</u>. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010

Part 1 (3 minutes)



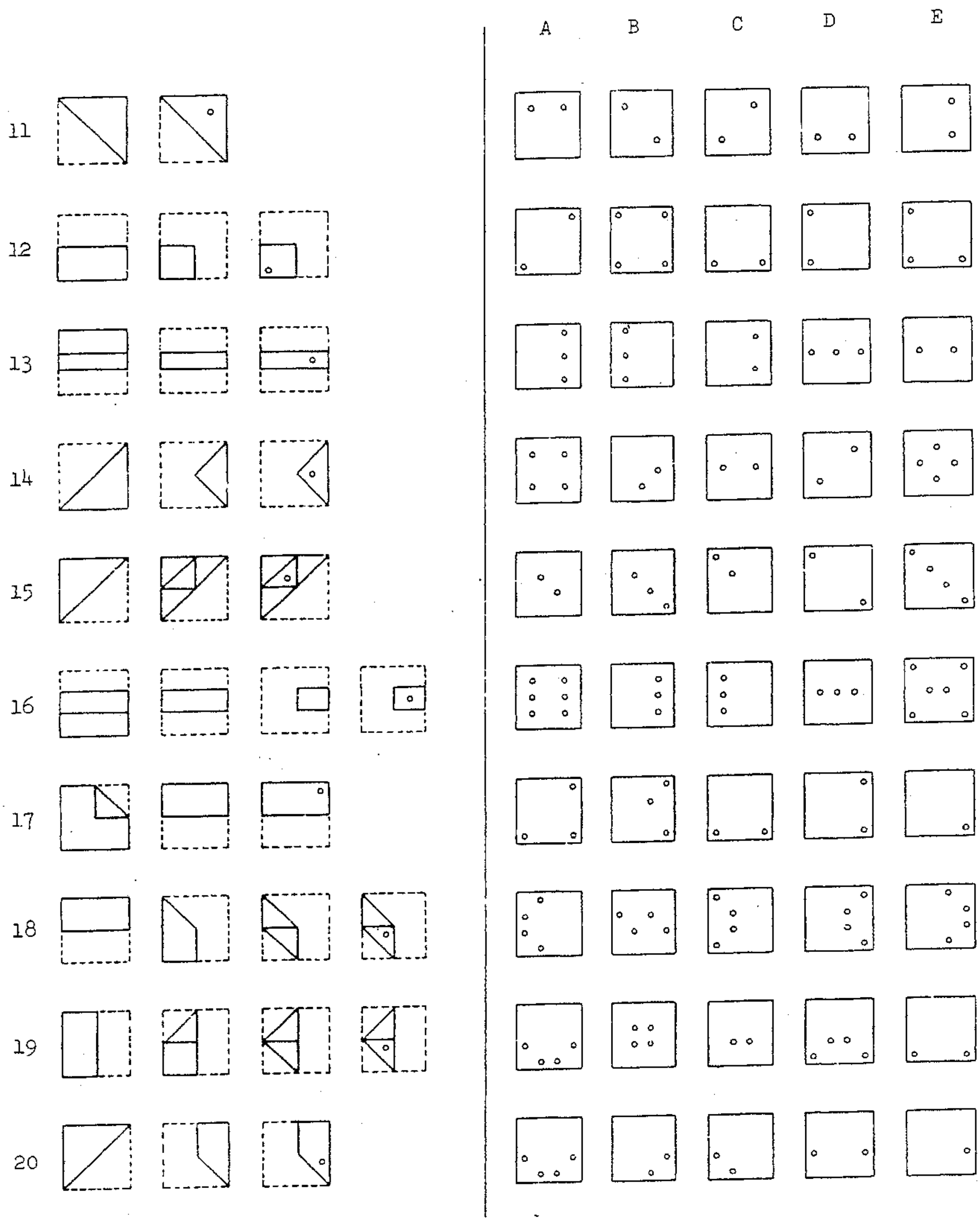
DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

Page 3

VZ-2

Part 2 (3 minutes)



DO NOT GO BACK TO PART 1, AND

DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. STOP.

MV MEMORY, VISUAL

Factor

The ability to remember the configuration, location, and orientation of figural material

There has been considerable debate as to whether or not this factor is due to test content. Thurstone (1946) thought that "the memorizing factor transcends the nature of the content" but more recent research has demonstrated the existence of iconic memory, which is used to store visual impressions. This suggests that visual memory is not simply the result of test content but involves cognitive processes different from those used in other memory factors.

There may be sub-factors of visual memory. Guilford describes six figural memory abilities. Petrov (1970) has found separate factors both for iconic memory and for short-term retention of visual material.

Identification: Guilford, MFU, MFC, and MFR, possibly others.

References: 21, 33, 55, 86, 91, 108, 109, 155, 164, 165, 174, and 179.

Building Memory -- MV-2

The subject is asked to indicate the location of a number of buildings seen on a previously studied map.

Length of each part: 12 items, 4 minutes for memorizing,
4 minutes for testing

Suitable for grades 6-16

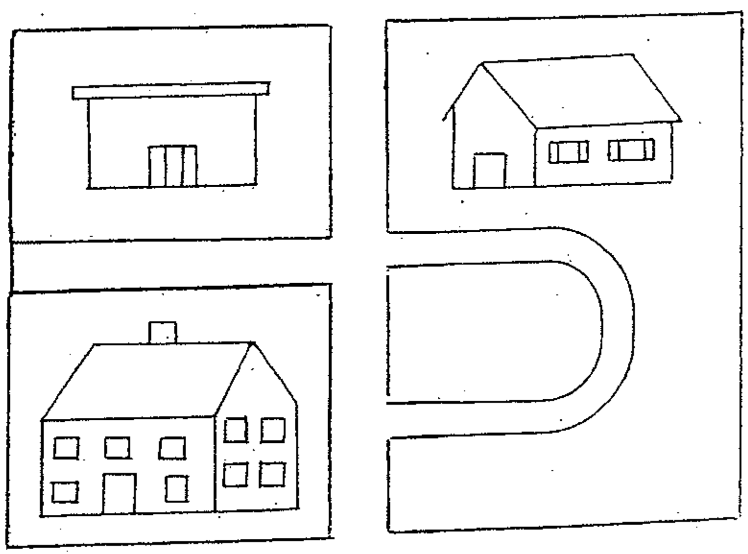
Name ____

BUILDING MEMORY -- MV-2

This is a test of your ability to remember the position of things on a street map.

You will be given a map with streets and buildings and other structures to study. After you have had some time to learn the street layout and the different kinds of structures, you will be asked to turn to a test page. On that page you will find the street map and numbered pictures of some of the structures. You will be asked to put an x on the letter that shows where each of the structures was located on the study map.

Now look at this simple and enlarged sample:



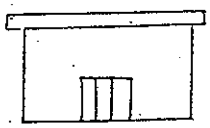
After you have studied the sample above for a minute, turn to the next page.

MV-2

Page 2

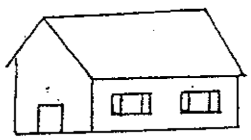
Look at the numbered houses on the left. For each item mark an X on the letter below each building that corresponds with where each house was located on the study map.

1.



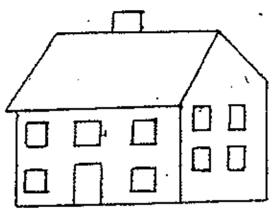
A B C D E

2.

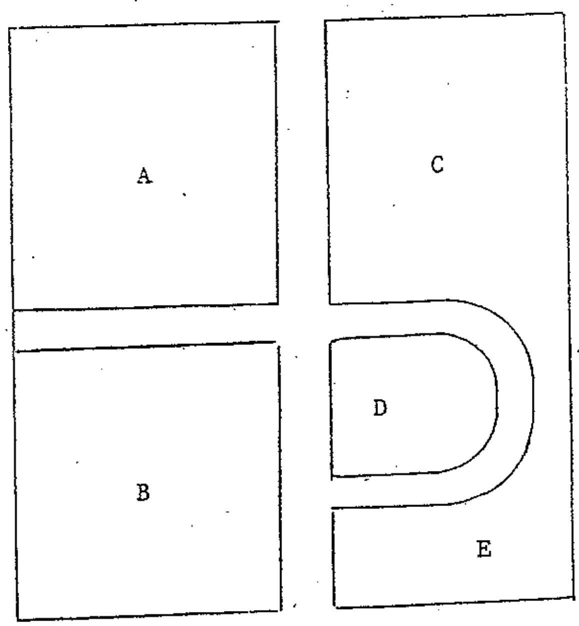


A B C D E

3.



A B C D E



Your answers for sample item 1 should be A, for 2, C, and for 3, B.

Your score on this test will be the number of buildings placed correctly minus a fraction of the number wrong. Therefore, it will <u>not</u> be to your advantage to guess unless you can eliminate some of the locations as definitely wrong.

There are two sections to each part of this test. The first section is a map which you will study for <u>4 minutes</u>. The second is the test section and contains 12 structures to be located on the map. You will have <u>4 minutes</u> to mark your answers. Mark A, B, C, D, or E for each building. In the test section, the buildings will be mixed up and not necessarily near the part of the map where you first saw them.

This test has two parts. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

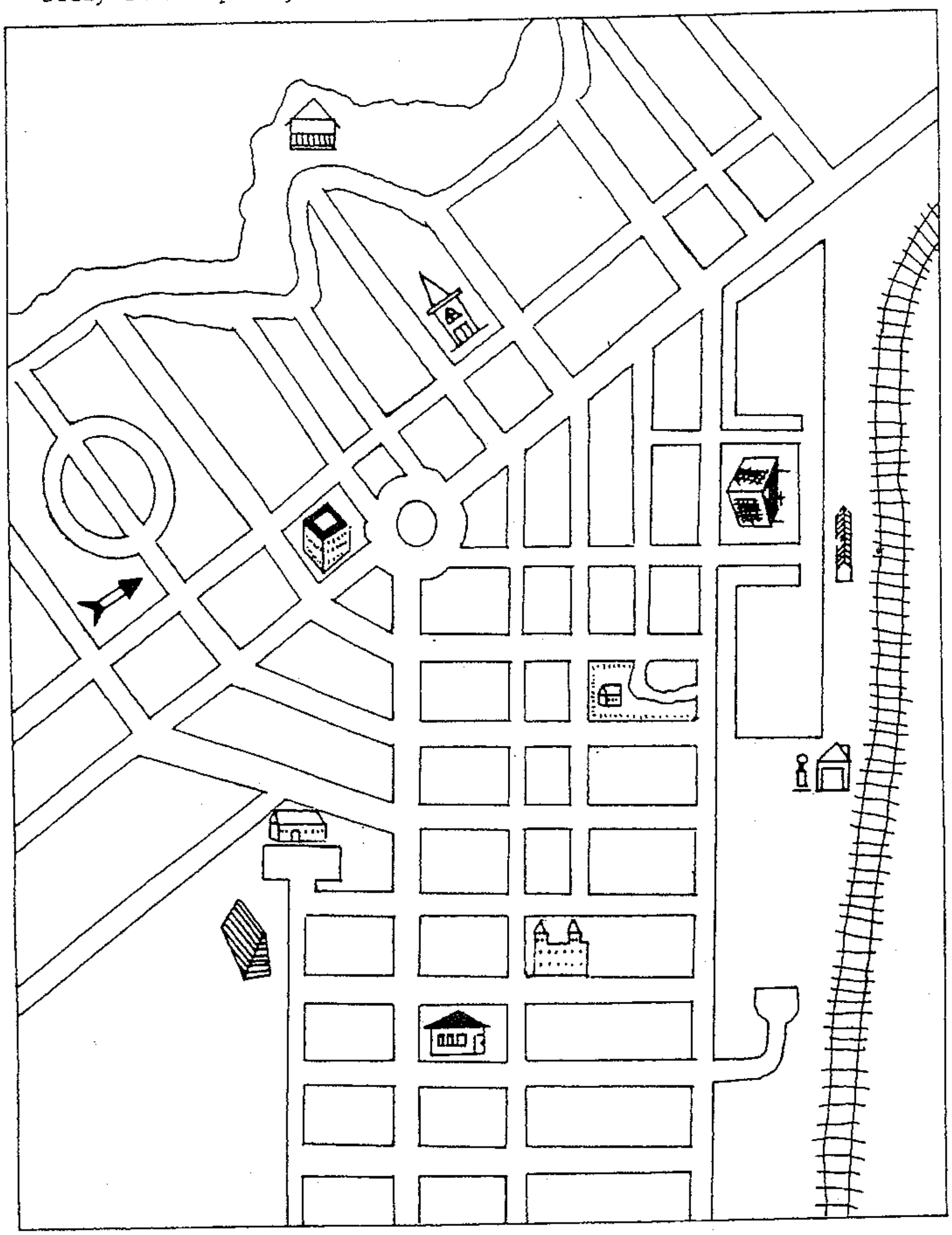
ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

MV-2

STUDY PAGE

Part 1 (4 minutes)

Study this map so you can remember where each building is located.



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP

TEST PAGE

Part 1 (4 minutes)

Mark an X on the letter below each building that shows where it was seen on the map.

1. 

A B C D E

2. 

A B C D E

3. 

A B C D E

4. 

A B C D E

5. 

A B C D E

6. 

A B C D E

7. 

A B C D E

8. 

A B C D E

9. 

A B C D E

10. 

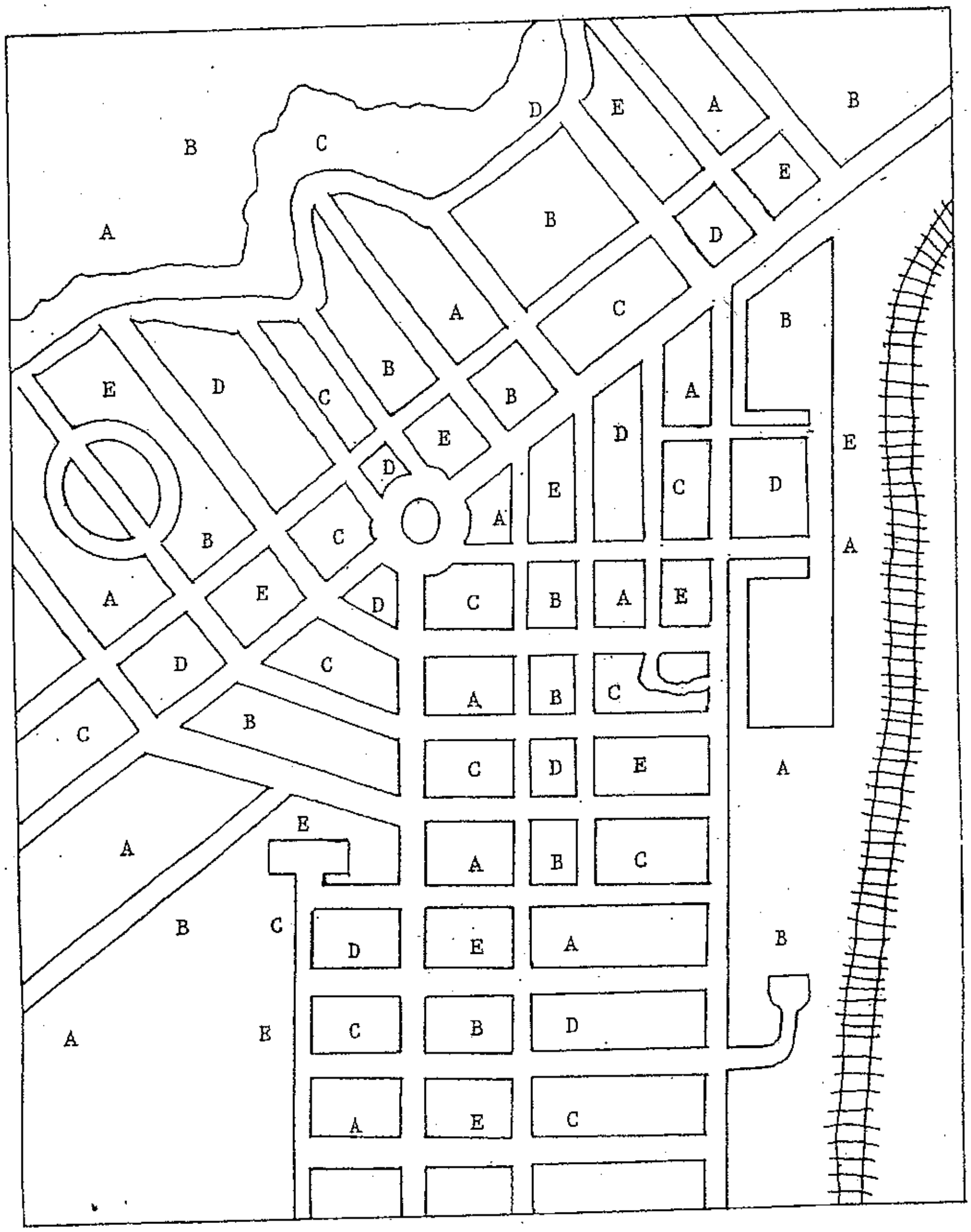
A B C D E

11. 

A B C D E

12. 

A B C D E



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

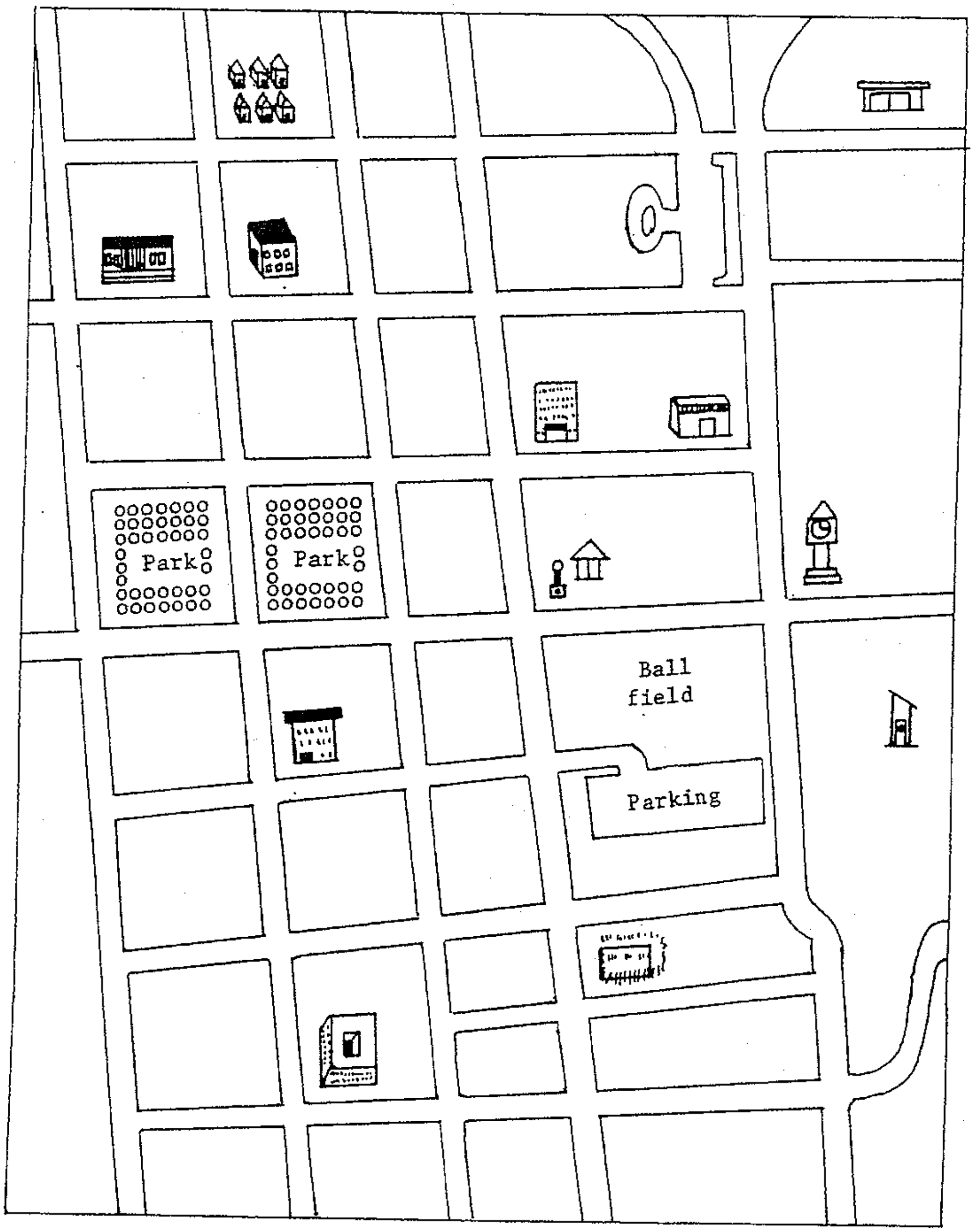
STOP.

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

STUDY PAGE

Part 2 (4 minutes)

Study this map so you can remember where each building is located.



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP.

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

TEST PAGE

Part 2 (4 minutes)

13. 

A B C D E

14. 

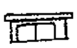
A B C D E

15. 

A B C D E

16. 

A B C D E

17. 

A B C D E

18. 

A B C D E

19. 

A B C D E

20. 

A B C D E

21. 

A B C D E

22. 

A B C D E

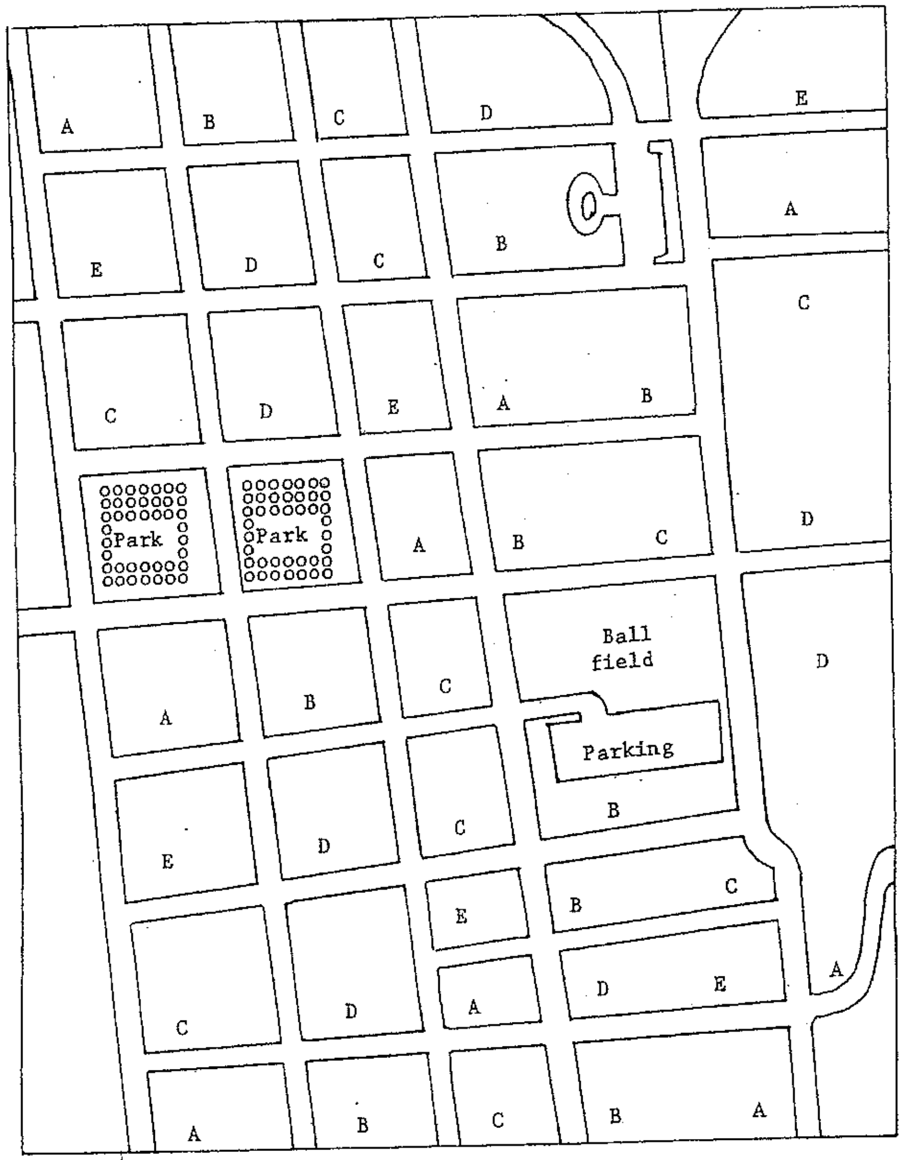
23. 

A B C D E

24. 

A B C D E

Mark an X on the letter below each building that shows where it was seen on the map.



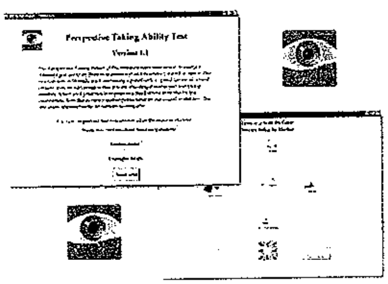
DO NOT GO BACK TO PART 1 AND DO NOT GO ON TO ANY OTHER TEST

STOP.

Perspective Taking Ability Test (PTA-Test)

Recent research of Prof. Maria Kozhevnikov (Department of Psychology, George Mason University, VA) has shown that there are two distinct abilities: mental rotation (an ability to imagine rotation of objects from a fixed perspective) and perspective taking (an ability to imagine a reoriented-self) [pdf]. The second skill (perspective-taking) is the skill, which is important for navigating in space.

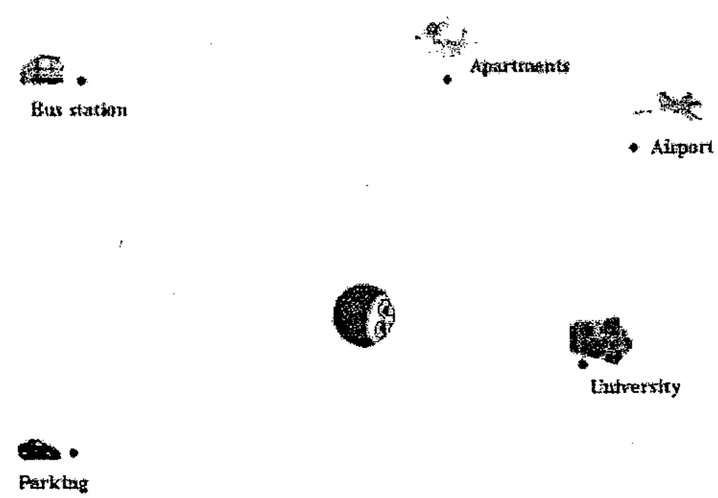
Up until now, the existing tests did not dissociate successfully between mental rotation and perspective-taking abilities, since most existing tests could be solved by using mental rotation as well as perspective taking strategy. As a result, all the existing commercially-available spatial tests measure mostly mental rotation ability (e.g., the ability to imagine rotating objects from a fixed perspective), which is a different ability not related to navigational skills.



We have developed the Computerized Perspective-Taking Ability (CPTA) test to measure spatial orientation ability. This new test was successfully validated and copyrighted jointly by MM Virtual Design, LLC and Rutgers University (see PTA test features). The results suggest that while solving this test, people in fact encode the objects shown on the display with respect to a body-centered coordinate system. It was also shown that while this test predicts reliably the spatial navigational abilities, mental rotation tests do not.

Our new Computerized Perspective-Taking Ability Test is the first valid measure of spatial orientation ability and could be successfully used for research as well as for training purposes and personnel selection in the professions that require high navigational abilities (e.g., astronauts, pilots, drivers).

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010



Imagine you are the figure.

You are facing the University.

Now point to the Airport

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

TRAINING SCRIPT

Participants will read along as this is dictated by the facilitator

You will be presented with the following training materials: (1) a map of the residential area and (2) a list of addresses.

You will be asked to determine whether the addresses are accurately reflected on the map. You may verify the list addresses in any order that you prefer. Four outcomes are possible during verification: (1) the ground situation is correctly reflected on the map; (2) the map erroneously displays a housing unit that is not on the ground; (3) the map erroneously displays a housing unit that is on the ground but incorrect; (4) the map does not display a housing unit appearing on the ground. Procedures for modifying your map will now be outlined by the facilitator.

A "think-aloud" method will be used during this exercise. You have been equipped with an audio recording device. You will be asked to verbalize your thoughts about performing the task throughout the exercise. These thoughts will be recorded to help us accurately recall your approach.

We ask that you perform this exercise using your typical practices for interpreting a map and identifying residential homes. It is important that you say aloud everything that you think or do. Do not feel uncomfortable or embarrassed about your approach or any of your thoughts. Even the minute pieces of information that you provide are important to us and all of your input will be held in the strictest of confidence.

Your facilitator's role will be only to observe and record your behavior. They will not interact with you other than to encourage you to think aloud or to get clarification on something that you said that cannot be verbally understood. We will begin by a mock exercise to acclimate you to the think-aloud process.

Once you have completed the training, you will then move on to the main exercise. It will be conducted in the same manner as the training—you should expect your observer to be completely passive at this point unless you have stopped verbalizing your thoughts. Before beginning the main exercise, please ensure that you are comfortable with the procedure and that all of your questions have been answered.

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

CODING SHEET

Date: __________ Observer: __________ ID: __________

Address: __________ Order: ____ Action: **V A M D** Incidental verification ☐ Time taken: _____

Map

Check: ______________________________

Rotation: ______________________________

Modification: ______________________________

Body

Heading: ______________________________

____

Body rotation: ______________________________

Street

Sign check: ______________________________

Numbering: ______________________________

Odd/even: ______________________________

Other: ______________________________

____

Miscellaneous

Confusion sources: ______________________________

____

List modification: ______________________________

____

NOTES

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

FIELD EXERCISE TESTS & QUESTIONNAIRE

Direction Test

A test of direction will occur just before the field exercise begins. The answer will be recorded by facilitator.

1. "Point due North."

Field exercise has been administered and is now complete.

Starting Point Test & Direction Test

A test to determine if the participant can locate their starting point will be administered followed by an additional test of direction.

1. "Point to your starting location."
2. "Point due North."

Post-Questionnaire

Questionnaire questions will be read to the participant. The answers will be recorded and later transcribed

1. How did you decide what address to start with?
2. How did you decide what address to do next? Did you change this approach for later addresses?
3. What features of the map were most helpful?
4. What features do you wish were on the map?
5. What features of the map could you have gone without?
6. Did the setting affect your approach to completing the tasks (e.g. weather, traffic, etc.)? If so, how?
7. How hard was it for you to find the addresses on the ground? Very Easy [1] – [5] Very Difficult
8. What address was the easiest to verify?
9. What address was the most difficult to verify?
10. What distracted you from the task?

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Resear
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

Date: 9/13/2010

To: Kofi Whitney
226 Atanasoff Hall

CC: Dr. Les Miller
112 Atanasoff Hall

From: Office for Responsible Research

Title: How do we use a paper map? An exploratory study of spatial ability and decision making.

IRB Num: 09-386

Approval Date: 9/13/2010 **Continuing Review Date:** 9/12/2011

Submission Type: Continuing Review / Modification **Review Type:** Expedited

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University. Please refer to the IRB ID number shown above in all correspondence regarding this study.

Your study has been approved according to the dates shown above. To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**
- **Obtain IRB approval prior to implementing any changes** to the study by submitting the "Continuing Review and/or Modification" form.
- **Immediately inform the IRB** of (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems** involving risks to subjects or others.
- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Research investigators are expected to comply with the principles of the Belmont Report, and state and federal regulations regarding the involvement of humans in research. These documents are located on the Office for Responsible Research website http://www.compliance.iastate.edu/irb/forms/ or available by calling (515) 294-4566.

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

LIST OF ATTACHMENTS FOR REVIEW Pg. 14 of Original IRB

IRB Modifications are in RED

1) **Study Advertisements** (for recruitment)
2) **Scheduling Scripts** (for scheduling participants)
3) **Informed Consent** (presented prior to Phase 1 screening)
4) **Background Information Questionnaire** (for Phase 1 screening)
5) **Cognitive Tests** (for Phase 1 screening)
6) **Training Script (for Phase 2 field exercise)**
7) **Coding Sheet** (for Phase 2 field exercise)
8) **Field Exercise Tests & Questionnaire** (for Phase 2 field exercise)
9) **Training/Field Exercise Map & Address List (for Phase 2 field exercise)**
   this was defined, but not included in original IRB—no changes to protocol have been made

ISU IRB # 1 09-386
Approved Date: 13 September 2010
Expiration Date: 12 September 2011

PARTICIPANTS NEEDED

to verify street addresses in a neighborhood for research study.

Participants must be ISU students, aged 18 or older, and fluent in English. Participants should have minimal exposure to Ames neighborhoods.

Participants will be offered compensation.

For more information or to schedule an appointment, please contact:

KOFI WHITNEY
Email: kwhitney@iastate.edu
Phone: (803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

ISU IRB # 1	09-386
Approved Date:	13 September 2010
Expiration Date:	12 September 2011

Newspaper advertisement

Participants Needed for Research Study

We are looking for participants to verify street addresses in a neighborhood. Participants must be ISU students, aged 18 or older, and fluent in English. Participants should have minimal exposure to Ames neighborhoods.

Participants completing the study will be offered compensation.

For more information or to schedule an appointment, please contact:
Kofi Whitney @ kwhitney@iastate.edu or (803) 546-0007.

SCHEDULING SCRIPTS

The Doodle Online scheduling tool will be used – http://www.doodle.com
After a student has been scheduled, a follow up email will be sent with appointment information

Screening – Email

Subject: Address Verification Study - Screening

Dear [*Student's Name*],

You have received this email because you have expressed an interest to participate in our Address Verification Study. The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1 hour to complete. You will receive a $10 gift card for participating in the screening.

You may be selected after this screening to later participate in a field exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. Compensation for the field exercise is a $20 gift card.

Click here [*Doodle Scheduling Link*] to schedule a screening appointment.

Thank you on behalf of my research group for participating in this study.

Sincerely,

Kofi Whitney
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Kofi Whitney @ 803-546-0007 or Dr. Les Miller @ 515-294-7934.

Screening – Phone

- "Hello, my name is [Scheduler's Name]. I am calling you because you have expressed an interest to participate in our Address Verification Study. I'd like to remind you that your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential."
- "May I continue?"
- "The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1hour to complete. You will receive a $10 gift card for participating in the screening. You may be selected after this screening to later participate in a field exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. Compensation for the field exercise is a $20 gift card."
- "May I email you the information that I have discussed along with a link that will allow you to schedule your screening appointment?"
- "Thank you for your time."

ISU IRB # 1 09-386
Approved Date: 13 September 2010
Expiration Date: 12 September 2011

SCHEDULING SCRIPTS (continued)

Field Exercise – Email

Subject: Address Verification Study - Field Exercise

Dear [*Student's Name*],

You have received this email because you participated in the screening portion of our Address Verification Study. We have reviewed your screening information and would like to invite you to participate in the field exercise portion of the study. This exercise will take approximately 2 hours to complete. You will receive a $20 gift card for your participation.

Please refer to your *Informed Consent* document for additional information regarding the study. You are welcome to contact us if you would like to receive another copy.

Click here [*Doodle Scheduling Link*] to schedule the field exercise.
Note: We may contact you to reschedule if weather conditions are not favorable.

Thank you on behalf of my research group for participating in this study.

Sincerely,

Kofi Whitney
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Kofi Whitney @ 803-546-0007 or Dr. Les Miller @ 515-294-7934.

FIELD EXERCISE TESTS & QUESTIONNAIRE

Direction Test

A test of direction will occur just before the field exercise begins. The answer will be recorded by facilitator.

1. "Point due North."

Field exercise has been administered and is now complete.

Starting Point Test & Direction Test

A test to determine if the participant can locate their starting point will be administered followed by an additional test of direction.

1. "Point to your starting location."
2. "Point due North."

Post-Questionnaire

Questionnaire questions will be read to the participant. The answers will be recorded and later transcribed

1. How did you decide what address to start with?
2. How did you decide what address to do next? Did you change this approach for later addresses?
3. What features of the map were most helpful?
4. What features do you wish were on the map?
5. What features of the map could you have gone without?
6. Did the setting affect your approach to completing the tasks (e.g. weather, traffic, etc.)? If so, how?
7. How hard was it for you to find the addresses on the ground? Very Easy [1] – [5] Very Difficult
8. What address was the easiest to verify?
9. What address was the most difficult to verify?
10. What distracted you from the task?

ISU IRB # 1 09-386
Approved Date: 13 September 2010
Expiration Date: 12 September 2011

INFORMED CONSENT

The purpose of this study is to gather information on how individuals compare addresses on a street with information on a paper map.

You will complete a brief background questionnaire followed by 3 cognitive assessments. This will take approximately 1 hour. Some subjects will be contacted at a later date to schedule an appointment for a field exercise. If selected for the field exercise, we will explain the task and train you on the procedures. You will then be transported, via CyRide, to an Ames neighborhood where you will practice the procedures. Next, we will give you a list of addresses to find and verify against a map that we provide. During the field exercise, you will be asked to think aloud as you reason through the task and this will be audio-recorded. This exercise will take approximately 2 hours to complete. There are no known risks to participation other than concerns that are normally associated with walking through a neighborhood.

There are no direct benefits to you as a participant other than the educational experience of being involved in an study. Your participation is helping us learn how we can improve address listing methods. By participating in this study, you will be offered a $10 gift card for completion of the background questionnaire and cognitive assessments. If you are selected for the field exercise, you will be offered a $20 gift card upon completion. You will need to sign a receipt for both gift cards.

Your participation in this study is completely voluntary and you may withdraw at any time. You may skip any part of this study that makes you feel uncomfortable or withdraw from the study at any time without penalty or loss of benefits to which you may otherwise be entitled.

Records identifying participants will be kept confidential to the extent permitted by applicable laws and regulations and will not be made publicly available. However, federal government agencies, the National Science Foundation, and the Institutional Review Board (a committee that reviews and approves human subject research studies) may inspect and/or copy your records for quality assurance and data analysis. These records may contain private information. To ensure confidentiality to the extent permitted by law, the following measures will be taken:

Data that identifies participants will be kept confidential. The information taken from this session will be assigned a unique code. Your name will not be associated with this information. Only researchers from Iowa State University working on this project will have access to data collected during this study. Study records will be kept confidential under password protected computer files. Data will be retained for two years and then will be destroyed.

The data results from this research may be used for educational or scientific purposes and may be presented at scientific and/or educational meetings or published in professional journals. Results will be released in summary form only with no personal identifying information.

For further information about the study contact Kofi Whitney at (803) 546-0007 or Dr. Les Miller at (515) 294-7588. If you have any questions about the rights of research subjects or research-related injury, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, Office of Research Assurances, (515) 294-3115, 1138 Pearson Hall, Ames, IA 50011.

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent.

Participant's Name (printed)____

______________________________ ____________________
(Participant's Signature) (Date)

INVESTIGATOR STATEMENT

I certify that the participant has been given adequate time to read and learn about the study and all of their questions have been answered. It is my opinion that the participant understands the purpose, risks, benefits and the procedures that will be followed in this study and has voluntarily agreed to participate.

______________________________ ____________________
(Signature of Person Obtaining Informed Consent) (Date)

THIS BOX IS FOR RESEARCHER USE ONLY	ID:________ Date:________

BACKGROUND INFORMATION

Please answer each of the questions by circling the appropriate response.

1. How often do you use maps?

1 = Never.
2 = Rarely.
3 = Sometimes.
4 = Frequently.
5 = All the time.

2. Student Status.

1 = Freshman.
2 = Sophomore.
3 = Junior.
4 = Senior.
5 = Graduate Student.

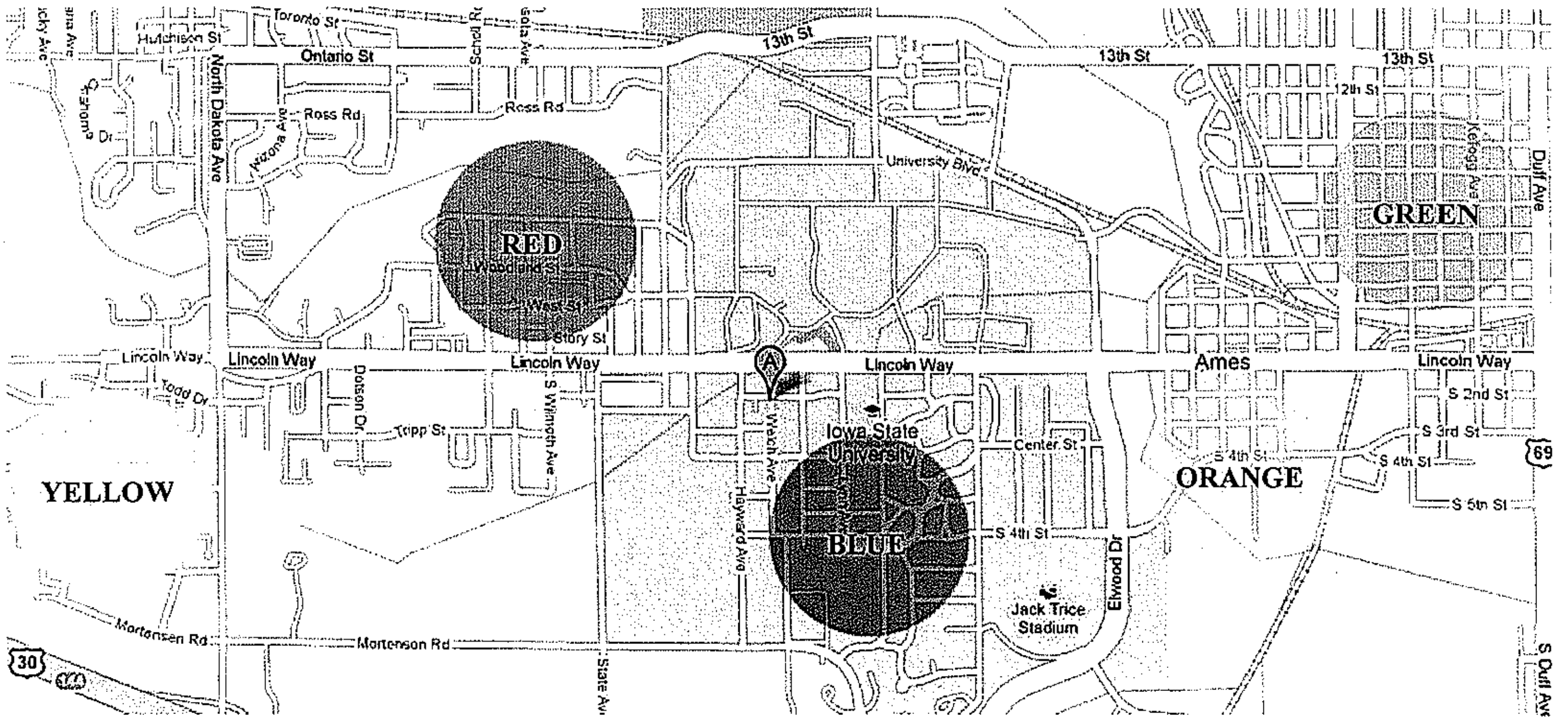
3. Gender.

1 = Male.
2 = Female.

4. How long have you lived in Ames?

1 = This is my first semester.
2 = 1-2 years.
3 = 2-3 years.
4 = 3-4 years.
5 = More than 4 years.

BACKGROUND INFORMATION (continued)



Please use the map of Ames above to answer the following questions.

5. How familiar are you with the neighborhood in yellow?

    1 = Never seen it.
    2 = Slightly familiar.
    3 = Moderately familiar.
    4 = Very familiar.

6. How familiar are you with the neighborhood in red?

    1 = Never seen it.
    2 = Slightly familiar.
    3 = Moderately familiar.
    4 = Very familiar.

7. How familiar are you with the neighborhood in blue?

    1 = Never seen it.
    2 = Slightly familiar.
    3 = Moderately familiar.
    4 = Very familiar.

8. How familiar are you with the neighborhood in orange?

    1 = Never seen it.
    2 = Slightly familiar.
    3 = Moderately familiar.
    4 = Very familiar.

9. How familiar are you with the neighborhood in green?

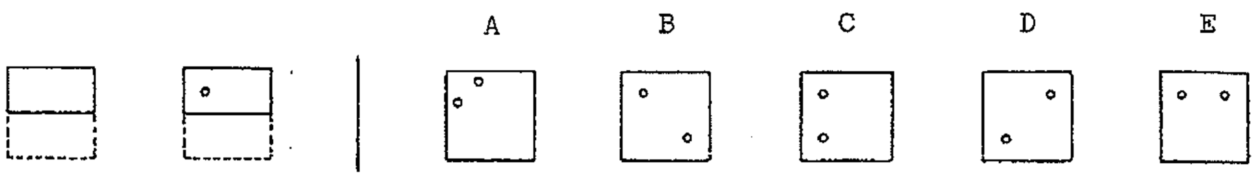
    1 = Never seen it.
    2 = Slightly familiar.
    3 = Moderately familiar.
    4 = Very familiar.

Name ______________________________

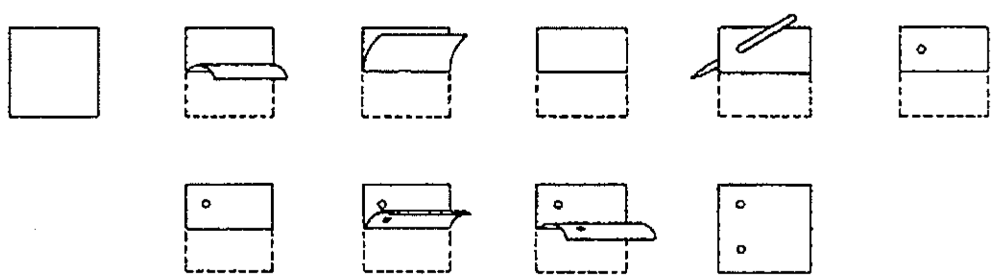
PAPER FOLDING TEST -- VZ-2

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures at the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper.)



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.



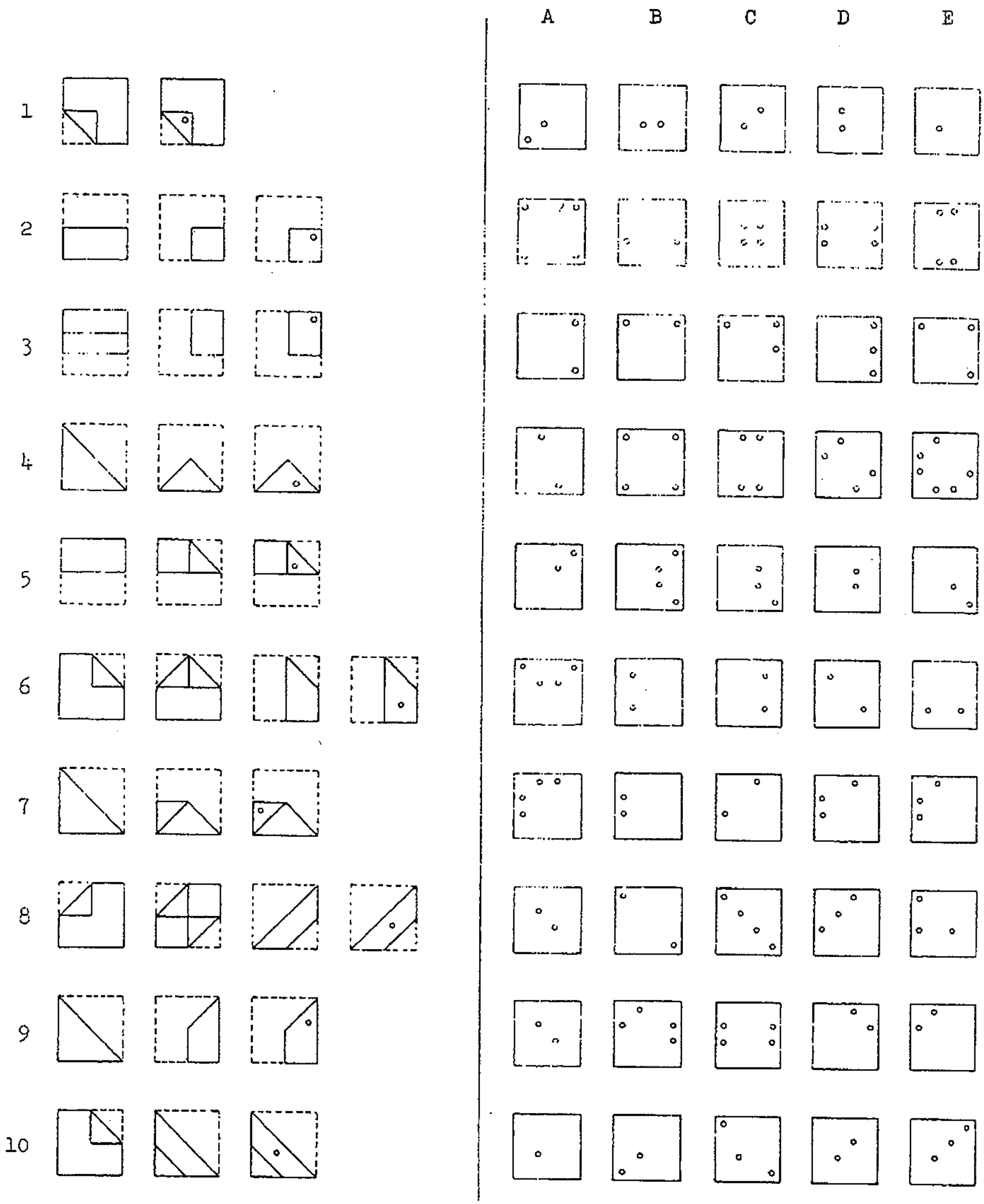
In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Part 1 (3 minutes)

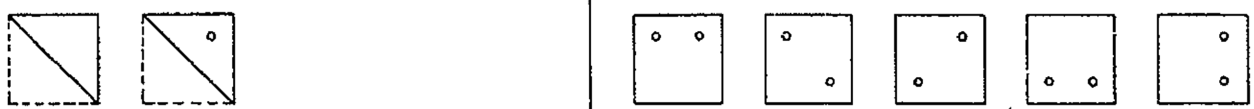


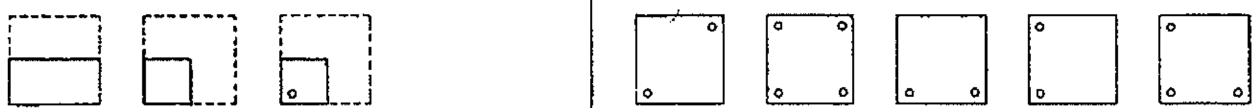
DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

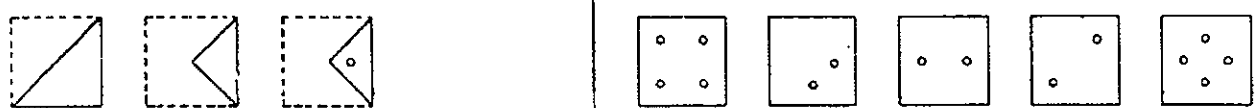
Part 2 (3 minutes)

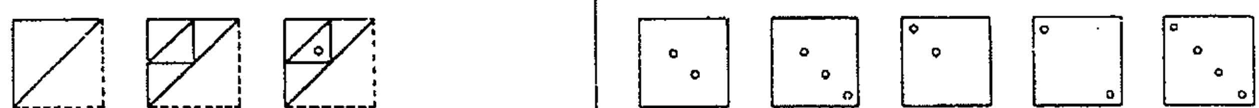
A B C D E

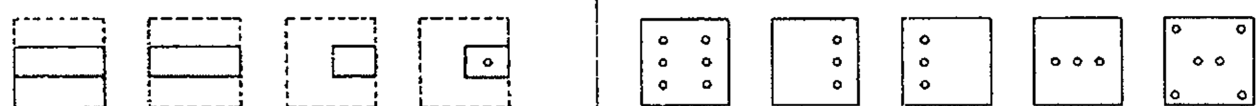
11 

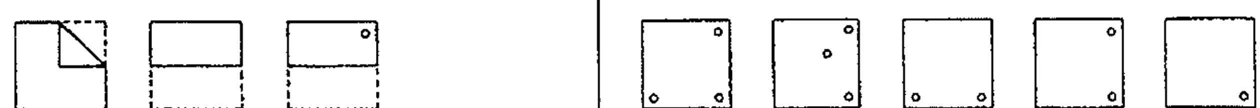
12 

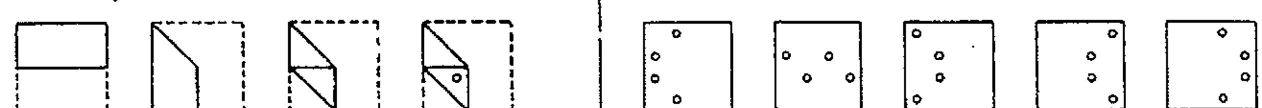
13 

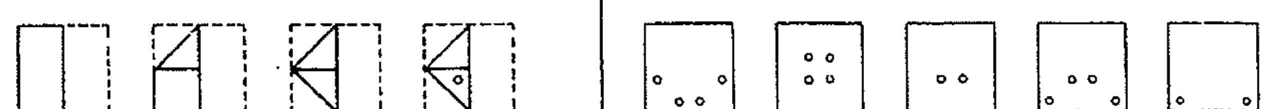
14 

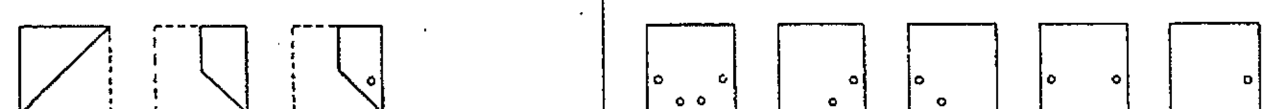
15 

16 

17 

18 

19 

20 

DO NOT GO BACK TO PART 1, AND

DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. **STOP.**

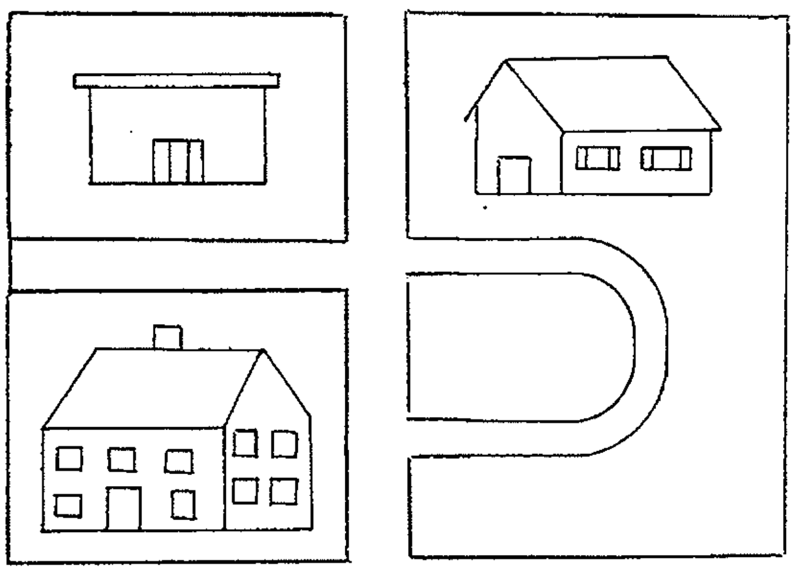
Name ______________________________

BUILDING MEMORY -- MV-2

This is a test of your ability to remember the position of things on a street map.

You will be given a map with streets and buildings and other structures to study. After you have had some time to learn the street layout and the different kinds of structures, you will be asked to turn to a test page. On that page you will find the street map and numbered pictures of some of the structures. You will be asked to put an x on the letter that shows where each of the structures was located on the study map.

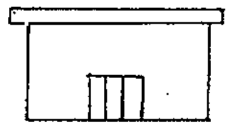
Now look at this simple and enlarged sample:



After you have studied the sample above for a minute, turn to the next page.

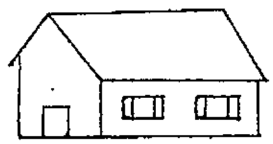
Look at the numbered houses on the left. For each item mark an X on the letter below each building that corresponds with where each house was located on the study map.

1.



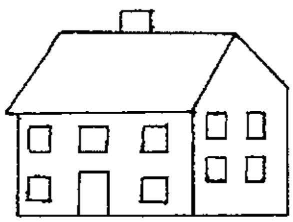
A B C D E

2.

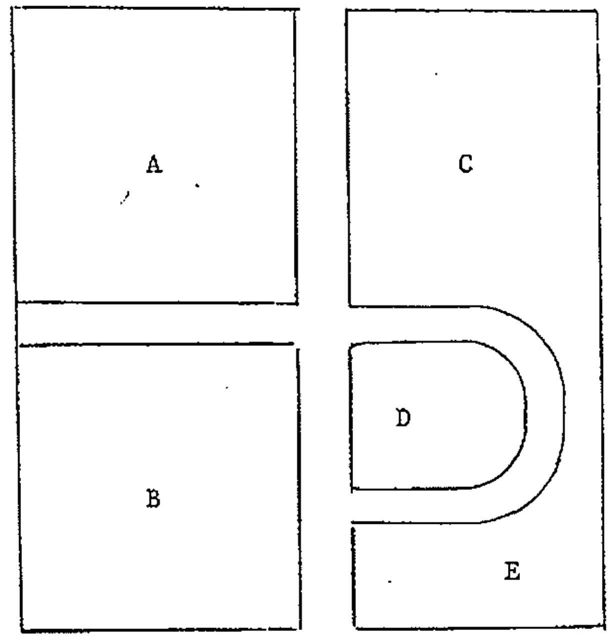


A B C D E

3.



A B C D E



Your answers for sample item 1 should be A, for 2, C, and for 3, B.

Your score on this test will be the number of buildings placed correctly minus a fraction of the number wrong. Therefore, it will <u>not</u> be to your advantage to guess unless you can eliminate some of the locations as definitely wrong.

There are two sections to each part of this test. The first section is a map which you will study for <u>4 minutes</u>. The second is the test section and contains 12 structures to be located on the map. You will have <u>4 minutes</u> to mark your answers. Mark A, B, C, D, or E for each building. In the test section, the buildings will be mixed up and not necessarily near the part of the map where you first saw them.

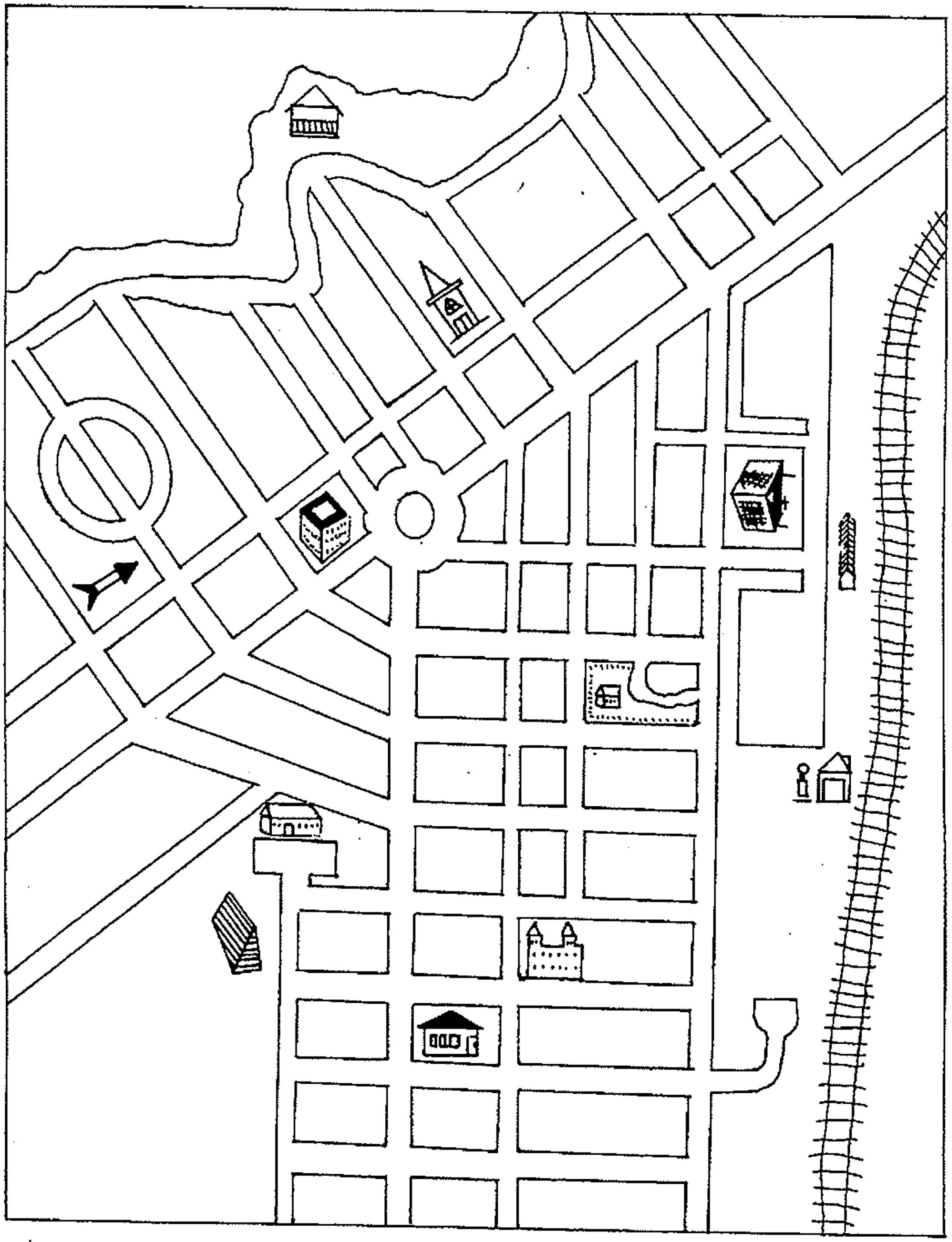
This test has two parts. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STUDY PAGE

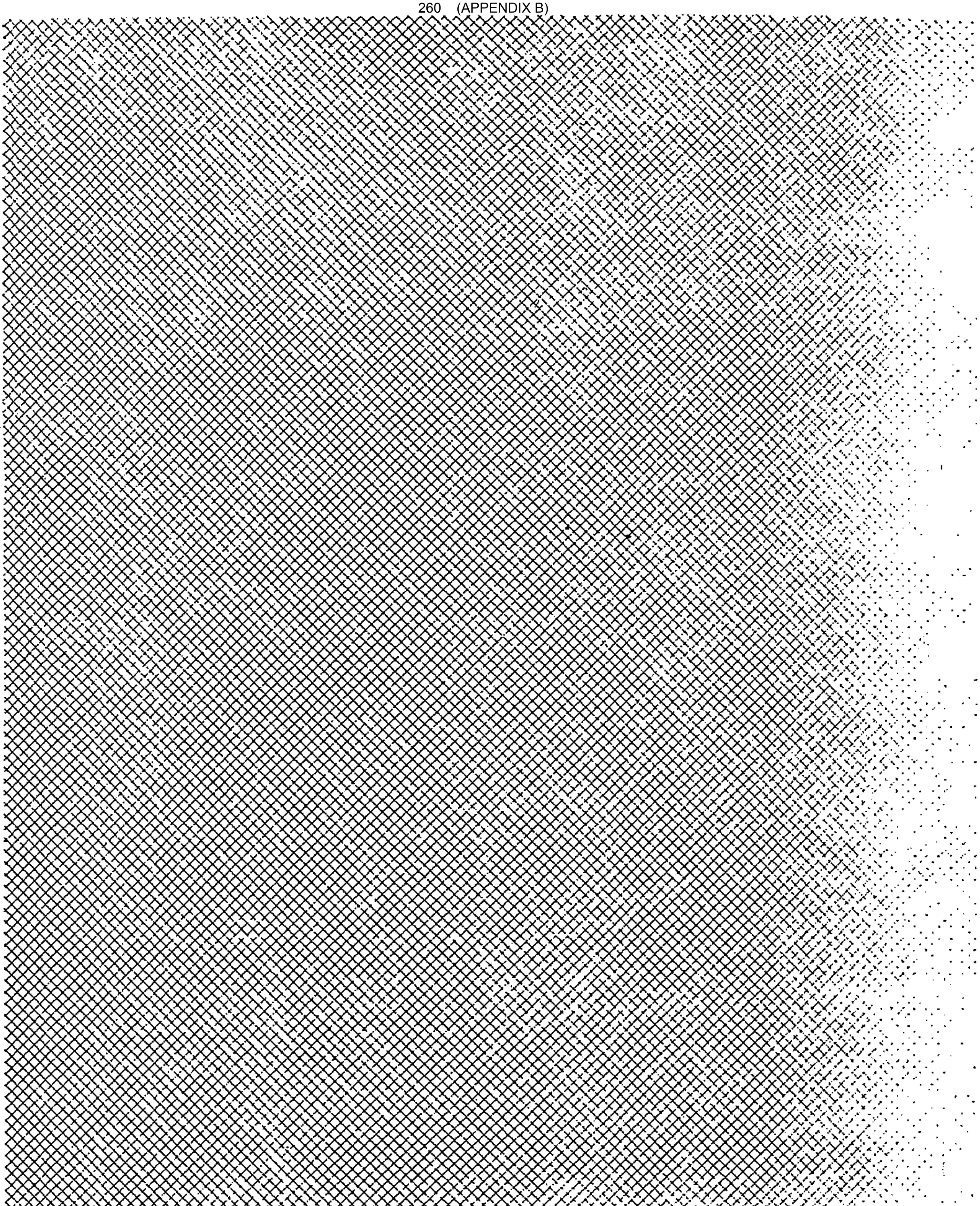
Part 1 (4 minutes)

Study this map so you can remember where each building is located.



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP



TEST PAGE

Part I (4 minutes)

Mark an X on the letter below each building that shows where it was seen on the map.

1. 

A B C D E

2. 

A B C D E

3. 

A B C D E

4. 

A B C D E

5. 

A B C D E

6. 

A B C D E

7. 

A B C D E

8. 

A B C D E

9. 

A B C D E

10. 

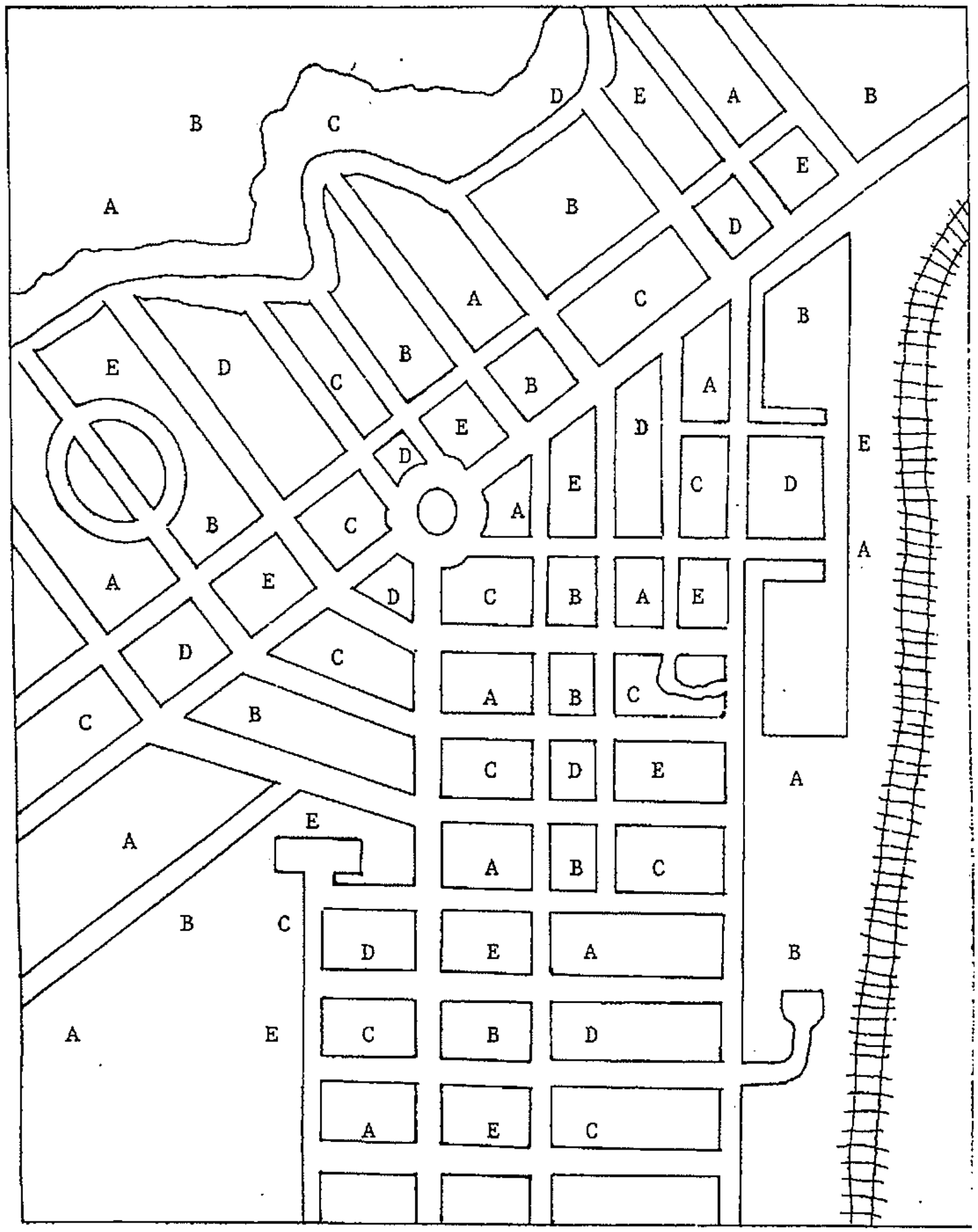
A B C D E

11. 

A B C D E

12. 

A B C D E



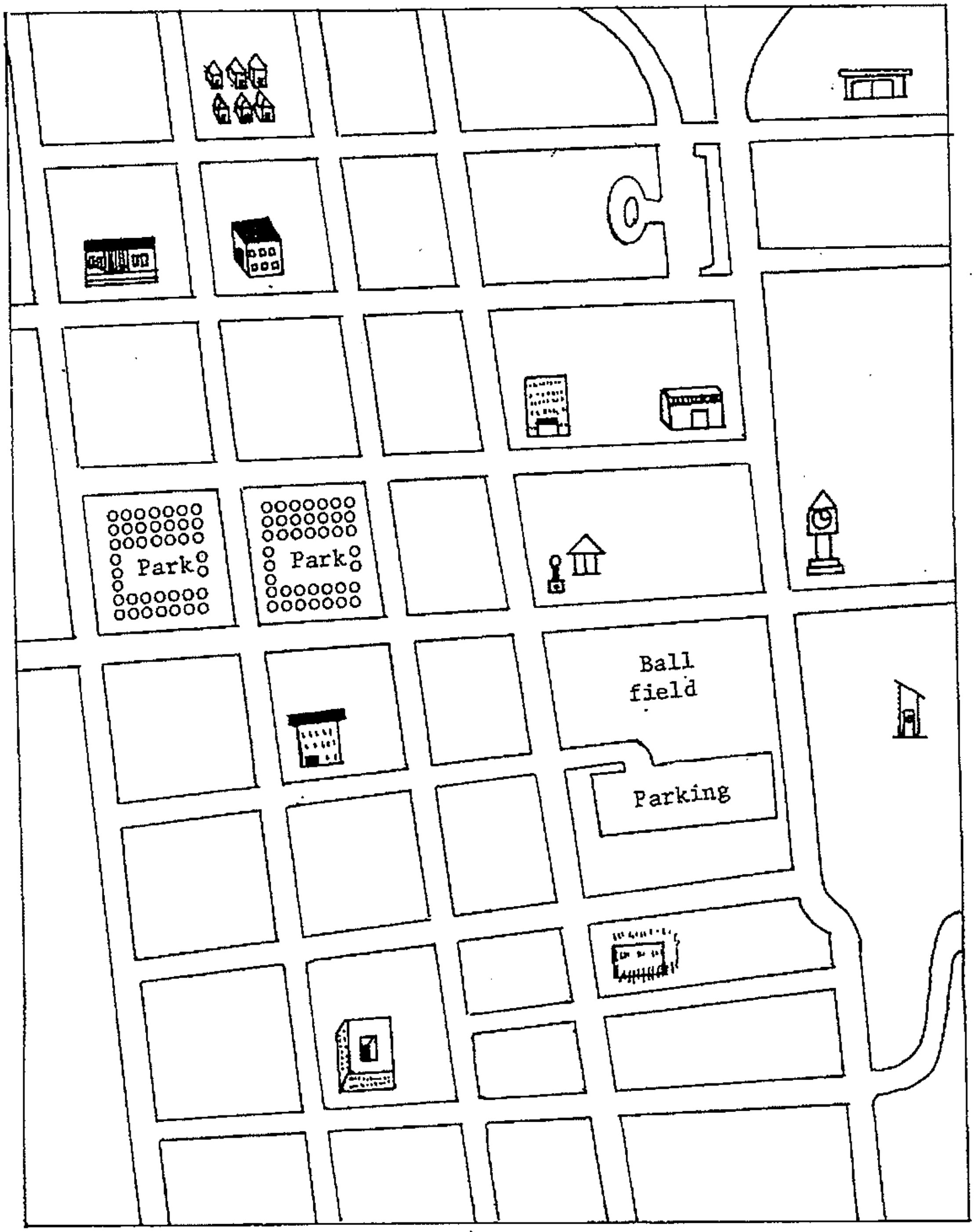
DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP.

STUDY PAGE

Part 2 (4 minutes)

Study this map so you can remember where each building is located.



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP.

TEST PAGE

Part 2 (4 minutes)

13. 

A B C D E

14. 

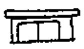
A B C D E

15. 

A B C D E

16. 

A B C D E

17. 

A B C D E

18. 

A B C D E

19. 

A B C D E

20. 

A B C D E

21. 

A B C D E

22. 

A B C D E

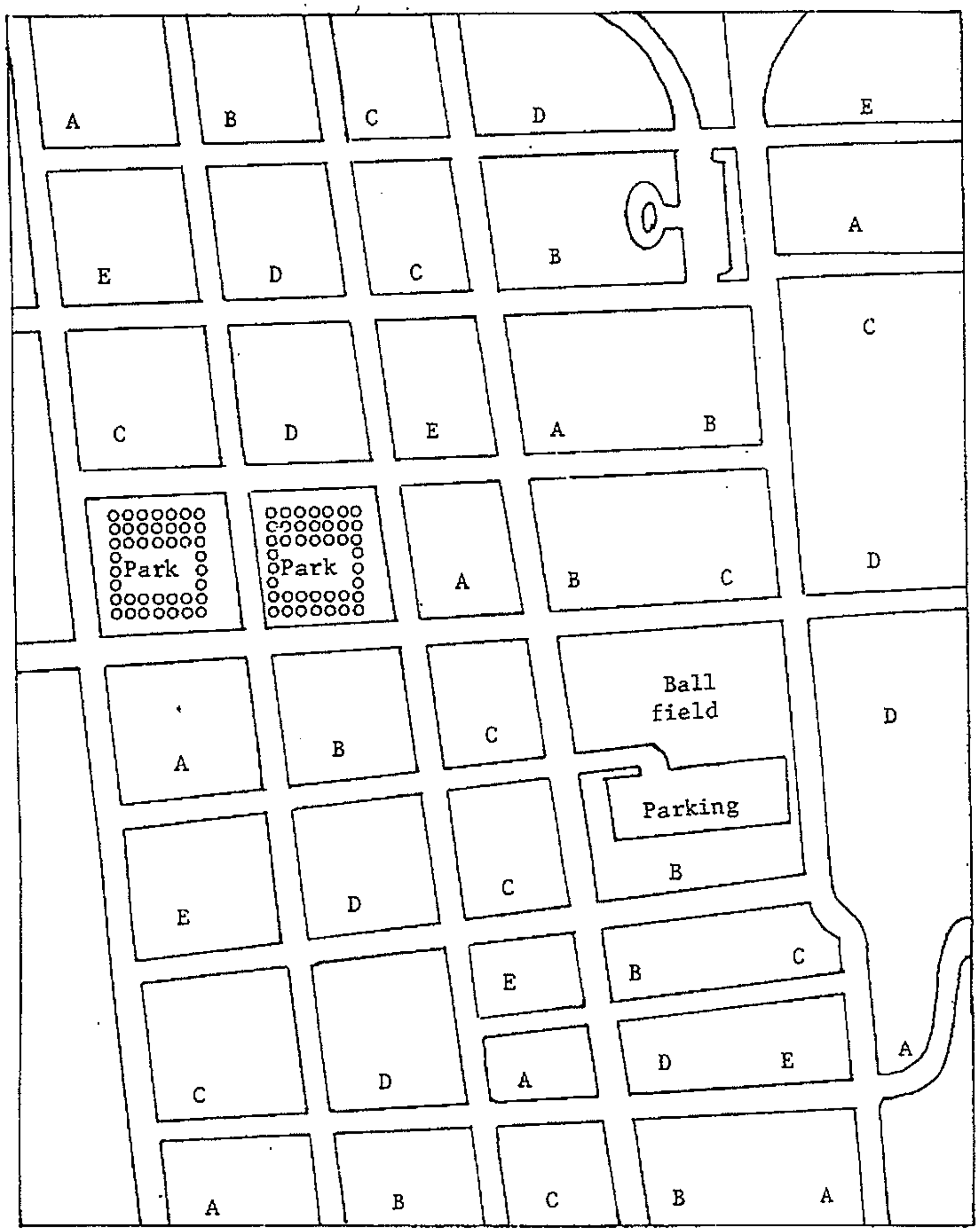
23. 

A B C D E

24. 

A B C D E

Mark an X on the letter below each building that shows where it was seen on the map.



DO NOT GO BACK TO PART 1 AND DO NOT GO ON TO ANY OTHER TEST

STOP.

TRAINING SCRIPT

Participants will read along as this is dictated by the facilitator

Kofi Whitney 8/19/10 11:53 AM
Comment: Training Script is similar to original. Edits were made for readability and to emphasize that participants focus on their map actions as they think-aloud.

I will give you a map and a list of three addresses.

Please determine if the addresses from the list are properly shown on the map. You may verify them in any order.

There are four possible outcomes for each address on the list:

1. The address on the list appears correctly on the map and in the neighborhood – **No Change;**
2. The address on the list does appear on the map but is located in a different location in the neighborhood - **Move.**
3. The address on the list does not appear on the map but does appear in the neighborhood - **Add;**
4. The address on the list does not appear on the map or in the neighborhood - **Delete;**

You may approach the task however you see fit. We ask that you think aloud about what you are doing and thinking as you perform the task so we can record you through the attached microphone. We are particularly interested in how you are using the map to support your work, solve problems you encounter, and make decisions about your task.

We want to know the WHAT, HOWs and WHYs of your actions. Do not feel uncomfortable or embarrassed about any of your thoughts. We want to know when you are confused and why, and we want you to tell us what you are thinking when things are going well. Even the little details are important to our understanding of how you use the map. Everything you say or do will be kept in the strictest of confidence.

My job will be only to observe and record what you do. I may say "what do you mean by that?" or "keep talking." Otherwise I will be completely passive.

The first three addresses are for training. While you work on them, I can answer your questions. Please ensure you are comfortable with the task before we move to the main exercise.

CODING SHEET

Date: ____________ Observer: ____________ ID: ____________

Address: ____________ Order#: ____ Action: **V A M D** Time taken: _____

Map

Check: ______________________________

Rotation: ______________________________

Modification: ______________________________

Body

Heading: ______________________________

____

Body rotation: ______________________________

Street Odd/Even

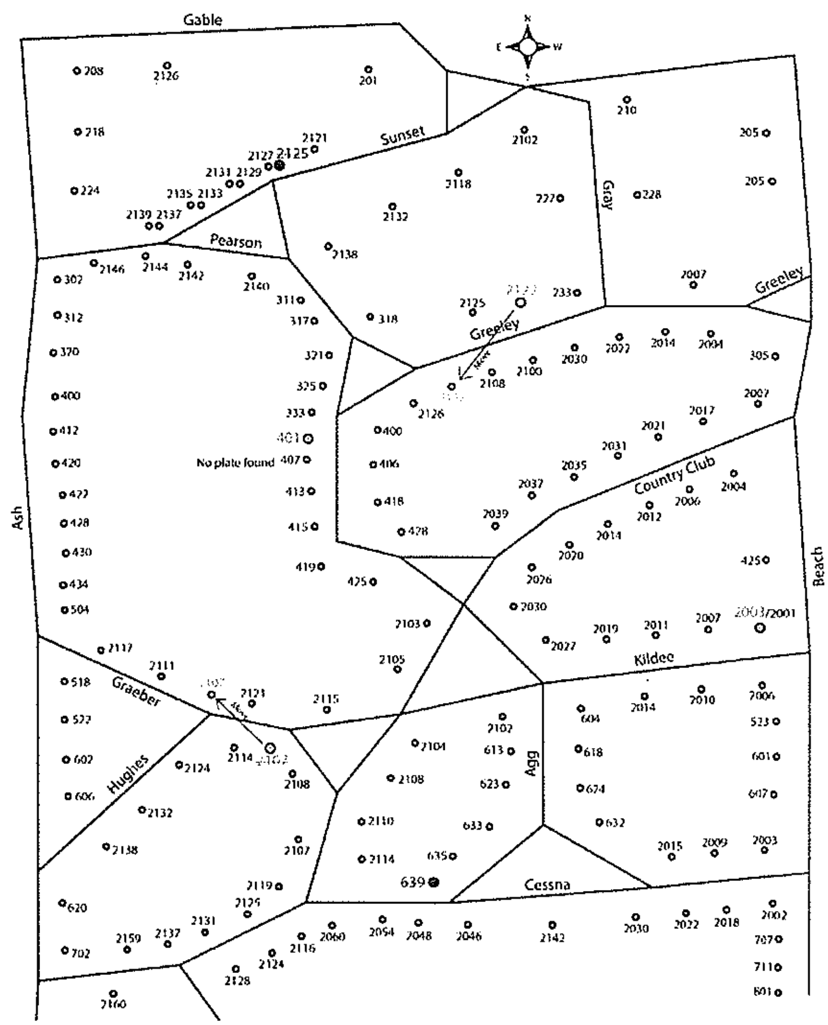
Sign check: ______________________________

Numbering: ______________________________

Odd/even: ______________________________

Other: ______________________________

NOTES



FIELD EXERCISE TESTS & QUESTIONNAIRE

Direction Test

A test of direction will occur just before the field exercise begins. The answer will be recorded by facilitator.

1. "Point due North."

Field exercise has been administered and is now complete.

Starting Point Test & Direction Test

A test to determine if the participant can locate their starting point will be administered followed by an additional test of direction.

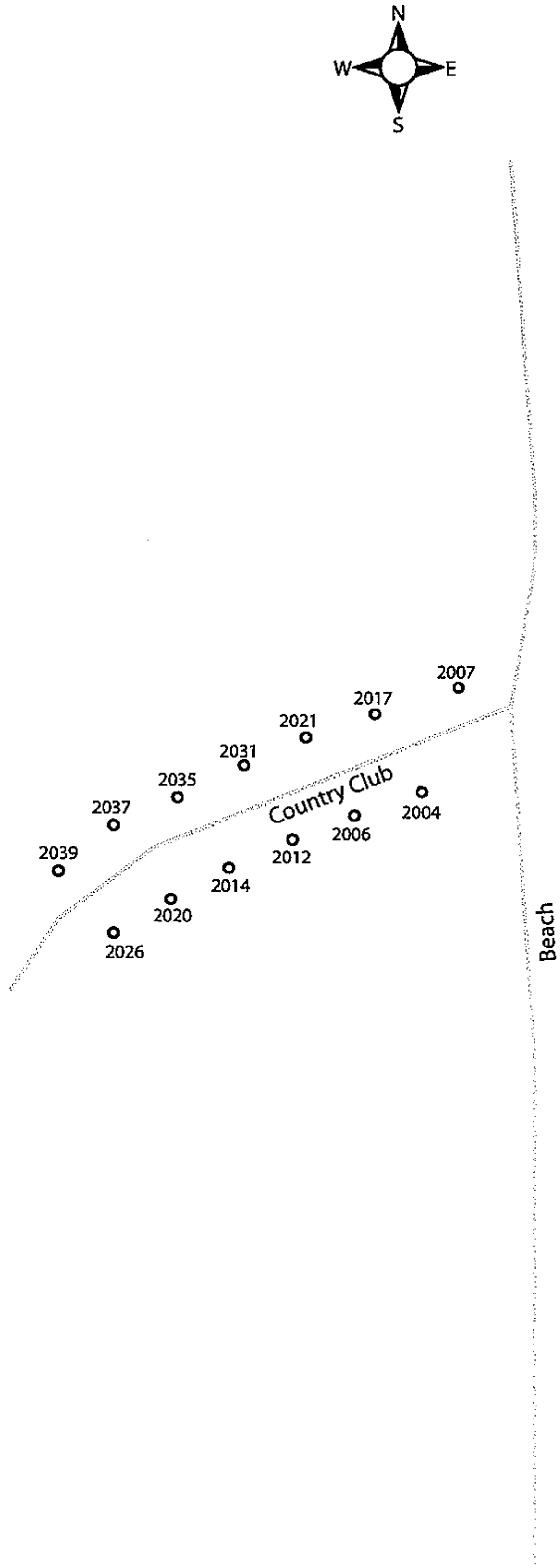
1. "Point to your starting location."
2. "Point due North."

Post-Questionnaire

Questionnaire questions will be read to the participant. The answers will be recorded and later transcribed

1. How did you decide what address to start with?
2. How did you decide what address to do next? Did you change this approach for later addresses?
3. What features of the map were most helpful?
4. What features do you wish were on the map?
5. What features of the map could you have gone without?
6. Did the setting affect your approach to completing the tasks (e.g. weather, traffic, etc.)? If so, how?
7. How hard was it for you to find the addresses on the ground? Very Easy [1] – [5] Very Difficult
8. What address was the easiest to verify?
9. What address was the most difficult to verify?
10. What distracted you from the task?

ID:____________



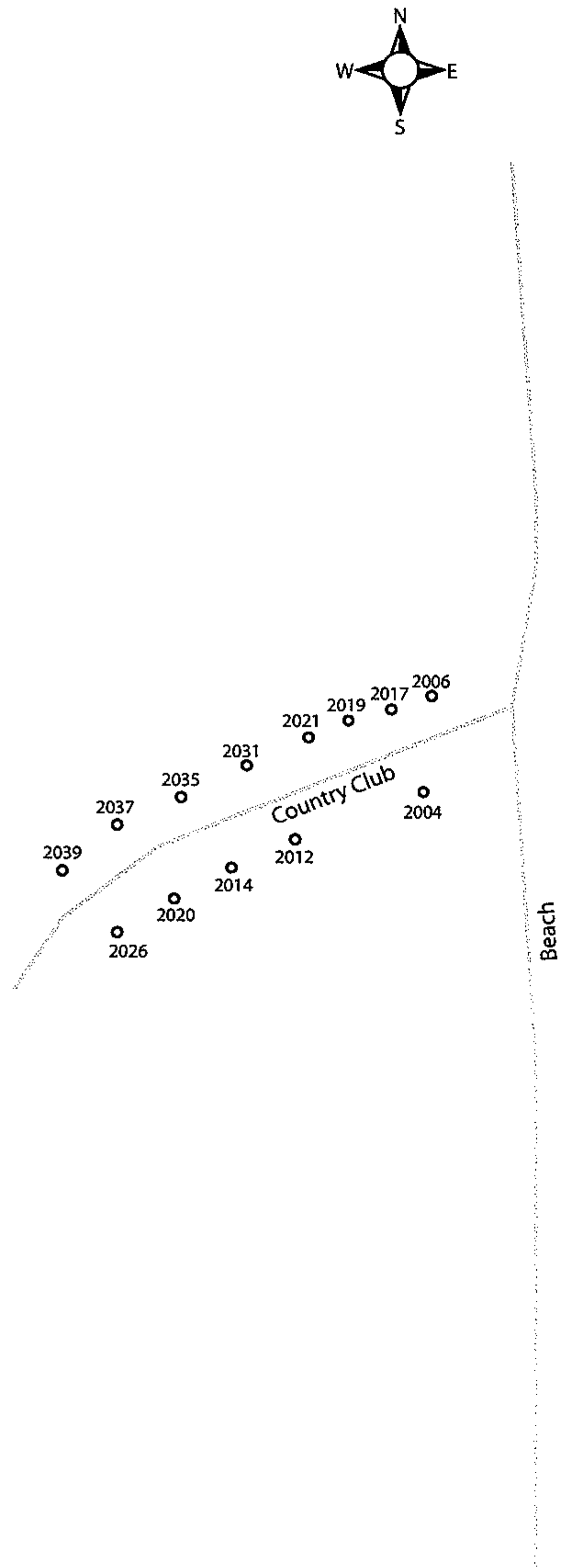
ADDRESSES TO VERIFY

2019 Country Club

2006 Country Club

2007 Country Club

ID:__________



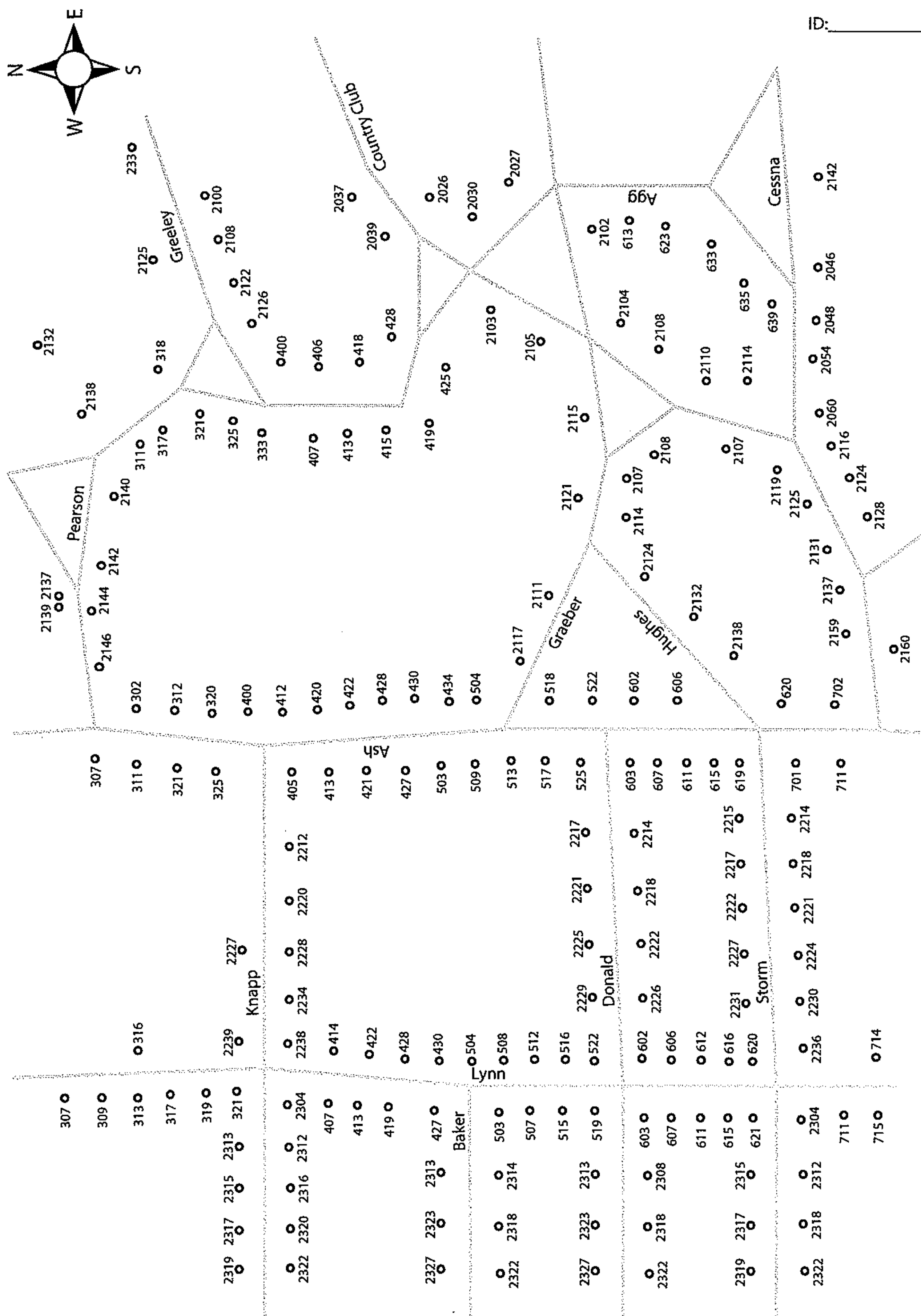
ADDRESSES TO VERIFY

2019 Country Club

2006 Country Club

2007 Country Club

ID:___________


N
E
S
W
Greeley
Country Club
Agg
Cessna
Pearson
Graeber
Hughes
Ash
Knapp
Donald
Storm
Lynn
Baker

Date: ____________ **Observer:** ______________ **ID:** ____________

2313 Knapp St.

2107 Graeber St.

639 Agg Ave.

509 Ash Ave.

401 Pearson Ave.

2221 Storm St.

524 Lynn Ave.

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Resear
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

Date: 3/30/2011

To: Georgi Batinov
226 Atanasoff Hall

CC: Dr. Les Miller
112 Atanasoff Hall

From: Office for Responsible Research

Title: Do Spatial Ability Differences Persist in a Virtual Environment

IRB Num: 10-075

Approval Date:	3/28/2011	**Continuing Review Date:**	3/30/2012
Submission Type:	Continuing Review / Modification	**Review Type:**	Expedited

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University. Please refer to the IRB ID number shown above in all correspondence regarding this study.

Your study has been approved according to the dates shown above. To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**
- **Obtain IRB approval prior to implementing any changes** to the study by submitting the "Continuing Review and/or Modification" form.
- **Immediately inform the IRB** of (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems** involving risks to subjects or others.
- **Stop all research activity if IRB approval lapses**, unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Research investigators are expected to comply with the principles of the Belmont Report, and state and federal regulations regarding the involvement of humans in research. These documents are located on the Office for Responsible Research website http://www.compliance.iastate.edu/irb/forms/ or available by calling (515) 294-4566.

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

LIST OF ATTACHMENTS FOR REVIEW

1) **Study Advertisements** (for recruitment)
2) **Scheduling Scripts** (for scheduling participants)
3) **Informed Consent** (presented prior to Phase 1)
4) **Background Information Questionnaire** (for Phase 1)
5) **Cognitive Tests** (for Phase 1)
6) **Research Participant Receipt Form**
7) **Training Script** (for Phase 2 field and virtual environment exercises)
8) **Coding Sheet** (for Phase 2 field and virtual environment exercises)
9) **Personal digital assistant storyboards** (for both Phase 2 treatments)
10) **Participant viewing perspective in the virtual environment**
11) **Phase 2 Exercise Tests & Questionnaire** (for Phase 2 field and virtual environment exercises)

ISU IRB # 1	10-075
Approved Date:	28 March 2011
Expiration Date:	30 March 2012

PARTICIPANTS NEEDED

to verify street addresses for a research study.

Participants must be 18 or older and fluent in English. Participants should have little exposure to Ames neighborhoods.

Participants in the study will be offered compensation.
For more information or to schedule an appointment, please contact:

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

Email: batinov@iastate.edu
Phone: (515) 450-543

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

Newspaper advertisement

Participants Needed for Research Study

We are looking for participants to verify street addresses for a research study. Participants must be 18 or older and fluent in English. Participants should have little exposure to Ames neighborhoods.

Participants in the study will be offered compensation.

For more information or to schedule an appointment, please contact **Georgi Batinov:** batinov@iastate.edu or (515) 450-5435.

Email advertisement

Subject: Participants Needed for Research Study

We are looking for participants to verify street addresses for a research study. Participants must be 18 or older and fluent in English. Participants should have little exposure to Ames neighborhoods.

Participants in the study will be offered compensation.

For more information or to schedule an appointment, please contact Georgi Batinov: batinov@iastate.edu or (515) 450-5435.

SCHEDULING SCRIPTS

The Doodle Online scheduling tool will be used – http://www.doodle.com

After a student has been scheduled, a follow up email will be sent with appointment information

Screening – Email

Subject: Address Verification Study - Screening

Dear [*Participant's Name*],

You have received this email because you have expressed an interest to participate in our Address Verification Study. The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1 hour to complete, after which you will receive a $10 gift certificate as compensation.

You may be selected after this screening to later participate either in a field exercise or in a virtual environment exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field and virtual environment exercises.
You will receive a $20 gift certificate for participating in the exercise. In the unlikely event that you are affected by virtual reality sickness, you will receive a $10 gift certificate and be excused from participating in the study.
Click here [*Doodle Scheduling Link*] to schedule a screening appointment.

Thank you on behalf of my research group for participating in this study.

Georgi Batinov
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Georgi Batinov@ 515-450-5435 or Dr. Les Miller @ 515-294-7934.

Screening – Phone

- "Hello, my name is [Scheduler's Name]. I am calling you because you have expressed an interest to participate in our Address Verification Study. I'd like to remind you that your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential."
- "May I continue?"
- "The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1 hour to complete, after which you will receive a $10 gift certificate as compensation. You may be selected after this screening to later participate in either a field exercise or a virtual environment exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. You will receive a $20 gift certificate for participating in the exercise. In the unlikely event that you are affected by virtual reality sickness, you will receive a $10 gift certificate and be excused from participating in the study.
- "May I email you the information that I have discussed along with a link that will allow you to schedule your screening appointment?"
- "Thank you for your time."

SCHEDULING SCRIPTS (continued)

Field Exercise – Email

Subject: Address Verification Study - Field Exercise

Dear [*Participant's Name*],

You have received this email because you participated in the screening portion of our Address Verification Study. We have reviewed your screening information and would like to invite you to participate in the field exercise portion of the study. This exercise will take approximately 2 hours to complete. You will receive a $20 gift certificate for participating.

Please refer to your *Informed Consent* document for additional information regarding the study. You are welcome to contact us if you would like to receive another copy.

Click here [*Doodle Scheduling Link*] to schedule the field exercise.
Note: We may contact you to reschedule if weather conditions are not favorable.

Thank you on behalf of my research group for participating in this study.

Georgi Batinov
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Georgi Batinov @ 515-450-5435 or Dr. Les Miller @ 515-294-7934.

Virtual Environment Exercise – Email

Subject: Address Verification Study - Field Exercise

Dear [*Participant's Name*],

You have received this email because you participated in the screening portion of our Address Verification Study. We have reviewed your screening information and would like to invite you to participate in the virtual environment exercise portion of the study. This exercise will take approximately 2 hours to complete. You will receive a $20 gift certificate for participating. In the unlikely event that you are affected by virtual reality sickness, you will receive a $10 gift certificate and be excused from participating in the study.

Please refer to your *Informed Consent* document for additional information regarding the study. You are welcome to contact us if you would like to receive another copy.

Click here [*Doodle Scheduling Link*] to schedule the virtual environment exercise.

Thank you on behalf of my research group for participating in this study.

Georgi Batinov
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Georgi Batinov @ 515-450-5435 or Dr. Les Miller @ 515-294-7934.

INFORMED CONSENT

The purpose of this study is to gather information on how individuals perform address surveying tasks in the field and in virtual environment.

You will complete a brief background questionnaire followed by 4 cognitive assessments. This will take approximately 1 hour. Some subjects will be contacted at a later date to schedule an appointment for either a field exercise or a virtual environment exercise. You will be given a $10 gift certificate for participation in this phase of the experiment, and you will need to sign a receipt for the gift certificate.

If selected for the field exercise, we will explain the task and train you on the procedures. You will then be transported, via CyRide, to an Ames neighborhood where you will practice the procedures. Next, we will give you a list of addresses to find and verify against a map that we provide on a personal digital assistant. This exercise will take approximately 2 hours to complete. There are no known risks to participation other than concerns that are normally associated with walking through a neighborhood. You will receive a $20 gift certificate for participating, and you will need to sign a receipt for it.

If selected for the virtual environment exercise, you will be asked to come to a virtual environment laboratory in Howe Hall. We will explain the task and train you on the procedures. Next, we will give you a list of addresses to find and verify in virtual environment against a map that we provide on a personal digital assistant. This exercise will take approximately 2 hours to complete. If you experience virtual reality motion sickness, you will be asked whether you want to continue the exercise. If you desire to stop, we will discontinue the experiment.

You will receive a $20 gift certificate for participating and will need to sign a receipt for it. In the unlikely event that you are affected by virtual reality sickness, you will receive a $10 gift certificate and be excused from participating in the study.

There are no direct benefits to you as a participant other than the educational experience of being involved in an study. Your participation is helping us learn how we can improve address listing methods.

Your participation in this study is completely voluntary and you may withdraw at any time. You may skip any part of this study that makes you feel uncomfortable or withdraw from the study at any time without penalty or loss of benefits to which you may otherwise be entitled.

Records identifying participants will be kept confidential to the extent permitted by applicable laws and regulations and will not be made publicly available. However, federal government agencies, the National Science Foundation, and the Institutional Review Board (a committee that reviews and approves human subject research studies) may inspect and/or copy your records for quality assurance and data analysis. These records may contain private information. To ensure confidentiality to the extent permitted by law, the following measures will be taken:

Data that identifies participants will be kept confidential. The information taken from this session will be assigned a unique code. Your name will not be associated with this information. Only researchers from Iowa State University working on this project will have access to data collected during this study. Study records will be kept confidential under password protected computer files. Data will be retained for two years and then will be destroyed.

The data results from this research may be used for educational or scientific purposes and may be presented at scientific and/or educational meetings or published in professional journals. Results will be released in summary form only with no personal identifying information.

For further information about the study contact Georgi Batinov at (515) 450-5435 or Dr. Les Miller at (515) 294-7588. If you have any questions about the rights of research subjects, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, 1138 Pearson Hall, Ames, IA 50011.

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent.

Participant's Name (printed)____

________________________________ ____________________

(Participant's Signature) (Date)

INVESTIGATOR STATEMENT

I certify that the participant has been given adequate time to read and learn about the study and all of their questions have been answered. It is my opinion that the participant understands the purpose, risks, benefits and the procedures that will be followed in this study and has voluntarily agreed to participate.

________________________________ ____________________

(Signature of Person Obtaining Informed Consent) (Date)

THIS BOX IS FOR RESEARCHER USE ONLY	ID:________ Date:________

BACKGROUND INFORMATION

Please answer each of the questions by circling the appropriate response.

1. How often do you use maps?

1 = Never.
2 = Rarely.
3 = Sometimes.
4 = Frequently.
5 = All the time.

2. Student Status.

1 = Not a student.
2 = Freshman.
3 = Sophomore.
4 = Junior.
5 = Senior.
6 = Graduate Student.

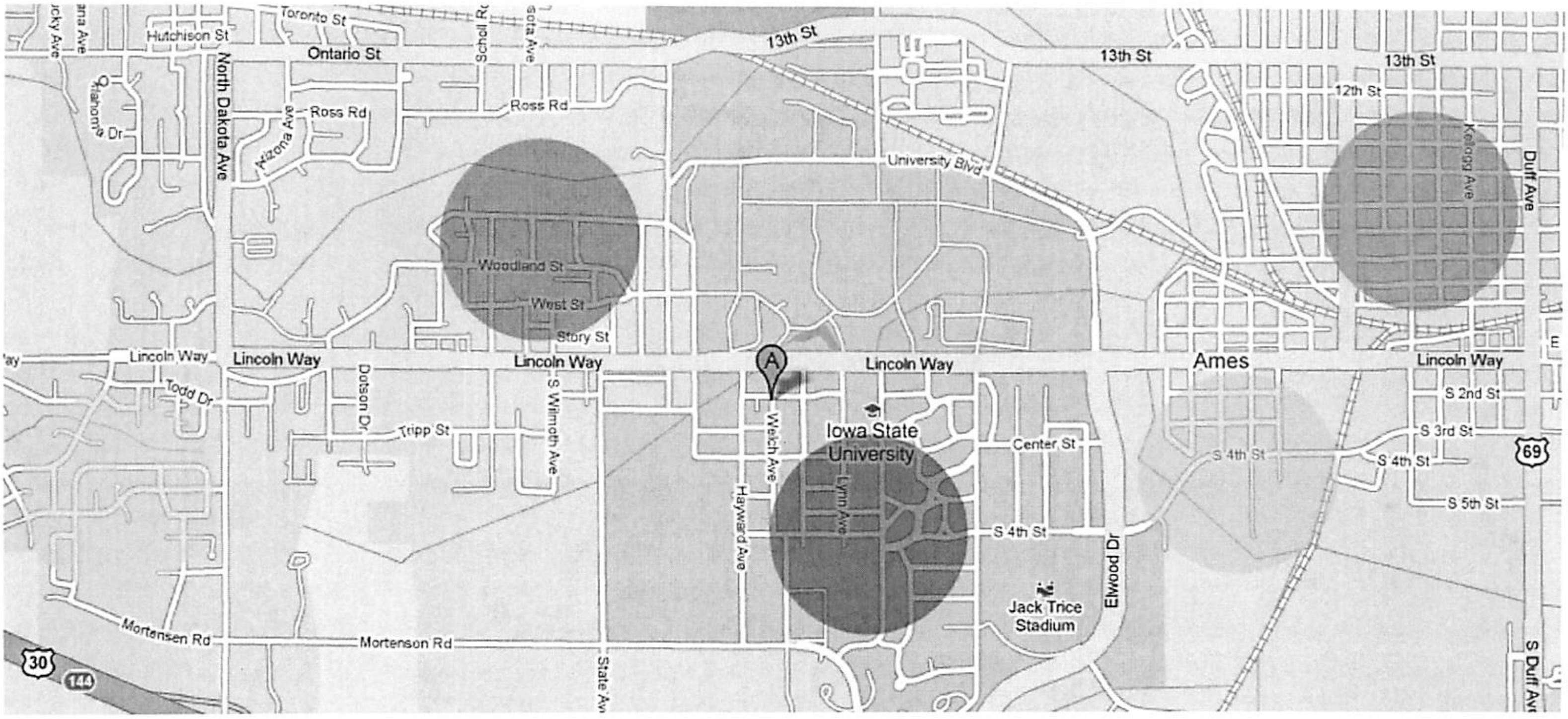
3. Gender.

1 = Male.
2 = Female.

4. How long have you lived in Ames?

1 = Less than 6 months.
2 = Less than a year.
3 = 1-2 years.
4 = 2-3 years.
5 = 3-4 years.
6 = More than 4 years.

BACKGROUND INFORMATION (continued)



Please use the map of Ames above to answer the following questions.

5. How familiar are you with the neighborhood in yellow?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

6. How familiar are you with the neighborhood in red?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

7. How familiar are you with the neighborhood in blue?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

8. How familiar are you with the neighborhood in orange?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

9. How familiar are you with the neighborhood in green?

1 = Never seen it.
2 = Slightly familiar.
3 = Moderately familiar.
4 = Very familiar.

COGNITIVE TESTS

Four cognitive tests will be administered to each participant in Phase 1:

1. Paper Folding Test (Ekstrom et al. 1976), measuring small-scale external-frame-of-reference mental rotation ability;
2. Building Memory Test (Ekstrom et al. 1976), measuring visual memory; and
3. Perspective Taking Ability Test (Kozhevnikov et al. 1976), measuring large-scale internal-frame-of-reference mental rotation ability.
4. Number Comparison Test (Ekstrom et al. 1976), measuring perceptual speed.

References

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service

Kozhevnikov, M., Motes, M., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, 20, 397-417.

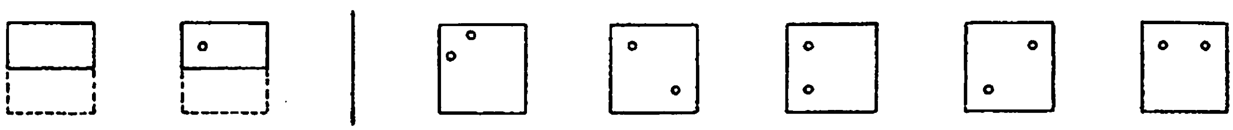
Name ____________________

PAPER FOLDING TEST — VZ-2

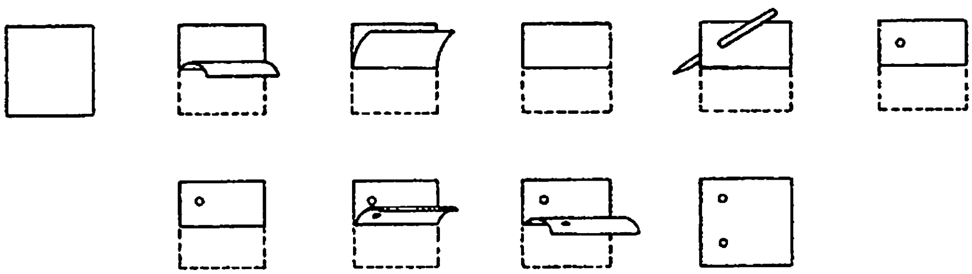
In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures at the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper.)

A B C D E



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.



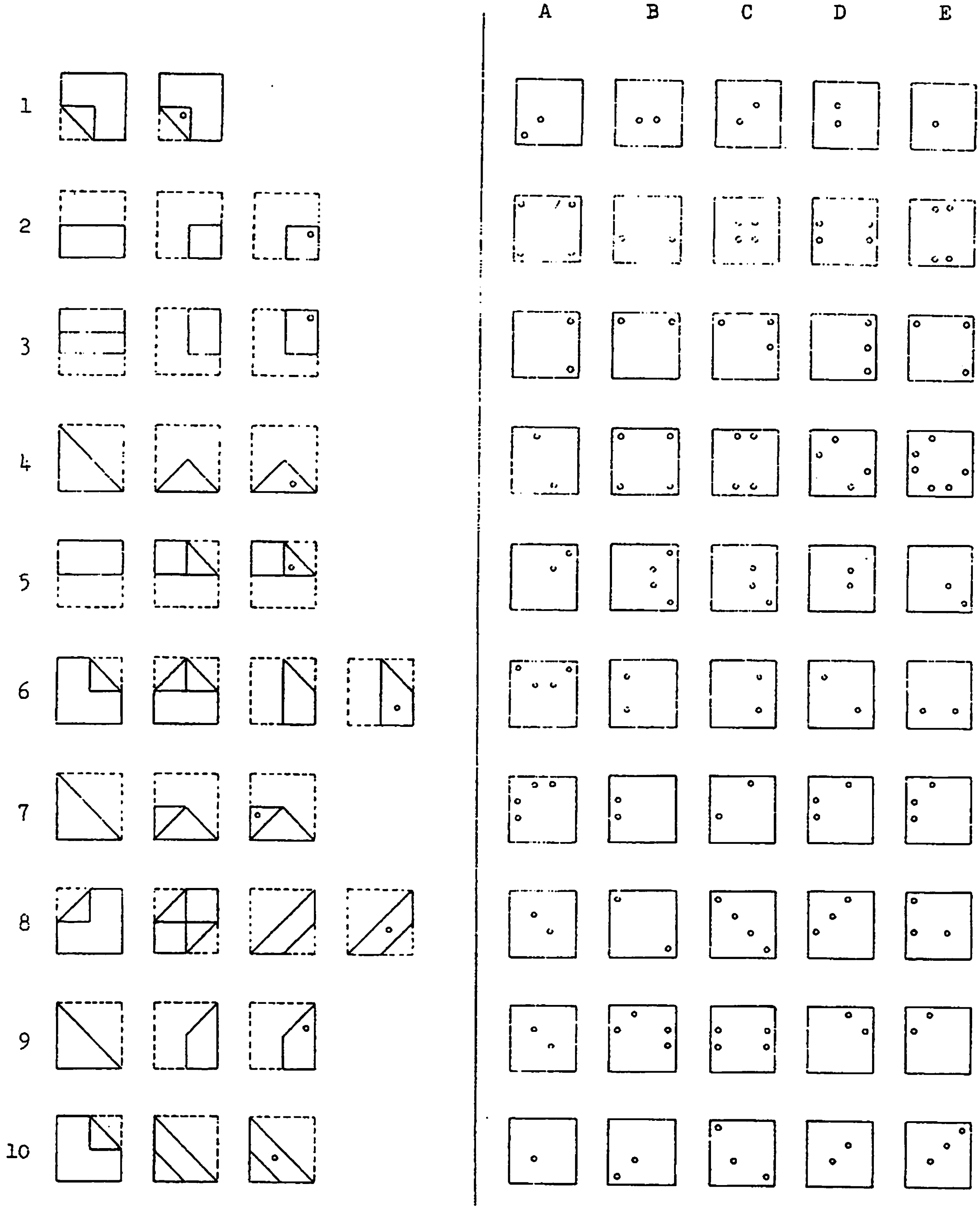
In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Part 1 (3 minutes)

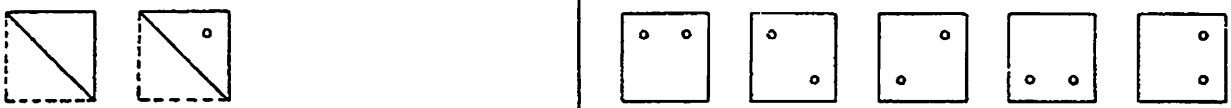


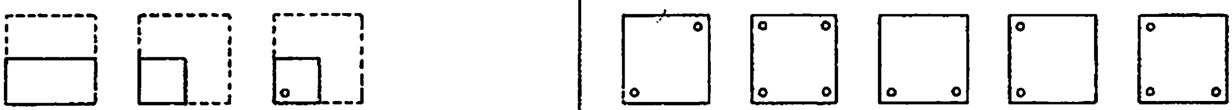
DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

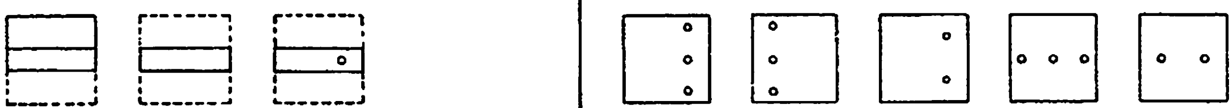
STOP.

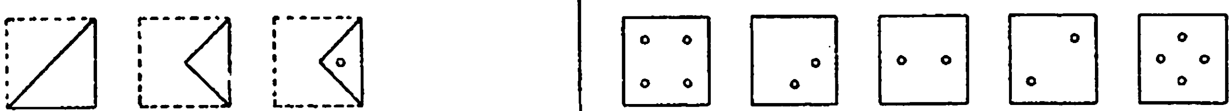
Part 2 (3 minutes)

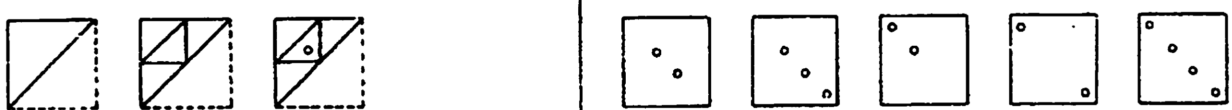
A B C D E

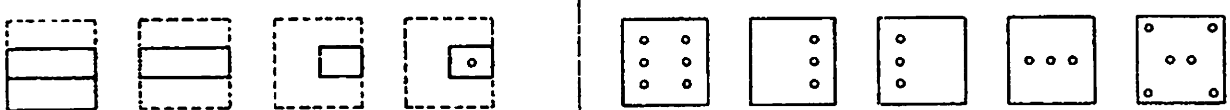
11 

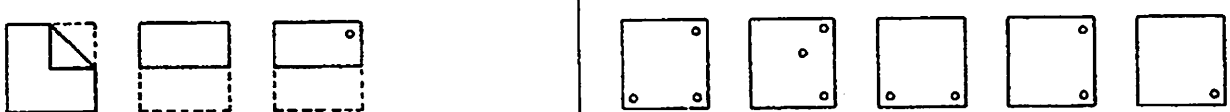
12 

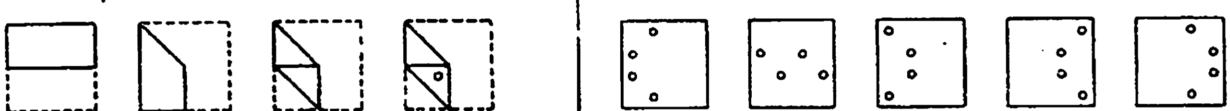
13 

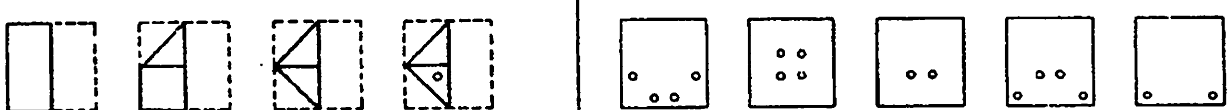
14 

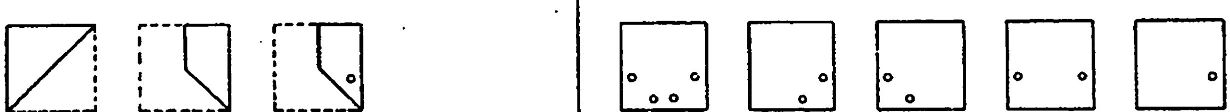
15 

16 

17 

18 

19 

20 

DO NOT GO BACK TO PART 1, AND

DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. **STOP.**

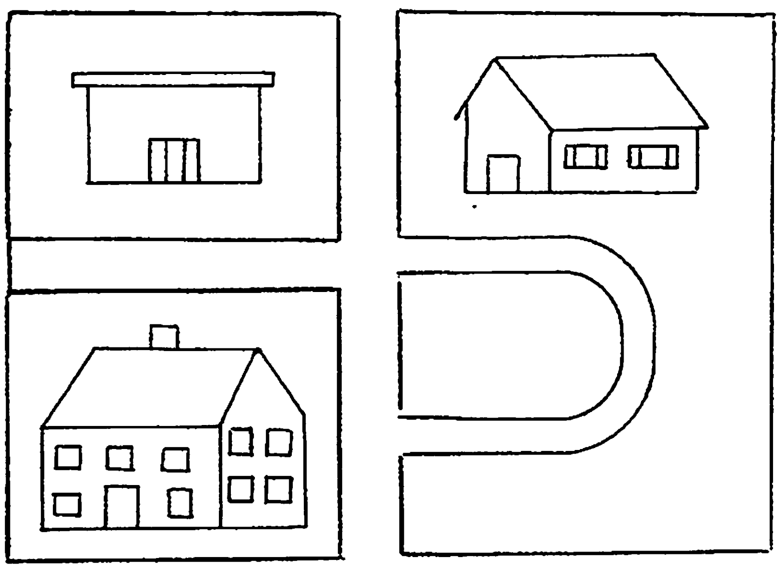
Name ______________________________

BUILDING MEMORY -- MV-2

This is a test of your ability to remember the position of things on a street map.

You will be given a map with streets and buildings and other structures to study. After you have had some time to learn the street layout and the different kinds of structures, you will be asked to turn to a test page. On that page you will find the street map and numbered pictures of some of the structures. You will be asked to put an x on the letter that shows where each of the structures was located on the study map.

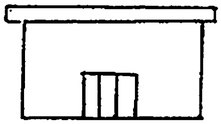
Now look at this simple and enlarged sample:



After you have studied the sample above for a minute, turn to the next page.

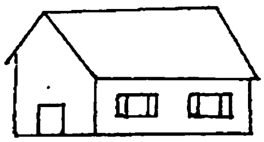
Look at the numbered houses on the left. For each item mark an X on the letter below each building that corresponds with where each house was located on the study map.

1.



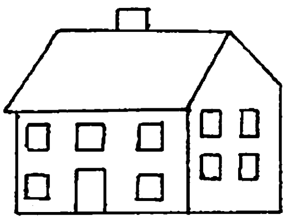
A B C D E

2.

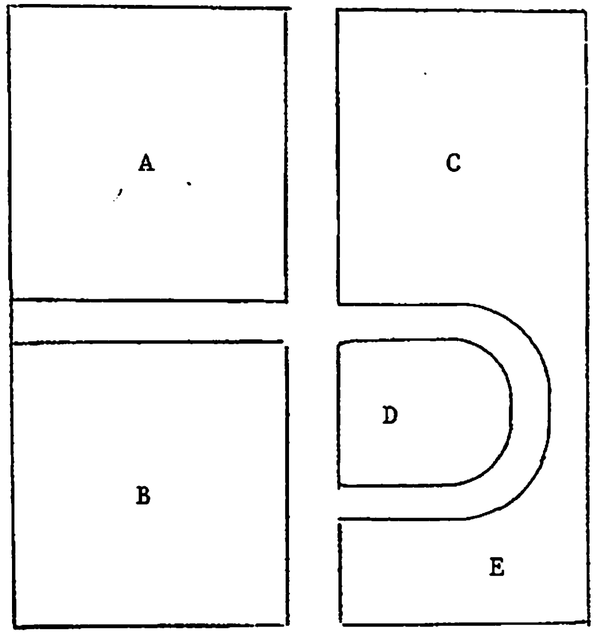


A B C D E

3.



A B C D E



Your answers for sample item 1 should be A, for 2, C, and for 3, B.

Your score on this test will be the number of buildings placed correctly minus a fraction of the number wrong. Therefore, it will not be to your advantage to guess unless you can eliminate some of the locations as definitely wrong.

There are two sections to each part of this test. The first section is a map which you will study for 4 minutes. The second is the test section and contains 12 structures to be located on the map. You will have 4 minutes to mark your answers. Mark A, B, C, D, or E for each building. In the test section, the buildings will be mixed up and not necessarily near the part of the map where you first saw them.

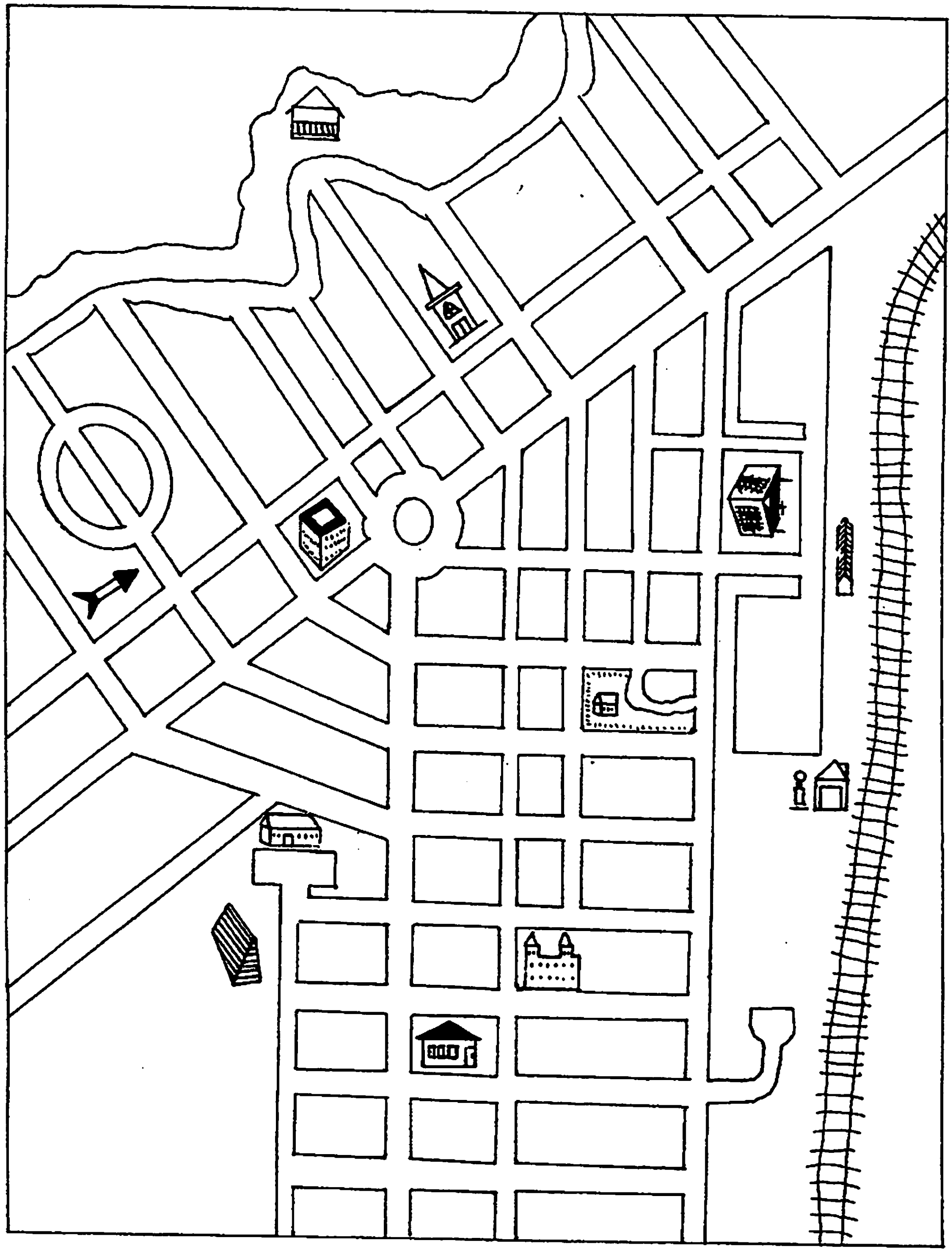
This test has two parts. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STUDY PAGE

Part 1 (4 minutes)

Study this map so you can remember where each building is located.



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP

Page 4 MV-2

TEST PAGE

Part 1 (4 minutes)

Mark an X on the letter below each building that shows where it was seen on the map.

1. 

A B C D E

2. 

A B C D E

3. 

A B C D E

4. 

A B C D E

5. 

A B C D E

6. 

A B C D E

7. 

A B C D E

8. 

A B C D E

9. 

A B C D E

10. 

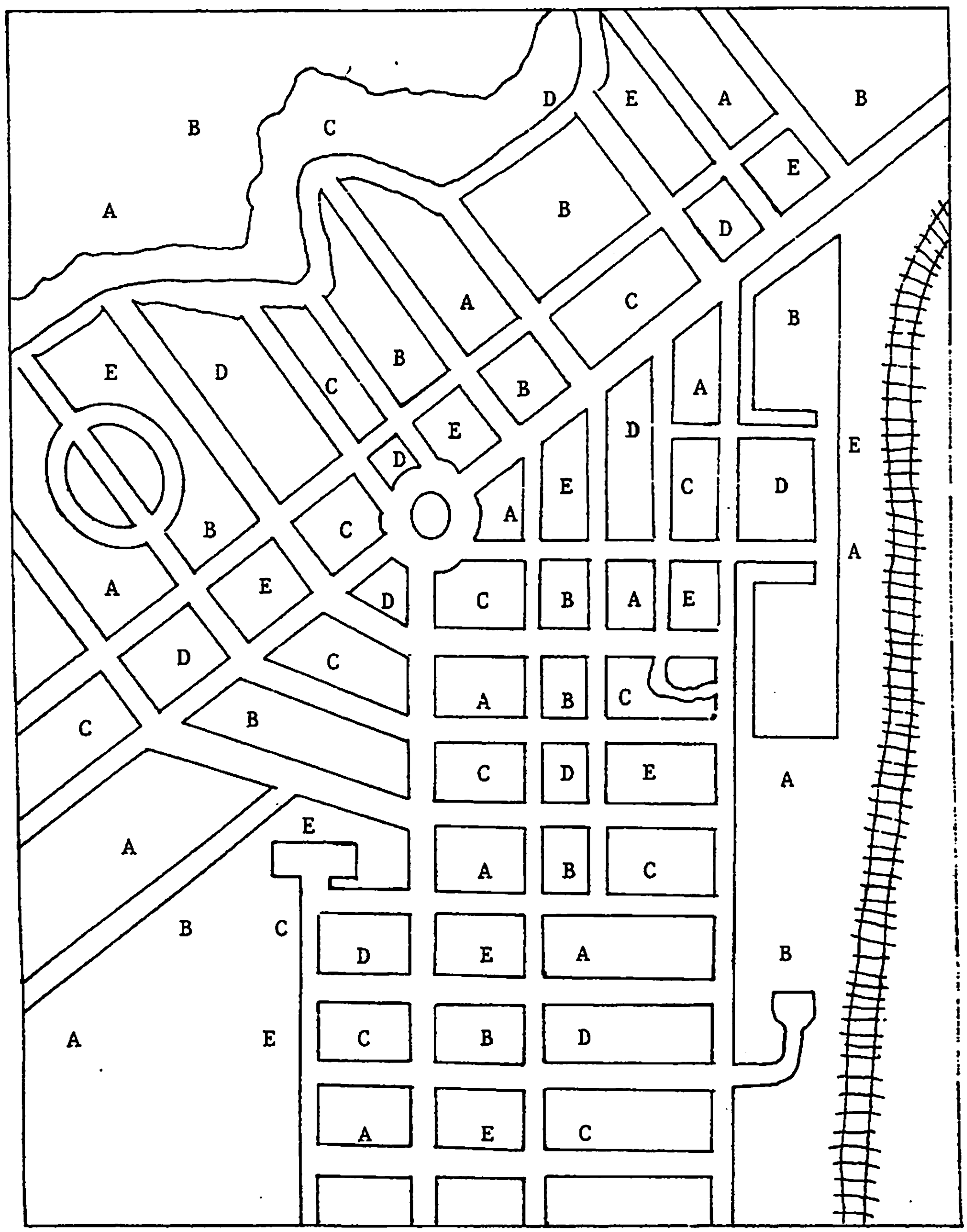
A B C D E

11. 

A B C D E

12. 

A B C D E



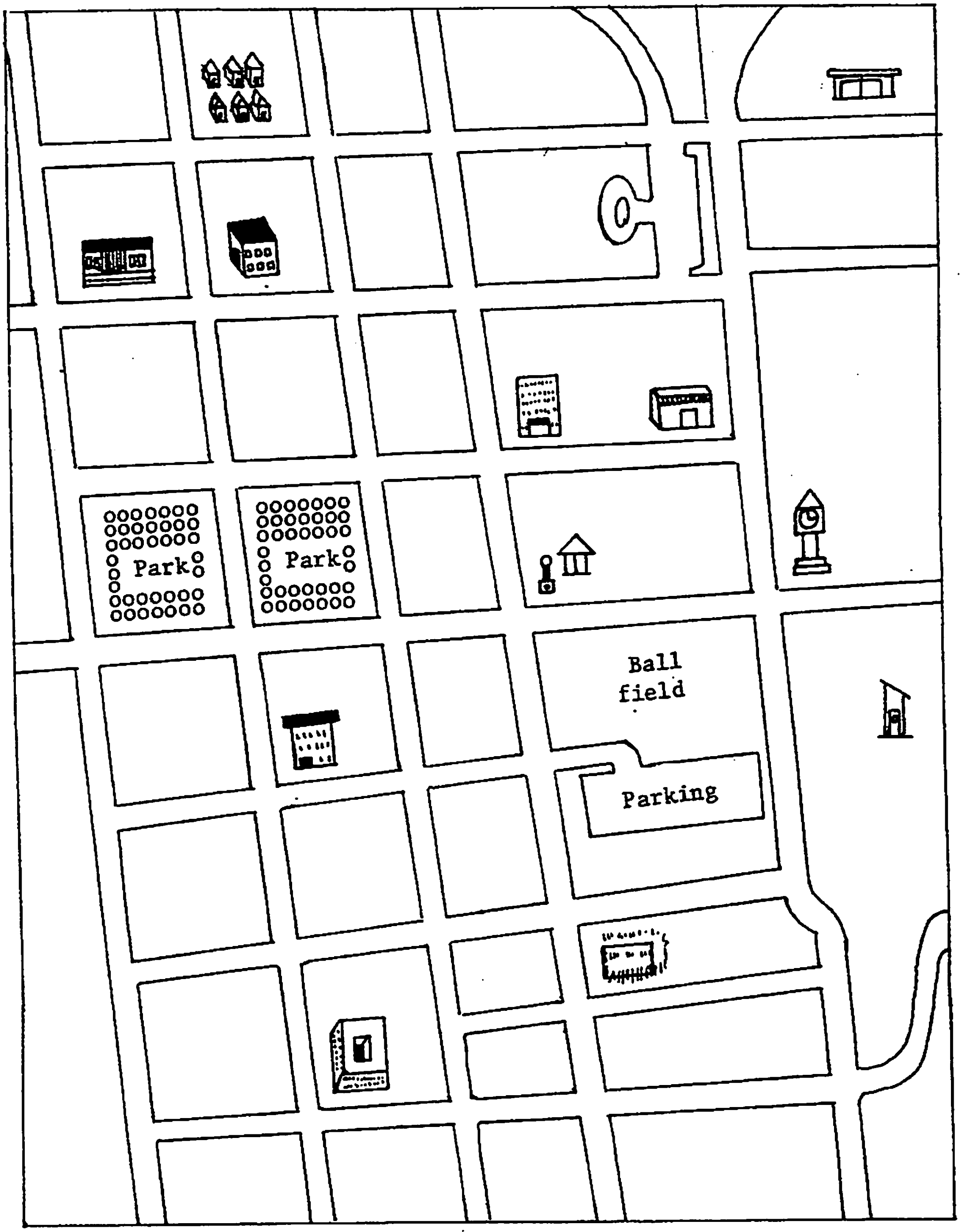
DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP.

STUDY PAGE

Part 2 (4 minutes)

Study this map so you can remember where each building is located.



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP.

TEST PAGE

Part 2 (4 minutes)

Mark an X on the letter below each building that shows where it was seen on the map.

13. 

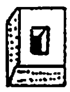
A B C D E

14. 

A B C D E

15. 

A B C D E

16. 

A B C D E

17. 

A B C D E

18. 

A B C D E

19. 

A B C D E

20. 

A B C D E

21. 

A B C D E

22. 

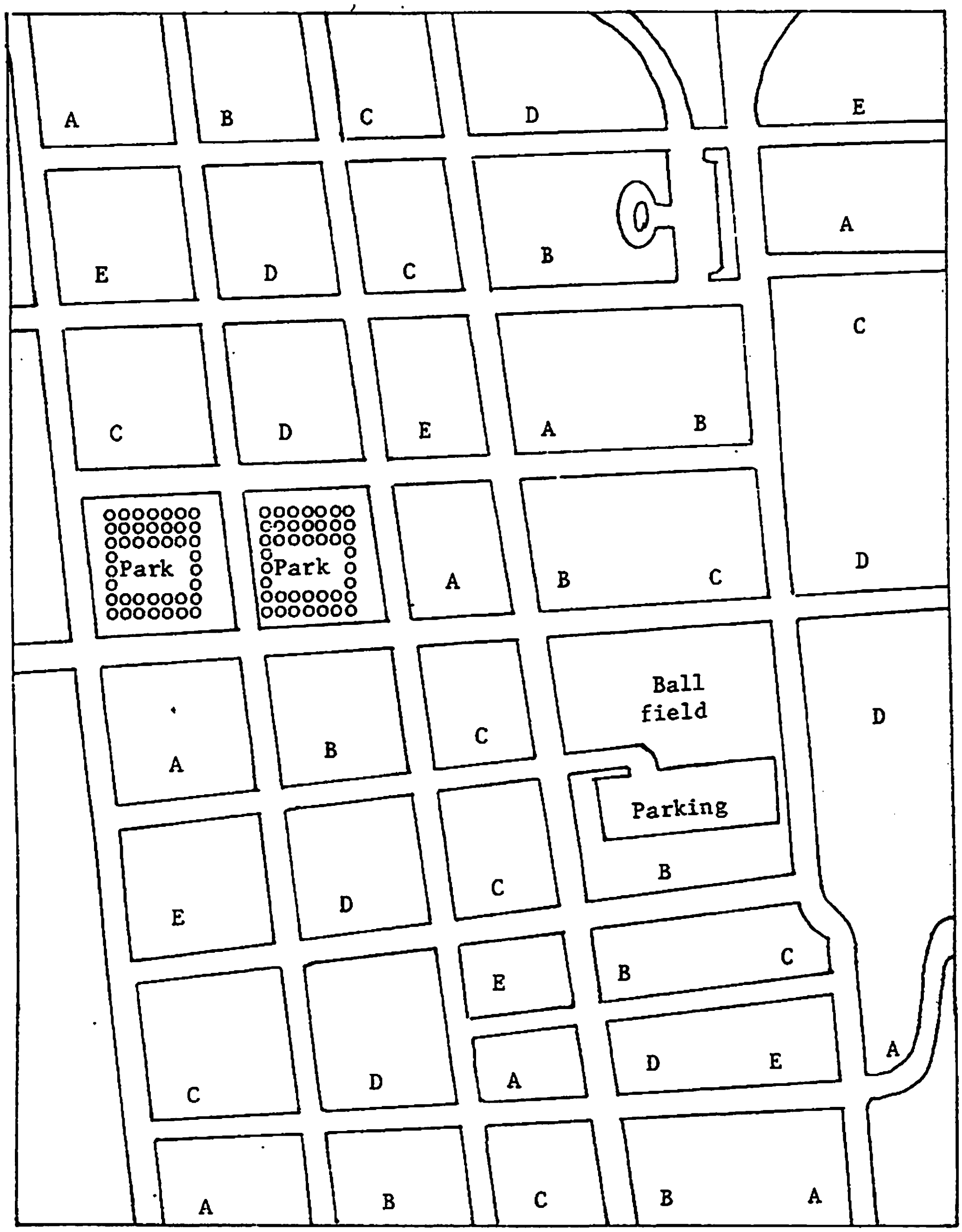
A B C D E

23. 

A B C D E

24. 

A B C D E



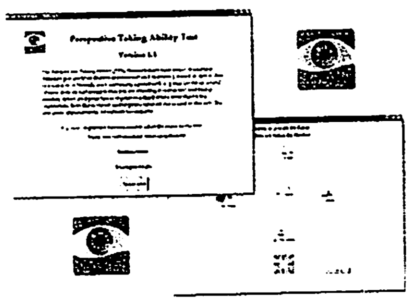
DO NOT GO BACK TO PART 1 AND DO NOT GO ON TO ANY OTHER TEST

STOP.

Perspective Taking Ability Test (PTA-Test)

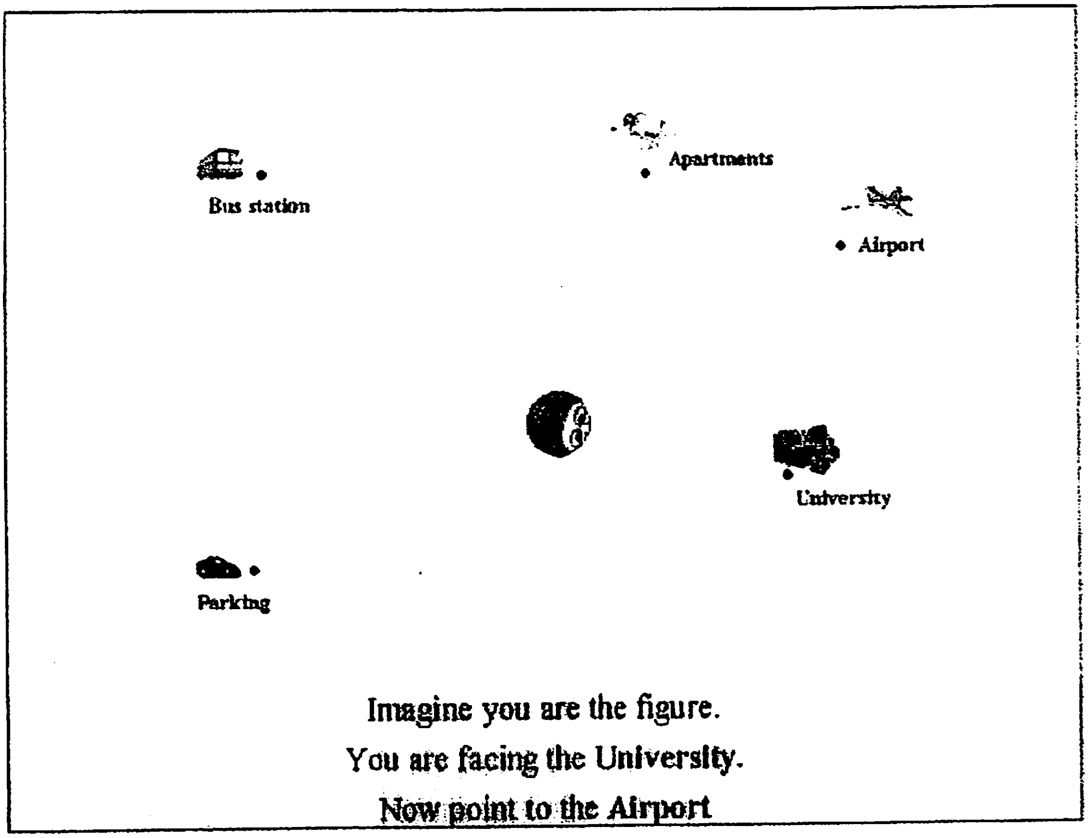
Recent research of Prof. Maria Kozhevnikov (Department of Psychology, George Mason University, VA) has shown that there are two distinct abilities: mental rotation (an ability to imagine rotation of objects from a fixed perspective) and perspective taking (an ability to imagine a reoriented-self) [pdf]. The second skill (perspective-taking) is the skill, which is important for navigating in space.

Up until now, the existing tests did not dissociate successfully between mental rotation and perspective-taking abilities, since most existing tests could be solved by using mental rotation as well as perspective taking strategy. As a result, all the existing commercially-available spatial tests measure mostly mental rotation ability (e.g., the ability to imagine rotating objects from a fixed perspective), which is a different ability not related to navigational skills.



We have developed the Computerized Perspective-Taking Ability (CPTA) test to measure spatial orientation ability. This new test was successfully validated and copyrighted jointly by MM Virtual Design, LLC and Rutgers University (see PTA test features). The results suggest that while solving this test, people in fact encode the objects shown on the display with respect to a body-centered coordinate system. It was also shown that while this test predicts reliably the spatial navigational abilities, mental rotation tests do not.

Our new Computerized Perspective-Taking Ability Test is the first valid measure of spatial orientation ability and could be successfully used for research as well as for training purposes and personnel selection in the professions that require high navigational abilities (e.g., astronauts, pilots, drivers).


Bus station
Apartments
Airport
University
Parking
Imagine you are the figure.
You are facing the University.
Now point to the Airport

Name ______________________________

NUMBER COMPARISON TEST — P-2

This is a test to find out how quickly you can compare two numbers and decide whether or not they are the same. If the numbers are the same, go on to the next pair, making no mark on the page. If the numbers are not the same, put an X on the line between them. Several examples are given below with the first few marked correctly. Practice for speed on the others.

659	___	659	7343801	___	7343801
73845	X	73855	18824	___	18824
1624	___	1624	705216831	___	795216831
438	X	436	971	___	971
4321459	___	4814259	446014721	___	446014721
658331	___	656331	5173869	___	5172869
11653	___	11652	6430017	___	6430017
617439428	___	617439428	516198045	___	518168045
1860439	___	1860439	55179	___	55097
90776105	___	90716105	65216067	___	65216057

Your score will be the number marked correctly minus the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you have some idea whether or not the numbers are the same.

You will have 1 1/2 minutes for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Part 1 (1 1/2 minutes)

Make an X on the line between the numbers that are not the same.

639	_____	639	414982	_____	415982
4714306	_____	4715306	60971	_____	60971
65382	_____	65372	16253948	_____	16253948
710	_____	710	42018591760	_____	43018591760
43210573	_____	43210573	647107569	_____	647107569
6182653905221	_____	6182653905221	721532992531	_____	721582992531
43270105338	_____	43276105338	341798301	_____	341798701
27109816843	_____	27109816853	80537051248	_____	80537051248
519605	_____	519605	5911306581491	_____	5911306581491
923452170687	_____	923452170687	83614081	_____	83614081
370543141	_____	310543141	49471307	_____	47471307
2570665292	_____	2570665292	6082649875	_____	6082647875
32018591670	_____	32018691670	5930582136	_____	5730582136
5471075693	_____	5471075683	236031794137	_____	236031294137
621532992531	_____	621582992531	805731195	_____	805131195
24179830	_____	24179830	48210435512	_____	48210435612
70537051248	_____	70537057248	405176841309	_____	405176841309
7361408	_____	7361708	80145349786	_____	80145349796
39471307	_____	39471507	53210573	_____	53210573
508264987503	_____	508264987503	718265390521	_____	718265390521
4930582136	_____	4930582136	5327010538	_____	5327010538
136031794137	_____	136031794137	37109816843	_____	37189816843
705731195	_____	705736195	619605	_____	619505
38210435512	_____	38210535512	123452170687	_____	123452190687

DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

Page 3 P-2

Part 2 (1 1/2 minutes)

Make an X on the line between the numbers that are not the same.

7573 _____ 7573

347820 _____ 349820

4951 _____ 4951

4573043 _____ 4571043

37501243 _____ 37501243

125093562816 _____ 125093562816

8350107234 _____ 8350107234

34861890172 _____ 3486170172

506915 _____ 596915

786071254329 _____ 786071255329

41345073 _____ 41345073

925660752 _____ 925660752

16719581023 _____ 16717581023

3965701745 _____ 3965701745

135299235126 _____ 135299235136

13897142 _____ 13897142

84215073506 _____ 84215073507

941856031194 _____ 941846031194

8041637 _____ 8071637

70317493 _____ 70317493

35789462805 _____ 35789462805

6312850394 _____ 6312850394

731497130631 _____ 731497130681

591137507 _____ 591127507

289414 _____ 289414

17906 _____ 17906

16719581024 _____ 16719581024

16719581024 _____ 16719581024

3965701746 _____ 3665701746

135299235127 _____ 135299235127

13897143 _____ 13897145

84215073508 _____ 84216073508

941856031195 _____ 941856431195

8041638 _____ 8041438

70317494 _____ 70317494

35789462806 _____ 35789562806

6312850395 _____ 6312850795

731497130632 _____ 731497130632

591137508 _____ 591167508

21553401284 _____ 21553401284

1251373807 _____ 1251373307

903148671504 _____ 903148671504

68794353108 _____ 68754354108

37501235 _____ 37501235

125093562817 _____ 125093562817

8350107235 _____ 8350107235

34861890173 _____ 34861840173

506916 _____ 506616

DO NOT GO BACK TO PART 1 AND DO NOT GO ON
TO ANY OTHER TEST UNTIL ASKED TO DO SO.

STOP.

CONFIDENTIAL CONFIDENTIAL CONFIDENTIAL CONFIDENTIAL

Iowa State University
Research Participant Receipt Form (RPRF)
Use if this payment is less than $75

Iowa State University (ISU) is required to maintain the confidentiality of information about research study participants while still complying with record keeping requirements of the State of Iowa, the Internal Revenue Service (IRS), and funding agencies. The purpose of this form is to serve as documentation of the receipt of compensation associated with participation in a research study conducted by ISU personnel.

I, ______________________________, have received/or am requesting compensation in the form and amount indicated below:
(Print Research Participant Name)

☐ Cash $___________

☐ Check $___________

☐ Gift Certificate/Card $___________

☐ Other Property – Describe:_________________________________

Value: $___________

____ ________________
Research Participant Signature Date

TO ISU PERSONNEL:
Research participants may be given the opportunity to participate without receiving payment if they choose not to complete this receipt form.

This form provides documentation for gift certificates/cards or other property purchased by ISU p-card--keep original form as part of your p-card documentation.

If an ISU check needs to be issued for payment, attach RPRF to completed honoraria voucher and submit to Accounting, 3606 ASB.

Date: 2/4/2011 **Observer:** ____________ **ID:** ___________

TRAINING SCRIPT

I will give you a map and a list of three addresses.

Please determine if the addresses from the list are properly shown on the map. You may verify them in any order.

There are four possible outcomes for each address on the list:

1. The address on the list appears correctly on the map and in the neighborhood – **No Change;**
2. The address on the list does appear on the map but is located in a different location in the neighborhood - **Move**.
3. The address on the list does not appear on the map but does appear in the neighborhood - **Add**;
4. The address on the list does not appear on the map or in the neighborhood - **Delete**;

You may approach the task however you see fit.

The first three addresses are for training. While you work on them I can answer your questions. Please ensure you are comfortable with the task before we move to the main exercise.

Practice Notes / Pointers

*** did any strategies/observations emerge (e.g. numbering patterns) / note places where you can complement and coach participant ***

CODING SHEET

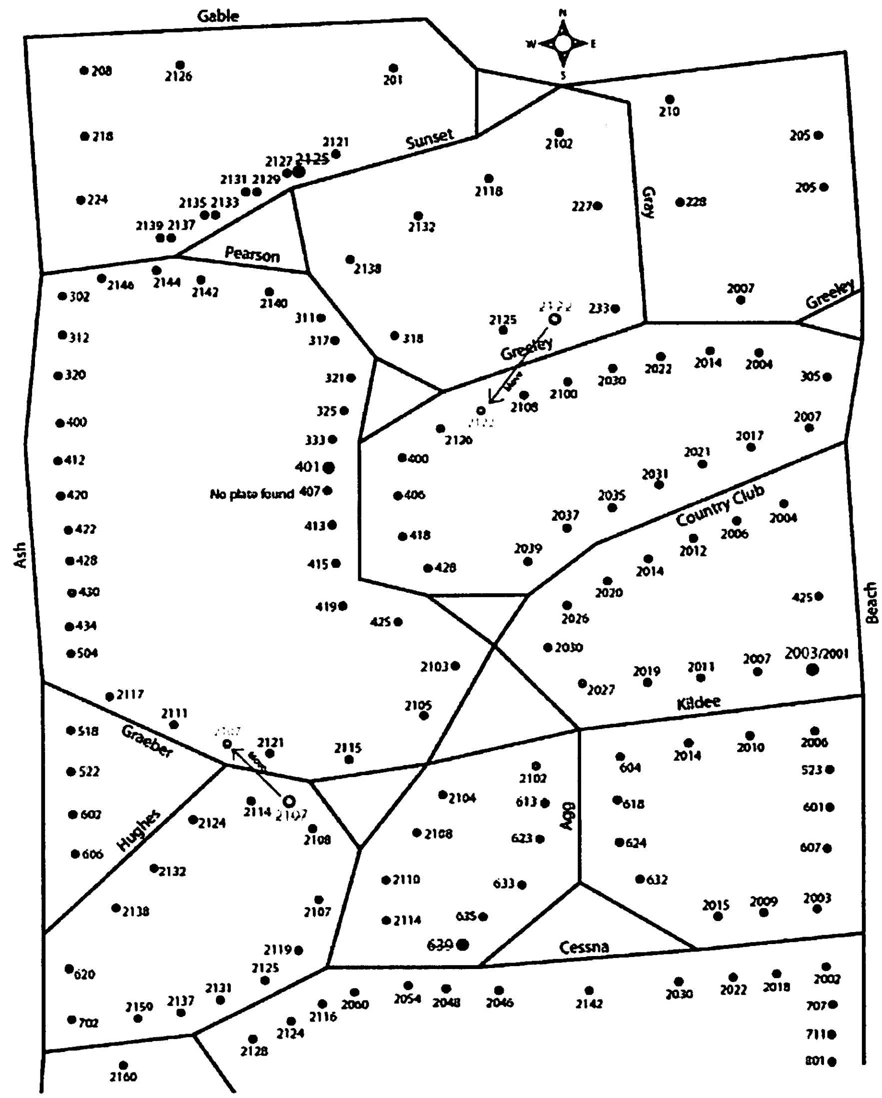
Date: 2/4/2011 Observer:______________ ID:____________

Address:______________ Order#:____ Action: **V A M D** Time of Verification:_____

Check	**Rotation**
Map:____________________	Map:____________________
List:____________________	Body:____________________
Sign:____________________	
Address:____________________	
MISC Lost/Confused:____________________	€ Odd/Even € Numbering Pattern

ADDRESS NOTES

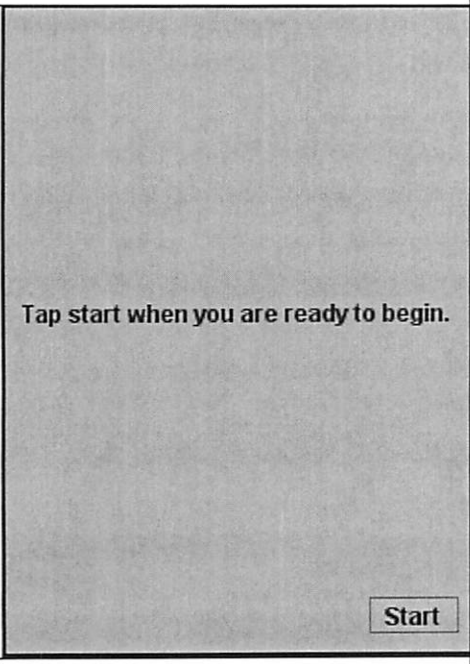
GENERAL NOTES



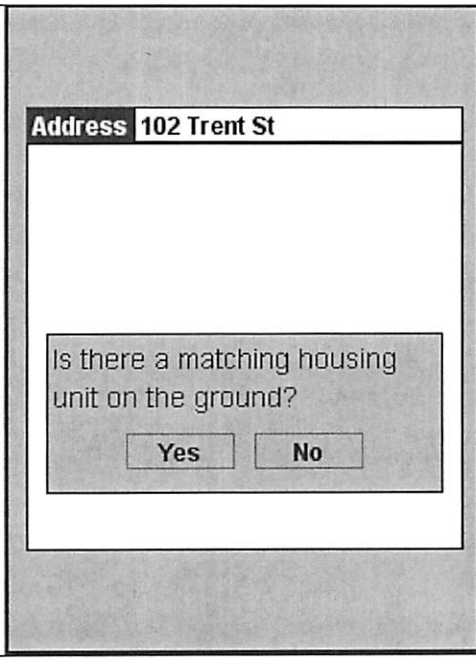
PERSONAL DIGITAL ASSISTANT STORYBOARDS

We will use the same test software on the personal digital assistant in both the field treatment and the virtual environment treatment.

Screen 0



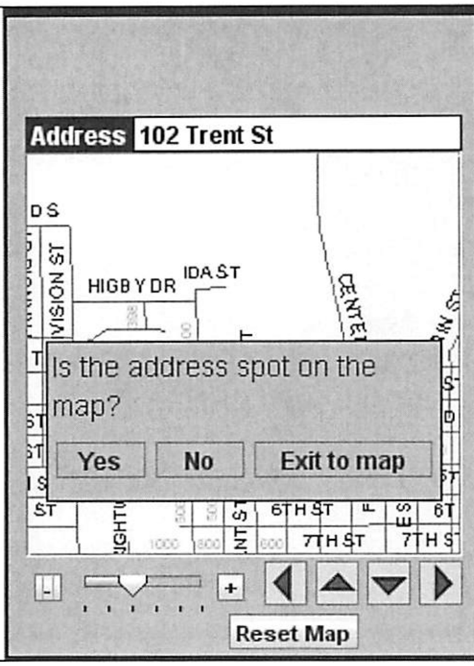
Screen 1



Screen 1a

Screen 2

Screen 3

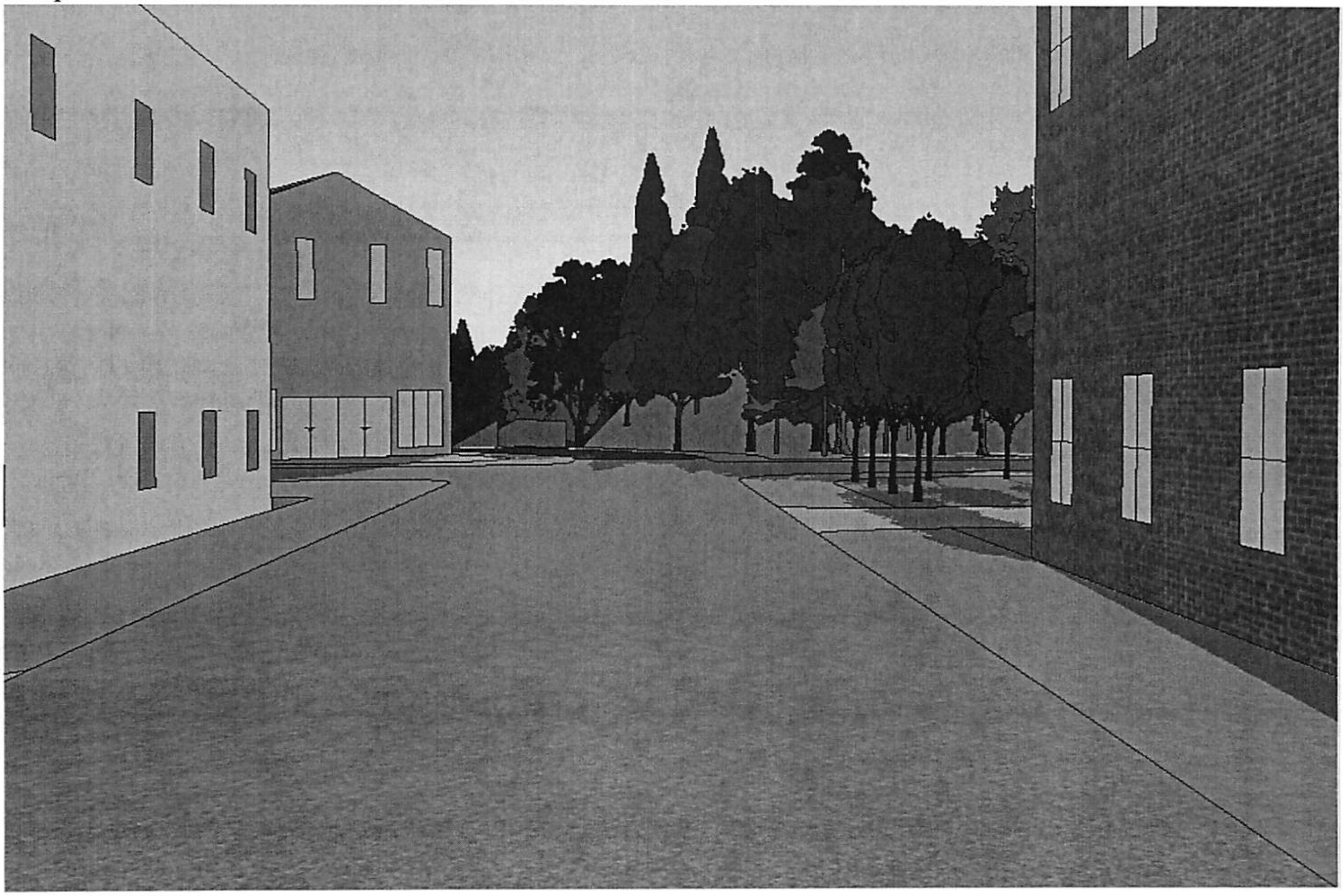


Screen 4

PARTICIPANT VIEWING PERSPECTIVE IN THE VIRTUAL ENVIRONMENT

The following snapshot shows the field of view of a participant assigned to the virtual environment treatment. The participant will be viewing the virtual environment through a first-person perspective.

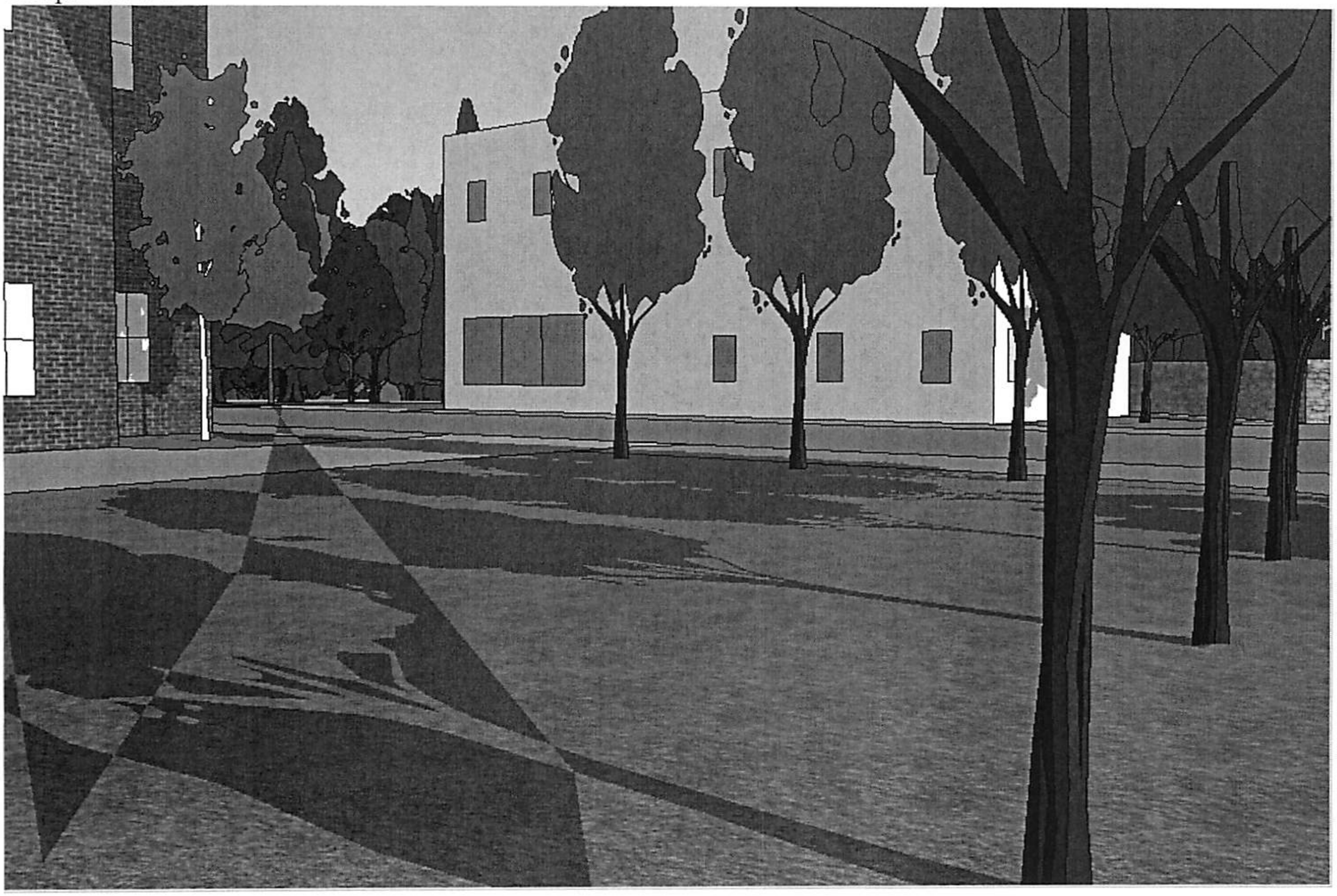
Sample view 1



PARTICIPANT VIEWING PERSPECTIVE IN THE VIRTUAL ENVIRONMENT (Continued)

The following snapshot shows the field of view of a participant assigned to the virtual environment treatment. The participant will be viewing the virtual environment through a first-person perspective.

Sample view 2



FIELD EXERCISE TESTS & QUESTIONNAIRE

Direction Test

A test of direction will occur just before the field and virtual environment exercises begin. The answer will be recorded by the facilitator.

1. "Point due North."

Participant has completed half of the field or virtual environment exercise

1. "Point to your starting location."
2. "Point due North."

Field exercise has been administered and is now complete.

3. "Point to your starting location."
4. "Point due North."

POST-QUESTIONNAIRE

*** give them the map as a reference for the questions ***

1. Did you have a plan before you began verification? What was it and did it help? If not, would you have planned if you had to do it again?
2. What features of the map were most helpful?
3. What features of the map could you have gone without?
4. What features do you wish were on the map?
5. With (1) being easiest and (5) being hardest, how hard was it for you to find the addresses?
6. What address was the easiest to verify?
7. What address was the most difficult to verify?
8. Did the setting affect your approach to completing the tasks (e.g. weather, traffic, etc.)? How?
9. Did anything distract you from the task? What?
10. Do you have any previous experience canvassing or verifying addresses with the aid of maps (e.g. political party volunteering, etc.)?
11. Do you have any prior knowledge of the neighborhood or area? What is it and did it affect your approach?
12. How far do you think you traveled in tenths of a mile (e.g. "1.2 miles")?

SUMMARY NOTES

*** here you can provide feedback on additional thoughts, overall style, etc. ***

APPENDIX D. VR/FIELD EXPERIMENT CODED OBSERVATIONS

Listed below are the coded observations that were recorded during the *VR/Field Experiment* (see *Chapter 4*). Two researchers shadowed each participant as they completed the task while following a think aloud protocol. In descending order, the table below lists observations that were found in both treatment environments (top; blue), observations that were exclusive to the real-world treatment (middle; red), and observations that were exclusive to the virtual reality treatment (bottom; green).

Codes Identified in RW & VR Observations (59 total; 52%)	
Administrative	Miscellaneous
Body Orientation (North-Up)	Planning (Local)
Body Orientation (Rotation)	Planning (No Plan)
Confusion (Intersection)	Planning (Pre-Plan)
Confusion (Numbering)	Planning (Target Address Change)
Cue (Ignores Address)	Planning (Target Address Selection: Nearby)
Cue (Number)	Plotted (Path)
Direction (Egocentric Reference Frame)	Real Error Discovered (Address)
Direction (World Reference Frame)	Skill (Poor Orientation)
Error (Direction)	Skill (Poor Route Planning)
Error (Navigation)	Skill (Poor Software)
Error (Verification)	Skill (Software)
Ground Search (Target Address Found)	Strategy (Checks Addresses Near Target)
Ground Search (Target Address Overlooked)	Strategy (Changed List Order for Route)
Ground Search (Target Street Found)	Strategy (Used List Order for Route)
Ground Search (Target Street Mismatch)	Strategy (North-Up During Planning)
Logged (Body Rotation)	Strategy (Navigates Without Map)
Logged (Lost/Confused)	Strategy (Not Thorough)
Logged (Map Rotation)	Strategy (Rotates Body During Verification)
Logged (Odd/Even)	Strategy (Rotates Map During Difficulty)
Logged (Verification Choice)	Strategy (Rotates Map During Orientation)
Logged (Verification Correctness)	Strategy (Rotates Map During Verification)
Logged (Verification Time)	Strategy (Thorough)
Map Orientation (No Rotation)	Strategy (Trusts Map Over Grounds)
Map Orientation (North-Up)	Strategy (Wayfinds Before Selection)
Map Orientation (Rotation)	Strategy (Zoom-In Verification)
Map Orientation (Track-Up)	Strategy (Zoom-Out Navigation)
Map Reset	Strategy (Zoom-Out Planning)
Map Search (Bearings)	Usability (Issue)
Map Search (Target Address)	

Codes Identified in RW Observations Only (15 total; 13%)	
Assistance Given	Plotted (Stop)
Confusion (Orientation)	Real Error Discovered (Street)
Cue (Ignores Landmark)	Strategy (Reset Map Before Began)
Logged (Numbering Pattern)	Strategy (Rotates Body During Orientation)
Planning (No Pre-Plan)	Strategy (Rotates Body During Planning)
Plotted (Disorientation)	Strategy (Verified Difficult Address Last)
Plotted (Move)	Strategy (Verified Targets Simultaneously)
Plotted (Pointing Test)	
Codes Identified in VR Observations Only (39 total; 35%)	
Confusion (Direction)	Plotted (Diagram)
Cue (Address)	Plotted (Map Checks)
Cue (Ignores Number)	Plotted (Lost)
Cue (Ignores Sign)	Plotted (Planning)
Cue (Intersection)	Plotted (Target Address Change)
Cue (Landmark)	Plotted (Verification Error)
Cue (Sign)	Skill (Navigation)
Ground Search (Alternate Target Address Found)	Skill (Orientation)
Ground Search (Alternate Target Street Found)	Skill (Route Memory)
Ground Search (Closer Target Address Found)	Strategy (North-Up During Navigation)
Ground Search (Target Address Mismatch)	Strategy (North-Up During Verification)
Ground Search (Target Address Not Found)	Strategy (Rotates Map During Navigation)
Ground Search (Target Location Found)	Strategy (Stylus as Compass)
Map Search (Target Address Found)	Usability (Feature)
Map Search (Target Address Missing)	Usability (Suggestion)
Map Search (Target Address Not Found)	Verification (Corrected)
Map Search (Target Address Overlooked)	VR (Advantage)
Map Search (Target Street Found)	VR (Disadvantage)
Planning (Target Address Selection)	VR (Observation)
Plotted (Bearings Regained)	