North Carolina Agricultural and Technical State University
Aggie Digital Collections and Scholarship

2015

A Comparative Study Of Indoor And Outdoor Scaled World Models In Residential Roofing

Tahmina Akter
North Carolina Agricultural and Technical State University

Follow this and additional works at: https://digital.library.ncat.edu/theses

Recommended Citation

Akter, Tahmina, "A Comparative Study Of Indoor And Outdoor Scaled World Models In Residential Roofing" (2015). *Theses*. 275.
https://digital.library.ncat.edu/theses/275

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at Aggie Digital Collections and Scholarship. It has been accepted for inclusion in Theses by an authorized administrator of Aggie Digital Collections and Scholarship. For more information, please contact iyanna@ncat.edu.

A Comparative Study of Indoor and Outdoor Scaled World Models in Residential Roofing

Tahmina Akter

North Carolina A&T State University

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department: Industrial and Systems Engineering

Major: Industrial and Systems Engineering

Major Professor: Dr. Tonya L. Smith-Jackson

Greensboro, North Carolina

2015

The Graduate School
North Carolina Agricultural and Technical State University
This is to certify that the Master's Thesis of

Tahmina Akter

has met the thesis requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2015

Approved by:

Dr. Tonya L. Smith-Jackson
Major Professor

Dr. Salil Desai
Committee Member

Dr. Daniel N. Mountjoy
Committee Member

Dr. Tonya L. Smith-Jackson
Department Chair

Dr. Sanjiv Sarin
Dean, The Graduate School

© Copyright by

Tahmina Akter

2015

Biographical Sketch

Tahmina Akter completed her Bachelor of Science degree in Industrial & Production Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh. She is now a Master of Science candidate in Industrial & System Engineering at North Carolina A&T State University.

Dedication

This thesis work is dedicated to my parents and my beloved family.

Acknowledgements

I would like to extend my sincerest gratitude to Dr. Tonya Smith-Jackson, for her guidance throughout my graduate experience and for giving me the opportunity to assist in the research that led to this project. I would like to thank Dr. Daniel Hindman of Virginia Tech who helped me with providing the data from the indoor study. Furthermore I would like to express deep appreciation for my other committee members: Dr. Salil Desai and Dr. Daniel Mountjoy, for their excellent advising throughout this project.

I wish to extend gratitude to my family members who helped me get through this difficult process. I want to acknowledge my co-researchers who helped me with data collection and analysis— Aric Slaughter, Jarrod Murphy, Evan Middlebrooks, Alea Singletary, Maria Perez-Suarez, Samuel Archer, Angelica Heard and Sefakor Fianu.. I want to thank undergraduate students who assisted in this research.

I acknowledge the National Institute for Occupational Safety and Health (NIOSH) for funding this project. I want to thank Dr. Taher Abu-Lebdeh, Dr. Sameer Hamoush, and NC A&T State University's American Society of Civil Engineers for their supports in building the outdoor scaled model. I am also grateful to WNAA Radio Station at NC A&T State University for their development of a public service announcement to recruit participants. Lastly, I want to give special thanks to individuals who participated in this project.

Table of Contents

List of Figures

List of Tables

Abstract

Falls from roofs are a serious problem in construction sectors. A significant number of workers experience fatal falls from residential structures while performing roofing tasks each year. According to a relatively new OSHA ruling for residential construction, fall protection equipment is mandatory when working at heights of six feet or more. However, lack of ease in using equipment, workers' attitude towards safety equipment and sometimes discomfort in usage of the equipment create obstacles in facilitating the use of fall protection equipment. The difficulty in adopting fall protection equipment has been studied by using scaled models to simulate the actual work environment. In this research, a comparative study was performed between indoor and outdoor scaled models to explore the ecological validity of the indoor scaled model by using the outdoor model as representative of a higher fidelity residential roofing context. The work procedures and task setups were kept similar between both contexts and subjects' behaviors were observed closely. Data were collected by questionnaire and observation techniques. The data were analyzed to examine the differences in both contexts. The primary hypotheses were focused on the presence or absence of differences between the indoor and outdoor models in terms of performance time, critical incidents and usability ratings. Fidelity of the scaled models was investigated based on the participants' perceptions. Mixed results were found. Findings supported the fidelity of the outdoor scaled model as a sufficient replication of the real world. The outdoor scaled model was perceived to be more similar to the actual construction site than the indoor scaled model. Guidelines about the applicability of the indoor and outdoor scaled models were provided based on the findings of this study.

CHAPTER 1

Introduction

1.1 Statement of the Problem

In the construction sector, falls are common causes of injuries and fatalities. Among the fatalities and injuries occurring on construction sites, many incidents and accidents were found on residential sites where fall heights are relatively low and personal protective equipment are not common. In 2013, a total of 699 workers died due to falls, slips and trips which was almost same as the previous year (.07%). Falls to a lower level accounted for 82% of those fatalities (BLS, 2013). A brief case description of some selected fall from elevation fatalities in the residential construction sites proves lower fall heights and severity level. According to Fatality Assessment & Control Evaluation (FACE) report, on August 12, 1992, a 31-year-old male roofer died of severe head injuries by falling from a 16 feet two-story roof and the roof pitch was 4:12 (NIOSH, 1992). On July 17, 1997, a 44-year-old male roofer (victim) died when he fell approximately 21 feet from a roof onto a driveway below. The victim lost his footing, slipped, and fell head-first from the roof onto the concrete driveway (NIOSH, 1997). On April 19, 2012, a 37-year-old Hispanic male laborer fell approximately 13.5 feet from a residential roof to a concrete driveway; he died immediately from his injuries (NIOSH, 2012).

A study by Huang and Hinze (2003) was conducted on the OSHA accumulated data of construction workers accidents involving falls from the period of January 1990 to October 2001. The study found that most fall accidents took place at elevations of less than 9.15m (30 ft) occurring primarily on new construction projects of commercial buildings and residential projects of relatively low construction cost. Lack of proper implementation of fall prevention techniques at the lower elevation construction projects is resulting the greater fall from the heights of less than 10 feet (> 20% of all falls). They found that falls are the leading causes of all the accidents (Figure1).The BLS data from 2006 to 2012 on fall related injuries demonstrates that accident rate due to falls has been increasing gradually each year. In 2012, the accident rate was increased by 14% compared to the previous year.

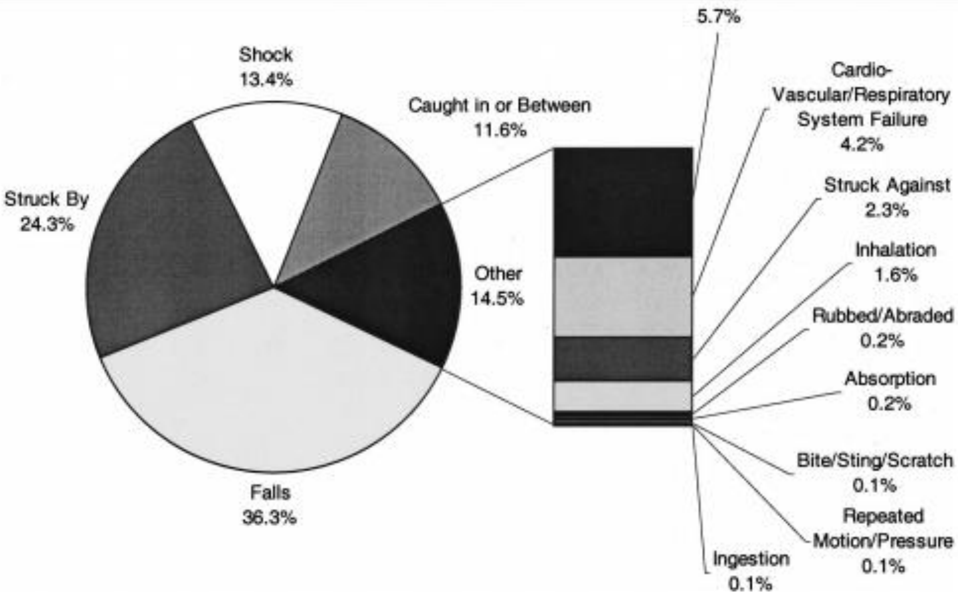

Figure 1: Causes of construction-fall accidents investigated by OSHA(1/90-10/01)(Huang & Hinze, 2003)

Falls in the construction industry represent a major safety hazard that must be addressed. According to BLS data for 2012 (BLS, 2012), about 12% of injuries occurred at the lower levels and 45% of the falls to a lower level involved falls of 20 feet or less. Although statistics have not been adequately established for falls in residential construction, the attributes of residential construction provide a relatively higher risk context compared to commercial construction. These attributes include the lower degree of regulation of residential construction, the likelihood of non-union companies with fewer opportunities for training, and the relatively rapid turn-around of a work project that allows little time for inspection or enforcement (Clark, 2014).

1.2 Use of Fall Arrest Systems

A risk assessment should be undertaken to determine the degree and duration of workers exposure to risk when selecting a fall protection system. The issues that should be considered for the selection of the fall arrest equipment for a particular environment are- potential fall height, frequency and time expenditure at that access, task type including ergonomics and proper equipment, workforce size, weather conditions and in roof access the presence of fragile areas, edge protection(Cameron, Gillan, & Roy Duff, 2007).

Recently, OSHA issued a new guideline pertaining to the use of fall protection in residential roofing (OSHA, 2015). According to the new rule (OSHA 1926.501(b) (2) (i)), each employee who is constructing a leading edge 6 feet (1.8 m) or more above lower levels shall be protected from falling by guardrail systems, safety net systems, or personal fall arrest systems(PFAS).

A PFAS is designed to safely stop a fall before the worker strikes a lower level. It includes three major components: A) An ***anchorage*** to which the other components of the PFAS are rigged, B) A full body ***harness*** worn by the worker, C) A connector, such as a ***lanyard or lifeline***, linking the harness to the anchorage. A rip-stitch lanyard, or deceleration device, is typically a part of the system.

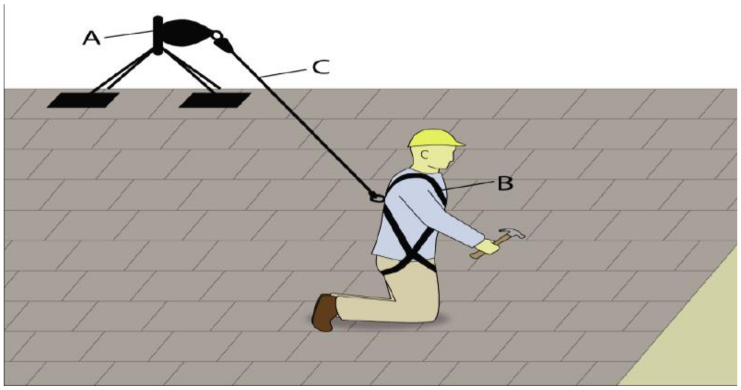

Figure 2: Personal Fall Arrest System (OSHA, 2015)

Research was carried out by Hung, Smith-Jackson, and Winchester (2011) where a number of construction workers were asked about the reasons for taking shortcuts and engaging in risky behavior. Workers were found over confident regarding their performance and since they perceived themselves as experienced they relied less on Personal Protective Equipment (PPE) and more on luck. Multiple task performance at the construction site is a common phenomenon for most of the workers. When they use PPE, their original workflow slows down (Hung et al., 2011).

In construction companies, different ethnicities of workers (Black American, European American, Asian, Hispanic and American Indian) perform construction work. Studies found that Hispanic workers are more prone to fatal injury compared to White, non-Hispanic workers (Dong, Fujimoto, Ringen, & Men, 2009).

In typical residential construction practices, workers do not use the appropriate anchorage point or typical equipment especially for a second story floor. For fall arrest systems (FAS), workers spend large amounts of time adjusting lanyards, which may decrease their productivity (Lederer, Choi, & Griinke, 2006). Moreover, major discomfort may be experienced when using personal protective tools while working. The discomfort can lead to other injuries. Performance degradation using the conventional FAS has been studied by Sa, Seo, and Choi (2009). This decreased performance can contribute to unwillingness to use PPE. All of these obstacles must be overcome to develop effective work systems to prevent falls and to increase user compliance.

1.3 Effectiveness of Scaled Model

Usability of FAS is a major factor in usage. Usability of fall arrest systems conducted on construction sites are problematic due to difficulties in measurement, interference of environmental variables, and bias of workers due to researcher presence. Usability is a multidimensional construct that ties different features of a product or system and considers the factors such as ease of use, flexibility, error handling, provision of help (Roy, 1999). There are many hazards at construction sites that could also place researchers at risk such as being struck by the falling objects, flying objects, swinging objects and objects on ground. Additionally, workers as well as the contractors may not like the presence of researchers and the feeling of being observed may lead them to other hazards.

These challenges on construction sites demand a solution that will support the realism of a field site as well as the control and safety of a laboratory setting. Scaled worlds are used to explore and acquire data to address complex problems. Scaled world models provide means to test fall arrest system design and investigate usability issues, which may not be feasible in the field setting due to ethics and risk (Angles, Trochez, Nakata, Smith-Jackson, & Hindman, 2012). Brehmer and Dörner (1993) described field research usability as complex where reaching any certain conclusion is difficult. In laboratory research, there is too little complexity to allow for any interesting conclusions that are specifically generalizable to the actual work system. The biggest advantage of a scaled model is the level of detail that researchers can get from the context. The scaled model can be used as many times as the researchers' desire and at any level of details. In the field level these details are absent. In field studies researchers can record data only one time where in the scaled models the data collection can be possible with many trials in a controlled environment (Elson, 2003). All these advantages make the scaled world popular to researchers.

However, it is important to have ecological validity of the scaled models as much as or close to what one would expect to have at an actual construction site. For ecological validity, the results obtained from the research must be representative of the conditions in the wider world.

1.4 Objectives

The objective of this research is to study whether any performance, usability or critical incidents differences exist between indoor and outdoor scaled model. The primary hypotheses that were chosen for the research were as follows:

1. H_o: There will be no differences between the indoor and outdoor models in terms of performance time
2. H_o: There will be no differences between the indoor and outdoor models in terms of critical incidents
3. H_o: There will be no differences between the indoor and outdoor models in terms of usability and safety ratings.

These hypotheses were constructed based on four performance times, usability ratings and critical incidents observed during the study. The hypotheses were used as a means to explore indirectly the fidelity of the two scaled models considering the actual construction site as high Fidelity. Fidelity of the scaled models was examined based on the perception of participants.

1.5 Purpose Statement

The purpose of the proposed study is to test the validity of an indoor scaled model by using an outdoor scaled model as the representation of the real world. The goal is to explore the range of variability between the two scaled worlds in an attempt to determine the validity of indoor scaled worlds as equivalent to outdoor scaled worlds. In this study, it will be analyzed whether any major differences exist in the roofers' performance time, critical incidents and usability ratings. Data was collected using observation, performance metrics and questionnaire techniques. The hypotheses are designed to test whether differences exist in performance times, critical incidents and usability ratings of the participants among the two different environments.

Finally, fidelity of both models was analyzed assuming the perception of an actual construction site as represented highest Fidelity.

CHAPTER 2

Literature Review

2.1 Prior Study regarding Fall Arrest System

Falls on construction sites are a noteworthy problem. Hu, Rahmandad, Smith-Jackson, and Winchester (2011) conducted a study based on 536 peer reviewed articles regarding causes of falls where they identified and modeled the responsible factors as work surface, workers' safety attitude and poor construction structure. These researchers provided some recommendations to reduce the risk of falls, including

- ensuring proper working surface,
- proper safety training with close supervision,
- proper fall arrest equipment,
- incorporating the safety culture,
- ensuring ergonomic worksite with soothing temperature,
- moderate humidity,
- proper lighting and low level of noise (Hu et al., 2011).

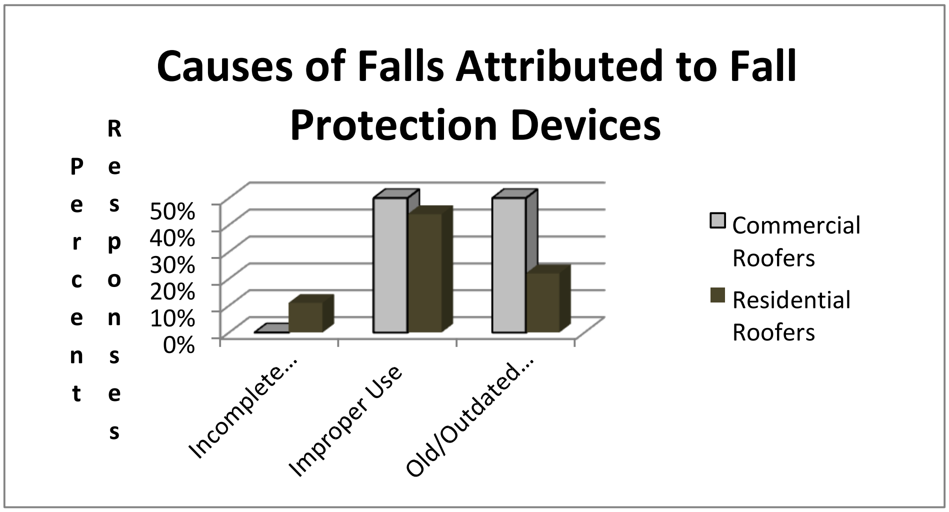

Figure 3 : Falls attributed to fall protection devices (Sa et al., 2009)

Ellis & Lewis (1993) developed a hierarchy of fall protection to protect workers from falls. The hierarchy starts with the elimination of fall hazards, then prevention by passive means like guardrails, restraining through Personal Fall Arrest System (PFAS) and finally warning workers about the fall hazards where nothing else can be done. This is similar to the well-known hazard prevention hierarchy— design out, guard against and warn.

In construction sites, the equipment commonly used for fall protection are safety nets, bracket scaffolds and PFAS. Fall protection equipment is largely used in commercial construction sites rather than the residential sites. Sa et al. (2009) explored the causes of falls in both commercial and residential sectors, and concluded that incomplete connection, improper use of the fall protections and old thereby unreliable equipment are mostly responsible (see Figure 3).

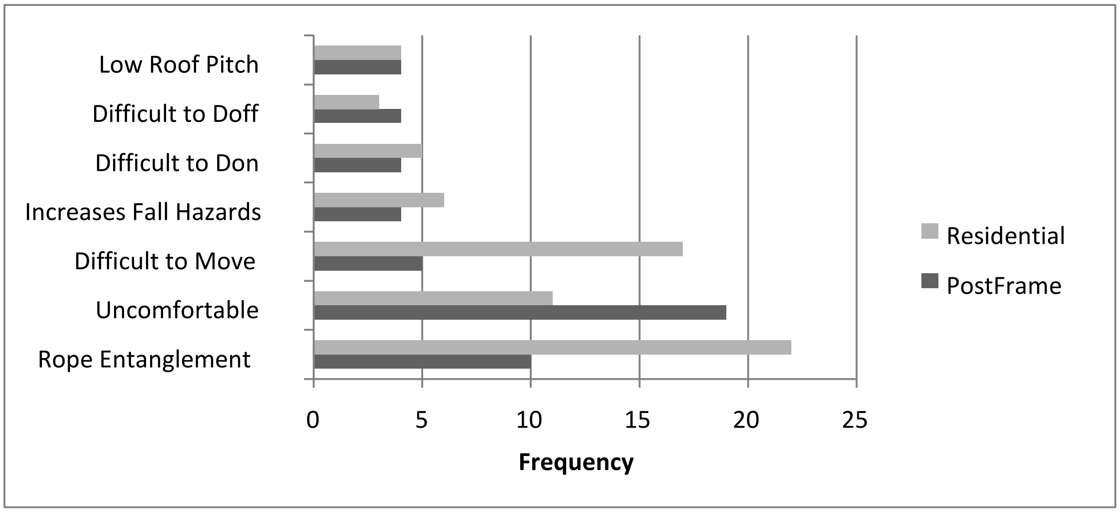

*Figure 4**:** Barriers related to design and usability of fall arrest systems (Smith-Jackson et al., 2011).*

Smith-Jackson et al. (2011) explored the barriers for not using the fall arrest systems among residential and post frame workers (see Figure 4). The factors that are significant are low roof pitch, difficulty in donning and doffing the harness, increased fall hazards, difficulty in movement, and discomfort after wearing the harnesses and rope entanglement with different parts of the body.

Angles (2013) noted significant constraints in the assembly of netting or guard rail systems regarding time, technical proficiency, material burdens, and costs incurred. Angles recommended a fall arrest system for personal protection which is composed of an anchor, lifeline and safety harness for residential construction works. The desirable characteristics of a personal fall arrest system is outlined by Ellis (2002). According to Ellis, the anchorage, the body support, the rope grab, the self-retracting lanyard and the life line are the main features to be considered for a fall arrest system. The Din 360 certification requirements outline design and ergonomic considerations of fall arrest systems such as materials and construction, static load bearing capacity and dynamic performance (Sharp, 2013).

As fall arrest systems are intended to protect the workers from injuries, it is crucial the system is made of strong and well fabricated material. Also, the size of the FAS is an important factor which needs to be considered. A detailed experimental study regarding development of sizing structure for fall arrest harness design was conducted by Hsiao, Friess, Bradtmiller, and Rohlf (2009). These researchers reported improved fall-arrest harness sizing system and strap configurations for men and women. If the workers need to adjust the FAS all the time, their performance can be hindered and they can face unwanted injuries.

There are several means to prevent falls from roofs. Johnson, Singh, and Young (1998) identified relative strength of several fall prevention techniques by using points assessed by multivariate analysis. These researchers found that prefabrication and PFAS variants can be two effective ways for fall prevention (see Figure 5). They mentioned PFAS as a highly feasible, protective, simple, economical and flexible system. However, frequent worker involvement and detraction from worker productivity make the system less popular.

Though OSHA has certain regulations regarding fall protection, compliance with those regulations are lower among the residential construction companies, especially small and medium sized companies. The reason behind the non-compliance is lack of proper resources. Hallowell, Roucheray, and Esmaeil (2012) mentioned that a proactive solution to this problem is to find out the appropriate fall prevention practices for specific task. They also established a framework to measure the effectiveness of common fall protection practices for reducing fall hazards in residential construction.

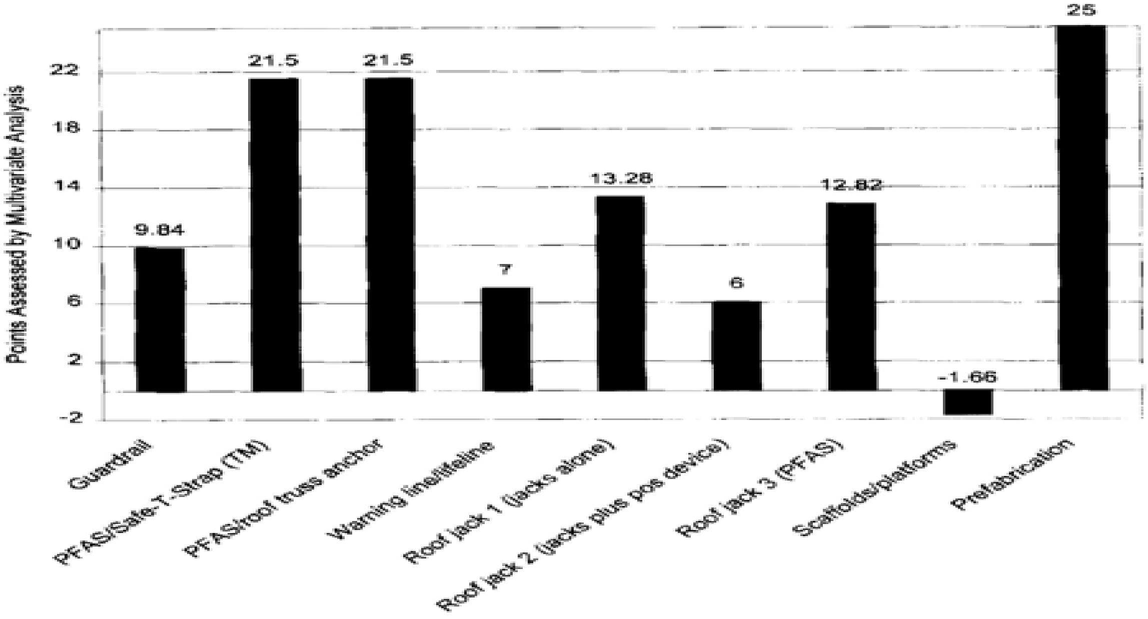

Figure 5: Relative strength of Fall Protection Systems (Johnson et al., 1998)

However, an engineering product requires technical validity for its acceptance. An innovative numerical analysis approach was developed by Drabble and Brookfield (2000) for predicting the forces occurring in each component of an FAS during a fall. These researchers compared results from the numerical analysis with results from experimental tests and with theoretical values, which were estimated by the energy balance method. An important consideration should be assessment of fall related injuries during fall arrest system design. A biodynamic simulation was performed by Kim and Ashton-Miller (2009) to identify the critical biomechanical factors involving falls related injuries. These researchers constructed a two degree of freedom discrete impact model in performed system identification and validation. To do so, they used experimental data to correlate the dynamic interactions of various biomechanical factors in bimanual forward fall arrests. If the body experiences sudden high impact during falls, it can lead to fatal injuries. The safe use of the upper extremity in forward fall arrests requires enough reaction times and synchronized defending movements of the upper extremity. The findings from the various research studies need to be incorporated in the conventional fall protection system.

Fall protection equipment is mostly used in commercial roofing tasks, where the elevation of the roof is relatively higher. The conventional fall protection devices that are used as commercial roofing are not well adapted to incorporate into residential construction (Kaskutas, Evanoff, & Miller, 2013). A detailed task analysis of fall protection for residential construction was performed by Angles et al. (2012) where worker's discomfort was found to be delineated on conventional fall arrest systems. A comparison of risk factors for falls from heights between commercial and residential roofers (Dong et al., 2009) shows the risk factor is significantly higher for residential roofers. The results clearly indicate the necessity of special design for residential fall arrest systems.

Angles (2013) conducted research on usability of harness systems with three types of FAS- 1) low end, no pad, one size fit to all and 2) mid-grade, pad in shoulder and universal size with some adjustable features and 3) high end with fully adjustable features, which was more expensive. He found that mid-grade and high end harnesses were better than the low-end counterpart. However, any distinct advantage was not found for the high grade harness over the mid-range harness. In the present research the low-end and mid-grade FAS has been considered for further study.

As the FAS has to be connected to a fixed point on the structure, prior to the selection of lanyard and its anchorage system, it is important to ensure the availability of anchorage points for the system. In commercial construction, concrete beam, column, and structural steel beam designs are used for anchoring lifelines as these points are strong enough compare to temporary structures, such as scaffolding. In residential construction where the structures are made of wood, the anchorage point can be trusses. According to Koch, Smith-Jackson, Morris, and Hindman (2013), connection of the fall arrest system to the truss members and the roof surface should be configured to allow the anchors to support the load and avoid a situation in which a single truss is used to support the load.

Measuring the effectiveness of a new device or apparatus in the construction arena is not a straightforward task. The most important fact is many external factors interfere with the independent variables and present difficulties in measuring the meaningful values with sufficient reliability (Bernold & Lee, 2010).There is no doubt that workers should use the fall arrest system with the proper anchorage point when they work at elevations to reduce the risk of falls. However, it is important to keep in mind that workers' decisions are not affected by a controlled environment. Tests in a protected environment with access to large, standardized testing apparatuses provide many advantages, however, have limited applicability to the real world of construction (Bernold & Lee, 2010).

When people work at a controlled environment, their responses may be affected. They may not behave the same way as they do in the actual work environment. This research will investigate the ecological validity of an indoor model compared to an outdoor model by using FAS while comparing some of the variables and responses of the participants. In studies like this, researchers need the natural behavior and spontaneous actions as the real work environment is being simulated in a controlled environment.

2.2 Simulated Task Environments

In previous research studies simulators are used to replicate some aspects of the working environment by producing a risk free environment where researchers can successfully obtain the desired flow of information. Simulation “refers to an activity that is designed to help participants acquire insight into the complex relationships and interconnected structures within a particular context. It is a way of preparing for (or reviewing) action in the real world (Leigh & Spindler, 2004, p. 54).” There are many advantages of the simulated environments which are identified by Maran and Glavin (2003). Some of the benefits are reduction of undesired interference, repetition of work procedure, alteration of tasks according to demand and improved accuracy.

A scaled world model is one type of simulated environment which has been used in this research and will be discussed later. There are other types of simulated task environments such as micro worlds, high- fidelity simulations and synthetic environments (Brehmer & Dörner, 1993). Micro worlds are dynamic computer generated environments where participants interact in the laboratory and simulate conditions encountered in the actual field. Micro worlds are becoming popular to many researchers because of the total control over the experiment and also accuracy and efficient data collection processes (Difonzo, Hantula, & Bordia, 1998).

High-fidelity simulation is defined as computer-driven techniques of human simulation that respond in real time to interventions (Hughes, Durham, & Alden, 2008). Fidelity may be defined as the degree to which the outcome of an action closely resembles to the convention or program model originally developed (Mowbray, Holter, Teague, & Bybee, 2003). High- fidelity simulation is an effort to design a training context that physically reproduces the actual performance environment to achieve psychological fidelity to the greatest extent possible (Kozlowski & DeShon, 2004). The equipment usage and controls, the reactions and behavior of the participants are made to be as realistic as possible. However, physical fidelity is absent in Kozlowski & DeShon's construct of high fidelity simulation method.

A synthetic environment serves as a tool for illustrating computer simulation where subjects are represented as real, can be viewed and sometimes touched (Weimer & Ganapathy, 1989). A taste of the physical world is not achievable by this environment. Space Flight projects often use synthetic environments as they give rapid, illustrative and detailed modeling of the critical elements of a mission by computing design accuracy, associated risk and mission utility (Gaskell, Husman, Collier, & Chen, 2007).

These simulated task environments differ from one another by three dimensions- tractability i.e, the researchers ability to collect the right amount of data at the right time, realism and engagement of the participants (Ehret, Gray, & Kirschenbaum, 2000)(see Figure 6).

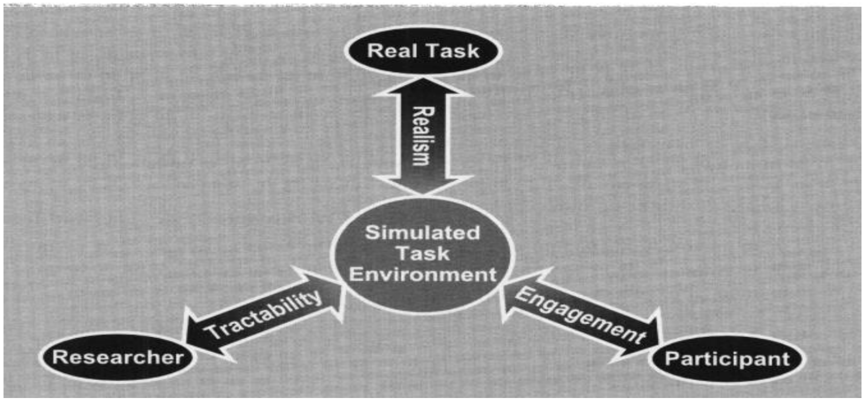

Figure 6: Relationships among three dimensions of simulated task environments (Ehret et al., 2000)

No single simulated model is better than another. Which simulated task environments are appropriate for the experiment should be identified by the research questions. In this research, it is very important that participants engage themselves in the same manner they do in the real environment. Moreover, simulation of the same environments to increase realism and the appropriate data collection in a time limit are also crucial. After studying different simulated task environments, it is identified that a scaled world model can be an effective measure to test the appropriateness of the fall arrest system as it has physical fidelity, participant engagement and realism. However, physical fidelity of both scaled models will be tested further.

2.3 Scaled World Model and Ecological Validity

Human Factors research focuses on human behavior on three different levels- cognitive, rational and social (Newell, 1994). As the natural work environment is more complex and dynamic, the turmoil that is obvious and an inherent part of the work environment can hinder the cognitive- level research. The researchers' expectations can be distorted for lack of suitable environments and therefore improper data collection. A scaled model can minimize this problem. Ehret, Kirschenbaum, and Gray (1998) identified one way to improve tracking of information flow is to collect data by using a scaled model where the task environments remain similar to the original environment. Therefore, participants show the same behavior of the regular environment.

A scaled model is most generally a physical representation of an object, which maintains accurate relationships between all important aspects of the model, although absolute values of the original properties need not be preserved. This enables scaled models to demonstrate some behavior or property of the original object without examining the original object itself. The reasons behind using the scaled models in research studies are accuracy with the data collection, multiple trials on the same platform, less cost and close supervision with less interruptions. Angles et al. (2012) mentioned one additional benefit of the scaled model is repeated audio and video record. Participants' responses can be tracked and further used by the researchers for coding. Figure 7 is a demonstration of the offset view of the scaled world roof which has been used in this study.

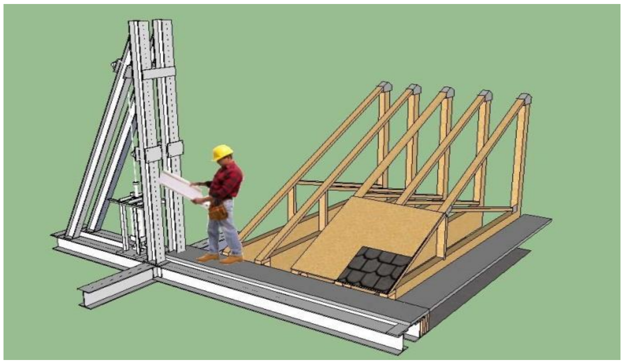

Figure 7: Offset View of Scaled World Roof (Angles et al., 2012)

Though scaled models are beneficial to many investigators, many researchers do not support the use of scaled models. Creating a scaled model is difficult as the originality and realism is absent because of excluding some aspects of the environment. During the design of a scaled model it is very important to properly identify which aspects of the environments need to be included and which to be excluded. The criteria for selecting the elements are identified through methodological and practical constraints (see Figure 8). In case of a complex task environment, the research questions identify the important functional relationships preserved in the scaled model while pairing away the others (Ehret et al., 2000).

Since simulation of original task environments using scaled models comes with cost, the elements that are retained should be carefully selected to possess the ecological validity. Participants' engagement within the scaled world model is also very important, as the participants' deep knowledge of tasks provide details that are missing from the real task environment and in this way, may enhance the realism of the scaled world.

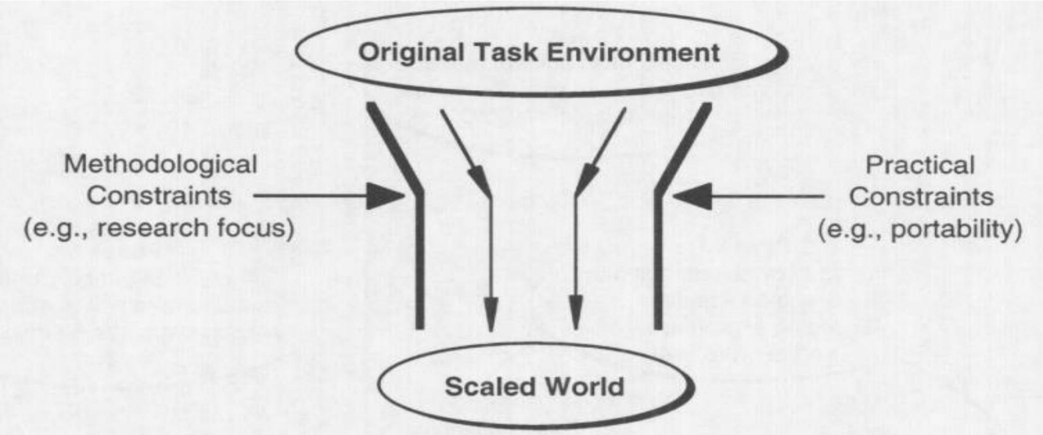

Figure 8: Schematic representation of scaled world development (Ehret et al., 1998)

2.4 Fidelity of the Scaled Model

Fidelity is defined as the degree to which the experience of utilizing the network equipment or simulation in the laboratory environment aligns with using it in an actual workplace environment. "Fidelity is central to the validity of any intervention study and is closely related to the statistical power of outcome analyses. . . . Failure to establish fidelity can severely limit the conclusions that can be drawn from any outcome evaluation (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001, p. 39)". Simply stated, fidelity means how real a model is.

Fidelity of simulation is studied in three general categories: equipment, environmental and psychological (Beaubien & Baker, 2004). The word fidelity is mostly used in health care sector. The Equipment fidelity refers to how closely a simulator resembles the actual roof structure. In health care, a body part is considered low fidelity whereas a complete, model-driven human body is considered high fidelity. Environmental fidelity refers to how closely the simulation location mimics the real world setting. Psychological fidelity refers to how closely the subject perceives the simulation approximates the reality of practice.

2.5 Assessment of Ecological Validity

Ecological validity refers to whether an observed behavior in a laboratory can represent the natural behavior in the world and the stimuli remain the same when removed from the natural context (Schmuckler, 2001). Bronfenbrenner (1977) defined ecological validity as “the extent to which the environment experienced by the subjects in a scientific investigation has the properties it is supposed or assumed to have by the experimenter (p. 516)”. Ecological validity represents the degree to which outcomes obtained from research and experiment are illustrative of surroundings in the wider world (VandenBos, 2007). In short, a model can be said to be ecologically valid if the experimental setup, work procedure, the responses of the subjects and also the results obtained from the settings represent the same as the real world. As in this research two scaled models have been used, the ecological validity of both models is very important.

There are some features, which have been identified as selection criteria of ecological validity. Weather conditions such as heat, lighting, noise level, participant selection criteria, work procedure, physical setup and equipment used are some of the important features that should be considered when choosing the appropriate model for the research. Reduced direct solar gain, privacy, working in quiet indoor conditions, or noisier outdoor conditions could influence performance and attitude outcomes (Clements-Croome, 2006). To simulate the real work environment, the outdoor scaled world model was constructed in a location adjacent to the roadside where the participants faced the same level of noise, work environment, temperature, light and view to provide an environment that is the same as the real work environment.

According to Shadish, Cook, and Campbell (2002) the participants should be chosen randomly from the universe, which the researchers wish to generalize for enhancing the validity.

However, they indicated that a more practical model is to use several different groups of people, settings and time and test the range of variability from where a causal relationship can be established. In this study, the participants of different age groups and ethnicities were selected and the experiments were administered at two different locations by different researchers.

A comparison of the attributes of the scaled world models to the real world environment are presented in Table 1. In the case of the outdoor model, the environmental attributes (i.e, temperature, humidity, light, noise level) are high Fidelity as these attributes closely match the real world environment. On the other hand, as these attributes are controlled by researchers in the indoor model, they are considered as low fidelity. Other attributes in both models are moderately Fidel compared to the real world. .

Table 1

List of attributes that are considered for ecological validity

Attributes	**Outdoor**	**Indoor**	**Real world**
Weather	High fidelity	Low fidelity	High fidelity
Light	High fidelity	Low fidelity	High fidelity
Noise Level	High fidelity	Low fidelity	High fidelity
PPE	Moderate Fidelity	Moderate Fidelity	High fidelity
Harness	Moderate Fidelity	Moderate Fidelity	High fidelity
Anchorage	Moderate Fidelity	Moderate Fidelity	High fidelity
Roofing Model	Moderate Fidelity	Moderate Fidelity	High fidelity
Work procedure	Moderate Fidelity	Moderate Fidelity	High fidelity
participant selection criteria	Moderate Fidelity	Moderate Fidelity	High fidelity

This research compares the indoor model to the outdoor model. The outdoor scaled model was built by the NC A&T State University's American Society of Civil Engineers using the replica of the indoor study at Virginia Tech. The outdoor model was constructed outside the laboratory beside 2105 Yanceyville Street, Greensboro, North Carolina to preserve the similar environmental conditions specified in Table 1. It is very crucial to ensure the validity and reliability of the experimental setup to the real environment, not only the setup itself but also with the information collected from the responses of the participants. A question may arise as to why the data was not collected from the actual work environment. There are many hazards present in the actual work environment with which the researchers are not familiar. These hazards can lead to injury. Sometimes the research cannot take place in the real worksite because of ethical reasons. Moreover, most of the time the construction companies are not willing to allow the researchers on a real world worksite because of low compliance with safety regulations. Another reason for not using the real world environment is interruption with the data collection. The researchers want the uninterrupted flow of the data which is not possible in a real work site.

Indoor experiments allow the greatest control over participants and the experimental conditions, and therefore have the highest internal experimental validity. On the other hand, the outdoor experiments have lower internal validity and the results are more broadly generalizable to the real world application of the experimental treatments (Abowitz & Toole, 2010).

Controlled variation is the foundation of empirical scientific knowledge. Where in the actual work environment, most of the factors cannot be controlled; the experiments in controlled variation may give the high command over the decision environment. The lab offers possibilities to control decision environments in ways that are hard to duplicate with the use of naturally occurring settings. In the laboratory, the experimenter knows the work setup, the order of the participants, the information that is sought and how the tasks are being performed- one shot or repeated measures. Participants are randomly assigned, since decisions are rewarded participants take their decisions seriously and wisely.

However, some researchers argued that laboratory experiments are not realistic and they proposed that the construction research should be conducted in the actual work site where environment and response of the participants are natural. One concern often raised in the laboratory experiment is scrutiny, that is, the possibility that participants in the lab behave differently because they perceive that they are observed (Falk & Heckman, 2009). This is known as Hawthorne effect, a form of reactivity in which subjects modify an aspect of their behavior, in response to their knowing that they are being studied.

The indoor model was built in the Wood Engineering Lab at Virginia Tech. Natural construction work systems tend to place workers in open spaces and expose workers to noise, heat, and daylight. It is very important that participants show the same attitude and do the same task as the real task environment. In research like this where physical setup, participants' actual behavior and also the environmental setup are very important, a slight distortion of these factors can distort the outcome greatly. By comparing the two scaled models based on the hypotheses, we will be able to find out whether any significant differences exist between the two models. If significant differences are found comparing the outdoor model to the indoor model, some conclusions may be drawn regarding the ecological validity of the indoor model. A fundamental assumption of this research is that the outdoor model is an ecologically valid proxy for the real world context. It is as close an approximation as possible. This research will be used as guidelines for the use of indoor models and which factors should be considered further to create an ecologically valid model.

CHAPTER 3

Methodology

3.1 Research Design

Previous research shows that an effective way to measure safety performance of a company is to conduct both quantitative and qualitative safety assessments (Jaselskis, Anderson, & Russell, 1996). In the current study, a mixed methods approach was used to collect quantitative data using ratings on questionnaires, qualitative-quantitative data using frequency counts of critical incidents based on observations, and qualitative data using open-ended responses and verbal protocols. Data were collected from two different sites where other constructs remained the same. These two groups were from North Carolina A&T State University (outdoor model) and Virginia Tech (indoor model).

Three hypotheses were explored—

1. H_o: There is no difference between the indoor and outdoor models in terms of performance time
2. H_o: There is no difference between the indoor and outdoor models in terms of critical incidents
3. H_o: There is no difference between the indoor and outdoor models in terms of usability and safety ratings.

3.2 Dependent and Independent Variables

Two (2) independent variables, each with two levels and fourteen (14) dependent variables were identified. The independent variables were-

- Harness type (low graded and medium graded)
- Context (indoor and Outdoor)

The variables considered as dependent measures were-

- Don Time
- Doff time
- Tar Time
- Shin Time
- Donning Harness Usability rating
- Doffing Harness Usability Rating
- Tar Paper Usability
- Shingle Usability
- Total critical incidents (CI) and
- Usability ratings for fall arrest system
- Perception of the scaled model (measured by feeling safer, cautiousness, resemblance and decision making practice)

Don time was defined as how much time each participant took to put on the fall arrest harness without any instruction. Doff time was the time required by each participant to take off the harness after finishing the task. Both the donning and doffing take place at the ground and time were measured by using the stop watch. Tar Time was defined as how much time a participant took to install tar paper on the roof wearing one of the harnesses. Shin Time was the time required by a participant to set two rows of shingles over the installed tar paper wearing one of the harnesses. All these times were measured in seconds. The other dependent variables are explained in the questionnaire section (3.4.3)

3.3 Design of Study

Each participant was studied against a specific context and harness type. Both context and harness type had two levels. The experimental design used a 2x2 between subjects design. There were four conditions to be tested and in each condition sixteen participants were tested. Total sample size was n= 64.

Context	Harness Type	
	Low-grade	Mid-grade
Indoor	16	16
Outdoor	16	16

Figure 9: Design of study

3.4 Apparatus

3.4.1 Roof Structure: The roofing apparatus was a 14 ft. X 10ft. X 8ft. wooden roof structure consisted of six trusses mounted on a cement foundation (see Figure 10). Trusses were spaced 2 feet on center with the final truss spacing slightly smaller to accommodate the 10 feet width. Bracing was installed as lateral (perpendicular) and diagonal (on back side and bottom). The platform was sloped in a similar way of an actual residential roof where the pitch was 6:12. A roof with a "6:12" pitch means it rises 6 inches for every 12 inches of horizontal roof run.

Figure 10: Indoor and Outdoor scaled models used for the study.

There are three types of roof pitches- low, medium and steeped. The pitch used in this study was a medium pitched roof which lies between the range of 3.5 and 7.5 inches. These are the most commonly observed pitched roofs in United States. These roofs are usually used for sheds and garages.

3.4.2 Fall Arrest System and Anchor: A fall arrest system consists of anchorage, tether, and fall arrest harness. In this study, two different types of harnesses were used- 1) low level harness with no padding and one size fits to all, 2) Mid- range harness where there are some pads over the shoulder area and some adjusting features (see Figure 11). Participants were asked to wear one of the harness .The harnesses were tethered to a self-retracting lifeline, which acted as a seat belt to protect the worker to go beyond a minimal distance (see Figure 12).

Protecta by DBI-SALA Full Body Harness (no padding, universal fit, ~$50)

DBI-SALA Delta Harness (thin shoulder padding, universal sizing, ~$100)

Figure 11: Fall Arrest Harness (Angles, 2013)

Figure 12: Self-retracting Lifeline (Angles, 2013)

Two anchorages— metal braced and tie-off, were used. The anchorages were tied with the trusses of the roof (see Figure 13).

Figure 13: Metal braced and tie-off anchorage

One part of the lifeline needed to be attached to the anchorage while the lower end was attached to the back of the harness. The participants were asked to do the roofing task wearing the harness and tied to the anchorage by the lanyard.

3.4.3 Questionnaires: Seven sets of questionnaires were used to collect data on demographics, harness usability and perceptions of fidelity of each scaled model. Using the scale from 1 (negative anchor) to 6 (positive anchor), each participant was asked to answer all of the questions other than the *Demographic Questionnaire*. These questionnaires are analyzed as follows:

Demographic Questionnaire: A demographic questionnaire was administered to all roofers to elicit the following information: age, weight, height, ethnicity, gender, years of construction and roofing experience and years of education. Each participant was required to complete the questionnaire before starting the roofing task (See Appendix A).

Donning Harness Usability Questionnaire: Right after the donning of the harness, participants were asked to answer a set of questions about the ease of putting on the harness, comfort and the difficulty of tying the lanyard to the anchor. Each scale was end-anchored with very poor on the left and very good on the right and was presented as a series of numbers spaced out across the page representing the response categories (See Appendix B).

Tar Paper Usability Questionnaire: After conducting tar paper installation on the roof structure, the participants were asked to answer another set of questions. Each participant was asked to answer the ease of working with the harness and lanyard (See Appendix C).

Shingle Usability Questionnaire: When the participants were done with the shingle installation task, the *Shingle Usability Questionnaire* (See Appendix D) was administered, which was similar in psychometric format to the *Tar Paper Usability Questionnaire*.

Doffing Harness Usability Questionnaire: The participants were asked to take the harness off when both of the installation tasks were accomplished. The time was measured and then the *Doffing Harness Questionnaire* was conducted. There were some open-ended questions here along with the rated questionnaire. The participants were encouraged to talk about the design of the harness at different locations on the body (See Appendix E).

Perception of the Scaled Model Questionnaire: Participants were asked about their feelings of the scaled model on which they had worked. Their perception about the resemblances of the model to the real roof and their comfort were also noted. They were also asked whether the scaled model affected their level of decision making (See Appendix F).

Usability Rating and Post-Task Interview Questionnaire: The participants were interviewed about the difficulty of performing the tasks wearing the harness and recommendations were elicited regarding design and costs (See Appendix G).

3.5 Participants

Total sample size was sixty four (n=64). Among them 32 participants took part in the indoor study and 32 in the outdoor study. The following criteria were used for the selection of the participants:

1) at least 18 years of age and weigh 310 lbs or less,
2) experienced no injury in the previous year related to falls,
3) had at least one year of roofing experience, and
4) had been employed with a construction company for at least one year at some point in time.

During the participant selection procedure the utmost effort was given to have individuals of different ages, ethnicities and genders. Participants were screened by telephone interview to confirm eligibility before taking part in the study. Among the participants two (2) were female and the rest were male. Both female roofers were from the outdoor study.

Participants were recruited by an advertisement provided on the Craigslist, Public Service Announcements on the radio, contacting roofing companies, and by distribution of flyers. The mean age of the participants was 36.9 years (*SD*= 10.46), mean height was 70.34 in. (*SD* = 3.37) and mean weight 188.69 lbs. (*SD* = 35.09). These participants had a mean roofing experience of 10 years (*SD* = 9.7). Among the 64 participants, 31% were African-American, 44% were European-American, 14% Hispanic, 3% Asian-American and 8% from other ethnic groups.

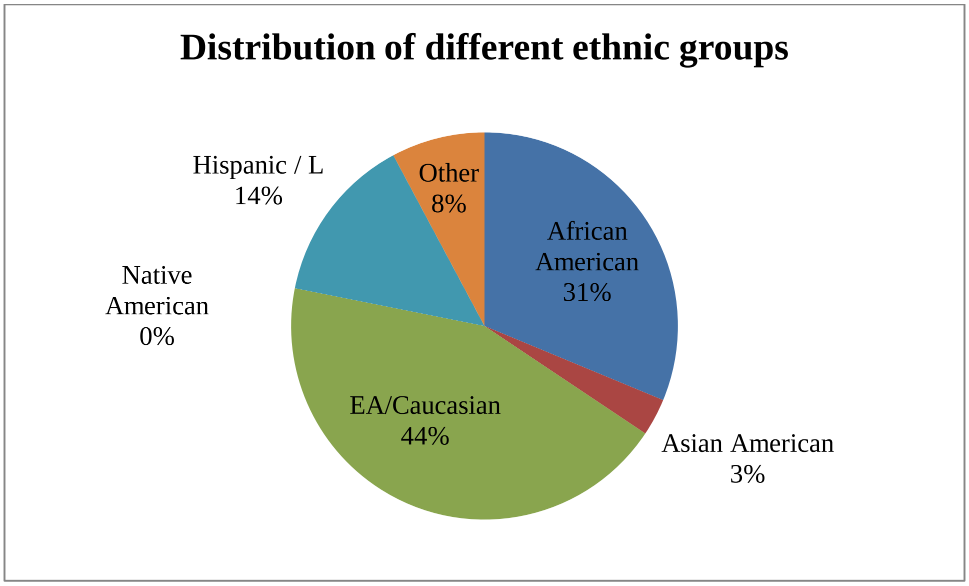

Figure 14: Distribution of different ethnic group

3.6 Procedure

As required by the Institutional Review Board (IRB), each investigator was certified by IRB before conducting the experiment. Three cameras were used to take the video footage of the task for later use of counting critical incident frequencies. Two were stationary (model-Canon XA10) and one was roaming (model- POV.1.5^{TM}) mounted on the helmet of one of the investigators. To get a better view of participants' tasks, one stationary camera was placed over an elevation of seven feet (7ft) and the other stationary camera was set on the ground at a minimum distance of three feet (3ft) from the roof model. The participants signed the consent paper allowing the use of video recorders on the eve of the experiment. After signing the consent paper, participants completed two set of questionnaires- Demographic and Safety Climate Questionnaires. Then they were asked to don one of the harnesses without any assistance from the investigators. Counterbalancing was used to determine the order of harness assignment. The whole task was videotaped and time was measured by a stopwatch. When they were done, the harnesses were adjusted to ensure they were wearing correctly. They were provided with a tool belt, hard hat, knee pads, gloves and goggles. Knee pads, goggles and gloves were optional. Participants were asked to complete two roofing tasks while wearing one of two fall arrest harnesses (low-cost version; medium-cost version). The two roofing tasks were to apply tar paper over the existing sheathing (oriented strand board/OSB) followed by laying two rows of shingles on the roof (see Figure 15). A follow-up questionnaire was administered after each of the task. They were asked to think aloud so that their attitude toward the harness and anchor can be understood.

Figure 15: Participant working at outdoor scaled model

The critical incidents were collected at the worksite by a Critical Incident Check Sheet which was further crosschecked with the video recording to count the critical incidents frequency and type of critical incidents. Ten a priori codes were used to identify critical incidents for each participant (Table 2). Each of the participants was compensated for their participation after completion of the study.

Critical Incidents Identification

Table 2

List of critical incidents a priori codes

Critical Incidents	**Description**
Slip on roof	Any part of the body that slips on the roof – hand, foot
Fall from roof	Full body falls from the roof
Loss of balance	Any gesture or movement indicating loss of balance such as flailing to regain balance or jerking the body to prevent falling
Extreme/ awkward posture	Bending, learning or reaching at an extreme angle forward, backward, right or left and doing so in such a way that there is a likelihood of losing balance
Harness Adjustment	Adjusting the body harness or shifting weight inside of the harness
Lanyard Adjustment	Adjusting or moving the lanyard because it is interfering with the task
Entanglement	Having the lanyard or the harness tangled around a tool belt or part of the body
Edge Violations	Balancing on the very edge of the roof without adjusting the body so that most of the weight is on the opposite side of the edge
Removal of PPE	Removing hard hat). In the protocol, gloves are optional. If they use gloves and then remove the gloves it is a critical incidents
Tool, Nail drops	Dropping tools or nails while working

3.7 Data Analysis

The data collected by different objective variables (i.e., performance time and critical incidents) and subjective variables (i.e., usability ratings and perception realism) were analyzed using SAS 9.4^{TM}. The objectives were to explore differences between the indoor and outdoor scaled model in terms of performance time, usability ratings, critical incidents and level of fidelity.

A Shapiro-Wilk test of normality was used to test the normality of the questionnaire data. Parametric tests were conducted to check if the distribution was normal or near normal; otherwise non-parametric (distribution free) tests were used. The significance level was set at α= .05. Correlation analyses were administered over all the dependent variables to check the correlation strength. Demographics of the participants were analyzed to see their influence over the different variables.

A GLM-based repeated measure ANOVA was used to examine significant differences among different performance times. Due to non-normality of these four variables a generalized linear model (glm) applying repeated measures was used. Akritas, Arnold, and Brunner (1997) mentioned that this modern version of ANOVA such as a repeated measure ANOVA, can be more robust than the classic ANOVA even if the normality assumption is violated. The contrast method was used as the post hoc test to examine the paired differences across all variables.

To test the hypothesis of no difference in critical incidents, the Wilcoxon ranked sum test was conducted by using critical incidents frequencies. Beforehand, participants' video footage was coded by the two independent coders. Coders' agreement level on the frequencies of the critical incidents was calculated. The criterion level for agreement reliability was set at .7 (Nunnally, 1978).

For testing the hypothesis of no difference in usability ratings, multiple analyses of variance were administered. Before MANOVA analysis, each usability construct was measured by using Cronbach's coefficient alpha to check the reliability. The criterion level for reliability was also set at .7 (Nunnally, 1978).

The fidelity of the two scaled models was measured based on four perceptual questions administered on participants of both contexts. Repeated measures ANOVA using the GLM procedure, was used to analyze the significant differences among these four variables in both scaled models. Finally, a contrast method was conducted as post-hoc test to identify any significant pairwise comparisons.

CHAPTER 4

Results

This chapter presents the results of the data analysis used to test hypotheses of the current research. Data analyses were conducted using SAS 9.4™ software. As mentioned in the design section, there were two contexts and two harnesses and therefore, four tasks conditions to be tested. In each task condition sixteen participants were studied.

Table 3

Dependent Measures and their derivations

Dependent Measures	**Description/Derivation**
Performance Time	
DonTime	Time in sec to don one of the harnesses
DoffTime	Time in sec to doff one of the harnesses
TarTime	Time in sec to install one tar paper on the scaled roof
ShinTime	Time in sec to install 2 rows of shingles over tar paper on the scaled roof
Usability Ratings	
DHUsab	Ratings on Usability questions after donning the harness
TPUsab	Ratings on Usability questions after tar paper installation
Susab	Ratings on Usability questions after shingle installation
DoffUsab	Ratings on Usability questions after doffing the harness
FASUsab	Ratings for Fall Arrest Systems Usability after completion the whole roofing task
CItotal	Total Critical Incident frequencies
Perception of Scaled Model	
Feeling Safe	Feeling of safety compared to the actual construction site
Cautiousness	cautiousness while working on respective scaled model
Resemblance	Physical representation to the actual construction site
Decision Making	Decision making practice compared to the actual construction work

Two independent variables were manipulated in each trial. They were: Context and Harness Type. A total of four performance measures, five usability measures, critical incident frequencies and four scaled model perception responses (dependent variables) were recorded for each participant from each of the trials which were used for this study. The dependent measures and their derivations are listed in Table 3.

In this study, actual construction site was considered as high Fidelity. Some of the environmental factors as temperature, noise level and light level were measured. A comparison of these factors is given in Table 4. The data displayed here are approximate and may vary.

Table 4

Comparison of the environmental attributes for contexts

Environmental Factors	Indoor Context	Outdoor Context	Actual Construction Site
Temperature	70°F	55-87 °F	55-94 °F (may change from state to state)
Noise Level	45 dB(considering no other distraction)	64-70 dB	70-80 dB
Light level	440 Lux	20,000-110000 lux	110000-120000 lux

The first hypothesis stated that there will be no differences between the indoor and outdoor scaled models in terms of performance time. The second hypothesis stated that there will be no differences between the indoor and outdoor scaled models in terms of critical incidents. The third hypothesis stated that there will be no differences between the indoor and outdoor scaled models in terms of usability ratings. For all statistical tests, an alpha level of .05 was used.

Demographic factors were explored to examine possible differential effects based on key demographics from the BLS that account for differences in injury and fatality rates. The following demographic factors were considered for further study: participant age, weight, height, roofing experience and ethnicity. A comparison of these demographics for indoor and outdoor context is shown in Table 5.

Table 5

Comparison of demographics for indoor and outdoor context

Characteristics	Mean (SD)/Total (%)	Indoor(n=32)	Outdoor(n=32)
Age	36.9(10.46)	36.8(10.7)	37(10.45)
Weight lbs	188.69(35.09)	181.66(27.89)	195.06(39.74)
Height(inch)	70.34(3.37)	70.89(3.80)	69.83(2.91)
Roofing Experience	10.06(9.7)	12.30(10.47)	7.82(8.14)
Ethnicity			
African American	20 (31.25%)	1(3.1%)	19(59.4%)
Asian American	2(3.12%)	1(3.1%)	1(3.1%)
EA/Caucasian	28 (43.75%)	21(65.63%)	7 (21.9%)
Native American	-	-	-
Hispanic	9(14.06%)	4(12.5%)	5(15.63%)
Other	5 (7.81%)	5(15.63%)	-

4.1 Hypothesis 1: Effect of Context on Performance Time

A Shapiro–Wilk test of normality was conducted on Don Time, Doff Time, Tar Time and Shin Time. All variables showed non-normality in the distributions with W values ranging from .36 to .84.

Descriptive Statistics

The descriptive statistics for this study are summarized in Table 6. The mean don time increased with a change of context from indoor to outdoor by 93%, while the time for tar paper installation increased by 47%. The increments of mean shin time and doff time were 39% and 57%, respectively, from indoor to outdoor context.

Table 6

Mean (standard deviation) of the all response variables

Dependent Variables	**Context**	
	Indoor (n=32)	**Outdoor (n=32)**
Don Time	67.55(32.08)	130.36(75.98)
Tar Time	238.53(101.55)	351.1(516.18)
Shin Time	356.63(149.94)	462.19(292.63)
Doff Time	16(6.98)	25.15(14.51)

Relationship between performance times

Spearman’s rho was computed to assess the relationship between four performance times as these variables were found non-normal. Significant p-values are shown in Table 7.

Table 7

Correlation Analysis for performance times

	Age	Weightlbs	RoofExp	DonTime	TarTime	ShinTime	DoffTime
Age	-						
Weightlbs	.48***	-					
RoofExp	.64***	.36***	-				
DonTime	-.06	-.11	-.21	-			
TarTime	.06	.11	-.15	.41***	-		
ShinTime	-.06	.06	-.28*	.36***	.59***	-	
DoffTime	.14	.13	-.12	.41***	.41***	.45***	-

* p<.05.

** p<.01.

*** p<.00.

Overall differences across performance times

It was hypothesized that measures of performance time, as indexed by don time, doff time, tar time and shin time, are not related to the context. A generalized linear model using repeated measures for times was administered. Results from the repeated measures ANOVA established that the mean performance time changed across different level of times, Wilks' λ= 0.18, *F* (3, 54) = 83.13, *p* <.0001. It was also found that changes of mean performance time across four different levels were significantly influenced by the change of context, Wilks'λ= 0.81, *F* (3, 54) = 4.15, *p* =.01.

A main effect for context was found. It was identified that indoor and outdoor context were significantly different for performance times, *F* (1, 56) = 3.73, *p*=.05. Univariate results for the four task times indicated that the effect of context was significant for mean don time, *F* (1, 56) =17.08, *p*=.0001, as well as mean doff time, *F* (1, 56) = 10.71, *p*= .001. Mean don time (*M*= 130.36, *SD*=75.9) in the outdoor context was significantly higher than the mean don time in the indoor context (*M*=67.55, *SD*=32.08). Also, mean doff time in the outdoor context (*M*= 25.15, *SD*=14.51) was significantly higher than mean doff time in the indoor context (*M*=16, *SD*= 6.98). Figure 16 represents context effects over mean don time and doffs time. In both cases significant effects have been found for outdoor context. Error bars represent standard deviations over the mean values.

A post-hoc test using contrast method identified that mean don time was significantly different from mean tar time (*p*<.0001), mean shin time (*p*=.001) and mean doff time (*p*<.0001). Mean tar time was found to be significantly different from mean shin time (*p*<.0001) and mean doff time (*p*<.0001). Mean shin time (*p*=.001) and mean doff time (*p* =.002) were also significantly different.

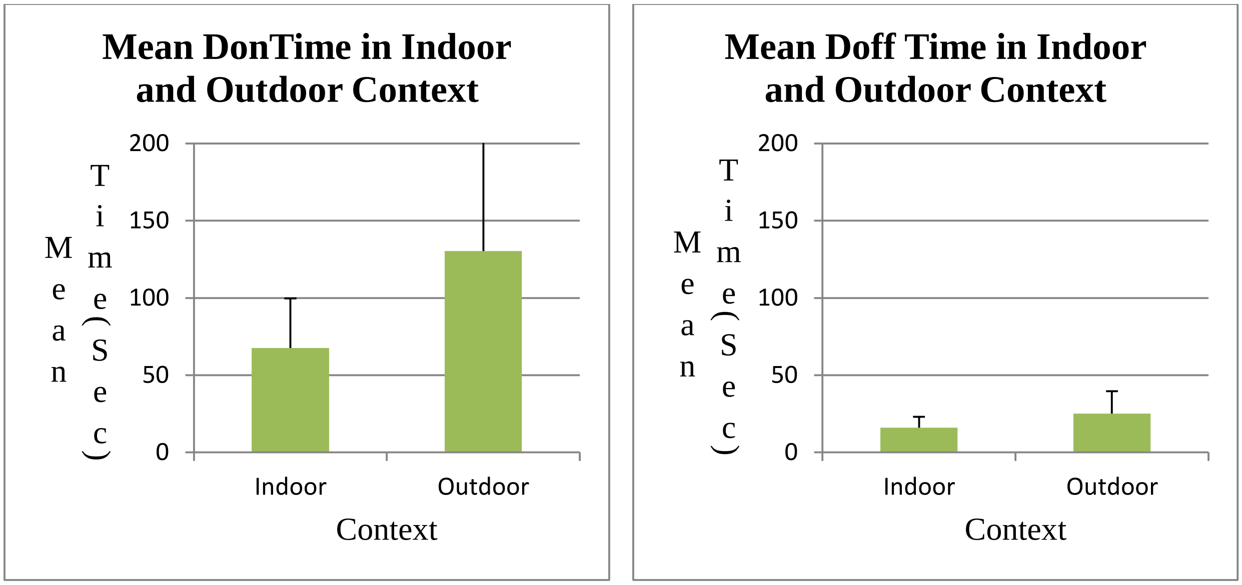

Figure 16: Comparison of (a) mean don time (b) mean doff time in indoor and outdoor context

<u>*Participant Demographics and Performance Time*</u>

A correlation analysis using Spearman’s rho was administered to examine the effect of age, height, weight and roofing experience on different performance times. A significant negative correlation was found between roofing experience and shingle installation time, r_s (61) = -.28, p= .02. This means that participants with high roofing experience took less time for shingle setting tasks. Other factors were found not to be significant.

For testing the ethnicity as an independent variable, six (6) ethnic groups were divided into two groups, White and Non-white, as Whites were prevalent in this study. A Wilcoxon ranked sum test administered for ethnic groups identified no differences between the performance times.

4.2 Hypothesis 2: Effect of Context on Critical Incidents

The second hypothesis was stated as: there would be no differences between the indoor and outdoor scaled models in terms of critical incidents. Among the ten critical incidents, six of them were found to be agreeable by both indoor and outdoor coders. Before coding the final video, each coder had to go through a trial coding, watching an online roofing task video and coding critical incidents. Once the initial frequencies for critical incidents were established, agreement was checked using correlation. Coders attended a resolution meeting to resolve any critical incident agreement percentage that were below .70. The agreement correlation among the critical incident frequencies in the indoor context was found to be .9 and in the outdoor context was found to be .99.

Table 8

critical incident frequencies observed in the indoor and outdoor context

	Context		
Harness Type	**Indoor**	**Outdoor**	**Total**
Low	74	45	119
Mid	52	32	84
Total	126	77	203

A total of two hundred and three (203) critical incidents were observed in both contexts. 62% of the critical incidents were observed in the indoor context and 38% were found in the outdoor context. Higher critical incident frequencies were found for the low-end harness compared to the mid-range harness (See Table 8).

Analyzing the raw frequencies for each of the critical incidents, it was found that awkward posture in indoor context (30%) and entanglement in outdoor context (26%) accounted for the highest frequencies among all (see Figure 17).

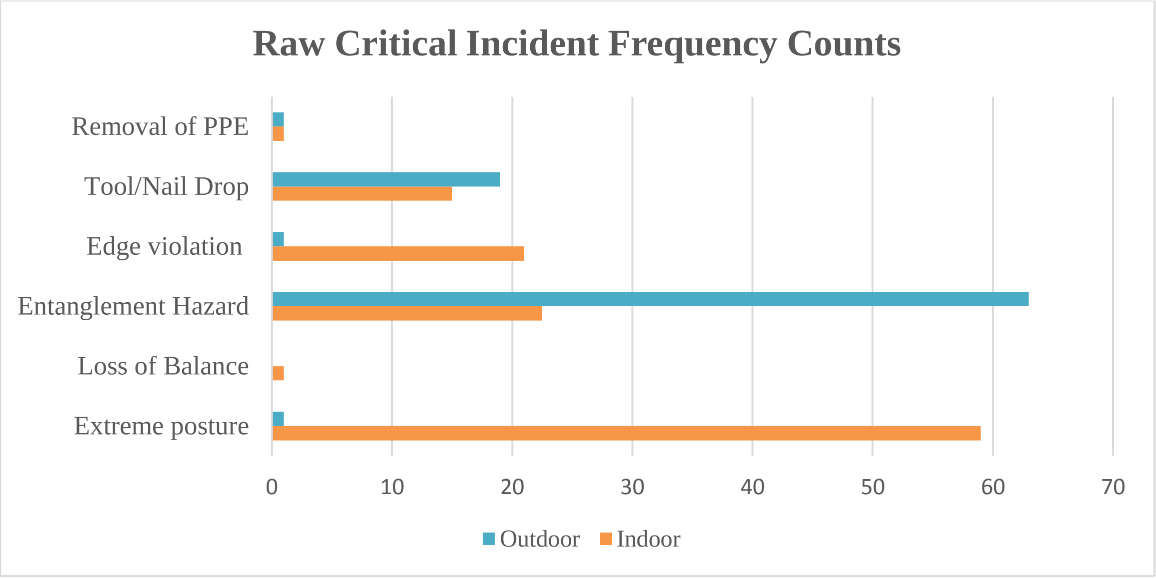

Figure 17: Raw critical incident frequency counts for indoor and outdoor context

Figure 18 demonstrates that among all of the critical incidents observed in the indoor context, awkward posture is the leading critical incident which accounts for 48% of all based on the relative frequency. In the outdoor context, 70% of critical incidents were due to entanglement and 25% due to the nail drops (see Figure 19).

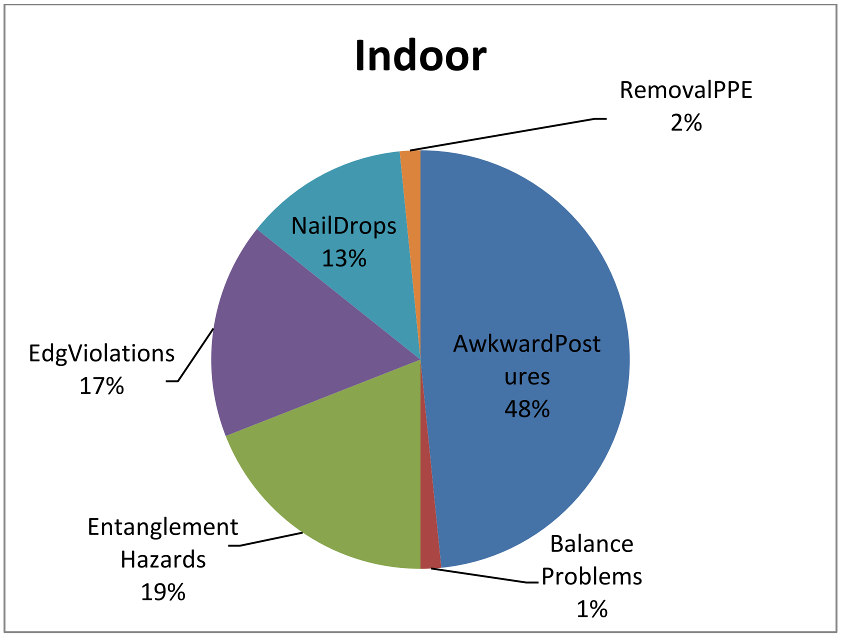

Figure 18: Distribution of critical incidents in the indoor context

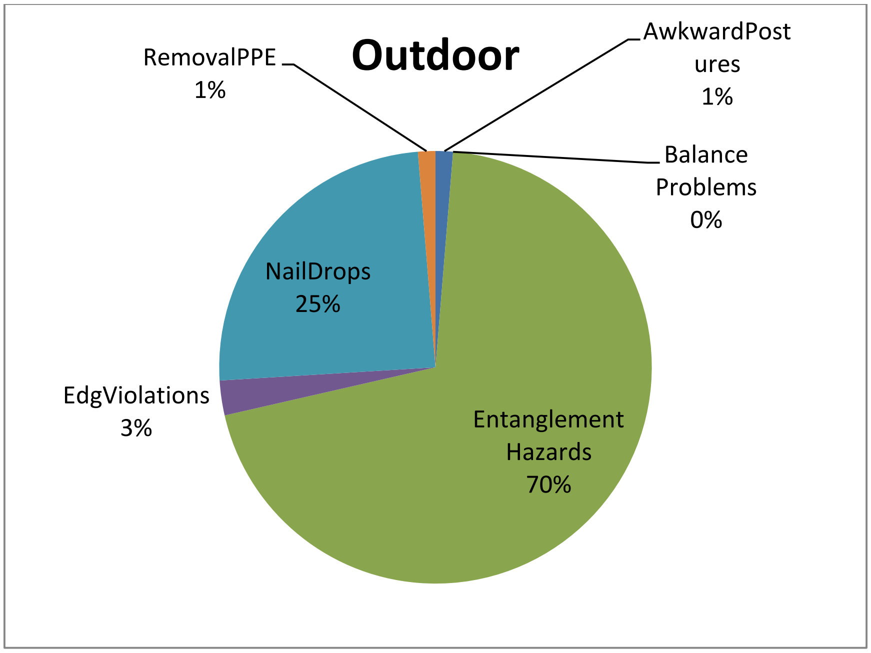

Figure 19: Distribution of critical incidents in the outdoor context

A correlation analysis was conducted with six variables developed on the basis of the quantized codes from the critical incidents. Table 9 shows a significant correlation between edge violations and awkward postures, *r* (10) =.82, *p*=.001. Also, edge violations and balance problems were found to be significantly correlated, *r* (10) = .57, *p*=.05.

Table 9

Correlation Analysis of mean critical incident frequencies

	Posture	Balance	Entangle	EdgeVio	NailD	RemoPPE
Posture	-					
Balance	.44	-				
Entangle	-.34	0	-			
EdgVio	.82**	.57*	-.34	-		
NailD	-.29	-.33	.33	-.12	-	
RemoPPE	.12	-.26	.05	.03	-.34	-

Posture= Awkward posture, Balance= Balance problem, Entangle= Entanglement, EdgVio= Edge Violation, NailD= Nail Drop, RemoPPE= Removal of PPE

* p<.05.

** p<.01.

*** p<.00.

Fisher’s exact test was administered with the total critical incident frequencies for the indoor and outdoor context. Primary results from the Fisher’s exact test indicated no significant difference in critical incident frequencies by using low and mid-grade harnesses when crossed with the indoor and outdoor contexts.

Participant Demographics on Critical Incidents

No significance was found between critical incidents participants’ age, weight, height and roofing experience. However, the Wilcoxon ranked sum approached significance in terms of ethnicity. The result from the Wilcoxon test indicated that critical incident frequency was comparatively higher among white participants (*Mdn*=22) than in non-white participants (*Mdn*=14), *Z*= 2.03, *p*=.06, *r*=.06. However, ethnicity may be confounded by context, since most of the non-white participants were in the outdoor study.

4.3 Hypothesis 3: Effect of Context on Usability Ratings

The third hypothesis was stated as indoor and outdoor models would not be different when usability ratings were considered. To test the hypothesis, five usability ratings were used.

- Donning Harness Usability
- Doffing Harness Usability
- Tar Paper Usability
- Shingle Usability
- Fall arrest system(FAS) Usability

All usability questions were measured on a six (6)-point scale (1= negative anchor, stated as strongly disagree, 6= positive anchor, stated as strongly agree). Cronbach’s Coefficient alpha was used to assess the reliability of the individual items by measuring internal consistency. The criterion value of Cronbach’s alpha was set at the .7 value (Nunnally 1978).

For each of the dependent variables, all usability ratings were investigated to see whether by deleting one or more items, the r-alpha value increased. If the deletion of items improved the Cronbach's alpha, a new summated rating value was calculated for the retained items.

Reliability of Dependent variables

Cronbach's Alpha was calculated as reliability statistics with five dependent variables separately. Cronbach's alphas for the four Doffing Harness Usability and six FAS Usability items were .76 and .75, respectively. For the other three variables, the Cronbach's Coefficient alphas approached the criterion value after dropping unreliable items (new alphas ranged from .62-.69) (see Table 10). After deleting items with lower reliability for each of the variables, a corrected usability rating was calculated by adding the remaining items for each of those variables.

Table 10

Cronbach's Coefficient Alpha for dependent variables

Variables	Cronbach's Coefficient alpha (raw)	Deleted Variables	Improved Alpha (raw)
Donning Harness Usability	.39	**DHUsab1**	.62
Tar Paper Usability Rating	.28	**TPUsab6**	.69
Shingle Usability Rating	.25	**SUsab9**	.66
Doffing Harness Usability	.76	-	-
Usability Rating for FAS	.75	-	-

Multiple Analysis of Variance (MANOVA) results

For Multiple Analysis of Variance (MANOVA), the items for each variable were added after deletion of the bolded items in Table 9. MANOVA analysis was conducted for context and harness type with five of the summated rating usability constructs. Results from the MANOVA demonstrated a significant multivariate effect for context, *Wilks' λ*=0.78, *F* (5, 50) =2.78, *p*=.02. No significant multivariate effect was found for harness type and context* harness type interactions.

Among all of the five variables, significant univariate effects were found on four of these corrected usability constructs. From the results (see Table 5), it was identified that mean Don Harness Usability, DHUsabT in the outdoor context (*M*=9.14, *SD*=2.69) was significantly different than the mean DHUsabT (*M*=7.09, *SD*=2.68) in the indoor context, *F* (1, 54) =8.08, *p*=.006. Also, a higher outdoor context effect was found for TPUsabT , *F*(1,54)=10.75,*p*=.001; SUsabT, *F*(1,54)=9.5, *p*=.003 and FASUsabT, *F*(1,5)=8.52, *p*=.005 (see Appendix H). Here TPUsabT, SUsabT and FASUsabT represent corrected total score on Tarpaper Usability, Shingle Usability and Fall Arrest System Usability constructs.

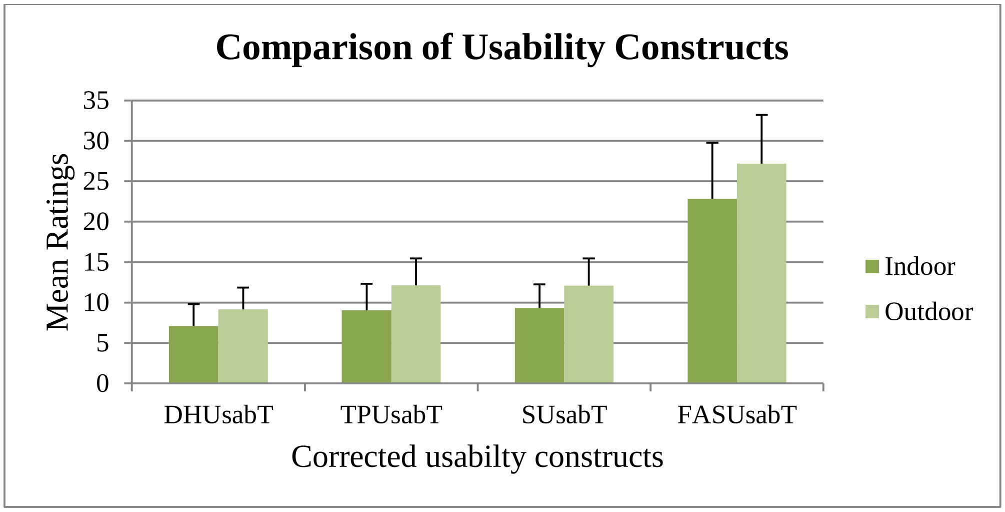

Figure 20: Effect of context on corrected usability mean

Significant context effects are shown in figure 20. It was found that the corrected means for these four usability variables were significantly higher in the outdoor context. This indicates that participants in the outdoor context rated usability to be higher in donning, tarpaper setting, shingle installation and fall arrest systems usability variables compared to the indoor context. Bars represent standard deviations above the mean values.

Participant Demographics and Usability

Table 11 shows the inter-correlations among age, weight, roofing experience and the corrected usability variables. Roofing experience was found to have a significant correlation with corrected don harness usability, r (62) =-.37, p=.004 and tar paper usability, r (62) =-.26, p=.03.

A Wilcoxon ranked sum test was conducted to evaluate whether usability variables were related by ethnicity. Results from the tests demonstrate no significant variation in all of the five usability variables between the two ethnic groupings: White and non-White.

Table 11

Correlation Analysis for Usability measures

	Age	Weightlbs	RoofExp	DHUsabT	TPUsabT	SUsabT	DoffUsabT	FASUsabT
Age	-							
Weightlbs	.48***	-						
RoofExp	.64***	.42***	-					
DHUsabT	-.2	-.24	-.37**	-				
TPUSabT	-.09	-.11	-.26*	.59***	-			
SUsabT	-.07	-.01	-.23	.48***	.85***	-		
DoffUsabT	.17	.07	-.008	.39***	.36***	.28*	-	
FASUsabT	.02	-.03	-.08	.29*	.57***	.63***	.19	-

DHUsabT=corrected don harness usability ratings, TPUsabT=corrected tar paper usability ratings, SUsabT= corrected shingle usability ratings, DoffUsabT= corrected doff harness usability ratings, FASUsabT= corrected fall arrest system usability ratings

* p<.05.

** p<.01.

*** p<.00.

4.4 Fidelity of the Scaled Worlds: Questionnaire Results

In this study, two scaled models were used and both of these scaled models were aimed to mimic the real life situation in terms of equipment usage, participant selections and roof model. Participants were asked to score their perception of realism based on a six (6)-point scale

(1 –strongly disagree and 6- strongly agree) over four questions designed to understand participants' perceptions of the scaled models (see Appendix G). They rated the scaled models, on which they worked, in terms of feelings of safety, cautiousness, resemblance and decision making criteria. While rating these questions, these participants compared the respective scaled model to actual construction sites which they had worked before.

A repeated measures ANOVA was used to analyze four responses. The results identified significant variations among the four fidelity variables, Wilks' λ=0.58, F (3, 59) = 14.34, p<.0001.

A main effect for context was found, F (1, 61) = 19.01, p<.0001. Univariate results further identified that cautiousness was significantly different across the context, F (1, 61) =13.86, p=.0004. Also, significant difference was found on resemblance for the context, F (1, 61) = 5.41, p=.02 (see Figure 21). Other constructs were found to be not significantly different.

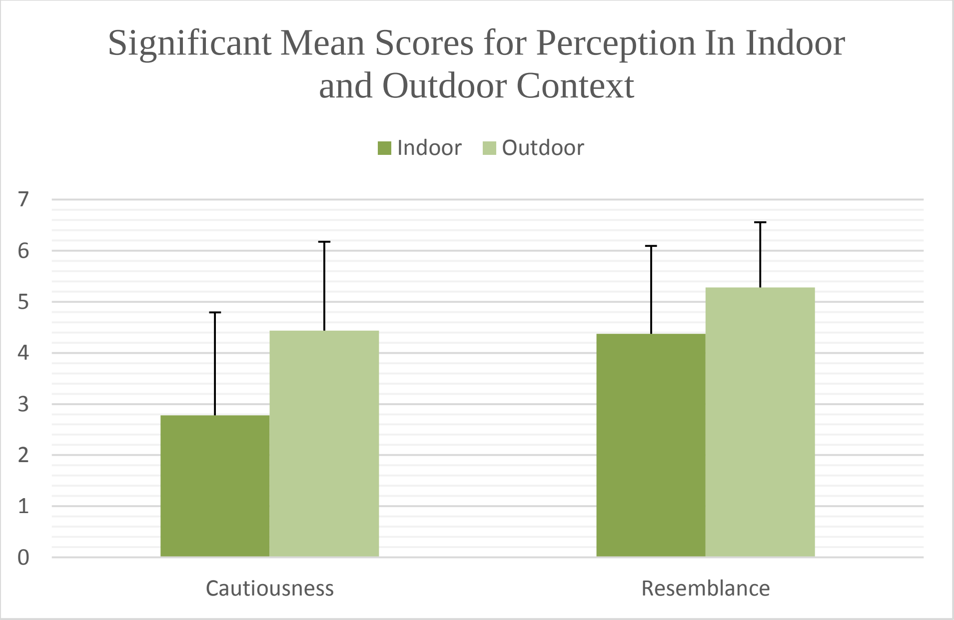

Figure 21: Significant mean scores on Cautiousness and Resemblance, based on the comparison to the actual construction site.

A post-hoc test was conducted using the contrast methods. Results demonstrated that feeling safe and cautiousness (p=.0006), feeling safe and decision making practice (p<.0001), cautiousness and resemblance (p=.0006), cautiousness and decision making practice (p=.02), resemblance and decision making practice (p<.0001) were significantly different.

Participant Demographics and Fidelity

A correlation analysis was administered using Spearman’s rho with the four responses and participants’ demographics (Table 12). From the Spearman’s rho correlation, it was found that participant’s years of experience on roofing was negatively correlated to the item measuring decision making practice , $r_s(62)=-.25$, $p< .05$. This means that when the participant was highly experienced, he/she reported they would make the same decision as if they were working at an actual construction site. Also, participants’ weight and perception of safety was found to be correlated, $r_s(62)=.27$, $p< .05$, meaning the more they weighted the safer they felt on the scaled model compared to an actual roof.

Table 12

Correlation Analysis on scaled model perceptions parameters

	SM1	SM2	SM3	SM4	Age	RoofExp	Weightlbs
SM1	-						
SM2	.34***	-					
SM3	-.09	-.01	-				
SM4	.24	.39***	-.15	-			
Age	-.12	-.13	.07	-.1	-		
RoofExp	-.17	-.27	-.07	-.25*	.64***	-	
Weightlbs	.27*	.1	.04	.05	.47***	.36**	-

SM1= Perception of the scaled model in terms of safety, SM2= perception of the scaled model in terms of cautiousness, SM3= Perception of the scaled model in terms of resemblance to the actual construction site, SM4= perception of the scaled model in terms of judgment/ taking decision.
* p<.05.

** p<.01.

*** p<.00.

A Wilcoxon ranked sum test identified that some of the participants' perception were influenced by ethnicity. The perception of feeling safe was reported higher among non-white participants compared to the white participants, $Z = -2.49$, $p < .05$, When the participants rated the scaled models in terms of cautiousness, it was found that non-White participants were less cautious than White participants, $Z = -2.94$, $p = .004$. Since most of the non-White participants were in the outdoor study, ethnicity may be confounded by context.

CHAPTER 5

Discussion

The purpose of this study was to compare the indoor scaled model to the outdoor scaled model considering the performance times, critical incidents and usability ratings reported by participants while wearing two different types of harnesses. As the outdoor scaled model was anticipated as representative of the real task environment, the effort of the current study was to measure the ecological validity of the indoor model. Three research objectives were considered to analyze the effect of context. The objectives are discussed respectively in sections 5.1- 5.3. The results for the fidelity of the scaled models are discussed in section 5.4.

5.1 Effect of Context on Performance Time

The first objective was to identify whether significant differences existed between indoor and outdoor scaled models when four of the performance times were considered. Significant positive correlations were found among these variables except in the case between don time and tar time.

A significant time effect was found across the four different performance times. These four times differed from one another without considering the effect of the context. These times were measured on four different tasks on the same participants. Because of the nature and complexity of the tasks, the required time for doing each of these tasks was expected to vary. Also, individual work skill might have some influence on the task time.

Further analysis identified that change of context also affected the performance times. As stated earlier, two scaled models were established in two places—one in a closed environment and another in an open environment. The environmental attributes have affected the four task times in this respect. A main effect for context identified that indoor and outdoor scaled models in some respect might differ.

However, no overall harness effect was found in the participants' performance time. This finding implied that no matter which harness was used, participants' performance would not significantly vary. In residential roofing, most of the small construction companies do not provide their roofers with harnesses although they are required as employers to do so. Thus, the participants may not be familiar with the use of harnesses and the complexity level may be similar for both of the tested harnesses.

Univariate results for the four times explored that mean don time and doff time in the outdoor context were significantly higher. Participants of the outdoor context took more time to don and doff their respective harnesses compared to the indoor participants. The donning and doffing time might be affected by the attributes of the environment. As the participants in the indoor scaled model donned and doffed the harnesses in a soothing environment with no sunlight or noise and fixed temperatures with low humidity, they might have taken less time in these tasks. However, participants' demographics were found not to be related to any of these performance times.

5.2 Effect of Context on Critical Incidents

Comparing the raw frequencies of the critical incidents, it was observed that indoor participants faced a higher number of critical incidents. People, when working in a closed environment under continuous supervision, find their tasks might be affected. They tried to work fast or slow and this deviation from the natural behavior may affect their overall critical incidents. Bentley et al. (2006) mentioned that behavioral and perceptional factors contribute to falls related to injuries in residential roofing. Moving too fast for a certain task or taking shortcuts and poor hazard perception can lead to critical injuries.

The correlation analysis demonstrated a high correlation between awkward postures and edge violations. Participants who had many incidents of awkward posture, also tended to commit more edge violations. Edge violations were also positively correlated with balance problems and almost 30% of the variability of the edge violation could be explained by balance issues. According to NIOSH (2000), loss of balance during work and unstable work surfaces are the leading risk factors in roof related injuries.

Any significant effect of context was not found on the critical incidents although higher overall critical incident frequencies were found in the indoor context. The Wilcoxon ranked sum test explored a slight variation of critical incident frequencies among the White and non-White groups. It was found that White participants had more critical incidents compared to non-White participants. Among the different critical incidents, awkward posture (21%) was the highest frequency for White participants and entanglement (23%) was the most common critical incident for non-White participants.

5.3 Effect of Context on Usability Rating

Participants' responses were collected using the scores on five usability constructs that were mentioned earlier. Internal consistency of the items for each usability construct demonstrated that the reliability of three constructs were less than the acceptance level, .7 (Nunnally, 1978). Kline (2013) mentioned that when dealing with psychological construct values, where diversity of the items to be measured is dominant, the alpha value may go below .7. According to Streiner (2003), the alpha value increases when items of the constructs are correlated to each other.

However, for higher internal consistency the length of the test is also important. If the construct contains more related items, the alpha value of that construct increases. In the current study, the reliabilities for the Doff Harness Usability scale with four items for one question and for the Fall Arrest System Usability scale with six questions were found to be .76 and .75, respectively. These results support the effectiveness of the questionnaires used.

MANOVA reported a context effect on the corrected usability measures. Univariate results demonstrated higher corrected usability means for all of the variables except doff harness usability in an outdoor context. Usability rating depends on the participants' satisfaction of the ease of use product. When participants are familiar with the overall settings, they tend to rate higher. Studies show that it is very common for people to become nervous when they are being tested. Often a relaxed atmosphere, an important condition for usability testing, can minimize the nervousness among people (Nielsen, 1994) (*p* 34). In this study, as the environmental factors such as temperature, noise level, humidity and light level were similar to the regular task environment where they regularly worked, participants may feel relaxed and provide a higher rating on these usability constructs. This indicates that participants in an outdoor context were more directed to the positive poles of the scales.

A correlation analysis explored high correlation between roofing experience and *Doffing Usability ratings* which indicates participants with high roofing experiences rated higher for Usability of doffing the harness. All of the corrected usability measures were found to be significantly correlated to each other. It implies that a score on one item also influences the score of other items. It can be inferred that based on the usability measures, the outdoor scaled model possessed the higher usability compared to the indoor scaled model.

5.4 Fidelity of the Scaled Worlds: Questionnaire Results

Participants' perceptions about the scaled world model in this study were captured on four dimensions— safety attitude, cautiousness, resemblance and decision making practice. A repeated measures ANOVA identified that mean scores of these four perception variables were significantly different. A post-hoc contrast further revealed that other than safety and resemblance issues, the remaining variables were significantly different from each other. Univariate results established that compared to the indoor participants, outdoor participants reported that they felt less cautious while working on the scaled model. The resemblance of the scaled model to the actual construction site was determined by eliciting perceptions from participants. The outdoor participants scored higher than the indoor participants in terms of resemblance or perceived fidelity. Finding also supports the fidelity of the outdoor scaled model as a sufficient replication of the real world. This means that the outdoor scaled model was perceived to be more similar to the actual construction site than the indoor scaled model. The similarity of the environment to the actual construction work site might have influenced the participant's response in this case.

Spearman's rho identified that safety and cautiousness were positively correlated. Participants worked on the scaled roof structure which was placed on the ground level. In an actual construction site, the roof structure is highly elevated and roofers have to work on steep roofs. Therefore, the feelings of safety among the participants might be greater in the scaled model in this case and they also felt less cautious when working on the scaled roofs.

Cautiousness and decision making practices were also found to be correlated. The participants, who were less cautious, also felt free in making decisions. Sutcliffe and McNamara (2001) elicited that decision practice is affected by two features— the characters of the decision and the context where the decision maker is situated. However, in this study, participants' decisions were found to be unaffected by the context. Therefore, based on the decision making criteria, the indoor and outdoor scaled model might be considered as similar.

Non-White participants reported higher safety feelings than the White participants. Also, the non-White participants remained less cautious while working on the scaled model than the white participants. However, these findings were confounded by the context, as most of the non-White participants were from the outdoor study.

CHAPTER 6

Conclusions

Based on the results found in this study, the indoor scaled model was as ecologically valid as the outdoor scaled model in terms of roofing tasks and critical incidents. However, donning and doffing the harness employed more time in the outdoor context compared to the indoor context. The times for donning and doffing procedures are a crucial concern for the usability of FAS. Outdoor participants gave higher ratings on all the usability items except for doffing. Also, for the perception of the scaled model questionnaire, participants rated the outdoor scaled model as having a higher fidelity. So it can be inferred that based on the usability issues, the outdoor scaled model possessed higher fidelity. These ratings are based on the participants' perceptions of their experiences. These perceptions are not concrete and they can be changed from person to person. In this study, different participants were used for the two contexts. In future same participant can be used for both contexts to test the influence of context and perceived behavior. In this research, we found some outcomes were influenced by the ethnicity. However, as most of the non-White participants were found in the outdoor studies, the ethnicity effect may be confounded by context.

Based on the perception of the results of this study, several recommendations can be made. A list of recommendations for the indoor and outdoor scaled model is shown in Table 13. When we talk about usability some concerns come to our mind; ease of use, satisfaction of the users, and overall performance of the system. When people rate the usability, they deliver their responses based on their cognition and judgment. Based on the findings of this study, the outdoor scaled model can be recommended for the usability studies where higher usability rating is desired. However, in many cases, especially in designing phase, the designer looks for the lower usability ratings. If the designer finds the lower ratings on the usability of a product, they try to find out causes and work for the improvement.

Table 13

List of recommendations

Problem Focus	**Indoor Model**	**Outdoor Model**	**Recommendations**
System Usability Study	√	√	• ***Indoor:*** When the designer or manufacturing needs precise, non-contextual data, especially when low-fidelity or moderate-fidelity prototypes are available. • ***Outdoor:*** For final prototype, when context attributes are known to influence behavior.
Monotonous Work Study	√	√	• Priority = ***Indoor:*** When the work is routinized and context will not change the work process.
Decision Intensive Study	√	√	• Priority = ***Outdoor:*** When time and accuracy both matter for complex decisions and when multiple contextual factors (i.e., environmental) are known to influence decision-making, outdoor testing is recommended.
Critical Incident Identification	√	√	• ***Indoor:*** When roofing product or roofing tasks are in need of study outside of other potential contextual effects. • ***Outdoor:*** When context influences use of the roofing product or roofing task.
Demographic Differences (i.e., gender, ethnicity, age, experience)	√	√	• ***Indoor:*** When initially exploring demographic differences in FAS compatibility and fit, impact of training, or performance • ***Outdoor:*** When context influences performance, such as examining the influences of thermal hazards on performance of roofing tasks by age or gender.

The indoor scaled model may be recommended in this scenario, as we found from our research indoor participants rated comparatively lower on the usability. The study which involves continuous and monotonous work with less variation, the indoor scaled model may be more appropriate. A task can be said as monotonous when (a) the task itself and the total work situation has little variation, (b) continuous attention of the individual is required so that no imagination can be developed, and (c) the work is well structured and preplanned with minimum cognition activity (McBain, 1961).

In studies where higher mental process; such as— problem solving or imagination is required, the outdoor scaled model is recommended. As people take decisions based on the situation, variation of contexts can affect the decision making practice. Studies, where researchers’ main focus is on critical incident identification, the indoor scaled model may be suggested based on the finding of this study. However, for each of these studies the other model can also be used. If researchers have some constraints such as—budget, time or weather; the indoor scaled model is the most appropriate solution.

The study only considered the regular weather condition. However, in many cases roofers need to work in adverse weather— rainy day, windy day or extreme cold weather. For future research these extreme weather condition should be considered, as they might have adverse effect on the performance and attitude outcome of participants. Some other factors that demand for future research are—time exposure on the roof, participant’s body dimension, roof pitch and the height of the scaled model. As the scaled models, on which this study was based were not very elevated, participants might have deemed the scaled model as safer and less risky.

Also, in the future, researchers should focus on other issues, such as certain aspects of cognition. In determining system usage, it is very important to understand how individuals employ cognition for decision making. Previous research explored that a person’s perception, evaluation and judgment about a target stimulus not only depend on the target itself but also the context where the target is integrated (Avramova, Stapel, & Lerouge, 2010; Biernat, 2012).

This study presents a novel method for understanding performance times, critical incidents, usability and worker attitudes associated with the adoption of FAS and PPT in both the indoor and outdoor scaled models. The results of this study will further help researchers to select the appropriate scaled model for construction research studies. The existing debate on construction research field environments will be impacted by these results. Also, cost-benefit tradeoff models could be developed from this research.

References

Abowitz, D., & Toole, T. (2010). Mixed Method Research: Fundamental Issues of Design, Validity, and Reliability in Construction Research. *Journal of Construction Engineering and Management, 136*(1), 108-116. doi: doi:10.1061/(ASCE)CO.1943-7862.0000026

Akritas, M. G., Arnold, S. F., & Brunner, E. (1997). Nonparametric Hypotheses and Rank Statistics for Unbalanced Factorial Designs. *Journal of the American Statistical Association, 92*(437), 258-265. doi: 10.1080/01621459.1997.10473623

Angles, J., Trochez, G., Nakata, A., Smith-Jackson, T., & Hindman, D. (2012). Application of Scaled World Model for Usability Testing of Fall Arrest System in Residential Construction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 2060-2064. doi: 10.1177/1071181312561430

Angles, J.(2013) Usability of Fall Arrest Harness.Retreived at 3/13/2015. https://vtechworks.lib.vt.edu/bitstream/handle/10919/23089/Angles_JN_T_2013.pdf?sequence=1

Avramova, Yana R, Stapel, Diederik A, & Lerouge, Davy. (2010). Mood and context-dependence: Positive mood increases and negative mood decreases the effects of context on perception. *Journal of personality and social psychology, 99*(2), 203.

Bentley, T. A., Hide, S., Tappin, D., Moore, D., Legg, S., Ashby, L., & Parker, R. (2006). Investigating risk factors for slips, trips and falls in New Zealand residential construction using incident-centred and incident-independent methods. *Ergonomics, 49*(1), 62-77.

Bernold, L., & Lee, T. (2010). Experimental Research in Construction. *Journal of Construction Engineering and Management, 136*(1), 26-35. doi: doi:10.1061/(ASCE)CO.1943-7862.0000085

Biernat, Monica. (2012). *Standards and expectations: Contrast and assimilation in judgments of self and others*: Psychology Press.

BLS 2012. Survey of Occupational Injuries and Illnesses data. Retrieved at 3/13/2015. http://www.bls.gov/iif/oshwc/cfoi/cfch0011.pdf

BLS, 2013. Census of Fatal Occupational Injury Summary. Retrieved AT 3/8/15. http://www.bls.gov/news.release/cfoi.nr0.htm

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior, 9*(2–3), 171-184. doi: http://dx.doi.org/10.1016/0747-5632(93)90005-D

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American psychologist, 32*(7), 513.

Cameron, I., Gillan, G., & Roy Duff, A. (2007). Issues in the selection of fall prevention and arrest equipment. *Engineering, Construction and Architectural Management, 14*(4), 363-374. doi: doi:10.1108/09699980710760676

Clark, S. (May 2008). Laborers' Health & Safety Fund of North America. *Residential Construction Presents Many Hazards. LifeLines Online, 4,* 1.

Clements-Croome, D. (2006). *Creating the Productive Workplace*: Taylor & Francis.

Difonzo, N., Hantula, D., & Bordia, P. (1998). Microworlds for experimental research: Having your (control and collection) cake, and realism too. *Behavior Research Methods, Instruments, & Computers, 30*(2), 278-286. doi: 10.3758/BF03200656

Dong, X. S., Fujimoto, A., Ringen, K., & Men, Y. (2009). Fatal falls among Hispanic construction workers. *Accident Analysis & Prevention, 41*(5), 1047-1052. doi: http://dx.doi.org/10.1016/j.aap.2009.06.012

Drabble, F., & Brookfield, D. J. (2000). Safety of fall arrest systems: A numerical and experimental study. *Proceedings of the Institution of Mechanical Engineers, 214*(10), 1221.

Ehret, B. D., Gray, W. D., & Kirschenbaum, S. S. (2000). Contending with complexity: Developing and using a scaled world in applied cognitive research. *Human Factors, 42*(1), 8.

Ehret, B. D., Kirschenbaum, S. S., & Gray, W. D. (1998). Contending with complexity: The development and use of scaled worlds as research tools. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting, 1*, 254.

Ellis, J. N. (2002). Personal fall arrest systems. *Professional Safety, 47*(12), 42-44.

Elson, J. L., II. (2003). *Expert and novice performance in an industrial engineering scaled-world simulation.* (3115740 Ph.D.), The Ohio State University, Ann Arbor. Retrieved from http://search.proquest.com/docview/305304349?accountid=12711

Falk, A., & Heckman, J. J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science, 326*(5952), 535-538. doi: 10.1126/science.1168244

Gaskell, R., Husman, L. E., Collier, J. B., & Chen, R. L. (2007). Synthetic Environments for Simulated Missions. *Aerospace and Electronic Systems Magazine, IEEE, 22*(7), 14-20. doi: 10.1109/MAES.2007.4285976

Hallowell, M., Roucheray, M., & Esmaeil, B. (2012). Developing a Framework for Measuring the Effectiveness of Common Fall Prevention/Protection Practices *ICSDEC 2012* (pp. 719-726).

Hsiao, H., Friess, M., Bradtmiller, B., & Rohlf, F. J. (2009). Development of sizing structure for fall arrest harness design. *Ergonomics, 52*(9), 1128-1143. doi: 10.1080/00140130902919105

Hu, K., Rahmandad, H., Smith-Jackson, T., & Winchester, W. (2011). Factors influencing the risk of falls in the construction industry: a review of the evidence. *Construction Management and Economics, 29*(4), 397-416. doi: 10.1080/01446193.2011.558104

Huang, X., & Hinze, J. (2003). Analysis of Construction Worker Fall Accidents. doi: 10.1061//asce/0733-9364/2003/129:3/262

Hughes, R. G., Durham, C. F., & Alden, K. R. (2008). Enhancing patient safety in nursing education through patient simulation.

Hung, Y. H., Smith-Jackson, T., & Winchester, W. (2011). Use of attitude congruence to identify safety interventions for small residential builders. *Construction Management and Economics, 29*(2), 113-130. doi: 10.1080/01446193.2010.521758

Jaselskis, E., Anderson, S., & Russell, J. (1996). Strategies for Achieving Excellence in Construction Safety Performance. *Journal of Construction Engineering and Management, 122*(1), 61-70. doi: doi:10.1061/(ASCE)0733-9364(1996)122:1(61)

Johnson, H., Singh, A., & Young, R. (1998). Fall Protection Analysis for Workers on Residential Roofs. *Journal of Construction Engineering and Management, 124*(5), 418-428. doi: doi:10.1061/(ASCE)0733-9364(1998)124:5(418)

Kaskutas, V., Evanoff, B., & Miller, H. (2013). Fall Prevention on Residential Construction Sites. *Professional Safety, 58*(7), 36-40.

Kim, K.-J., & Ashton-Miller, J. A. (2009). Segmental dynamics of forward fall arrests: A system identification approach. *Clinical Biomechanics, 24*(4), 348-354. doi: http://dx.doi.org/10.1016/j.clinbiomech.2009.01.007

Kline, P. (2013). *Handbook of psychological testing*: Routledge.

Koch, L. M., Smith-Jackson, T., Morris, J., & Hindman, D. P. (2013). Wood Truss Installation Loads Due to OSHA Fall Protection *Structures Congress 2013* (pp. 1255-1264).

Kozlowski, S. W., & DeShon, R. P. (2004). A psychological fidelity approach to simulation-based training: Theory, research and principles. *Scaled worlds: Development, validation, and applications*, 75-99.

Lederer, M., Choi, S. D., & Griinke, D. (2006). Fall protection equipment effects on productivity and safety in residential construction. *Journal of Construction Research, 07*(01n02), 149-157. doi: doi:10.1142/S1609945106000578

Maran, N. J., & Glavin, R. J. (2003). Low- to high-fidelity simulation – a continuum of medical education? *Medical Education, 37*, 22-28. doi: 10.1046/j.1365-2923.37.s1.9.x

McBain, William N. (1961). Noise, the" arousal hypothesis," and monotonous work. *Journal of Applied Psychology, 45*(5), 309.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation, 24*(3), 315-340.

Newell, A. (1994). *Unified theories of cognition*: Harvard University Press.

Nielsen, J. (1994). *Usability engineering*: Elsevier.

NIOSH (1992). Alaska Face: Fatality assessment & Control Evolution. Retrieved at 3/13/2015. http://www.epi.hss.state.ak.us/injury/occupation_injury/reports/docs/niosh91ak53.htm

NIOSH, 1997. West Virginia State Face Report: Roofer Dies in West Virginia Following a Fall from a Roof. Retrieved at 3/13/2015. http://www.cdc.gov/niosh/face/stateface/wv/97wv055.html

NIOSH, 2012. Hispanic worker falls from residential roof – North Carolina. Retrieved at 3/13/2015. http://www.cdc.gov/niosh/face/in-house/full201202.html

Nunnally, J. (1978). Psychometric methods. *New York: McGraw*.

OSHA, 2015. Safety and Health Regulations for Construction. Retrieved at 3/13/2015. https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_id=10757&p_table=STANDARDS

Roy, S. (1999). VRUSE—a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems. *Appl Ergon, 30*(1), 11-25. doi: http://dx.doi.org/10.1016/S0003-6870(98)00047-7

Sa, J., Seo, D. C., & Choi, S. D. (2009). Comparison of risk factors for falls from height between commercial and residential roofers. *J Safety Res, 40*(1), 1-6. doi: 10.1016/j.jsr.2008.10.010

Schmuckler, M. A. (2001). What Is Ecological Validity? A Dimensional Analysis. *Infancy, 2*(4), 419-436. doi: 10.1207/S15327078IN0204_02

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*: Wadsworth Cengage learning.

Sharp, M. (2013). Arresting Developments. *ISHN, 47*(11), 32-32,34.

Smith-Jackson, T., Hindman, D., Shields, L., Thomas, B., Gallagher, C., & Koch, L. (2011). Safety climate and use of fall arrest systems. *Proceedings of CIB W099 2011, Prevention: Means to the End of Construction Injuries, Illnesses and Fatalities*, 24-26.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment, 80*(1), 99-103.

Sutcliffe, Kathleen M., & McNamara, Gerry. (2001). Controlling Decision-Making Practice in Organizations. *Organization Science, 12*(4), 484-501.

Appendix A

Questionnaire1: Demographic Questionnaire

Please read each question carefully and answer each to the best of knowledge.

1) Size of Company (# of employees including yourself) -----------------------------
2) How long in years have you worked in construction? -----------------------------
3) How much experience in years do you have as a roofer (must have at least 1 year)? -------
4) How long in years have you worked for THIS Company (minimum 1 year)? --------------
5) Age (must be at least 18 years) -----------------
6) Gender: --------------------Male ---------------------Female
7) Racial/ Ethnic Group(circle one):

   ☐ African- American/Black (not of Hispanic origin) ☐ Asian/Pacific Islander

   ☐ American-Indian/ Native American ☐ Hispanic/Latino

   ☐ European-American/Caucasian ☐ Other----------

8) Education (circle one):

   ☐ Less than 12 years ☐ Vocational/trade school degree or certification

   ☐ 12 years (graduated) ☐ College degree (B.S., B.A.)

   ☐ Some college or trade school ☐ Graduate+ (Advanced degree)

9) Have you ever fallen from a roof while working in construction? ---------Yes ---------No
10) If yes, did it happen in the past year? (If yes, we cannot include you in this study)

   ---Yes ----No

We would like to take your weight and height.

Weight ___________________

Height ___________________

Appendix B

Questionnaire 3 Donning Harness Usability

1. How do you rate the ease of putting on the harness?

Very Difficult					Very Easy
1	2	3	4	5	6

2. How do you rate ***wearing the harness***?

Very Comfortable					Very Uncomfortable
1	2	3	4	5	6

3. How do you rate the difficulty of ***tying the lanyard to the anchor***?

Very Difficult					Very Easy
1	2	3	4	5	6

Appendix C

Questionnaire 4 Tar Paper Usability

4. How do you rate ***working with the harness on***?

Very Difficult					Very Easy
1	2	3	4	5	6

Very Comfortable					Very Uncomfortable
1	2	3	4	5	6

5. How do you rate ***working with the lanyard present***?

Very Difficult					Very Easy
1	2	3	4	5	6

6. ***How often*** was the lanyard in your way while working?

Never					Always
1	2	3	4	5	6

Appendix D

Questionnaire 5 Shingles Usability Questionnaire

7. How do you rate ***working with the harness on***?

Very Difficult					Very Easy
1	2	3	4	5	6

Very Comfortable					Very Uncomfortable
1	2	3	4	5	6

8. How do you rate ***working with the lanyard present***?

Very Difficult					Very Easy
1	2	3	4	5	6

9. ***How often*** was the lanyard in your way while working?

Never					Always
1	2	3	4	5	6

Appendix E

Questionnaire 6 Doffing Harness Usability Questionnaire

10. How do you rate **taking off the harness**?

Very Comfortable					Very Uncomfortable
1	2	3	4	5	6

Very Difficult					Very Easy
1	2	3	4	5	6

Very Fast					Very Slow
1	2	3	4	5	6

Very Inconvenient					Very Convenient
1	2	3	4	5	6

Appendix F

Perceptions of Scaled Model Questionnaire

1. I felt safer working on this roof than an actual residential construction site

Strongly Disagree					Strongly Agree
1	2	3	4	5	6

2. I was less cautious while working on this roof than on one at a real residential construction site

Strongly Disagree					Strongly Agree
1	2	3	4	5	6

3. The structure that I worked on today accurately represented a roof at a real residential construction site.

Strongly Disagree					Strongly Agree
1	2	3	4	5	6

4. I made decisions that I would not normally make while conducting the same roofing tasks on an actual residential construction site.

Strongly Disagree					Strongly Agree
1	2	3	4	5	6

Appendix G

FAS Usability Rating and Post-Task Interview Questionnaire

Thank you for participating in our study. Your input will be very much appreciated. Feel free to let us know if you have any questions.

5. How do you rate wearing the harness from performing a task without wearing the harness?

Very Comfortable					Very Uncomfortable
1	2	3	4	5	6

Very Difficult					Very Easy
1	2	3	4	5	6

Very Inconvenient					Very Convenient
1	2	3	4	5	6

6. After wearing the harness would you recommend it to another company?

Highly Disagree					Highly Agree
1	2	3	4	5	6

7. Do you feel like your age or weight made it difficult to move around?

Very Difficult					Very Easy
1	2	3	4	5	6

8. After working in the harness would you agree that $200 is a reasonable price for the harness?

Highly Disagree					Highly Agree
1	2	3	4	5	6

Appendix H

Significant Univariate Effects for Context (at alpha= 0.05 level)

Dependent Variable	df	df error	F	p-value	Context	Means
Donning Harness Usability	1	54	8.08	0.0063	Indoor	7.09375
					Outdoor	9.14285714
Tar Paper Usability Rating	1	54	10.75	0.0018	Indoor	9.03125
					Outdoor	12.125
Shingle Usability Rating	1	54	9.57	0.0031	Indoor	9.32258065
					Outdoor	12.09375
Usability Rating for FAS	1	54	8.52	0.0051	Indoor	22.8064516
					Outdoor	27.1875